分类号：　　单位代码：10422

密　级：公　开　　学　　号：202020910



山东大学
SHANDONG UNIVERSITY

博士学位论文

Dissertation for Doctoral Degree
(学术学位)

论文题目：**The effect of users' stereotypes on perception bias in interpreting quality**
刻板印象驱动的口译质量感知偏差研究

作者姓名　李天韵
培养单位　翻译学院
专业名称　外国语言文学
指导教师　崔　英

2025 年 6 月 2 日

分类号：　　单位代码：10422

密　级：公　开　　学　　号：202020910



山东大学
SHANDONG UNIVERSITY

博士学位论文

Dissertation for Doctoral Degree
(学术学位)

论文题目：**The effect of users' stereotypes on perception bias in interpreting quality**
刻板印象驱动的口译质量感知偏差研究

作者姓名　李天韵
培养单位　翻译学院
专业名称　外国语言文学
指导教师　崔　英

2025年6月2日

The effect of users' stereotypes on perception bias in interpreting quality

Li Tianyun

Supervised by Prof. Cui Ying

A dissertation submitted in conformity with the requirements

for the degree of Doctor of Philosophy

in Foreign Language and Literature

School of Translation Studies, Shandong University

2nd June, 2025

原 创 性 声 明

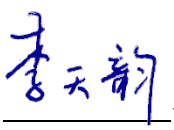
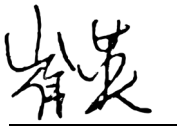
本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的科研成果。对本文的研究作出重要贡献的个人和集体，均已在文中以明确方式标明。本声明的法律责任由本人承担。

论文作者签名： 日　期：2025 年 6 月 2 日

关于学位论文使用授权的声明

本人同意学校保留或向国家有关部门或机构送交论文的印刷件和电子版，允许论文被查阅和借阅；本人授权山东大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或其他复制手段保存论文和汇编本学位论文。

(保密论文在解密后应遵守此规定)

论文作者签名： 导师签名： 日　期：2025 年 6 月 2 日

Contents

Tables II
Figures III
Abstract IV
摘　要 VIII
Chapter 1 Introduction 1
1.1. Background of the research 1
1.2. Objectives of the dissertation 4
1.3. Research questions 6
1.4. Significance of the research 7
1.5. Dissertation structure 9
Chapter 2 Literature review 10
2.1. Interpreting quality 10
2.1.1. Criteria for interpreting quality assessment 11
2.1.2. Definition of interpreting users 22
2.1.3. Users' expectations of interpreting quality 23
2.1.4. Users' overemphasis on delivery in perceiving interpreting quality 27
2.2. An interpreter's social categorical information affecting users' perceptions of interpreting quality 29
2.2.1. Gender 31
2.2.2. Race 32
2.2.3. Accent 33
2.2.4. Professional title 35
2.2.5. Age 36
2.2.6. Summary 36
2.3. Stereotypes on gender, ageing and professional title 37
2.3.1. Social categorization 38
2.3.2. The automaticity of social categorization and stereotype 42
2.3.3. Gender stereotypes 46
2.3.4. Stereotypes of aging 51
2.3.5. Stereotype of professional titles 54
2.4. Interpreter's social categorical information 59
2.4.1. Gender ratio of the interpreters' group 59
2.4.2. Age levels of interpreters 60

2.4.3. Interpreters' socioeconomic status 61
2.5. Summary 61
Chapter 3 Exploring the stereotype of an interpreter in China 64
3.1. Methods for stereotypes measurement 64
3.1.1. Open-ended and semi-open feature collection methods 65
3.1.2. Binary-structure measurement methods 67
3.1.3. Explicit-implicit measurement methods 69
3.1.4. The rationale for adopting the free association 70
3.2. Exploring the public stereotype of an interpreter 72
3.2.1. Design of the questionnaire 72
3.2.2. Distribution and collection of questionnaires 74
3.2.3. Data cleaning 75
3.2.4. Demographic statistics of respondents 78
3.2.5. Major findings 80
3.2.6. Minor findings 85
3.3. Discussion 89
3.4. Summary 92
Chapter 4 Exploring users' stereotyped interpreting quality perception 93
4.1. Experiment design 93
4.1.1. Experimental framework and variable control 93
4.1.2. Participants 95
4.1.3. Independent and dependent variables 95
4.1.4. The experimental stimulus 97
4.1.5. Preliminary set-up 99
4.2. The pilot study 101
4.2.1. Participants 101
4.2.2. Experimental procedures 102
4.2.3. Quantitative data analysis 103
4.2.4. Qualitative data analysis 107
4.2.5. Summary 108
4.3. The formal study 110
4.3.1. Hypotheses 110
4.3.2. Participants 111
4.3.3. Procedure 112

4.3.4. Analysis of expert assessment 113
4.3.5. Descriptive analysis of users' perception results 114
4.3.6. Significance analysis of users' perception results 115
4.4. Discussion of the formal study 119
4.4.1. Severity or leniency in expert assessors' ratings of interpreter performance ... 120
4.4.2. Gender stereotype on quality perception 120
4.4.3. Professional titles mitigate the negative stereotypes of natural cues 121
4.4.4. Ageing stereotype on quality perception 122
4.5. Summary 124
Chapter 5 A correction model for users' stereotyped interpreting quality perceptions 125
5.1. The dynamic interactive model of person construal explained 125
5.2. Towards an interactive quality perception model in interpreting 129
5.2.1. Explaining users' stereotypical quality ratings through the dynamic interactive model 130
5.2.2. Constructing the interactive model of interpreting quality perception 132
5.3. Correction of users' biased quality perception ratings 135
5.3.1. Calibration of users' quality perceptions through higher-level input 135
5.3.2. Manual correction of ratings based on an interpreter's perceived social categorical information 136
5.4. Summary 137
Chapter 6 Conclusion 138
References 141
Appendix 1. The questionnaire on the public stereotype of a professional interpreter in China 170
Appendix 2. Informed consent form for the collection of interpreting video materials beforehand recording 187
Appendix 3. Speech transcription content of experimental video 189
Appendix 4. The speech transcript of the interpreter's rendition 193
Appendix 5. Test for users' listening comprehension 198
Appendix 6. Questionnaire on users' perceptions of interpreting quality 200
致谢 205
攻读学位期间发表学术论文目录 208
学位论文评阅及答辩情况表 209

"For the most part we do not first see, and then define. We define first and then see".

(Lippmann, 1922, p. 81)

Tables

Table 2.1 A selected summary of studies on the criteria for interpreting quality assessment . 16
Table 2.2 Criteria on interpreting quality assessment by Liu (2015) 22
Table 2.3 A summary of studies on users' expectations of interpreting quality 23
Table 2.4 Briefing the different dimensions of social categorical information, take the interpreter Zhang Jing (张京) as an example 41
Table 2.5 Gender stereotypes and gendered expectations by Deaux & Lewis (1983), translated from Zuo (2015). 48
Table 2.6 The illustrative examples of age-stereotypical attributes and characteristics summarised by Rothermund & De Paula Couto (2024) 53
Table 3.1 Other demographic information of respondents 78
Table 3.2 The ten most frequently assigned stereotypical traits of the interpreting profession by the Chinese public. 83
Table 3.3 Sentiment analysis results of the descriptors submitted by respondents through free association regarding the interpreter group. 87
Table 4.1 Variables, content, and methods of measurement or control in this study..... 97
Table 4.2 Overall independent sample t-test analysis between the experimental group and control group across the four categories 104
Table 4.3 Independent-sample t-test analysis of high-proficiency foreign language users in the experimental and control groups 104
Table 4.4 Comparison of mean expert ratings with experimental and control group score differences 106
Table 4.5 Sorted results of qualitative data..... 107
Table 4.6 Effective sample size, mean age, and CET-4 pass rate for each participant group 112
Table 4.7 Assessment scores of experts. 113
Table 4.8 Summary of users' stereotypical perceptions of interpreting quality across conditions 116

Figures

Figure 1.1 Structure diagram for this dissertation 9
Figure 3.1 The word cloud of the most frequently mentioned descriptors for the stereotype of interpreters. The top ten descriptors are outlined in the first paragraph of this subsection. 86
Figure 3.2 The sentiment analysis of the descriptors reveals that the majority of respondents hold positive emotional perceptions of the interpreter group. 87
Figure 4.1 Photos presented in the control group, gender-swapped female group, and age-progressed group 100
Figure 4.2 Independent-sample Kruskal-Wallis test results of the three group means, group 1 represents expert ratings, group 2 represents the experimental group, and group 3 represents the control group. 107
Figure 4.3 The group relations between Y-M-NS as the control group and the other five experimental groups. 111
Figure 4.4 An interaction plot of mean scores by conditions. 114
Figure 5.1 A general diagram of the dynamic interactive model of person construal from Freeman and Ambady (2011). 126
Figure 5.2 Construct of the interactive model of quality perception (iMQP). 132
Figure 5.3 An example of the content within the higher-level cognitive states 134

Abstract

An interpreter's rendition is the direct product of an interpreting assignment, and accordingly, research on interpreting quality has consistently occupied a central position within interpreting studies. Since Bühler (1986) first proposed linguistic and extra-linguistic criteria for interpreting quality assessment, numerous interpreting scholars have subsequently advanced various evaluative frameworks. Most of these frameworks involve objective ratings provided by professional assessors according to established criteria. In this context, interpreting quality assessment can be conceptualized as a laboratory-based evaluation method, wherein objectivity and independence are preserved by isolating the assessment from contextual factors and subjective influences inherent to interpreting settings. However, from the perspective of interpreting as an interactive communicative process, interpreting involves tripartite communication among speakers, interpreters, and users (also termed "listeners"). Interpreters are thus not merely "signal transmitters" (Shannon & Weaver, 1998) or "signal relay stations" (Pöchhacker, 2004), but rather pivotal agents capable of shaping communicative dynamics (Downie, 2020, p. 15; Liu, 2021; Mason & Ren, 2012; Ren & Mason, 2011). During interpreting assignments, an interpreter's personal attributes are explicitly perceivable by both speakers and users. Furthermore, studies found that an interpreter's social categorical information which is not directly associated with linguistic competence may implicitly influence users' perceptions of interpreting quality.

As the direct recipients of interpreting services, users also possess the right to evaluate an interpreter's performance (Bühler, 1986). Moreover, user feedback regarding interpreting quality can substantially influence interpreters' professional advancement (Cheung, 2020). Studies indicate that users expect interpreters to deliver high-quality renditions (Kurz, 2002), with their expectations regarding fidelity, stylistic expression, and terminological accuracy closely aligning with those of expert assessors. However, given their typically limited proficiency in comprehending the source language, they tend to place greater emphasis on perceptible features such as fluency and vocal quality when perceiving interpreting quality. Consequently, significant discrepancies exist between users' quality perceptions and experts' quality assessments. Additionally, prior research has revealed that users exhibit biases favouring certain social categorical information of interpreters, even when such attributes have no direct relevance to interpreting quality. For example, Asian users have demonstrated greater tolerance for a female interpreter's errors; American users express lower trust toward Black interpreters

compared to their white counterparts; Hong Kong users rate an interpreter with a non-native accent with less tolerance; corporate interpreting users implicitly demonstrate negative biases toward older interpreters; and users from Mainland China attribute higher domain competence scores to interpreters with professional titles. Thus, users' perceptions of interpreting quality are not only subjective but also demonstrate predictable patterns of bias. Exploring such biases not only elucidates the underlying reasons for deviations in users' quality perceptions but also offers potential avenues for correction of these deviations.

Such perceptual biases exhibited by users based on an interpreter's social categorical information can be theoretically grounded in the concept of "stereotype" within social psychology. A stereotype refers to the fixed, generalized, and oversimplified perceptions held by social groups toward certain categories of individuals, events, or objects. By reviewing three prevalent stereotypes commonly associated with interpreting quality perception, this dissertation observes that public stereotypes regarding gender, age, and socioeconomic status align consistently with perceptions of an interpreter's corresponding inherent social categorical information. For example, females are perceived as being more adept at expression and having greater aptitude in foreign language acquisition compared to males; elderly individuals are viewed as conservative and cognitively sluggish; professional titles reflecting high socioeconomic status invoke a halo effect, leading to a generalized positive impression. Consequently, users' perceptions of interpreting quality are stereotype-driven, resulting in biased quality ratings.

To address the stereotype-driven biases in users' quality ratings, this study starts by designing and distributing a questionnaire to the general Chinese public. It aims to collect prevalent stereotypes concerning interpreters' gender, age, and socioeconomic status in mainland China. The survey also utilized the typical traits assignment methods from social psychology to gather representative occupational characteristics attributed to the interpreting profession. The findings revealed that the Chinese public stereotype an interpreter as a female aged between 25 and 35 years (young to middle-aged adults). Interpreters' socioeconomic status was perceived as moderate, with their political status analogous to that of journalists, economic status similar to programmers, and social status comparable to musicians. The top five stereotypical characteristics identified were high professional ability, strong psychological resilience, good disposition, fluent expression and high IQ. Sentiment analysis, based on the Emotion Vocabulary Ontology, indicated that public sentiment toward interpreters was primarily positive, with the majority reflecting praise.

Based on the public stereotype of an interpreter as "young to middle-aged females with

moderate socioeconomic status", this study designed a set of controlled experiments using age, gender, and socioeconomic status as independent variables. The dependent variables comprised four evaluative dimensions: fidelity, delivery, domain knowledge, and trust. Among these, fidelity and delivery were derived from the interpreting quality assessment criteria summarized by Liu (2015), aiming to measure perceptual differences between user groups and expert raters on traditional interpreting quality dimensions. Domain knowledge and trust, on the other hand, represent interdisciplinary dimensions that integrate interpreting quality assessment with stereotype testing. To reduce confounding variables, this study edited voice tones through audio software, and image-editing software was used to alter the gender and age of a perceived interpreter's CV photo. A general linear model was employed to test each dependent variable separately, it found that a perceived ad hoc interpreter possessing higher professional titles (e.g., physician) significantly enhanced users' perceptions of interpreting quality in terms of fidelity, delivery, domain-specific terminology, and trust. Additionally, a perceived male interpreter received higher quality perception scores compared to his female counterpart, and a younger interpreter was rated more positively than an aged interpreter. Subsequently, another general linear model was applied to examine the combined effects of multiple variables, demonstrating that professional titles exerted the greatest influence on quality perception, followed sequentially by gender and age. Consequently, within the context of this study, young male ad hoc interpreters with professional medical titles were more likely to receive favourable quality ratings from users, despite no substantive differences in the interpreted content compared to outputs perceived as originating from older male or young female full-time interpreters.

Based on the empirical findings, this study proposes an interactive Model of Interpreting Quality Perception (iMQP) to explain stereotype-induced biases observed in perceptions of interpreting quality. Drawing on Freeman and Ambady's (2011) Dynamic Interactive Model of Person Construal, iMQP posits that interpreting quality perception unfolds in a bottom-up manner, progressing through four hierarchical levels: cue, category, stereotype, and higher-order cognitive processing. At the cue level, both auditory and visual cues are processed; thus, users perceive not only the interpreter's audio rendition but also visual attributes such as facial features and physique. These visual and auditory cues subsequently enter the category level, enabling users to categorize interpreters according to their social categorical attributes and activating associated stereotypes at the stereotype level. According to the proposed model, even expert quality assessments will pass through the entire processing chain from cue level to higher-order cognitive level, which implies that expert assessors are not inherently objective or impartial. This explains observed tendencies among expert evaluators toward overly lenient or

overly strict scoring. However, compared with general users, experts maintain an advantage at the higher-order cognitive level, where their source-language comprehension skills, professional interpreting ethics, and cross-cultural communicative competence collectively mitigate stereotype effects. The model also proposes two methods for correcting users' deviations in quality perceptions. The first is pre-emptive intervention at the higher-order cognitive level, emphasizing the interpreter's professional titles to counteract negative stereotype effects related to gender and age. The second is a post hoc correction, adjusting user ratings based on empirically demonstrated bias tendencies. Lastly, this study acknowledges existing limitations and provides recommendations for future research.

Keywords: interpreting users, quality assessment, quality perception, stereotype, bias correction

摘　要

口译译文是口译任务的直接产品，因此译文质量研究一直是口译研究的重点领域之一。自 Bühler (1986)提出口译质量评估的语言和语外评估标准以来，已有数十位口译研究者相继提出口译质量的评估（assessment）标准。口译质量评估可认为是基于实验室设定的评分法，由专业评分人独立于现场环境设定、剥离影响评分的主观因素，以保证评分的独立性及客观性。但从口译的互动属性来看，口译是由讲话人、口译员与用户（也称“听众”）构成的三方交际活动，口译员不仅仅是“信号传输器”(Shannon & Weaver, 1998)，也是交际中能影响交际走向(Mason & Ren, 2012; 任文和伊安·梅森, 2011)的主要角色(Downie, 2020, p. 15; 刘益光, 2021)。口译员的自身属性不仅能清晰地被讲话人和用户所感知，其非语言能力相关的社会分类属性亦在无形中影响到用户对其口译质量的感知。

作为口译服务最直接的接收方，用户有权对口译员的表现进行评价(Bühler, 1986)，甚至用户的质量反馈影响到口译员的职业发展(Cheung, 2020)。虽然用户同样期待口译员能产出高质量的口译译文(Kurz, 2002)，他们对译文的信息还原、表达风格、术语准确度方面的期待与专家评分人基本一致。然而，由于用户理解原语的水平通常较为有限，在感知口译质量时明显更侧重听得见的流利度、音色等表达层面的内容。因此，用户的口译质量感知（perception）与口译质量评估存在较大出入。另一方面，已有研究发现用户对口译员的社会分类属性存在偏好，即便此类属性与口译质量不存在关联。例如，亚洲用户对女性口译员的译文中的错误更为宽容、欧美用户对黑人口译员的信任度显著不如其白人搭档、我国香港用户对非本地口音的口译员评分更苛刻、企业口译用户对年长的口译员表现更为消极、我国大陆用户对具有职业头衔的口译员赋予了更高的行业能力评分等。因此，用户的口译质量感知不仅是较为主观的，同时具有某种规律的倾向性。通过探索此类倾向性，不仅有助于论证用户在口译质量感知中形成的偏差，也为修正此类误差提供解决思路。

用户由于口译员的社会分类属性引起的感知误差，可以从社会心理学领域的“刻板印象”寻找理论基础。刻板印象是社会群体对某些特定类型人、事或物固定、概括而笼统的看法。通过梳理三类常见且与口译质量感知有关的刻板印象，本文发现，公众对性别、年龄、社会地位的刻板印象与其对口译员的社会分类属性的感知方向一致。例如：女性相较男性被认为更擅长表达、更具有外语习得天赋；年长人士被认为是保守的、迟钝的；而职业头衔折射出的高社会经济地位通过激发晕轮效应，使积极刻板印象泛化至整体层面。因此，用户对口译质量的感知受到刻板印象驱动，并由此造成感知评分上的误差。

为纠正刻板印象驱动的评分误差，本研究首先设计并发放了面向我国社会公众的调查问卷，旨在收集公众对于口译员的性别、年龄和社会经济地位的刻板印象。问卷亦通过社会心理学中的典型特征指派法收集了对于口译员职业群体的典型职业特征。研究发

现:中国公众对口译员的刻板印象的性别为女性、年龄分层为25到35岁之间的中青年。口译员的社会经济地位整体居于中等，其中政治地位与记者类似、经济地位与程序员类似、社会地位与音乐家类似。位列前五的典型特征分别为专业能力高、心理韧性强、仪表端庄、口才好、智商高。特征词通过基于情感词汇本体库的情感分析发现，公众对口译员群体的情感极性以正面为主，其中绝大多数为赞扬。

基于公众对口译员群体持有的"中等社会经济地位的中青年女性"的刻板印象，本研究进一步设计了以年龄、性别及社会经济地位为自变量，以准确（fidelity）、表达（delivery）、行业术语（domain knowledge）、信任（trust）四项评分为因变量的系列控制变量的实验。在因变量中，准确和表达基于刘敏华(2015)总结的口译质量评估标准，旨在考察用户组相对专家组在传统口译质量维度上的感知差异，而行业术语和信任两重维度是结合口译质量评估和检验刻板印象的跨学科维度。通过使用软件修改同一份音频的音色、同一张照片的性别或年龄以减少混淆变量。首先使用一般线性模型对单个因变量进行检验，发现非全职口译员拥有的更高职业头衔（例如医师）显著提升用户在准确、表达、行业术语和信任度上的质量感知，同理男性口译员的质量感知评分超过女性口译员、年轻口译员的感知评分超过年长口译员。之后使用一般线性模型检验多重变量叠加时的影响作用，发现职业头衔对质量感知的影响程度最高，依序为性别和年龄。因此，在本研究场景设定下，由年轻男性医师临时担任的口译员更容易获得用户对其口译质量的积极评价，即便译文内容与感知到的年长男性或年轻女性全职口译员的产出没有区别。

在实证发现的基础上，本研究提出了口译质量感知互动模型（interactive model of interpreting quality perception）以解释口译质量感知中出现的刻板印象偏差。本模型基于Freeman 和 Ambady(2011)的人际建构互动交互模型（the dynamic interactive model of person construal），主张口译质量的感知自下而上经过线索、类别、刻板印象和高阶认知四个层面。线索层面由听觉线索及视觉线索构成，因此用户不仅接收到译文音频，也接收到口译员的面部及体型等视觉线索。视觉线索和译文音频中的听觉线索进入类别层面，使用户形成对口译员社会分类属性的认识，并在刻板印象层面激活刻板印象。本模型认为，即便是专家质量评估也需要经过从线索到高阶认知的完整过程，因此专家评分亦不是完全客观平等的，可以解释专家评分中过于宽容或严苛的倾向。但与用户相比，专家在高阶认知方面具备优势，其原语理解能力、口译职业准则、跨文化交际能力等认知基础有助于减轻刻板印象。本模型同时提出两种修正用户质量感知偏差的方式，例如前期通过高阶认知层面强化对口译员职业头衔的认识，以积极刻板印象抵消性别、年龄的劣势影响；后期则可基于现有实证数据的偏差倾向调整评分数值以进行修正。本研究最后指出了现存的不足之处，为后续研究给出了建议。

关键词：口译用户；口译质量评价；口译质量感知；刻板印象；偏差修正

Chapter 1 Introduction

1.1. Background of the research

As an essentially oral occupation, the interpreting profession is characterized by the three-party interaction among the speaker, the interpreter and the user *(also known as "listener" or "audience")*. Throughout an interpreting assignment, especially consecutive interpreting, the interpreter is positioned between the speaker(s) and the user(s) as the third role. The primary responsibility of interpreters is to meet the interpreting service users' expectations to the best of their abilities (Kurz, 2002). Users refer to the individuals who received interpreting services throughout interpreting assignments, such as conference participants and delegates (Bühler, 1986), they are the most direct-end listeners of an interpreter's production. In other words, interpreters serve users by providing interpreting services.

According to the classical conduit model of interpreting (Anderson, 1976), the interpreter is expected to function as a neutral intermediary between the two parties, with the primary role of literally conveying their utterances in respective languages. Under such a concept, the goal of facilitating mutual understanding between the interactants is largely neglected. Thus, another model of the concept, the interaction model of interpreting views this assignment as a multifaceted interactive process involving three parties: the speaker, the interpreter and the user (Wadensjö, 2005). This model emphasizes the complex exchanges on social, cultural, emotional and cognitive levels through communicative processes (Pöchhacker, 2004). Since the interpreter takes over the speaker's original utterance, they also take on functional roles through interpreting for users, such as its vocative and pragmatic functions. Through the lens of the interaction model, an interpreter works *with* the speaker and users rather than simply *for* them (刘益光, 2021, 第 33 页). Interpreters are more than "signal transmitters" (Shannon & Weaver, 1998) or "switching stations" (Pöchhacker, 2004), they can be "active participants" (刘益光, 2021) and even "parents" (Downie, 2020, p. 15).

Users possess the right to judge an interpreter's performance, as they are the most direct recipients of interpreting services (Bühler, 1986). What users expect from an interpreter's production is also quality interpreting (Kurz, 2002), as quality should "begin with customer needs and end with customer perception" (Kotler & Armstrong, 1999, p. 568). However, the definition of quality in the ears of users may quite differ from the standard interpreting quality assessment criteria. The concept that users understand and feedback regarding interpreting

quality is generally referred to as “quality perception”, as it differs from the concept of “quality assessment” (Cheung, 2013, p. 29). Interpreting quality assessment is the comprehensive evaluation of an interpreter’s rendition by experts, who typically consist of professional interpreters, interpreting trainers at universities or members of industrial organizations such as the International Association of Conference Interpreters (AIIC). Experts’ scoring on renditions should be based on established industry standards, for example, the test rating rubrics of the China Accreditation Test for Translators and Interpreters (CATTI) (黄敏和王斌华, 2024). Throughout the assessment process of an interpreter’s performance, raters should remain objective, impartial and thus untainted by subjective preferences (Gile, 2009; Pöchhacker, 2004; Setton & Dawrant, 2016). This perspective is consistently emphasized and reinforced in the codes of conduct prescribed by national and international professional interpreting organizations (Mason & Ren, 2012).

Given that in the vast majority of interpreting scenarios, the interaction is among the speaker, users, and interpreter but with the absence of expert assessors to evaluate interpreting quality. In this case, the role of perceiving and assessing interpreting quality is generally undertaken by the users because they are the most direct recipients of renditions. However, users typically have not undergone formal interpreter training, they also lack the ability to simultaneously monitor both the source text and the target texts. Since users rely on the interpreter’s production to understand the speaker, in normal circumstances, they are not able to thoroughly understand the original text produced in the source language. This reliance limits their ability to accurately evaluate the extent to which the interpreter’s rendition maintains a “sense consistency with the original message” and achieves “completeness of interpretation” (Bühler, 1986, p. 231). Due to general users’ limited capacity in monitoring the source speech, they may turn to focus on delivery components of the interpreter’s production – hence users often mix “delivery-related criteria” (for example, self-corrections) with “content-related criteria” (for example, completeness of information) (Bühler, 1986; Kurz, 2002). Hence, even when significant portions of content were distorted or missing, yet the delivery was smooth, users may still form false impressions of a high-quality interpretation. Conversely, they may mistakenly a highly accurate rendition as flawed when the shortcomings originate from the source itself (Shlesinger, 1997). Thus, users’ subjective perception of interpreting quality may hardly be a reliable indicator.

Referring back to the triangular framework of the interaction model, users also interact with the interpreter through their feedback, questions and queries while receiving interpreting services (Wadensjö, 2005). Users, in general, are placed in a higher position than interpreters –

not only because of their "field of power" in communication dynamics (Bourdieu *et al.*, 1984; Mason & Ren, 2012), but also due to the importance of users for speakers. From the perspective of the field of power, interpreters are particularly vulnerable to users' dissatisfaction if the interpreting service is perceived as falling short of expectations (Hong, 2020). In some cases, they may even be scapegoated for issues unrelated to the actual quality of the interpretation (Cheung, 2015). Even when users are merely event participants, they may still lodge complaints with event organizers or language service providers for an interpreter's underperformance – as they perceived, thus potentially impacting the interpreter's career development (Chiaro & Nocella, 2004; Kurz, 2002; Yang, 2023). Under other circumstances, users themselves are the direct employers of interpreters. For example, full-time staff interpreters work for large international enterprises and organizations, or when a company temporarily hires an ad hoc interpreter to host foreign guests, users hold significant authority over the professional trajectory of interpreters. As users lack the capacity for objective assessment of interpreting quality, their perceptions of an interpreter's performance might be heavily influenced by subjective preferences and assumptions. This, in turn, affects the professional competence of practitioners in the interpreting market. From such a perspective, users' perceptions of interpreting quality may constitute an important social issue, as it directly pertains to the fairness of assessments regarding an interpreter's professional performance.

Users' perceptions of interpreting quality may also be influenced - and potentially biased - by aspects of an interpreter's personal attributes, which are not necessarily linked to their linguistic competence. Such biased quality perceptions may be triggered by an interpreter's perceived gender, race, locality and professional titles, as will be further discussed in the literature chapter. Interpreters who are male instead of female (Hong & Choi, 2020), Black instead of Caucasian (Obasi, 2013), non-local speaking with a non-native accent instead of local (Cheung, 2013), and full-time practitioners working for language service providers instead of ad hoc interpreters with highly-reputed professional titles (李天韵和崔英, 2022) are more likely to receive more stringent criticisms from users for their interpreting abilities. In another word, users have stereotyped perceptions of interpreting quality. Stereotypes are pervasive, oversimplified, and generalized beliefs or assumptions about a specific group or category of individuals, often grounded in attributes such as gender, race, or social status (Beeghly, 2015). While they can be positive, negative, or neutral, stereotypes shape perceptions, judgments and behaviours of users in a partial manner. Such biased perceptions may result in inequitable treatment of interpreters based on their affiliation with a particular group, leading to more partial and less accurate judgments of interpreting quality by users. In this regard, users' perceptions

of interpreting quality are most likely to be biased, as they may either overrate or undervalue an interpreter's performance based on certain non-linguistic attributes.

An interpreter may receive unjust treatment from users, not because of their rendition throughout an interpreting assignment, but due to whose personal attributes that evoked users' stereotypes. Given the significant impact of users' responses on interpreters' performance, research featuring interpreting quality should also incorporate such stereotypes inherent in user's perceptions. We contend that research on interpreting quality should not be confined to abstract, laboratory-based expert assessments. Instead, it should also take into account the societal issues inherent in the triadic interaction among the speaker, the user and the interpreter. While efforts have been made for a holistic assessment scale of interpreting quality (Lee, 2015, 2019) by assessors and automated quality assessment methods (Han & Lu, 2021, 2023), rarely there are studies by interpreting researchers venturing to address such practical quality perception issues in realistic interpreting scenarios. Consequently, it is challenging to apply the findings of experts' interpreting assessments to authentic interpreting scenarios. As a result, research on interpreting quality struggles to move beyond laboratory or classroom settings and fails to align with actual quality perceptions in practical interpreting assignments.

In real-world interpreting scenarios, feedback on an interpreter's performance mostly comes from users rather than expert assessors, as such performance is influenced by users' stereotypes associated with an interpreter's personal attributes. In other words, such performance perception ratings are often "contaminated". However, such "contaminated" data is not without value. By incorporating information about interpreters' social categorical information, it becomes possible to adjust users' quality perception ratings based on their stereotypical biases. This adjustment allows for a relatively fair re-evaluation of an interpreter's performance, even when there is a lack of expert assessment scores. It functions by correcting the users' ratings to account for the influence of interpreters' personal attributes. Therefore, the main objective of this dissertation is to explore the effect of users' stereotypes on perception bias in interpreting quality, and if possible, with a further exploration of establishing one correction model for users' stereotyped perceptions of interpreting quality. This model aims to adjust users' quality ratings of an interpreter's performance by accounting for the tendencies and degrees of their underlying stereotypes.

1.2. Objectives of the dissertation

The first focus of the present dissertation is to explore the generally held public stereotype of interpreters. Given that stereotypes are standardized and simplified conceptions or images

commonly held by members of a certain group, as they lead to oversimplified opinions and prejudiced attitudes for those groups. Stereotypes can be based on noticeable characteristics such as race, age, gender, social class and occupation. Before we perform data analysis to construct this quality correction model, it is necessary to first establish the reference level, that is the generally held stereotype of a "typical" interpreter. Therefore, the first focus of the present dissertation is to identify the public users' general occupational stereotype of a full-time interpreter in China. Research on stereotypes typically falls within the domain of social psychology, with related studies often concentrating on three primary categories, or the "big three": gender (or sex from a psychological perspective), age, and race (Cale, 2006; Eysenck, 2013). Gender, age and race are powerful social cues of social categorization (Weisman *et al.*, 2015), as individuals automatically rely on these visually salient cues to make rapid categorical judgments, which subsequently lead to stereotyped behaviours (佐斌等, 2019). However, since the author of this dissertation is of Chinese nationality and the major research was conducted in China, while China is a country with a relatively homogeneous population structure (The ChinaMAP Consortium *et al.*, 2020) and a small proportion of foreigners among its permanent residents (宋全成, 2013; 宋全成和熊方洲, 2024), race may hold limited research value in this dissertation. We turn to select an ad hoc interpreter's professional title(s) as the third category for experiments, not only because status is also an important social cue that can directionally influence individual stereotypes (佐斌等, 2019), but also due to the complexity of interpreters' titles as there are full-time freelance interpreters, full-time staff interpreters and ad hoc interpreters working part-time. If we consider another widely researched and commonly agreed occupation, such as the nurse, it becomes apparent that the general public holds a common occupational stereotype about nurses: a typical nurse is perceived as a young (Calvo-Calvo, 2014) female (Clow *et al.*, 2014; Feng *et al.*, 2016) whose social status is considered inferior to that of physicians (Wilbourn & Kee, 2010; 李强, 2000; 李强和刘海洋, 2009; 田志鹏等, 2013).

The second focus of the present dissertation is to investigate how an interpreter's personal attributes, to be specific, gender (male or female), age (young or aged) and professional titles (with or without additional title information) triggered users' stereotypes, and in turn, affect their perceptions of interpreting quality. After we accomplished the first objective of the dissertation, a stereotypical image of a full-time interpreter in China, for example, a young female with a moderate occupational status can be identified. Since not all active interpreters in the market conform to this stereotype, by controlling for equivalent interpreting quality, we can explore how an atypical interpreter's personal attributes triggered stereotyped quality

perceptions of users. These findings will allow us to further understand how interpreters with non-typical identities influence users' perceptions of interpreting quality. This focus will provide the dissertation with experimental data, allowing for the statistical determination of whether the stereotyped perceptions are positive or negative. Furthermore, this approach enables quantification of the specific extent of these biases, providing a measurable basis for their potential correction.

Since it becomes adjustable for corrections to be made with to interpreter's performance, the third focus of this dissertation is to explore a possible correction model for users' stereotyped perceptions of interpreting quality. Based on the statistical analysis outcomes from the second focus, which identified the appropriate directions and degrees for correction, this model will incorporate the interpreter's three specific identity tags, those are, the three primary categories examined in this dissertation: gender (male versus female), age (young versus old), and professional status (full-time versus ad hoc). After inputting the mean value of users' quality rating scores, this model will make according adjustments and then output the corrected score to reflect an interpreter's actual level of performance. This model would allow for the adjustment of users' perception ratings of an interpreter's performance in the absence of an expert assessment score, thereby producing a comparatively more objective assessment score that approximates expert assessments.

1.3. Research questions

The aim of the present dissertation is to explore the possibility of developing a correction model for addressing users' stereotyped perceptions of interpreting quality. An interpreter's perceived social categorical features, specifically the slightly modified "big three" categories: gender, age and socioeconomic status reflected through professional title will serve as the stimulus or stimuli for a set of experiments before we utilize the experimental outcomes for developing the model. We will first conduct a questionnaire survey targeting the general Chinese public to identify the public stereotype of an interpreter in China (Q1). This allows us to set up a baseline stimulus to be used for the control group. We will then examine each individual variable to determine the directionality in users' stereotypes – if they were positive or negative (Q2). Following this, we aim to obtain a comprehensive understanding of how these stereotypes interact across multiple variables, specifically by analysing professional title + gender, title + age, and the combination of title + gender + age (Q3). This correction model itself will specify the direction and degree of adjustment for users' quality ratings of an interpreter's performance, thereby correcting the perception bias caused by their stereotype of the interpreter (Q4). Taken

together, the following four research questions will be answered:

Question 1: What is the stereotype of an interpreter by Chinese users?

Question 2: Does an interpreter's perceived social categorical information activate users' stereotypes, thus biasing their perceptions of interpreting quality? If so, in which direction?

Question 3: How do multiple variates (gender, age and professional title) interact with each other, and does one category exert a major influence?

Question 4: Is it feasible to develop a correction model to modify the average score of users' quality ratings?

1.4. Significance of the research

An interpreter was expected to always remain impartial, neutral and invisible, functioning as a conduit for both parties being engaged in a conversation. The reality goes the opposite: interpreters were found as active participants (Mason & Ren, 2012; 任文和伊安·梅森, 2011), public figures (Delgado Luchner, 2019; Du, 2021; Hong & Choi, 2020) or scapegoats in some other accounts (Cheung, 2015). The notion of interpreting quality should be assessed in a very objective manner, though widely accepted among researchers and professional organizations, such assessments are often conducted in laboratory settings but not authentic interpreting assignments. In real-world contexts, users instead of expert assessors function as primary evaluators of interpreters' renditions and exert a decisive influence on the interpreter's career development. Users are not formally trained interpreting assessors, which is a phenomenon that has been reiterated in previous studies – their perceptions of interpreting quality can be influenced by an interpreter's non-linguistic factors such as whose accent (Cheung, 2013), gender (Hong & Choi, 2020), race (Cheung, 2020; Obasi, 2013), professional titles (李天韵和崔英, 2022) and other personal attributes. This discrepancy between inside-laboratory quality assessment framework and practical conditions opens up new research perspectives from the standpoint of interpreting service users. Apart from rediscovering what constitutes good interpreting quality through users' eyes, there is also a need to investigate the biases inherent in users' quality perceptions. Such efforts could possibly enrich the research perspective from the subfield of interpreting quality studies.

The research significance of this dissertation goes as follows: with regard to theoretical

accounts, the present dissertation introduced stereotype, the key concept in the research field of social psychology to interpreting research. Although previous interpreting researchers have recognized the potential biases present in users' perceptions of interpreting quality, some have also employed psychological concepts to explain their experimental findings, such as reverse linguistic stereotyping (Cheung, 2020; Kang & Rubin, 2009) and stigma (Cheung, 2013; Goffman, 1969), rarely have they interpreted their experimental findings from a more comprehensive, macro-level perspective. Therefore, the present dissertation sought to explain its experimental outcomes with one single yet systematic social psychological concept: stereotype. With a history spanning over one entire century (Gilbert *et al.*, 1969), in social psychology there have developed relatively mature theoretical frameworks, research paradigms, and experimental methodologies exploring social biases. Among these, stereotype is a well-established and evergreen research topic. Stereotypical researches into the "big three" categories, namely race, gender and age has been extensively studied. To name a few, there is substantial literature on the social division of labour work between men and women, including their representation in leadership and dominant occupational roles. Studies have also thoroughly explored both positive stereotypes of aged individuals, such as positive stereotypes like being knowledgeable and wise, and negative stereotypes such as slower reaction times and memory decline. Those theories and experimental design pathways may well organize the previous interview-based and experiment-based research findings, and reorganize them under the social psychological theoretical account.

The present dissertation also improves both the scope of data collection and the experimental designs. The previous studies exploring stereotypes in interpreting quality perception through users' lenses, as they were conducted by different scholars, may hardly achieve consistency in experimental design and testing of variables. For example, among the limited number of users-based quality research, Hong & Choi (2020) and Obasi (2013) employed open-ended interviews as their qualitative method of research instead of quantitative means, which lack statistical significance to support their findings. Although Cheung conducted a series of experiments on accent (2013, 2020) and race (2020), and Chiaro (2002) explored the impact of professional titles on quality and trust, there remain some issues of their experiments that warrant further scrutiny. This dissertation employed a unified set of experimental settings, allowing for consistent control of experimental variables measuring age, gender and professional titles, thus ensuring uniformity in both experimental design and data outcomes. Based on the research outcomes of single variable examinations, this dissertation further conducted bivariate examinations to investigate which variable assumes a more dominant role

when two factors are simultaneously considered. The final three-variable analysis examines the hierarchical relationship of multiple influencing factors, concluding that *professional titles* > *gender* > *age* in terms of impact. This finding provides crucial data support for the subsequent model development. The proposed model will facilitate the adjustment of users' perception ratings of an interpreter's actual performance when experts are not available, and thus yield a comparatively more objective evaluation that approximates expert assessment scores. This represents the last research significance of this dissertation.

1.5. Dissertation structure

This dissertation consists of seven chapters. Apart from the Introduction, the present chapter, the main body of the dissertation starts with the Literature review (Chapter 2), followed by the next three chapters (Chapter 3 to Chapter 5) to address the four research questions listed above, concluded by the Conclusion (Chapter 6), see Figure 1.1:

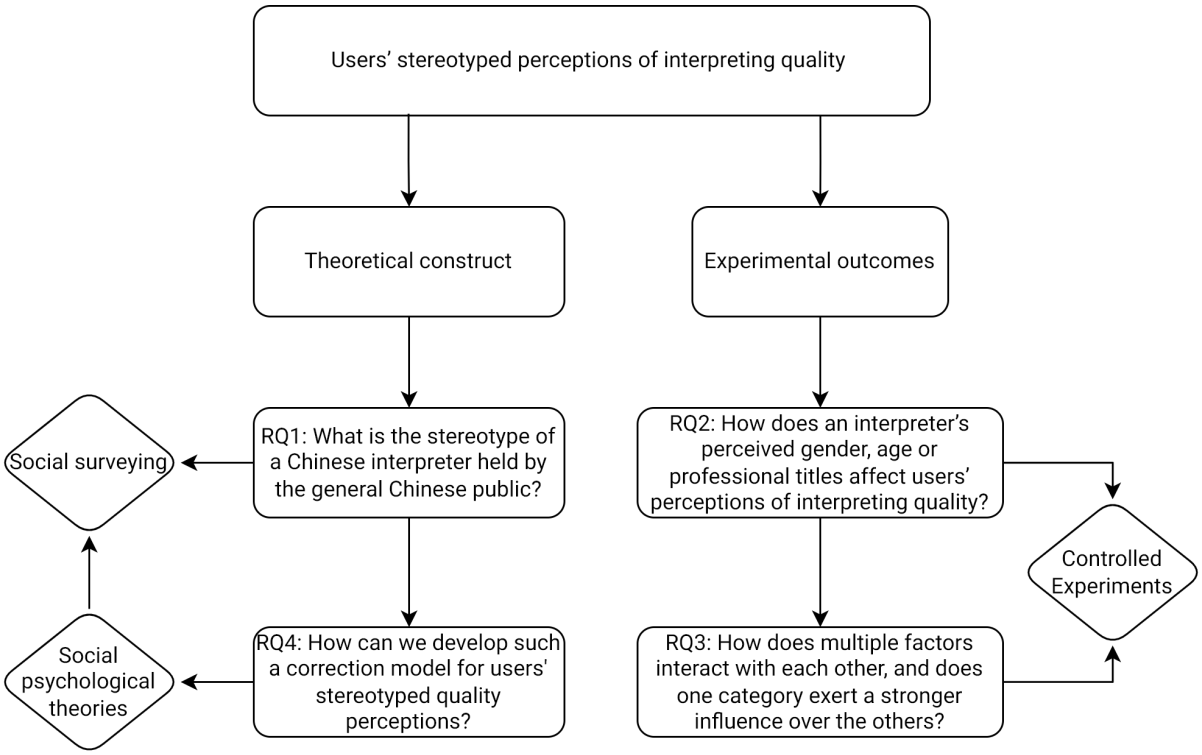


Figure 1.1 Structure diagram for this dissertation

Chapter 2 Literature review

The current chapter reviews relevant literature on interpreting quality, its assessment criteria from the perspectives of both experts and users, users' stereotyped perceptions of interpreting quality and their causes, stereotypes related to gender, age and professional titles and their manifestations in interpreting. The objective of this chapter is to provide theoretical support for understanding interpreting quality, and such factors that may bias users' perceptions from standardized quality assessment criteria.

2.1. Interpreting quality

Being a communicative activity that facilitates interaction and understanding between two individuals and/or grouped users who speak different languages, interpreting is a dynamic, context-dependent process that goes beyond linguistic translation (Bühler, 1986; Wang, 2023). As such, interpreting serves as a key mechanism in fostering social cohesion, inclusion, and dialogue in multilingual and multicultural environments. This socially-situated characteristic underscores the need for interpreters to process not only advanced linguistic proficiency, but also with the awareness of all users' social roles, expectations, and the overall communicative context (Anderson, 1976).

Quality is defined as the complete set of attributes of an entity that influence its capacity to fulfil both explicit and implicit needs (Hoyle, 2017). On a rather superficial basis, interpreting quality signifies the degree to which an interpreter's production meets the demand of end users and clients. However, as the notion of quality is not a fixed entity that can calculated from a very quantitative basis, interpreting quality is now generally regarded as a "complex and multifaceted concept that does not easily lend itself to empirical measurement" (Grbić, 2015, p. 334) after decades of researches delving into this concept. In other words, interpreting researchers have not, as yet, constructed a unanimously accepted definition of quality that is applicable to all interpreting contexts, even after decades of research (Collados Aís & Becerra, 2015; Grbić, 2015). In other words, the concept of quality in interpreting does not possess a single, absolute definition, as Moser-Meser (1996) already indicated nearly three decades ago. Among the underlying reasons for this phenomenon, one is that an interpreter's performance is not only perceived by users (as the recipient) and language service providers (as the hirer), but also by peer interpreters when working in a team of two interpreters or above (Hoza, 2022). Often being paired with a boothmate, an interpreter's performance is also to be evaluated by this counterpart (as the peer), either consciously or subconsciously. An expert interpreter, who

may occasionally hold positions as the team leader of interpreters, or be invited by the event organizer also supervise over and assess the team's performance (Obasi, 2013). Without doubt, users and peer interpreters differ markedly in terms of foreign language competencies and levels of interpreter training, their respective understandings, definitions and judgement criteria for quality itself also vary considerably. It brought about divergent considerations if the two groups assessed an interpreter's performance simultaneously: for example, studies as indicated in subsection 2.1.2 have found users' overemphasis on paralinguistic and communicative features in renditions, which confounded their judgements in information completeness and fidelity.

Thus, it is interesting to investigate what users perceived of an interpreter's production to diverge from their original expectations before an interpreting assignment starts. Such perceptions may also possibly differ from the assessment results by peers and expert interpreters. The present section offers a review of previous studies into interpreting quality assessment by experts, users' expectations of interpreting quality and their actual perceptions of it. We aim to delineate the typical characteristics of users' perceptions of interpreting quality. For example, due to their limited proficiency in foreign languages, users may place excessive emphasis on the paralinguistic features of the interpretation. Additionally, visible elements, such as first impressions of the interpreter and the interpreter's personal attributes, are often given significant weight in their overall evaluation of quality.

2.1.1. Criteria for interpreting quality assessment

The traditional perspective considered experts more qualified assessors than users to evaluate interpreting quality, as such quality assessment frameworks including ISO 18841 (ISO/TC 37/SC 5, 2018) were developed and revised by expert interpreters. Given the decades of scholastic contributions to interpreting quality that has brought a large number of research papers, theories and concepts, this subsection provides a concise summary of interpreting quality assessment from the experts' viewpoints. The aim is to delineate the typical characteristics of expert assessments from users' quality perceptions.

To establish an idealized benchmark for interpreting quality assessment, Déjean Le Féal (2008) constructed the concept of a perfect rendition, defining it as "zero defects". As illustrated by Déjean Le Féal (2008), a perfect rendition is expected to be presented with maintains identical cognitive content, being delivered with equivalent clarity and precision in the same language style. The linguistic and oratorical quality of the interpretation should match, if not exceed, that of the original speech (Déjean Le Féal, 2008, p. 155). The notion of "optimum quality" proposed by Moser-Merser (1996, p. 44) echoes zero defects, which posits that in an

ideal scenario, an interpreter is capable of delivering a complete and accurate rendition of the original message without any distortion. Furthermore, the interpreter strives to capture all extralinguistic information that the speaker may have provided, while adhering to the constraints imposed by external conditions. In summary, scholars including Déjean Le Féal and Moser-Merser proposed such perspectives that conceptualize quality as perfection and assert that perfection can be achieved through personal commitment and sustained through consistent and continuous effort. Contemporary translation and interpreting theories generally reject the perfectional quality concept which emphasizes the precise reproduction of the source text (Grbić, 2008): a perfect interpretation with zero defects is unattainable due to the inherent limited cognitive capacity for language processing (Dong & Li, 2020; Gile, 2009, p. 189; Seeber, 2011), and the unpredictable nature of real-time communication. Besides linguistic non-equivalence between source and target languages and divergence among different cultures (a typical example can be found in the English-Chinese language pair) (Chen, 2004; Cui & Liu, 2024), other frequent external conditions such as unfamiliar non-local accents (Albl-Mikasa, 2010; Cheung, 2013; Hale *et al.*, 2011), technical errors (Mouzourakis, 2006; Peng & Dai, 2024), improvised speeches deviated from the speech draft (Seeber, 2017) also restrained such an idealized construct within the pedagogical settings.

On the premise that ideal interpreting quality does not equate to perfection or zero defect, such a standard for the assessment of interpreting quality should be proposed from an operationally feasible approach. That is to say, such criteria for assessment are less “abstract” (Grbić, 2008, p. 245) but more measurable, pragmatic and applicable within real-world contexts, thereby effectively guiding interpreter training and accreditation. Since the term assessment generally refers to a “systematic method of obtaining information, used to draw inferences about characteristics of people, objects, or programs” (American Educational Research Association *et al.*, 2008, p. 216), in this regard, we may at first revisit such related principles in the most standardized documentation of interpreting, the *Interpreting services - General requirements and recommendations* of the International Organization for Standardization, or ISO 18841:2018 in abbreviation (2018). ISO standards are developed with the intention of being adopted as national standards (Timmermans & Epstein, 2010). The compiler of ISO 18841:2018, the Translation, Interpreting and Related Technology subcommittee, consists of experts in translation and interpreting, including Verónica Pérez Guarnieri. In the subsection of competencies in ISO 18841:2018 (ISO/TC 37/SC 5, 2018, p. 8), a competent interpreter is expected to convey the substance of all statements with accuracy, fidelity, and impartiality, ensuring that no additions, omissions, or misleading elements are introduced that could distort

the speaker's intended meaning. The general competence of an interpreter can be further divided into six sub-competences: linguistic competence, intercultural competence, interpersonal competence, technical competence and competence in research and in information acquisition. The linguistic competence of an interpreter refers to such proficiencies in production skills, listening and reading comprehension. This includes the ability to understand diverse regional accents and dialectal variations, and to recognize different language registers, such as formal and informal speech, specialized terminology, idiomatic expressions, colloquialisms, and slang. An interpreter is also expected to be capable of transitioning between formal and informal speech levels and accurately interpreting for individuals from varied educational and cultural backgrounds. For the intercultural competence, it denotes such awareness of being well-versed in cultural differences, thereby addressing cultural misunderstandings that may hinder communication and create biases. Additionally, the interpreter's professional consciousness and comprehension of relevant cultural factors should encompass the ability to interpret the significance of participants' behaviours, tones, and gestures within the communicative event, as well as to identify and navigate linguistic differences and similarities. Interpersonal competence is an interpreter's capacity to develop robust communication and relational skills necessary for effectively navigating demanding situations characterized by cultural and linguistic diversity. This competence encompasses abilities such as making introductions, establishing and maintaining appropriate positioning, managing turn-taking, and discerning the appropriate moments and methods for intervention, whether to seek clarification or for other pertinent reasons. An interpreter's capability in operating on-site language distribution systems and online remote interpreting platforms is whose technical competence. The competence in research and in information acquisition, include expertise and experience in utilizing terminology management tools, databases, specialized software and online search engines to develop effective strategies for the efficient exploitation of available information sources to prepare for interpreting assignments. The final competency of a proficient interpreter, known as domain competence, requires the interpreter to be thoroughly prepared with relevant domain-specific knowledge pertinent to the assignment and to possess a strong command of the specialized terminology and its functional equivalents in the working languages. Such a standard in defining qualified interpreting services is also applicable in China: T/TAC 3-2018, namely *Translation services – Requirements for interpreting services*, serving as the national standard for interpreter performance assessment in China, is the Chinese translation of ISO 18841:2018 (中国翻译协会, 2018). Other national standards such as D 1202 in Australia, DIN in Germany and UNI 10574 in Italy will not be

further discussed in this dissertation.

In addition to international and national standards, interpreting qualification certification examinations designed for interpreter accreditation also delineate criteria for assessing interpreting quality (Hlavac, 2013). Notable examples include the National Accreditation Authority for Translators and Interpreters (NATTI) Accreditation Test in Australia, the Language Training and Testing Centre (LTTC) Test in Taiwan province of China and the China Accreditation Test for Translators and Interpreters (known as CATTI) administered by the Foreign Language Publishing and Distribution Administration Bureau in mainland China. For illustration, the consecutive interpreting examination of CATTI Level 2 outlines quality parameters such as accurate pronunciation, clear enunciation, normative language usage, fluent speech flow, and an appropriate speech rate. It also emphasizes the proficient application of interpreting techniques, the complete and accurate rendition of the original content, and the absence of mistranslations and omissions (全国翻译专业资格（水平）考试办, 2004). This rubric can generally be categorized into informational completeness and delivery competencies. Parallels that of CATTI, the NAATI rubric encompasses transfer competency including the faithful transfer of meaning, the skills associated with the consecutive interpreting model, smooth delivery, and language competency, which refers to the idiomatic use of the target language[1]. Rubrics of the LTTC test also generally consist of the assessment criteria on content (complete and accurate delivery) and delivery (smooth and idiomatic production). To sum up, the aforementioned conference interpreting certification examinations exhibit significant overlaps. The primary parameters can be concluded into the interpreter's accurate comprehension of the source text and the fluent production of the target text, as summarised in a meta-analysis by Wang (2017).

In the same study, Wang (2017) also reviewed quality assessment parameters of interpreting proficiency examinations from seven conference interpreting programs across five countries and regions within translation education institutions. For example, the Faculty of Translation and Interpreting at the University of Geneva categorizes interpreting quality parameters into five dimensions: diction (appropriateness, naturalness, accuracy, diversity, correct register), vocal attributes (pitch, timbre, accent, dynamic intonation), expression (fluency, stability, clarity), and microphone usage (distance from the microphone, microphone noise). Excluding the technical dimension, the content dimension aligns with the previously

[1]. see https://www.naati.com.au/wp-content/uploads/2023/07/certified-interpreter-assessment-rubrics.pdf

mentioned criteria on content, while the dimensions of diction, vocal attributes, and expression correspond to the aforementioned criteria on delivery. Assessment criteria proposed by the School of Modern Languages at the University of New Castle consists of two general categories: fidelity and completeness and the target language usage & public speaking skills. In mainland China, the assessment rubric proposed by the School of Interpreting and Translation Studies of the Guangdong University of Foreign Studies (王巍巍, 2017) covers three major categories: completeness, fluency and effectiveness. The completeness rubric refers to an interpreter's ability to faithfully and precisely convey the source language information without distorting the original meaning, thereby maintaining the communicative intent. Fluency involves an interpreter's ability to appropriately utilize the target language based on their understanding of the source language, including the selection of suitable language and style for specific contexts, and the seamless transition between different languages. Effectiveness pertains to an interpreter's proficiency in utilizing vocal elements, gestures, eye contact, and other presentation techniques to achieve the same communicative effect in the interpretation as in the source language. To summarize, aside from a few minor differences (such as microphone usage and personal appearance), the majority of standard parameters "largely converge" on information, faithfulness and information (王巍巍, 2017, 第 51 页).

Beyond international and national standards, as well as the assessment criteria for interpreter accreditation and training established by various institutions, interpreting researchers have dedicated several decades to exploring and developing standards for interpreting quality. In an initial empirical effort to operationalize the construct of quality, Bühler (1986) surveyed AIIC members to identify the criteria they utilize in assessing the performance and behaviour of their professional peers. This study deconstructed the notion of interpreting quality into more concrete components, resulting in the development of sixteen specific survey items. These items were categorized into two primary dimensions: linguistic-semantic and extra-linguistic. The linguistic-semantic criteria included sense consistency with the original message, logical cohesion of utterances, correct grammar usage, completeness of interpretation, fluency of delivery, native accent, use of appropriate style, and use of correct terminology. Conversely, the extra-linguistic criteria comprised pleasant voice, thorough preparation of conference documents, endurance, poise, pleasant appearance, reliability, ability to work in a team, and positive feedback from delegates. Subsequent to Bühler's foundational work, interpreting researchers such as Schjoldager (1996), Kurz (2002), Kalina (2002), Pöchhacker (2002), Riccardi (2002), Bartłomiejczyk (2007), Lee (2008), Choi (2013), Han (2015, 2016a; 2021) and Chen et al. (2022) have developed analytic rating scales based on

rubrics for assessing interpreting quality. Based on the comprehensive overview by Lee (2015), minor modifications were implemented to create Table 1 below. This table summarizes a selection of assessment criteria for interpreting quality as identified by various researchers:

Table 2.1 A selected summary of studies on the criteria for interpreting quality assessment

Researcher	**Who to assess**	**Assessment criteria for interpreting quality**
Bühler **(1986)**	Conference interpreters	Linguistic-semantic: • Sense consistency with the original message • Logical cohesion of utterance • Correct grammar usage • Completeness of interpretation • Fluency of delivery • Native accent • Use of appropriate style • Use of correct terminology Extra-linguistic: • Pleasant voice • Thorough preparation of conference documents • Endurance • Poise • Pleasant appearance • Reliability • Ability to work in a team • Positive feedback of delegates

Schjoldager **(1996)**	Interpreter trainer	Quality as exception (point deduction): • Comprehensibility/delivery (irritating outbursts, unpleasant voice, excessive repairs, unfinished sentences, etc.) • Language (mispronunciation, unidiomatic language, inappropriate register, etc.) • Coherence/plausibility (abrupt endings/beginning, etc.) • Loyalty (omissions, unjustified changes/additions, etc.)
Kurz **(2002)**	Conference interpreters	• Accent • Voice • Fluency • Logical cohesion • Sense consistency • Completeness • Grammar • Terminology
Pöchhacker **(2002)**	Not mentioned	• Accurate rendition • Adequate target language expression • Equivalent intended effect • Successful communication interaction

Kalina **(2002)**	Not mentioned	Semantic content • Consistency • Logic and coherence • Completeness • Accurateness • Unambiguity • Clarity • Reliability Linguistic performance • Grammatical correctness • Adherence to target language norms • Comprehensibility • Stylistic adequacy • Terminological adequacy • Discretion • Lack of disturbances Presentation • Voice quality • Articulation • Public speaking • Discipline • Simultaneity • Technical mastery • Conduct

		Equivalence • Phonological deviations • Prosody deviations • Logical/semantic deviations • Lexical deviations • Morphosyntactic deviations • Reformulation Accuracy • Omissions • Additions • Incomplete sentences (simultaneous)
Riccardi **(2002)**	Not mentioned	Appropriateness • Register • Successful solutions Usability • Production deviations • Pauses • Technique • Overall performance • Eye contact (consecutive only) • Hand control and/or gesticulation and/or posture (consecutive only)

Lee **(2008)**	Professional interpreters	Accuracy • Deviations • Intended effects • Faithful rendering • Coherence/logic Target Language quality • Grammaticality • Phonology • Morphology • Syntax • Naturalness • Register • Style Delivery • Articulation • Pauses • Hesitation • False starts • Fillers • Excessive repairs • Frequent self-corrections • Voice
Choi **(2013)**	Professional interpreters	Accuracy • Opposite sense • No sense • Imprecision Expression • Terminological errors • Grammatical errors Presentation • Speed • Overuse of pause fillers/backtracking

Han **(2015, 2016a)**	Expert raters	Information completeness • Percentage of source text propositional content delivered in target text Fluency of delivery • Disfluencies that impede comprehension Target language quality • Instances of unnatural and incorrect usage
Chen et a. **(2022)**	Expert raters	Information completeness • Deviations • Inaccuracies • Minor/major omissions Fluency of delivery • (Un)filled pauses • Long silence • Fillers • Excessive repairs Target language quality • Unnatural expressions • Grammatical errors

To sum up, the above-mentioned criteria generally concur on three major categories: content, delivery, and language quality. A few criteria among them may still fall into more than one category (Lee, 2015), for example, an interpreter's self-correction may hamper not only delivery-related aspects, but also target language quality. Therefore, Liu (2015) further concluded this construct into two major categories: *content fidelity* and *output quality*. Content fidelity includes accuracy and completeness, which may map those sub-criteria related to information completeness and/or errors in the above-mentioned table. The remaining sub-criteria which were not directly related to information, for example, fluency of delivery and target language quality can be concluded into the other category of output quality. The two

major criteria were labelled as *accuracy* and *delivery*. Based on the two "most important criteria documented in the interpreting literature" (Liu, 2015, p. 165), Table 2 presents the assessment criteria consisting of *accuracy* and *delivery*.

Table 2.2 Criteria on interpreting quality assessment by Liu (2015)

Researcher	Who to assess	Assessment criteria for interpreting quality
Liu **(2015)**	Professional interpreters	Accuracy • Major errors • Minor errors Delivery • Hesitation • Repetition • Self-correction • Redundancy • Usages of grammar or terms • Coherence

2.1.2. Definition of interpreting users

To remove possible ambiguity, "users" in this dissertation refers to the person or group of persons who "ultimately uses the interpreting service delivered" (ISO/TC 37/SC 5, 2018, p. 2). In other words, users refer to the very end users (to say listeners or audiences) who directly listened to the rendition of the interpreter(s) throughout conferences or community activities. The general notion of users also includes clients and customers, who are the persons or parties that "pay for the interpreter's services" (ISO/TC 37/SC 5, 2018, p. 2) through formal agreements for the provision of interpreting. However, they are not that related to the subjects who are eligible for interpreting quality perceptions, as clients may not finally listen to the rendition through their ears. Besides the participants who attend on-site, remote audiences or listeners can also be classified as users, primarily attributable to the advancements in remote interpreting technologies in the digital internet era (Mouzourakis, 2003, 2006). It differs from the traditional notion that users have to be on-site.

2.1.3. Users' expectations of interpreting quality

As previously discussed, expert interpreters expect an interpreter to achieve ideal interpreting quality throughout interpreting assignments. Although errors, mistakes and omissions are inevitable due to the cognitive limitations of humans, the ideal performance of a capable interpreter should minimize these flaws to the extent that such errors remain invisible to the users (Grbić, 2008). Thus, an ideal rendition could be expected by expert assessors. On the other hand, there are also other stakeholders who naturally have a vested interest in expecting quality interpreting services, such as clients, peers, interpreting trainers and users (Grbić, 2008). Users expect high-quality interpreting because they are the target group relying on interpreters to fulfil their needs (Pöchhacker, 2002). Quality is critically important for effective communication from the users' perspective. In the absence of a demand for high-quality renditions, the necessity for interpreters would be eradicated, and the profession as a whole would cease to exist (Park, 2014). While the common definition of quality expectations refers to the desires or wants of customers, specifically what they believe a service provider ought to deliver rather than what the provider actually would offer (Teas, 1993), users' expectations of interpreting quality refers to what users believe they would like to get on the interpreter-mediated event (Pöchhacker, 2019), and to the extent that they are satisfied with it (Kurz, 2002). While quality in interpreting is increasingly regarded as a relative rather than an absolute concept (Macías, 2006), it may be contingent upon factors such as the mode of interpreting, its situational context, and the stakeholders who have access to it (Grbić, 2008). In this regard, quality in the ears of users may quite differ from the generally concurred assessment criteria proposed by accreditation authorities and interpreting scholars. Table 2.3 presents a summary of the studies on users' expectations of interpreting quality:

Table 2.3 A summary of studies on users' expectations of interpreting quality

Researcher	User type	Expectations
Bühler **(1986)**	Conference participants (through interpreters' perspectives)	Linguistic criteria • Native accent • Logical cohesion of utterance • Correct grammatical usage • Use of correct terminology • Fluency of delivery • Completeness of interpretation • Use of appropriate style

		Extralinguistic criteria • Reliability • Thorough preparation of conference documents • Ability to work in a team • Pleasant voice • Pleasant appearance
Marrone **(1993)**	Lecture attendees	More important • Substance • Fidelity • Completeness of information Less important • Linguistic quality • Prosodic features of interpretation (good voice, pleasant delivery)
Kurz **(1993)**	Conference participants	• Sensen consistency • Logical cohesion • Use of correct terminology • Completeness of interpretation • Fluency of delivery
Kopczyński **(1994)**	Conference participants and speakers	More important • Terminological precision • Detailed content • General content • Grammaticality Less important • Fluency • Diction • Pleasant voice
Moser **(1996)**	Conference participants and speakers	Faithfulness • Accuracy of content

		Synchronicity • Short time lags Rhetorical skills • Regular delivery • Absence of hesitation • Complete, grammatically correct sentences • Clarity of expression Quality of the interpreter's voice • A lively, non-monotonous voice • Clear enunciation
Kurz **(2002)**	Delegates	More important (above 3 points) • Sense consistency • Logical cohesion • Terminology • Completeness • Fluency Less important (below 3 points) • Grammar • Voice • Accent
Russo **(2005)**	Media audience members	More important • Rendition of general dialogue content • Fluency of delivery • Word/image synchronization Less important • Pleasant voice
Christensen **(2011)**	Court attendees	• Knowledge of relevant terminologies

- Knowledge of the legal system
- Style restoration
- Impartiality
- Grammatical correctness
- Fluency

Drawing on the criteria of quality expectations from genuine users, it is evident that they largely align with the concerns prioritized by expert assessors. Users of interpreting services concerns both content fidelity, such as information completeness (Bühler, 1986; Kurz, 1993, 2002; Moser, 1996), correct use of terminologies (Bühler, 1986; Christensen, 2011; Kurz, 1993, p. 2002, 2002, p. 2011), error suppression (Christensen, 2011; Moser, 1996) and output quality, such as fluency of delivery mentioned in all above-mentioned studies and proper language style (Bühler, 1986; Christensen, 2011; Moser, 1996). Due to prolonged time lags that may lead to a disruption in synchronicity, conference participants (Moser, 1996) and court attendees (Christensen, 2011) also expect interpreters to promptly produce renditions, thereby safeguarding overall operational efficiency.

As mentioned before, interpreting users constitutes a broad category. As per the definition of interpreting users in ISO 18841 (ISO/TC 37/SC 5, 2018), any individual or group that is capable of using an interpreting service qualifies as a user. In the digital era of remote interpreting, this definition extends to include remote television audiences who rely on interpreter-mediated content (Mouzourakis, 2006; Pöchhacker & Liu, 2024; Saeed *et al.*, 2022; Seeber & Fox, 2021). For example, young viewers on online video platforms such as Bilibili are also considered interpreting users. Such an expansive group can hardly be inherently homogeneous. Survey-based studies setting genuine conference users as interview subjects have underscored the heterogeneous nature of this population (Christensen, 2011; Diriker, 2004; Moser, 1996). Despite the heterogeneity in population distribution, users' expectations regarding content fidelity remain generally consistent, with the primary variations occurring in output quality. Experienced users, especially those who attend formal conferences (Marrone, 1993; Moser, 1996), symposiums (张威和柯飞, 2008) and court sessions (Christensen, 2011) tend to focus much greater attention on content match than the delivery factors, while general conference participants (Kurz, 2002) and media audiences (Russo, 2005) also considered an interpreter's voice less important than content accuracy.

In summary, despite variations within the user group, their expectations for a "good interpretation" (Kurz, 2002, p. 394) largely align with those of expert assessors. Experienced

users tend to supersede content fidelity over output quality, which is consistent with the quality assessment guidelines proposed by interpreting scholars (Gieshoff & Albl-Mikasa, 2024; Han, 2022; Lee, 2008; Zwischenberger, 2010). In other words, general users also consider factors related to language delivery, such as fluency, rhetorical effectiveness and lively intonation to be less important than content accuracy. In practice, however, there is a significant divergence between their actual quality perceptions from their expectations they expressed through interpreting researchers' surveys and interviews. This discrepancy is evident not only in their overemphasis on paralinguistic features such as fluency, and silence pauses, but also in their perception of an interpreter's non-linguistic personal attributes, which are unrelated to any criteria for interpreting quality assessment. Such discrepancies in these aspects will be further elaborated in the next two subsections.

2.1.4. Users' overemphasis on delivery in perceiving interpreting quality

Since users rely on interpreters as intermediates to access the speaker's content and also to convey their own messages, the completeness and accuracy of the information to be interpreted are the core to their needs of accepting interpreting services. This focus is consistent with their feedback regarding quality expectations, as elaborated in the previous subsection. However, some existing empirical research on users' reactions has indicated that they tend to disproportionately emphasize the interpreter's paralinguistic features of delivery in perceiving interpreting quality. Paralinguistic features refer to the non-lexical elements that accompany spoken language and convey meaning beyond the literal interpretation of words (Chen, 2009). One of the earliest empirical studies in this area is the investigation of Berk-Seligson (1988) into the use of polite forms by court interpreters. In this study, 551 participants comprising both monolingual English speakers and bilingual Spanish/English speakers served as mock jurors. Half of the participants listened to a Spanish witness testimony interpreted into English that included polite address terms, while the other half heard the same testimony without these politeness markers. The findings indicated that participants exposed to the rendition with politeness markers perceived the Spanish-speaking witness as significantly more convincing, competent, intelligent, and trustworthy. Consequently, the use of polite language was seen as enhancing the perceived competence of the interpreter. Gestures and facial expressions as non-verbal information, were also found as facilitators for stronger user reactions (杨承淑等, 2011).

Users' overemphasis on delivery, on some occasions, also affects their perception of content fidelity. For example, Collados Aís (2002) invited 42 legal experts to rate the interpreting quality of three videos of simultaneous interpreting: one in monotonous intonation

but with no content errors, another one in lively intonation but with some content errors, and the last one in lively intonation also without content errors. The participants rated the two lively-spoken renditions higher in intonation, as could be expected, however, the monotonously spoken video with no content errors was rated slightly lower in logical cohesion and sense consistency with the original message than its lively-spoken counterpart. Significant differences were also observed between these two in regard to overall quality. This finding demonstrates that lively intonation can significantly affect logical cohesion and sense consistency, even though intonation is merely a paralinguistic feature. Collados-Ais's subsequent research (2016) further corroborates this effect: renditions delivered with a melodious delivery, even when they contained errors, were generally rated more favourably than renditions being delivered in a monotonous tone, despite it conveyed complete sense consistency. In China, the same phenomenon was also observed by Zhang & Ke (2008): users seeking general information tended to have higher expectations regarding deliverable factors such as grammatical accuracy, pronunciation, intonation, vocal appeal, and style (张威和柯飞, 2008).

In the context of interpreting quality, fluency refers to the smoothness of delivery, it includes both the stability of the speech flow and the frequency of self-repairs, filled/unfilled pauses and slips (Macías, 2015). Some previous studies have also found the effect of fluency interfering with users' perceptions of interpreting quality. A rendition delivered fluently may create a false impression of high quality, even when significant portions of the message are distorted or omitted. Conversely, listeners may erroneously deem a faithfully rendered interpretation as deficient if the limitations stem from the source text itself rather than the interpreter's performance (Shlesinger, 1997). For example, after manipulating a video of German-Spanish interpretation by adding additional silent pauses, Macías (2006) then invited 43 expert users to perceive its interpreting quality. A replica of the control video interpretation, with only an additional 18 seconds of silent pauses, was rated lower in terms of logical cohesion, correct rendition of sense and completeness. Silent pauses also were found to hamper users' impressions of the interpreter's professionalism. This finding was also supported by the study of Park (2014), in which participants assigned lower ratings to linguistically identical interpretations in which one version incorporated inserted hesitations while the other contained self-corrections. The results indicate a clear, albeit modest, preference for interpretations with hesitations over those with self-corrections, as the latter had a more pronounced negative impact on listeners' perceptions. Specifically, self-corrections adversely affected judgments of the interpreter's accuracy, logical cohesion, and overall professionalism. Through a controlled study measuring hesitations, pauses and audible breathing, participants in the study of Rennert

(2010) also reported a negative impact of lower fluency on the perceived general interpreting quality. Rennert (2010) also suggested the link between perceived fluency and users' perception of the interpreter's accuracy.

To conclude, although information completeness and accuracy do not equate paralinguistic features such as pauses or self-corrections, users' perceptions of an interpreter's less fluent delivery can lead to their erroneous impressions of the overall quality of the interpretation, which leads to their doubt about the interpreter's professionalism. Interpreting scholars contend that firstly, limited foreign language proficiency among users hampers their ability to monitor the source text adequately, thus preventing them from discerning differences in accuracy between the source and the interpreted text (Cheung, 2020; Liu, 2023). In addition, users may possibly have only access to the interpreters' vocal features instead of from the source speaker (Becerra, 2016), especially in simultaneous interpreting when the interpreter's audio channel takes over the ground, thus when listeners encounter an interpreter's unnatural speech flow, they instinctively perceive it as the most salient deviation from normal speech delivery (Park, 2014, p. 20). These observations also indicate that even those users with legal expertise are not necessarily qualified to serve as reliable assessors of interpreting quality (Collados Aís, 2002). The next subsection will summarize how users' perceptions of an interpreter's social categorical information affect their overall perception of quality, despite these features were not classified as communicative acts and thus bear no intrinsic relation to interpreting quality.

2.2. An interpreter's social categorical information affecting users' perceptions of interpreting quality

In the research field of social psychology, "social category" refers to the grouped representation that individuals were classified into different categories (Klapper *et al.*, 2017). These categories, each sharing similar perceived characteristics and group attributes (佐斌等, 2019; 温芳芳和佐斌, 2019), provide certainties for structured social information (Rule & Sutherland, 2017). Since it is more than often that a target individual possesses multiple social identities (Roccas & Brewer, 2002), perceivers tend to rely on general automaticity mechanisms to infer the categories to which the individual belongs. For instance, during initial encounters, people quickly categorize strangers based on identity-related attributes (Krueger, 2001). The mental process of classifying the target individual into some (usually more than one) social categories is social categorization. Social categorization simplifies perception and cognition by either identifying inherent patterns of similarity within the social environment, imposing an organizational structure upon it, or both (Krueger, 2001). The process of categorizing others is

often being processed subconsciously, which means that such categorization is automatically being activated yet without having an awareness of such activation (Macrae & Bodenhausen, 2000; 佐斌等, 2019). An individual's facial feature is an easy-accessible cue used to identify the other as a member of a certain social category (Mason *et al.*, 2006; Rule & Sutherland, 2017). Certain social categories based on physiological attributes, such as age, race, and gender (known as the "big three"), are characterized by highly discernible markers, enabling individuals to classify members of these groups with near-perfect accuracy (Rule & Sutherland, 2017; Weisman *et al.*, 2015). Non-physiological social categories such as political affiliation, religious belief, and sexual orientation, although not very highly visible, also can be notified with minimal cognitive effort and without conscious awareness (Rule & Sutherland, 2017).

Also often in unintentional ways, social categorization triggers stereotypes (Macrae & Bodenhausen, 2000; Stolier Ryan M & Freeman, 2016). Such stereotypes, once activated, may elicit evaluative biases and related attitudes to the target individual. For example, Schaller (1991) found that the mere act of categorizing individuals into social groups automatically initiates a cognitive mechanism that influences the perception and processing of group-relevant information in a biased manner. This perception of a correlation between the group and certain behaviours may often be erroneous, such as the prevailing stereotype of female drivers (廖熠佳, 2019; 李莲莲, 2016) and female PhD graduates (李廷茹, 2008; 李莲莲, 2016) in China. Such concepts of formation and presentation of stereotypes will be elaborated in the next section. Since individuals of social categories are defined by attributes such as gender, race and age in specific social roles, over time, the characteristics associated with these roles become linked to the category itself, thereby reinforcing stereotypes. In return, these stereotypes restrict category members to their current roles, and thereby form a cycle known as a "vicious cycle" (Eagly & Koenig, 2021, p. 343) that hinders category members' movement into other roles with alternative demands.

Such stereotypes of interpreters' perceived social categorical information also bring about erroneous judgments of themselves. When put into the scenarios of interpreting assignments, it indicates users' stereotyped judgments on interpreting quality. The nature of interpreting service as a face-to-face communicational activity emphasized interpreters' socially situated properties. Either attending on-site or on-cloud, an interpreter cannot possible to completely hide his or her social categorical information, especially for those traits that can be the most easily captured by users through speaking, such as gender, age and so forth (Wadensjö, 2005).

An interpreter's social categorical information was found to affect users' perceptions of interpreting quality. In addition to the "big three", an individual's social markers such as whose

nationality, accent and socioeconomic status (Roccas & Brewer, 2002) can also be perceived at ease (Rule & Sutherland, 2017). As a pioneer in this subfield, Strong and Rudser (1986) collected subjective feedback from 25 sign language interpreting users on a couple of tape-recorded renditions. Sign language users, particularly those of community interpreting assignments were found to place a special preference on those interpreters who have deaf parents than hearing parents, as they expect that interpreters with such family environments perform better. Some experienced user often relates older-sounding interpreters with having deaf parents, and thus assume older-sounding sign language interpreters with higher interpreting competencies. This assumption could nevertheless be misidentifying: some interpreters with deaf parents in fact perform the worst (Strong & Rudser, 1986, p. 309). In this regard, using this subjective judgment basis about an interpreter's family background contributed to their stereotyped perceptions (Grbić, 2008), leading to their skewed perceptions of interpreting quality and biased preferences for interpreter hiring.

There is a scant amount of literature addressing users' perceptions of interpreter's social categorical information affecting interpreting quality perception, with even fewer studies employing experiment-based quantitative methodologies. Despite this scarcity, existing researches nevertheless unveil that interpreters' social categorical information is primarily reflected in age, race, gender and accent, with some studies also noting the influence of professional titles. A summary of these findings is provided below:

2.2.1. Gender

An interpreter's gender is found to affect users' perceptions of interpreting quality. Such effects, however, demonstrate inconsistent effects varying between East Asia and North America. In North America, female interpreters were found to be perceived as less capable of undertaking an interpreter's role than their male counterparts. The only quantitative study to date on this topic is by Christen y Gracia (2023). The experiment in this study involved 156 native Spanish-speaking participants who were evenly distributed by gender. The participants rated renditions of four TED videos being rendered by either male or female interpreters using a five-point Likert-scale questionnaire. Despite the experimenter having carefully selected the interpreters to guarantee the same level of interpreting quality, users still assigned significantly higher scores to the male interpreters in three key categories: general interpreting quality, rhythm and delivery than their female counterparts. The other interview-based qualitative study by Artl (2015) reached similar findings. Female interpreter respondents reported that their appearance and gender frequently attracted users' attention, which eclipsed their professional competence.

Users harboured biases about their expertise based on the interpreters' appearance and "the concept of femininity" (Artl, 2015, p. 50). In certain circumstances, their gender and outward appearance rendered them overly conspicuous from the users' perspective, which led to heightened scrutiny of the female interpreters' competence or professionalism. In contrast, none of the male interviewees indicated experiencing similar impacts related to their physical appearance in the course of their professional roles. However, this situation in East Asia appears to be reversed. Existing literature indicates that female interpreters receive greater leniency and recognition for interpreting quality compared to male interpreters, and their female gender affords them distinct advantages in the interpreting market. In an interview-based study, Hong and Choi (2020) noted users' "extreme gender bias and stereotype" of male interpreters (p. 674). Even if they are as competent as their female counterparts, seasoned male interpreters are still less likely to be acknowledged by user clients. On the contrary, users exhibited strong preferences for female interpreters in both consecutive and simultaneous conference settings. This imbalanced treatment shaped the phenomenon that male interpreters in South Korea should struggle to acquire a proper recognition of their professional identity. In terms of interpreting quality, Yang (2023) analysed online audiences' responses to an ad-hoc female interpreter's performance through video-time-based immediate subtitles (*Danmu* or *Danmaku*). Although there is no direct comparison with a perceived male interpreter in this study, it found that audiences elicited tolerant and appreciative responses for an ad hoc female interpreter's lapse. Instead of ascribing these shortcomings to the interpreter's incompetency, users tend to attribute them to the inherent complexity of the source text, and exhibit strong sympathy, empathy and tolerance to this female ad-hoc practitioner for such flaws. Besides the above studies directly into an interpreter's gender and users' perceived interpreting quality, some other studies also indicated no direct yet interesting findings: female interpreters are more likely to capture media exposure (Du, 2021; Hong & Choi, 2020) and being granted with positively gender-stereotyped nicknames, such as "女神 (goddess)" (Du, 2021), "the beautiful interpreter" (Hong & Choi, 2020) and "小姐姐 (Miss interpreter)" in both China and South Korea.

2.2.2. Race

There are three studies indicating that interpreters of colour are subject to users' differentiated perceptions of interpreting quality based on their race. This differentiation also extends to users' trust in their interpreting abilities. Obasi (2013) used questionnaires, telephone interviews, and in-depth interviews to conduct an interview-based study involving Black sign language interpreters in the United Kingdom. Twelve black interpreters participated in this research. The

findings revealed that these interpreters frequently encountered lower quality expectations from British users in their working environments. Renditions produced by Black interpreters were often "implicitly assumed" to be of lower quality due to users' racial stereotypes (Obasi, 2013, p. 107). Even participants who had obtained accreditation from the Register of Sign Language Interpreters reported facing rejection from clients, which limited their professional opportunities. Parallels to the findings of Obasi (2013), Nakahara (2016) also identified biased perceptions among American users regarding the interpreting quality of interpreters of colour. However, in Nakahara's study, the interpreters in question were Asians (including Filipino, Japanese American, and Chinese from mainland China, Hong Kong SAR and Taiwan province) instead of Blacks. Asian interpreter interviewees often experience assumptions about their qualifications, being mistaken for students or junior interpreters despite their acquired experience and expertise. They also reported facing users' consistent doubt require themselves demonstrate professional qualifications and competence. Some participants recounted encounters with less respectful users who explicitly stated that white interpreters were superior to their Asian counterparts, even requesting a transfer to a white interpreter. In contrast to the stereotypical quality perceptions observed among British and American users in the studies by Obasi (2013) and Nakahara (2016), Cheung (2020) found that Chinese users generally exhibited more lenient and tolerant attitudes toward an English-Chinese consecutive interpreter with a Caucasian appearance. In this study, two groups of participants in mainland China were shown a video of former American President Obama's speech being interpreted into Chinese by an interpreter speaking with a non-native accent. One group of participants was presented with a photo of a Chinese-looking male alongside the video, while the other group was shown a photo of a Caucasian-looking male. Results from independent sample t-tests indicated that the Caucasian-looking interpreter was rated significantly higher in terms of comprehensibility, fluency, idiomatic usage, clarity of accent, and professionalism. On the contrary, the Chinese-looking interpreter speaking with a foreign accent received negative comments and speculations from users due to his substandard pronunciation.

2.2.3. Accent

An accent is an important social cue which conveys information about a speaker's geographical (Bestelmeyer, 2024), ethnic (Louw & de Wet, 2007; Paladino & Mazzurega, 2019) and socio-economic background (Stewart *et al.*, 1985; 纪莉和吴逸悠, 2015). It is more than merely a byproduct of coalitional categorization, but also one dimension of social categorization (Pietraszewski & Schwartz, 2014). The accent is a comparatively frequently studied factor in

interpreting quality perception, as accent is also a salient linguistic feature (Petrochuk, 2021). Evidence from the four experiment-based studies suggests that an interpreter peaking with a standardized accent is generally perceived as more competent than speaking with a substandard accent. For instance, in Taiwan province of China, Zhang (2009) conducted a study in which four professional interpreters, including one who also worked as a broadcaster, recorded a passage in three pronunciation styles: Standard Mandarin, a natural guise, and a regionally accented guise. The content of the renditions was identical and carefully controlled. A total of 121 participants rated these four matched guises. The findings revealed that renditions delivered in Standard Mandarin were perceived by users as more professional and of higher quality compared to those delivered with regional accents, despite the content and overall quality of the renditions being identical. This result echoes the finding in the study of Lim & Fan (2023), who conducted a controlled experiment involving four hundred participants in the Taiwan province of China. The participants being randomly assigned in equal numbers listened to one of four simultaneous interpreting versions, each distinguished solely by accent: Standard Mandarin, Mandarin with a local Taiwan provincial accent, Malaysian-accented Mandarin, and American-accented Mandarin. Analysis of participants' quality ratings revealed that the interpreting quality was perceived less favourably in terms of comprehensibility and performance satisfaction as the accent's markedness increased. The Standard Mandarin accent was deemed the most appropriate stylistic choice, attributed to its standard prestige, while American-accented Mandarin received the lowest ratings for comprehensibility. Likewise, interpreting users in Hong Kong tend to underrate Cantonese interpretations delivered with non-local accents. Cheung (2013) conducted a controlled experiment to examine native Hong Kong Cantonese speakers' perceptions of the quality of three different simultaneous interpreting renditions into Cantonese. These renditions differed solely in the interpreters' accents: native-accented Cantonese, Mandarin-accented Cantonese, and English-accented Cantonese. Results showed that participants who listened to renditions with non-native accents consistently assigned lower quality ratings across all survey items including comprehensibility, delivery, error frequency, pacing, fluency, and the interpreter's professionalism compared to those who listened to native-accented Cantonese renditions. An exception to these findings is presented in the study by Hale et al. (2011), which examined the impact of interpreters' foreign accents on witnesses' testimony. The results indicated that participants rated accented versions of testimony as more credible, honest, trustworthy, and persuasive compared to non-accented versions. This phenomenon was explained by the unique context of court interpreting, where accented renditions may function as outgroup categorical information in social psychology,

thereby minimizing potential biases or favouritism typically associated with ingroup dynamics. Consequently, this study differs in functional nature from the three previously discussed studies.

2.2.4. Professional title

A professional title is a formal designation that signifies an individual's professional role, rank, or status within a specific field, often reflecting their expertise, qualifications, or occupational achievements (Bass, 1986; Jones, 2018). Thus, the title serves as a social marker conveying information about possible social hierarchy (Beike & Sherman, 1994).

The interpreting profession occupies a unique position in the labour market, as a significant proportion of practitioners work on a part-time basis, commonly referred to as ad hoc or lay interpreters (Giustini, 2022; Katan, 2009). For instance, the annual industrial development report by the China Academy of Translation and the Translators Association of China indicated that over half of interpreters in the field are part-time practitioners (p.69). In addition to their professional identity as an "interpreter", ad hoc interpreters often hold multiple professional identities, such as university professors, lawyers, or physicians. When an ad hoc interpreter's primary professional title is advantageous for certain interpreting tasks, such as a practising physician taking the role of the interpreter for an international medical seminar, the interpreter himself or herself can strategically highlight his/her primary professional title to users and/or clients. This emphasis on their specialized expertise can, in turn, shape users' perceptions of the interpreter's quality and competence.

Two experiment-based studies have found such a facilitatory effect of the interpreter's perceived professional title on users' quality perceptions. Chiaro (2002) conducted a study comparing the quality ratings of genuine TV audience users with those of interpreting undergraduates. The experimental stimulus was rendered by an ad hoc interpreter who, despite lacking formal interpreter training, was a highly reputed TV presenter widely recognized by Italian television audiences. Since interpreting students have undergone formal interpreter training, they were assumed to be more competent assessors of interpreting quality compared to the general TV audience. The findings indicated that TV audience users were influenced by their trust in the TV presenter's reputation, they were assigned significantly higher quality ratings in fidelity, lexis, syntax, and style than the interpreting students. Chi-square analysis indicated a strong association between TV users' trust and quality perception, particularly for the variables of lexis, fidelity, and style. This suggests that the trust inspired by the TV presenter's professional title had a notable impact on audience perceptions. In contrast, interpreting students placed significantly less trust in the TV presenter's rendition, resulting in

lower quality ratings. In line with the research conclusion of Chiaro (2002), also as one of the author's research achievements during the PhD stage, Li & Cui (2022) experimented with users' quality ratings of an ad hoc medical interpreter's performance. This study involved experimental and control groups, both of which assessed the same interpretation video. The only variable was the presentation of the interpreter's perceived professional title as a highly educated physician and medical scholar prior to the video. Quantitative and qualitative analyses revealed that the interpreter's professional titles have generated both positively biased quality ratings in fidelity, fluency and domain knowledge, but also significantly enhanced users' trust in the professional competence of the interpreter.

2.2.5. Age

Ageing is an inevitable process for all humans (North & Fiske, 2012), interpreters as well. According to the AIIC lifespan study, some active interpreter interviewees were 70 years or older at the time of the interview, as there is no such mandatory retirement requirement for professional interpreters (Liu *et al.*, 2020). In other words, interpreting theoretically can be a lifelong profession. Aged interpreters may face negatively biased perceptions in interpreting quality. The same experiment by Christen y Gracia (2023), as previously mentioned, found that older female interpreters were rated significantly lower in interpreting quality compared to younger interpreters of both genders. Christen y Gracia attributed this result to stigmatizing attitudes rooted in patriarchal norms that negatively impact perceptions of older women. This negative stereotype of aged interpreters also manifests in clients' decisions to discontinue hiring aged interpreters, as reported by interviewees in the lifespan study by Liu et al. (2020).

2.2.6. Summary

In conclusion, the findings from the above-mentioned literature suggest interpreters' social categorical information including age, race, gender, accent, and professional titles can alter users' perceptions of an interpreter's performance. When contextualizing these findings within specific cultural and regional settings, such as the United Kingdom or the Americas, users tend to favour interpreters who align with the social categorical profile of a young, local Caucasian male interpreter. While in East Asian countries like China, users exhibit a stronger preference for highly-educated female young interpreters of the same nationality. As these studies were conducted by different researchers under varying contextual frameworks and methodological designs, their ecological validity is inherently limited. Consequently, the findings are context-specific and may not be generalizable to more diverse cultural settings. The variations in research scope, participant demographics, and cultural contexts restrict the applicability of the

conclusions to broader, multicultural scenarios. While the aforementioned experimental studies examined individual traits as isolated variables, we argue that a more comprehensive approach is necessary to account for the multifaceted nature of social categorical information of an interpreter, such as whose age, gender, and professional titles, in altering users' perceptions of interpreting quality. Specifically, it is essential to explore the effects of multiple variables and their potential interactions, as individuals often simultaneously possess multiple social identities. This complexity is particularly evident when these identities are perceived as interconnected and interdependent, rather than as distinct or singular constructs (Roccas & Brewer, 2002).

Future research on users' perceptions of interpreting quality can benefit from adopting a more interdisciplinary approach. As mentioned in previous literature, users are typically laypersons who lack the ability to comprehend the speaker in the source language. Their perceptions of interpreting quality are often skewed, not only due to an overemphasis on delivery but also because of their perceptions of an interpreter's social categorical information. This phenomenon is consistent with research topics in social psychology. It is therefore advisable to draw on social psychological theories and experimental findings to address the complexities of users' quality perception in interpreting. In the next section, we will briefly introduce the concept of social categorization in social psychology and then explore the impact of stereotypes related to gender, ageing, and professional titles on quality perceptions.

2.3. Stereotypes on gender, ageing and professional title

Due to its significant influence on social cognition, behaviour and intergroup relations, stereotype is one of the key research topics in the field of social psychology. The concept of stereotype was first formally defined in the 1920s by Walter Lippmann (1922) in his monograph *Public Opinion*, where Lippmann described it as a mental paradigm that shaped people's preconceived impressions before they established contact with the target person. Stereotype is defined as a cognitively structured oversimplified belief about the characteristics, attributes, or behaviours of members of some certain social category (Kanahara, 2006). These beliefs are widely held by the general social public (Czopp *et al.*, 2015) and are gradually shaped through cultural narratives, social interactions, and media representations (Seiter, 1986; Ward & Grower, 2020). They act as cognitive shortcuts that allow humans to save repetitive cognitive resources (Carlston & Skowronski, 1994; Macrae C Neil *et al.*, 1994). Through this means of shortcuts, stereotypes help humans to seek the optimum option while avoiding the unpleasant sensations associated with ambiguous uncertainty, and steer individuals toward rapid response (Fiske &

Neuberg, 1990; Ford, 2000).

As stereotype helps to provide a simplified framework for understanding complex social information, it facilitates mental energy conservation through cognitive shortcuts (Macrae C. Neil *et al.*, 1994). In general, stereotypes provide individuals with cost-effective judgments that can be accurate in certain contexts; for instance, the assumption that professors are knowledgeable (Anderson, 2010). However, the formation of stereotypes often relies on limited and indirect exposure to real-world examples, with significant influence from educational systems (Brey & Pauker, 2019) and media narratives (Ward & Grower, 2020). Such an oversimplified method of categorization can lead to overly generalized and rigid prejudices. For example, non-Asian citizens in the United States may assume that all Asian American students should excel in mathematics (Cvencek *et al.*, 2015), illustrating how stereotypes can perpetuate broad and inaccurate generalizations. Moreover, as such heuristic information processing exaggerates group uniformity and obscures differences between individuals within a particular group, it might impose detrimental consequences on such counter-stereotypic individuals, whose personal traits and behaviours may deviate from such generalized expectations associated within their social group (Wen *et al.*, 2020; Zuo *et al.*, 2019).

The primary research areas in social psychology concerning stereotypes include age stereotypes, gender stereotypes, occupational stereotypes, national stereotypes, class stereotypes, and racial stereotypes (Fiske, 2017). Among which, factors like gender, age and race (which are also the "big three" of the social psychology research area (Weisman *et al.*, 2015)) belong to physiological clues, and factors like occupational identity, socioeconomic status and political affiliation belong to social clues (温芳芳和佐斌, 2019). Building upon the discussion in the preceding section regarding interpreters' social categorical information, including gender, race, age, accent and professional titles, this section will provide a concise overview of the stereotypes associated with these factors and their potential implications on the social psychological perspective: what are these stereotypes and they ways they manifest to skew individuals' judgements.

2.3.1. Social categorization

Humans perceive the world through the filter of categorization, distinguishing between natural kind categories such as animals and artificial categories such as occupations (Krueger, 2001). According to the *Oxford Dictionary of English*, a category constitutes a "class or division of people or things regarded as having particular shared characteristics" (Stevenson, 2010, p. 273). Categorization emerges when distinct objects or individuals are cognitively treated as

equivalent through the identification of inherent similarities (Mervis & Rosch, 1981; Roets *et al.*, 2015). Human cognition relies on categorization as a basic mechanism (Freedman *et al.*, 2001) to organize complex information, reduce ambiguity, and optimize decision-making processes (Jenkins, 2000; Krueger, 2001). This adaptive process constitutes a dynamic interaction between perceptual systems and environmental stimuli, involving the grouping of phenomena according to culturally and cognitively salient attributes (Cohen & Lefebvre, 2005, p. 21). As a fundamental cognitive tool, categorization shapes language use, influences knowledge and identity construction, and informs evaluations of others and subsequent decision-making (Conradie et al., 2017).

In the field of social psychology, categorization is understood as the mental grouping of objects or individuals based on shared characteristics or functions deemed significant within a given cultural, cognitive, or institutional context (Bodenhausen *et al.*, 2012; Jenkins, 2000; Krueger, 2001). Social categorization specifically addresses how individuals classify themselves and others (Krueger, 2001). In social categorization, these similarities, differences, or shared attributes may include race, gender, age, occupation, or cultural affiliation (佐斌等, 2019; 温芳芳和佐斌, 2019). This cognitive framework organizes complex social knowledge, structures past experiences into coherent patterns (Fiske & Neuberg, 1989), and generates predictive expectations about future interactions (Lerner & Lomi, 2018). As a crucial psychological mechanism for human adaptation, social categorization provides an organizing principle that meets individuals' need for structured information to aid decision-making in novel contexts. The efficiency of social categorization stems from its minimal cognitive demand. This fundamental heuristic facilitates navigation of intricate social environments by simplifying information processing, largely through cognitive shortcuts (Pattyn *et al.*, 2015; Stolier Ryan M. & Freeman, 2016). By categorizing others, individuals rely more on memory-based rather than data-driven processing, which demands greater cognitive effort. Given individuals' extensive prior experience with these categories, social categorization persists even under time pressure. Under cognitive load, individuals increasingly rely on these pre-existing categorical frameworks rather than engaging in detailed information processing (Taylor, 1981). For instance, an ERP-based experiment with faces as the priming stimulus found that gender categorization occurs approximately 195 milliseconds after viewing a face stimulus (Zhang *et al.*, 2016). Hence, social categorization functions as an efficient cognitive mechanism, allowing individuals to swiftly access previously stored knowledge (along with any implicit attitudes) to make rapid categorical judgments.

People rely on perceptual cues to cognitively group others. According to the interactive

theory of person construal (Freeman & Ambady, 2011, 2014), social perception operates through a system integrating low-level sensory processing of facial, vocal, and bodily features with higher-order cognitive interpretations. This process of social categorization hinges on various cues, which can be divided into two broad types: natural cues and social cues (佐斌等, 2019). Natural cues refer to biologically rooted, genetically stable characteristics that exhibit minimal intra-individual variation over time. These include phenotypical attributes such as height, gender, race (manifested through facial morphology and skin tone), and age. Thus, its features are largely determined by genetic inheritance and resistant to modification through personal effort. By contrast, social cues are more complex and context-dependent (Rule & Sutherland, 2017), reflecting socially constructed indicators used to classify individuals into particular social groups. These may include abstract or invisible markers such as linguistic accents, professional titles, political affiliations, and socioeconomic status. Unlike natural cues, social cues are mutable and often attainable through personal achievement or contextual shifts. For instance, nationality or occupational roles can be altered through immigration or career advancement. Furthermore, social cues are inherently dynamic, as their meaning and significance shift across different social settings. In sociological terms, these cues align with Linton's (1936) concepts of "ascribed status," referring to characteristics assigned without regard to individual merit or abilities, and "achieved status," referring to those that require specific qualities or accomplishments and are open to individual attainment.

As individuals inherently possess overlapping identity markers across various social dimensions, social categorization operates through multidimensional cognitive frameworks (Wattenmaker *et al.*, 1995). This process inherently involves multiple categorization, which is a dynamic interaction in that distinct social cues intersect rather than remaining singular and isolated (Crisp & Hewstone, 2007; Mullen *et al.*, 2001). Multiple categorization occurs when individuals are simultaneously classified into distinct social groups, often intersecting social categories across different dimensions. For example, a person might be categorized such as gender (female), race (Mongoloid), age (middle-aged), and occupation (scientist). Those cues may collectively form a composite identity like a "middle-aged East Asian female scientist" (Crisp & Hewstone, 2007). This subsection will not elaborate extensively on the psychological mechanisms underlying multiple categorization formation, in turn, we will centre on the centralised cue effect.

Under conditions of multiple social categorizations, individuals tend to prioritize one centralised cue to guide their holistic judgments and attitudes toward others. The centralised cue effect emerges when a particular categorical cue demonstrates greater cognitive salience or

sociocultural significance within composite prototypes. Take Prof. Yan Ning (颜宁) as an example. Prof. Yan was born in 1977, and is aged 48 in the year 2025. The period of middle age ranges from 45 to 65 years (Stevenson, 2010, p. 1110). Yan possesses both natural cues (female, Mongoloid, middle age) and social cues (Shandong origin, an academician of the Chinese Academy of Sciences, a professor at Tsinghua University). Public perception in China predominantly categorizes Prof. Yan as a "female scientist", a "scientist" in particular. Media interviewers generally emphasise the scientific achievements of Prof. Yan to reinforce the salience of her professional identity (藻灵, 2014), rather than foregrounding whose physical appearance (Cho, 2017; Du, 2021) or domestic roles (Auspurg & Düval, 2024, 2024; de Lemus *et al.*, 2014) that align with traditional gender stereotypes towards women. It reveals how occupation (scientist) supersedes gender (female) in shaping holistic evaluations (Settles *et al.*, 2009), so as the salience of the identity centrality.

To conclude this subsection, we outline the different dimensional social categorical information of an interpreter in China. Using the well-reputed public figure Zhang Jing (张京) as an illustrative case, according to publicly available online sources[2] and media reports (张东亮, 2021), Zhang Jing serves as a consecutive interpreter at the Department of Translation and Interpretation under the Ministry of Foreign Affairs of the People's Republic of China. Following the analytical framework outlined in the preceding subsection, we outline the social categorical information related to Zhang Jing in Table 2.4:

Table 2.4 Briefing the different dimensions of social categorical information, take the interpreter Zhang Jing (张京) as an example

Dimension	Category	Content
Natural cues	Race	Mongoloid
	Gender	Female
	Ethnicity	Majority (Han ethnicity)
	Age	Adult (born in 1985)
	Height	Average
Social cues	Place of origin	Hangzhou, Zhejiang Province
	Accent	Putonghua (standardized Mandarin)
	Nationality	People's Republic of China

[2]. An online report by Zhejiang Online (浙江在线):《两会美女翻译走红，初中起立志做外交官》(The beautiful "Two Sessions" interpreter goes viral, has aspired to become a diplomat since middle school), http://finance.people.com.cn/n/2013/0314/c1004-20787883.html

Native language	Chinese
Marriage	Married
Education	Master's degree
Professional title	Conference interpreter
Affiliation	The Ministry of Foreign Affairs of the People's Republic of China
Popularity	Well-known

2.3.2. The automaticity of social categorization and stereotype

Allport (1954) noted early in the 1950s that "categorical thinking is a natural and inevitable tendency of the human mind" (p. 171). When encountering targets with multiple social dimensions (e.g., gender, race, age, occupation, social status), the human brain spontaneously organizes these stimuli into social categories through rapid, effortless cognitive operations. This process of categorical information processing can be highly automatic, particularly for low-level or reflexive cues. Substantial evidence from cognitive neuroscience supports the notion of automatically activated social categorization. For instance, event-related potentials (ERPs) have shown that race can be detected very early in processing (approximately 100ms after stimulus onset), as reported by Ito and Urland (2003). Their findings also indicated that attention to gender emerges about 50ms later than race but still occurs at an early stage. Additional evidence points to gender detection arising only slightly later than age detection, yet both typically occur within 100 ms (Dobs *et al.*, 2019). In terms of age detection, while young and older adults may differ in the response time needed, both generally complete the process prior to 170ms (Fernandes *et al.*, 2023; Wiese *et al.*, 2008). This aligns with the N170 ERP waveform, which is associated with the early detection of basic structural encoding of facial recognition, including cues such as race, age, and gender (Kuefner, 2010; Rossion & Jacques, 2012). Likewise, although certain social cues emerge later than age and gender in the categorization process, they may still be processed automatically. For instance, identity recognition (e.g., identifying a well-known celebrity) follows age and gender detection yet typically occurs within 100 ms (Dobs *et al.*, 2019). Accent recognition, on the other hand, often arises around 400 ms (corresponding to the N400 waveform) and is generally viewed as exceeding the automated processing threshold, thus falling under controlled processing (Goslin *et al.*, 2012). However, research using mismatch negativity (MMN) paradigms has demonstrated that early sensitivity approximately to abnormal speech features set off at approximately 200ms. For example, incongruent cues (e.g., a German accent paired with a Turkish face) elicit a more negative N2 component from 200ms to 280ms (Hansen *et al.*, 2016). In this case, it suggests that accents can also be categorized automatically. While advanced

social categorization involving narrative or affective dimensions necessitates deliberate cognitive engagement. Magnetoencephalography analyses reveal that complex social judgments (e.g., distinguishing cinematic character archetypes like heroes versus villains) activate a distributed neural network comprising the dorsal medial prefrontal cortex (dmPFC), precuneus, temporoparietal junction (TPJ), and left occipital face area (OFA), as corroborated by fMRI evidence (Ron *et al.*, 2022). This higher-order processing requires the deliberate cognitive engagement of neural resources compared to automatic category detection.

The automaticity of social categorization is a product of the human brain's evolution, which provides advantages for individuals engaged in cognitively demanding tasks (Jeon & Friederici, 2015). First, automation conserves cognitive resources. The brain's processing capacity is limited, yet the influx of new information is vast on a daily basis. If the brain were to devote excessive time to gathering and analysing all possible information, communication and decision-making would become lengthy and inefficient. In response to the overwhelming amount of information, the brain must adopt a cognitive miser approach, expending minimal cognitive resources to achieve maximal cognitive efficiency (Corcoran & Mussweiler, 2010; Fiske & Taylor, 2020). Social categorization provides a framework for preliminary assumptions which allows individuals to quickly assign others to a specific group through cognitive shortcuts (Logan, 1988). This facilitates the formation of basic expectations about others' behaviour and interaction patterns, thereby reducing uncertainty and complexity in social contexts (FeldmanHall & Shenhav, 2019). Meanwhile, the simple yet effective categorization enables individuals to allocate their preserved cognitive resources to situations that demand more conscious processing, such as discerning others' true intentions and planning future actions (Moors & Houwer, 2006). Thus, in a fast-paced and information-laden social environment, automatic social categorization allows individuals to navigate interactions with others more smoothly. Second, automaticity also enhances group identity. Upon rapidly categorizing others or themselves into a particular group, individuals subconsciously create distinctions between the ingroup and outgroups (Gong *et al.*, 2024; Jenkins, 2000; Otten & Wentura, 1999; Spears *et al.*, 2004). This division not only aids in the quick identification of those who are more trustworthy or likely to cooperate with them but also fosters the formation of shared goals, values, and emotional connections among group members. When individuals perceive themselves as part of a group, they are more likely to demonstrate greater reliance on and support for their peers (Baysu *et al.*, 2011), which reflects the psychological mechanism of in-group bias (Berzonsky, 2008; Schaller, 1991). In-group bias facilitates more effective collaboration and mutual assistance among group members, thereby strengthening the group's

stability (Scheepers *et al.*, 2006; 张莹瑞和佐斌, 2006).

Such automaticity in social categorization simultaneously triggers the activation of stereotypes. As previously discussed, a stereotype is a set of highly generalized and simplified beliefs or expectations about a particular social group, accompanied by assumptions of broad applicability (Kanahara, 2006). For instance, American students may hold the stereotype that their Chinese classmates are surely good at mathematics (Cvencek *et al.*, 2015). Stereotypes typically lie dormant in long-term memory but are rapidly activated for use when individuals engage in ongoing cognitive tasks (Mason *et al.*, 2006; Spencer *et al.*, 1998; Stangor, 2009). This process of invoking stereotypes is defined as "stereotype activation", which refers to the moment when a perceiver comes into contact with a member of the stereotyped group and relevant cues render the group's trait constructs and behavioural characteristics accessible in the perceiver's mind (Kunda & Spencer, 2003). Since human relies on activated stereotypes nearly continuously in their perception of others, stereotype activation is considered one of the most fundamental and central processes in social perception (李欣璐等, 2023). It is also seen as an inevitable outcome of social categorization (Macrae & Bodenhausen, 2001; Macrae C. Neil *et al.*, 1994; 杨亚平等, 2015).

Stereotype activation is generally regarded as an automatic process that typically escapes conscious regulation. For example, the widely accepted dual-process model by Brewer (1988) posits that individuals engage in social cognition and impression formation via two distinct pathways: one that is rapid and category-based, and another that is more deliberate and individuated. Initially, when encountering a new person, perceivers rely on easily recognizable natural and social cues to generate snap judgments. These automatic categorizations simplify complex social information, helping individuals orient themselves in social contexts with minimal cognitive effort. However, when the context or perceiver's goals demand more precision, this process shifts toward individuation. In this secondary, more deliberate phase, the perceiver collects and assesses detailed, person-specific data, thereby refining or even overriding the initial category-based impression. Another widely recognized theoretical framework, the Continuum Model proposed by Fiske and Neuberg (1990), also posits that an individual's impression formation is not a binary process but rather unfolds along a continuum. Impression formation gains through rapid, automatic categorization to more deliberate, individuated processing. At the initial stage, perceivers quickly assign others to social categories based on salient features and stereotypes, which serves as an efficient heuristic to generate preliminary impressions. However, when the situation demands greater accuracy or when the perceiver is both motivated and equipped with sufficient cognitive resources, this

initial categorization can be superseded by a more controlled process that involves the systematic integration of individuating information.

This unconscious and automatic processing, while quick and efficient, also gives rise to negative social biases such as prejudice and discrimination (Lee *et al.*, 2015; Nelson, 2015; Stangor, 2009). First, stereotypes contribute to cognitive biases. First, stereotypes result in cognitive biases. Once activated, stereotypes trigger a process of information filtering. Individuals tend to focus only on cues that align with the stereotype while remaining insensitive to contradictory or counterexample information. This filtering mechanism leads to a biased and distorted collection of information, further reinforcing the original stereotype. Correspondingly, this cognitive bias also results in memory biases, where individuals are more likely to remember information that aligns with established stereotypes when recalling past events or evaluating others, while disregarding or forgetting contradictory or challenging information. For instance, under the pervasive stereotype that Black Americans are lazy (Reyna, 2000; Sigelman & Tuch, 1997), White individuals who hold such a stereotype may show little interest in positive news about hard-working Black Americans, instead focusing on and magnifying negative stories about them (Jelsma *et al.*, 2022). As a result of cognitive biases, activated stereotypes lead to biases in behavioural decision-making, resulting in discriminatory decisions. Because stereotypes foster general prejudice or discrimination towards certain groups, individuals may unconsciously draw on these stereotypes to assess the degree to which candidates possess the specific abilities required for a task (Lee *et al.*, 2015). Influenced by social norms and group affiliations, decision-makers tend to overlook individual differences but over-rely on group characteristics when making decisions. For example, in China, the stereotype that boys are better at mathematics than girls is a typical gender stereotype (Li *et al.*, 2022). Parents who hold this stereotype tend to allocate more family educational resources to their daughters and are more likely to enrol them in math tutoring programs. However, the increased likelihood of the parents' choice for their daughters to attend math tutoring classes is driven primarily aimed at remedial teaching, rather than enrolling on self-achieving Olympiad math competitions (刘劼成和赵西亮, 2024). Second, due to the bias of in-group versus out-group relationships, individuals tend to show more positive attitudes and preferences toward their own group (in-group) while exhibiting more exclusion or hostility toward members of other groups (out-group). These discriminatory decisions lead to inequities in resource distribution and career advancement. While in China, this in-group favouritism is reflected in workplace networks. Researchers have reported a strong emphasis on regional connections within the workplace, with an over-allocation of resources to individuals from the same hometown (Li, 2018; 滕越

洋, 2019).

In the following three subsections, we will provide a brief overview of the representation of gender stereotypes, stereotypes of ageing, and social class stereotypes from the perspective of social psychology research. As elaborated in the previous sections, the differentiation in how users perceive the performance of interpreters through these three categories is clearly manifested. Age, gender and professional title are also the key foci of the present dissertation. Regarding accent, we do not conduct extensive research on it, as previous studies by researchers like Cheung (2013, 2020), Zhang (2009), Lim & Fan (2023) have already provided substantial experimental outcomes on this topic. In terms of an interpreter's perceived race, although prior research has highlighted users' stereotyped perceptions of interpreting quality based on an interpreter's racial characteristics, we will not delve into the effect of racial stereotypes on interpreting quality. Although the Chinese population consists the Han ethnic group and other fifty-five minority ethnic groups, the sole Caucasian ethnic group in China, the Tajik ethnic group consisting of approximately 50,000 individuals, representing less than 0.001% of the population (国务院第七次全国人口普查领导小组办公室, 2022). Consequently, the overwhelming majority of Chinese residents identify as part of the broader Mongoloid racial group. Besides permanent local residents, there are approximately 710,000 foreign residents in China which accounts for only 0.05% of the country's population (方彬楠和冉黎黎, 2024). As a result, it is reasonable to assume that most interpreting users in mainland China share the same racial background, rendering racial stereotypes less relevant in this specific context. Most interpreters working in mainland China are also native Mandarin speakers of local origin, as reflected in the survey conducted by the Translators Association of China (中国翻译协会, 2019) and researchers Setton and Guo (2009) on professional interpreters in Shanghai. Therefore, within the context of the demographic and sociocultural landscape in mainland China, an interpreter's race is of minimal relevance as our considerations of users' perceptions of interpreting quality.

2.3.3. Gender stereotypes

Before we proceed to the key features of the stereotypes of a certain gender or sex, it is necessary to define the concepts of "sex" and "gender" in the present dissertation. These terms have been utilized in interpreting research, with some scholars such as Christen y Gracia (2023) using "sex", and some using "gender" in their studies (Acosta Vicente, 2024; Bartłomiejczyk, 2019; Cho, 2017; Giustini, 2021), and some used both expressions (Defrancq *et al.*, 2021; Gentile, 2018; Maryns, 2013). Sex refers to a person's biological sex, that is the anatomy of his

or her reproductive system and secondary sex characteristics (Stevenson, 2010, p. 1618). Sex is mostly considered to be a male-or-female binary, but the issue of gender is a bit more complex as it can refer to two different concepts: either a person's gender refers to the social role expected by society, based on the sex of the person, or to their personal identification of one's own gender based on an internal awareness (Kaiser, 2012; Torgrimson & Minson, 2005). While the female sex is traditionally assigned to the gender female, a person's sex and gender do not always align, and the person may be transgender, i.e. have a gender identity or gender expression that differs from the sex that was assigned to them at birth. Some people do not identify as one of the two traditional genders, but consider themselves nonbinary, bigender or genderfluid among others. An example can be found in the chief editor of the *Journal of Language and Social Psychology*, Prof. Nicholas A. Palomares, who self-identified as a genderfluid or nonbinary (preferred pronouns: they/she/he)[3]. In the current dissertation, we still prefer to use "gender" instead of "sex", not only because of the conventional use of gender stereotypes instead of sex stereotypes in the field of social psychology (Biernat, 1991; Eagly & Steffen, 1984; Ellemers, 2018), but also due to the scarcity of literature on such novel non-binary constructs of genders in the field of translation and interpreting studies. Thus, it makes the exploration of gender effects on interpreters much simpler and more straightforward.

Gender stereotypes refers to socially constructed beliefs that assign specific roles and attributes deemed appropriate for males and females (Deaux & Lafrance, 1998; 刘晅和佐斌, 2006). The psychological mechanisms and historical origins of gender stereotypes are widely studied topics in stereotype research. However, they offer limited theoretical relevance to this dissertation. We instead focus on the two primary dimensions, that is the descriptive and prescriptive dimensions of gender stereotypes (López-Sáez & Lisbona, 2009). The descriptive dimension dictates assumptions about how men and women typically behave, for example, it assumes men are perceived as rational (Renner & Masch, 2019), while women are seen as emotional (Barrett *et al.*, 1998; Renner & Masch, 2019). The prescriptive dimension, on the other hand, establishes norms regarding how they should behave, such as the expectation that men should appear rational and women should appear emotional. For the prescriptive features of gender stereotypes, Deaux & Lewis (1983) proposed a comprehensive multicomponent view of the structure of gender stereotypes which incorporates information on gender labels, gender

[3]. From whose open biography at https://web.archive.org/web/20240618014250/https://commstudies.utexas.edu/faculty/nicholas-palomares

role behaviours, and one aspect of physical appearance. Since its original document is not available in electronic format, this study references the translation by Zuo (2015, 第 95 页) of the work, see Table 2.5.

Table 2.5 Gender stereotypes and gendered expectations by Deaux & Lewis (1983), translated from Zuo (2015).

Categories	**Male**	**Female**
Gender label	Independent and proactive, competitive, decisive, persistent, confident, resilient, with a sense of superiority	Emotionally rich, wholeheartedly caring for others, gentle, kind, empathetic, understanding, friendly, helpful
Gender role behaviours	Head of the household, the primary source of family income, leader, responsible for home repairs	Source of emotional support, managing household chores, taking care of children, doing laundry and cooking, responsible for home decoration
Occupations	Truck driver, insurance agent, telephone installer, chemist, urban planner	Career counsellor, telephone operator, teacher, nurse
Physical appearance	Tall, strong, broad-shouldered	Soft-spoken, delicate, graceful, tender

The multi-component view of the gender stereotype structure by Deaux & Lewis (1983) decomposed gender stereotypes into four dimensions: gender labels, gender role behaviours, occupations, and physical appearance. We explain this framework together with the summary of gender stereotypes concluded by Ellemers (2018). Among these, gender labels represent stereotypical perceptions of personality traits. Men are generally characterized as agentic, meaning their traits are oriented toward goal achievement and task functioning. Thus, typical male labels are associated with attributes such as competence, assertiveness, and decisiveness. In contrast, women are perceived as communal, which emphasises the maintenance of relationships and social functioning. As a result, female labels are linked to traits such as benevolence, trustworthiness, and morality (Abele & Wojciszke, 2014; Sczesny *et al.*, 2018). Males and females are also assigned with given gender roles. Males are anticipated with more work-oriented priorities, such as the primary source of household income. Females are expected

to be responsible for family-oriented assignments, such as household management and child-raising. The occupational division of the male and the female reflects competence and warmth respectively. Warmth and competence are the two universal dimensions proposed by Fiske et al. (2002), as warmth refers to the degree of perceived friendliness, trustworthiness, morality, and sociability, which signal whether a group is likely to act in ways that benefit or harm others. While competence reflects the perceived intelligence, skill, efficacy, and ambition, which indicate the group's capacity to achieve goals or exert influence. Thus, females are expected to be more capable in interpersonal connections, while not encouraged for professional achievements. Males on the contrary.

Given that four decades have passed since Deaux & Lewis (1983) proposed the typical stereotypical occupations for both genders, some of the occupations such as telephone installer and telephone operator largely faded from contemporary career landscapes (郑世林等, 2014). Consequently, the occupations associated with traditional gender stereotypes have also undergone substantial changes. According to the contemporary literature on such stereotypical occupations for both genders, pro-male occupations primarily concentrated in fields such as science, engineering, applied technologies and mathematics (Charlesworth & Banaji, 2019). Pro-female occupations, on the contrary, are more commonly associated with auxiliary administration (Riach & Rich, 2006), caring (Wilbourn & Kee, 2010) and elementary education (Carrington, 2002; Sebastian *et al.*, 2022). In other words, such STEM-related occupations such as scientists (Carli *et al.*, 2016; Makarova *et al.*, 2019), IT technicians (Ensmenger, 2015), engineers (Angouri, 2011) and mathematicians (Makarova *et al.*, 2019) are seen as masculine occupations, while receptionists (Silva & Couto, 2023), elementary level teachers (but not for higher education) (Carrington, 2002; Sebastian *et al.*, 2022) and nurses (Wilbourn & Kee, 2010) are widely treated as feminine occupations.

There is no direct literature stating that the language service industry is stereotyped as a feminine field or neutral. The only scattered evidence we may acquire from Hong & Choi (2020) and Du (2021) is that clients and media reporters treat an interpreter as a female by default. Based on the above-mentioned gender stereotype structure (Deaux & Lewis, 1983; Ellemers, 2018) and empirical shreds of evidence, we may assume that our translation and interpreting profession is a pro-female discipline. First, an interpreter serves as a supportive communal (House, 2013) but not an agentic role, which indicates that patience, meticulousness, and emotional labour are required from practitioners. The interpreting profession is comparable to those traditional communication-oriented or assistant-type roles such as teachers, nurses and social workers, which has been reflected in the survey-based studies by Gentile (2013, 2016,

2018). These traits also align with the stereotypically feminine qualities mentioned above. According to an English Interpreter Demand Survey (吴泽萍和青立花, 2015) in the Pearl River Delta, which is one of the most developed metropolitan areas in China, the ratio of male to female in-house interpreters employed by state-owned enterprises, public institutions, private companies, and Chinese-foreign joint ventures was 11:27. Although there is a quantitative disadvantage for male practitioners, this ratio may reflect the enterprise employers' favouritism towards females instead of males, as Wu and Qing (2015) also noted that "female interpreters are perceived as more detail-oriented, adaptable, and patient, with better memory and communication skills, thereby making them more suitable for interpreting than their male counterparts" (p.67). A second contributing factor is that foreign language learning itself is a gender-stereotyped domain. In university-level foreign language and translation programs, female students significantly outnumber male students, creating a gender imbalance at the entry point to the profession. For example, Cheng, Huang, and Zhang (2014) reported a 1:10 male-to-female ratio at the Central China Normal University in Wuhan; in Zhejiang, Ying and Jiang (2002) found that from 1996 to 2001, the ratio at Zhejiang Normal University expanded from 1:5.3 to 1:9.6; and among eight Schools of Foreign Languages in Shaanxi, this figure is 1:5.8 (姚艳梅, 2011). In terms of academic performance, female students demonstrate a consistent advantage over their male counterparts from middle school through university. A nationally representative survey of Chinese middle school students by Xu (2020) found that girls outperform boys in English by 61.5% of a standard deviation, which is one of the largest recorded gender gaps worldwide. Luo (2021) corroborated that female middle school students outperform boys in English by around 50%. At the high school level, Zhang and Tsang (2014) investigated the gender achievement gap in the National College Entrance Exam within a typical Chinese municipality and found that female students performed significantly better in English. Meanwhile, at the university level, Cui (2015) conducted a random sampling of foreign language majors' examination results, revealing that male students not only had lower pass and excellence rates across all language courses but also displayed less enthusiasm for learning. These findings have reinforced the stereotype that "girls are better suited for foreign language study, whereas boys are not" (Cheng Y *et al.*, 2022, p. 608; 马莉萍等, 2016). As a consequence, parents may actively discourage boys from choosing foreign language majors (Cheng Y *et al.*, 2022), and boys under stereotype threat may exhibit reduced interest in foreign language learning (Chaffee *et al.*, 2020) and be hampered in English learning (Li & McLellan, 2021). Finally, media coverage often amplifies evidence supporting this bias while overlooking or downplaying contradictory cases (Cho, 2017; Du, 2021). For instance, although there are males

serving as press conference interpreters, very few images (3 out of 234) featuring them appear in media reports (p. 7), thereby restricting the visibility of male interpreters. It is important to note, however, that this lack of representation should not be conflated with professional incompetence, as studies have found no statistically significant differences in the capabilities of male and female interpreters (Christen y Gracia, 2023).

From this gender-stereotype perspective, people generally expect men and women to exhibit appearance, behaviour, and thought processes consistent with the norms associated with their respective gender groups (秦启文和余华, 2001). When individuals deviate from these expectations, they are considered examples of gender counter-stereotypes (Wen *et al.*, 2020). Gender counter-stereotypes refer to instances where men's or women's behaviours contradict commonly held gender-role expectations. With regard to career choices, since men and women are often presumed to follow distinct occupational paths, individuals pursuing careers not typically associated with their gender are labelled as gender counter-stereotypes (Wilbourn & Kee, 2010; 刘晅和佐斌, 2006). Male nurses, for instance, are viewed as men in women's professions (Kadushin, 1976), whereas female mathematicians are perceived as entering masculine disciplines (Smeding, 2012).

Because gender stereotypes exert a pervasive and powerful influence on individuals' thoughts and behaviours, the conflict between one's existing stereotypes and new, contradictory information triggers cognitive dissonance (Tsang, 2017). To maintain cognitive balance, individuals tend to preserve cognitive consistency and avoid uncertainty. Consequently, when confronted with counter-stereotypical information, they naturally adopt a defensive stance, reinforcing existing stereotypes and resisting disconfirming evidence (Plaks *et al.*, 2001). This motivation to uphold one's stereotypes is often referred to as the backlash effect (Rudman & Fairchild, 2004; Rudman & Phelan, 2008; Wen *et al.*, 2020), wherein individuals not only decline to revise their preconceived biases but may even intensify them. Thus, individuals who deviate from socially prescribed gender norms either by displaying counter-stereotypical traits or by conducting counter-stereotypical behaviours may face punitive responses from others (Wen *et al.*, 2020). For example, managers' discrimination of a more agentic and less communal female applicant (Rudman & Glick, 2001), pressure from peer males for a modest male to behave confident and ambitious (Moss-Racusin *et al.*, 2010) and expressing negative attitudes about stay-at-home fathers (Brescoll & Uhlmann, 2005).

2.3.4. Stereotypes of aging

Through the general age stratification, individuals aged 65 or above are classified as older adults

(Denton & Spencer, 1999). The secondary data analysis of the Sample Survey on the Aged Population in Urban/Rural China found the threshold for men entering old age at 63.70 years, and 59.95 years for women (Liang, 2014). Age consists of one of the major three natural categories in the research field of stereotypes, together with gender and race. However, the key characteristic of age is that age is a continuously acquired, unavoidable natural attribute. This distinguishes it from other social categories that are relatively fixed and resistant to change (Lamont *et al.*, 2015), such as race, gender, and height. For instance, an average-height female Chinese citizen cannot naturally transform into a tall Black male over her lifetime, yet she will inevitably progress through infancy, childhood, adolescence, adulthood, middle age, and eventually old age. Consequently, age represents the only social category that "everyone may eventually join" (North & Fiske, 2012, p. 982). Unlike other classifications, age cannot be reversed, and individuals cannot skip or revert to earlier stages, which means all individuals acquire each age identity in turn until reaching old age (温芳芳和佐斌, 2021，第 31 页).

Age stereotypes refer to the attitudes and expectations that society holds toward the aged as a distinct social group (Hummert, 1999; Levy *et al.*, 2000). Because all individuals may eventually reach aged, it means that the age stereotype is not characterized as a single but rather a "complex, multi-faceted, and domain-specific" construct (Rothermund & De Paula Couto, 2024, p. 1). Early research on age stereotypes was grounded in a sociocultural perspective, when related studies emphasized the identification and analysis of their content and structural characteristics (Hummert, 1999). Questionnaires and adjective checklists were designed evaluate prevailing beliefs about older adults. For example, one of the most representative studies by Schmidt & Boland (1986) invited one group of young adult participants to generate a list of traits they associated with older adults. A second group was then asked to categorize these traits into distinct clusters, defining different subtypes of older individuals. Three primary trait clusters associated with older adults were identified: a general category comprising eight traits that signified chronological ageing, such as having grey hair, being retired, and having wrinkled skin; a negative category containing 59 traits such as sick, senile, sad, complaining, and poor; and one positive category with 32 traits that depicted older adults as wise, capable, healthy, wealthy, and patriotic. One of the most up-to-date illustrative examples of the content dimensions of age-stereotypical attributes and characteristics was summarised by Rothermund & De Paula Couto (2024), see Table 2.6:

Table 2.6 The illustrative examples of age-stereotypical attributes and characteristics summarised by Rothermund & De Paula Couto (2024)

Life domain	Positive	Negative
Family, social relations	Integrated, connected	Lonely, conflicted
Leisure	Active, ample opportunities to enjoy leisure activities	Inactive, apathetic, listless
Personality	Wise, tolerant, calm	Stubborn, rigid, anxious, intolerant
Work	Experienced, reliable	Unable/unwilling to learn, inflexible
Finances	Generous, supportive, affluent	Poor, miserly
Autonomy	Autonomous, free from obligations	Dependent, needing assistance
Spirituality, religion	Faithful, religious	Cynical, disillusioned
Appearance	Elegant, kempt, charismatic	Unattractive, ugly, wrinkled, bald, humped
Health, physical fitness	Spry, hale	Frail, sick/ill, feeble
Mental fitness	Sharp, clear, up to date	Forgetful, demented, senile

As shown in the table, stereotypes about the aged contain both positive and negative aspects. In terms of work and cognitive functions, they are often described as wise, tolerant, calm, experienced, and reliable, as well as mentally sharp, clear, and up-to-date on the positive side. Such traits are frequently attributed to the extensive social and professional experiences accumulated over their lifetimes. Conversely, negative stereotypes portray older adults as stubborn, rigid, anxious, and intolerant in personality; unable or unwilling to learn and inflexible at work; and forgetful, demented, or senile in mental functioning. These negative traits could be attributed to both biological and psychological factors. Ageing is associated with the decline of the hippocampus and prefrontal cortex, which leads to memory decline and cognitive deterioration (Morrison & Baxter, 2012; Snytte *et al.*, 2024). From a cognitive standpoint, older adults may rely heavily on familiar problem-solving strategies, leading to what is known as cognitive rigidity (Christoph & Li, 1985; Kirby *et al.*, 1979). This habitual reliance makes it more difficult to incorporate new ideas and adapt their cognitive schemas, resulting in more rigid patterns of thought and behaviour than young individuals.

Although societal stereotypes about older adults encompass both positive and negative

attributes, negative elements tend to be more prominent in real-world interactions and social evaluations. For instance, Zuo et al. (2007) employed the implicit association test alongside explicit measures to examine university students' implicit and explicit age-related stereotypes. While participants did not express explicit negative attitudes toward older adults, their implicit responses revealed a stronger association between old age and negative descriptors. Such negative attitudes may result in discrimination based on a person's age, which is commonly referred to as ageism. It often manifests in interactions with the aged group, such as the lower motivation from social work students to work with aged populations compared to other age groups (Weiss, 2005) and employment discrimination that elder people may often face in hiring and workplace settings (Arvey, 1979; Liu *et al.*, 2020).

Studies have also found more negative evaluations towards older voices than their younger counterparts. In an overview of a set of studies of the sociopsychological implications of elder people's language and communication, older speakers are often evaluated more negatively than younger speakers, frequently being described as "doddery", "vague", and "rambling" (Williams & Giles, 1991, p. 112). Their voices are either commented on less or perceived as more distressed and weak. Young listeners, in particular, exhibit lower attentiveness to older voices, leading to significantly poorer recall rates for messages delivered by older speakers compared to those conveyed by younger individuals. Ryan and Capadano (1978) also found elder female speakers being perceived as more reserved, passive, disengaged, and inflexible than their younger counterparts, while male speakers with older-sounding voices were viewed as less flexible. The other study by Goy et al. (2016) invited 40 young participants to listen to speech samples from 30 older speakers (mean age=71.3) and 30 younger speakers (mean age=19.0). Participants in this study perceived older voices as rougher, coarser and less pleasant compared to younger voices, even when the elder speakers were in good health. Despite only minimal acoustic differences between the two age groups, listeners consistently judged both the acoustic and social qualities of older speakers more negatively than those of younger speakers. Moreover, older speakers were also attributed with negative social qualities, being perceived as less powerful and less engaged in communication.

2.3.5. Stereotype of professional titles

A title is a descriptor placed before a person's name, signifying aspects such as marital status, academic qualifications, high social standing, or official rank (Stevenson, 2010, p. 1851). Titles can generally be categorized into four types: marital status titles, honorary titles, professional or professional titles, and academic titles. In European and English-speaking countries,

honorary titles are traditionally based on lineage or accumulated wealth, as exemplified by titles such as Dame and the King. While in socialist countries such as China and the former Union Republics of the Soviet Union, hereditary titles have been abolished and replaced by honorary distinctions based on labour contributions. Examples of such titles include the Hero of Socialist Labour (Kamushadze, 2016) and the Hero Taikonaut (新华社, 2008).

Among these titles, professional titles refer to formal honorifics that denote an individual's occupational position, such as chairman, director-general and president. While a professional title typically signifies individuals with managerial expertise, teamwork capabilities, and institutional reform competencies, it also overlaps with academic titles which emphasize knowledge and research (Jaffe, 2017). This overlap is understandable, as professor and academician are both professional roles within the research field. Dr can also be considered a professional title, as it serves as an identifier of expertise and represents the advanced qualifications attained through rigorous academic and professional training (Jones, 2018).

Since professional titles mostly indicate high-ranking positions rather than lower-ranking ones (Stevenson, 2010, p. 1851), they play a prominent role in signalling the high-status identity of their holders (Jones, 2018). According to Bourdieu, professional titles (including educational titles) function as a form of symbolic capital which can be converted into economic capital (higher salaries), social capital (enhanced professional status), and cultural capital (greater professional recognition) (Borthwick *et al.*, 2015; Bourdieu & Wacquant, 2013). Scarcity intensifies value (Sehnert *et al.*, 2014), titles held by fewer individuals tend to carry greater prestige, whereas an oversaturation of titleholders can lead to inflation, diminishing its perceived worth (Martinez *et al.*, 2008; 宋强和林新奇, 2013). Thus professional titleholders actively work to maintain the boundary of their credentials (Abbott, 1988; Borthwick *et al.*, 2015), which is the exclusivity of their credentials. Such strategies employed to maintain the value of professional titles are not only reflected in the restriction of the number of titleholders within a group but also in the exclusion of marginalized professional practitioners. For instance, the Chinese Academy of Sciences imposes strict entry requirements and a fixed quota on the number of academicians to regulate access to this prestigious title (中国科学院, 1994; 李志民, 2015). While in the United Kingdom, surgeons have restricted podiatrists from using the title podiatric surgeon, allowing only the designation of podiatric surgical practitioner to preserve the symbolic capital associated with mainstream surgical professionals (Borthwick *et al.*, 2015). When a professional group experiences a saturation in titleholders, those holding higher-status titles may deliberately downplay the recognition of individuals with lower-ranking titles. For example, in a promotional video for a Leaders Fellow Program at a U.S.

university, administrative leaders such as the president, provost, and vice presidents were explicitly referred to by their formal titles. However, academic personnel below the rank of professor, such as associate professors and lecturers were merely addressed as colleagues or fellows (Jaffe, 2017). This deliberate omission may reinforce the awareness of differentiated elitists from the ordinary.

Professional titles serve as markers of an individual's position within the social hierarchy, directly linked to their social class, socioeconomic status, and overall societal standing (Weeks & Leavitt, 2017). The social class itself is a stratification system that orders individuals according to their unequal access to material, social, and cultural resources (Durante *et al.*, 2017; Moya & Fiske, 2017). An individual's placement in, and perceptions of different social stratifications exert significant psychological and practical effects. Through an examination of 127 professional titles in the United States, Weeks and Leavitt (2017) found that variables such as income, educational attainment, and social prestige are significantly correlated with social class and status. Occupations characterized by higher income, greater educational requirements, and elevated prestige are typically regarded as having higher social standing, take the dentist group for example. In contrast, individuals without professional titles or those holding occupations with lower socioeconomic status are viewed as occupying lower positions in the social hierarchy, such as gas station attendants (Weeks & Lupfer, 2004). Thus, professional titles frequently serve as a control variable for social psychological experiments to operationalize differences in social stratification (Weeks & Leavitt, 2017; Weeks & Lupfer, 2004). These titles provide easily accessible, highly generalized information that functions as a salient heuristic cue during cognitive processing. Where observers were required to conduct complex information processing, they often relied on prominent attributes such as professional titles, particularly in contexts with limited additional information. That is to say, what would otherwise require a more nuanced, multidimensional evaluation may become overly reliant on this single piece of information, potentially fostering unwarranted inferences about other, unverified qualities of the individual. Because professional titles symbolize elevated social standing, they tend to carry positive connotations, leading observers to exhibit favourable perceptual stereotypes toward those who hold such titles.

Multiple experiments have demonstrated that perceptions of socioeconomic status can influence decision-making, even when such a perceived trait has rare connections with another attribute. For instance, Goodman and Gareis (1993) designated a lawyer as a high-status profession due to its high income and educational prerequisites, and a gas station attendant as a low-status one. This study investigated how the requester's occupational status, either a

lawyer, a gas station attendant, or unspecified, influenced the likelihood of helping. Using a "wrong number" paradigm involving 240 urban residents in the United States, the results revealed that lawyers received assistance more frequently than gas station attendants. Notably, these differences did not vary according to the helper's own socioeconomic status or gender. Research also suggests that socioeconomic factors and educational attainment can affect sentencing severity for criminals. In an analysis of drunk driving cases, Nienstedt et al. (1988) employed event-history analysis and modelling analysis to examine the role of education in sentencing, and found that higher education levels significantly reduced the probability of receiving jail time or combined sentences, instead shifting outcomes toward probation. Each additional year of education lengthened the probation period by 59.56 days while shortening jail terms by 30.54 days. Freeman (2006) corroborated this finding by exploring how professional title (businessman) and educational qualification (MBA degree) as indicators of high socioeconomic status influenced sentencing in a simulated felony case. Among 273 participants, those who read about the criminal's titles were assigned more lenient punishments compared to the control group.

The psychological mechanism driving such decision-making arises when individuals holding professional titles demonstrate outstanding academic or occupational qualifications, prompting others to automatically generalize these specialized competencies to broader abilities and even personality traits. From a stereotype research perspective, this positive cognitive bias is attributable to the halo effect (Forgas & Laham, 2017). Initially proposed by Thorndike (1920), the halo effect is defined as a cognitive bias in which raters frequently allow their overall perception of a person's merit to inflate or diminish evaluations across unrelated dimensions. For example, in teacher evaluations, the merit of a teacher often shows strong correlations with unconnected attributes (e.g., 0.63 with voice quality, 0.79 with disciplinary power), whereas objective indicators (standardized intelligence tests) tend to exhibit minimal convergence with subjective ratings. These inconsistencies were ascribed as halo-driven errors, positing that raters subconsciously conflate holistic judgments with particular trait assessments, even when explicitly instructed to evaluate them separately. Subsequent theoretical work by Fisicaro and Lance (1990) subdivided the halo effect into three models: the general impression model, the salient dimension model, and the inadequate discrimination model. We contend that the positive stereotypical errors stemming from professional titles are principally rooted in the salient dimension model, which postulates that a rater's evaluation of a target's performance on a salient attribute directly biases judgments of other, less focal attributes. In contrast to the general impression model, which emphasizes the rater's holistic view of the target, the salient

dimension model highlights explicit cross-dimensional dependencies engendered by the prominence of particular qualities. Thus, when an individual excels in a key dimension such as research productivity, this standout attribute may exert influence on other conceptually-independent dimensions such as teaching effectiveness.

The halo effect stemming from professional titles can extend to the perception of attributes unrelated to professional competence. For example, as the level of an academic title increases, there is a growing tendency to perceive its holder as physically taller. In a study by Wilson (1968), 110 undergraduate students were divided into five groups, each evaluating the same unfamiliar man randomly assigned an academic rank ranging from student to Cambridge University professor. Another course administrator who was familiar with the student groups served as the control group. An ANOVA analysis indicated that higher academic status was associated with increasingly inflated height estimations: the perceived professor was rated as an average of 4.9 inches taller than the perceived student, while no significant difference emerged among the control groups. Dannenmaier and Thumin (1964) reported similar findings by administrative titles: 46 nursing students were asked to estimate the heights of four individuals with varying given professional titles, ranking from a classmate to the assistant director of the school. The results also revealed a statistically significant link between authority status and perceived height distortion. Additionally, professional titles can influence the likelihood of securing loans. Zhang and Cai (2018) found that borrowers using more prestigious titles had significantly higher loan approval rates, which effect was especially pronounced for those from smaller-sized enterprises. Interestingly, this higher approval rate indeed led to an even higher default rate from those smaller enterprises that employed more impressive titles. Professional titles can also affect voter intentions. In an experiment by Boas (2014), participants recruited via Facebook advertisements were randomly assigned to different conditions, each exposing them to a candidate's name either with or without a medical doctor title. Respondents then rated their voting intentions toward the candidate on a 1–7 scale. The findings indicated that candidate perceived as a medical doctor received significantly higher voting intention scores compared to those who did not use any title.

Boas's research also found that using a medical doctor title enhanced voters' assessments of a candidate's personal characteristics. The candidate who employed the medical doctor title received significantly higher ratings on perceived intelligence and ability, although no significant effect were found on perceptions of honesty. From a stereotypical perspective, society commonly attributes to medical doctors positive traits such as high intelligence, strong capability, and reliability (吕小康和刘颖, 2018). In fact, high-status professional groups

represented by medical doctors are often subject to societal stereotypes emphasizing well-being, health, and intelligence (Durante *et al.*, 2012). This positive appraisal of ability corresponds to what Fiske et al. (2002) term competence, which refers to the extent to which individuals or groups are seen as capable of effectively performing tasks and meeting social expectations for success. Durante et al. (2017) measured perceptions of 37 societal groups across 27 regions including the Hong Kong SAR of China, and observed a consistent stereotype whereby high-status groups are seen as more competent. These perceptions stem from the social perception that such groups cross-nationally possess the resources, education, and expertise necessary to thrive in competitive settings. A similar concept introduced by Nier et al. (2012) is agency, which is the perceived competence, assertiveness, and dominance of individuals or groups to act independently, exert control, and influence outcomes. High-status groups are often stereotypically regarded as highly agentic, setting them apart from lower-status groups.

2.4. Interpreter's social categorical information

As elaborated in subsection 2.2.2, the vast majority of interpreting users and practitioners in mainland China share the same racial background (Mongoloid). Consequently, this dissertation does not present an overview of the ethnic composition of interpreters in China. Given the scarcity of directly relevant data on Chinese interpreters, we instead draw upon existing global literature to summarize interpreters' gender ratios, age distribution, and socioeconomic status, thereby providing a theoretical foundation for Chapter 3 and the subsequent sections.

2.4.1. Gender ratio of the interpreters' group

Multiple sources indicate that women account for a substantial proportion of registered interpreters worldwide. Take the world-level AIIC as an example, it consists of 2,191 female interpreters and 740 male interpreters in the year 2010, thus approximately 75% of AIIC members are women (Albl-Mikasa, 2010; Zwischenberger, 2017). Among other survey-based studies, female interpreters consist of 66% (Albl-Mikasa, 2010), and 75.7% (Gentile, 2013) of the total population. At the national level in Europe, this ratio can be even higher; survey-based studies reveal that 86% of registered interpreters in Germany's VKD and 90.5% in Italy's Assointerpreti are female (Zwischenberger, 2017).

Although there are not yet available public open statistics, smaller-scale investigations offer insight into the gender distribution among Chinese interpreters. Han (2016b) conducted a survey of English-Chinese simultaneous interpreters in mainland China by reaching out to AIIC and the Shanghai Interpreters' Association; of the 89 respondents, 51 (approximately 57%) were

women. Similarly, Wu and Qing's (2015) study in the Pearl River Delta found that the male-to-female ratio among full-time interpreters was 11:27, while part-time interpreters showed an 8:15 ratio. Beyond China, other East Asian countries such as Japan and South Korea exhibit a similar trend of a female-dominated interpreter population, as noted in the interview-based studies of Giustini (2021) and Hong & Choi (2020).

Overall, data consistently suggest that women form roughly two-thirds of the interpreting workforce, a finding that aligns with the earlier observation that a stereotyped interpreter is presumed to be female.

2.4.2. Age levels of interpreters

Survey results from various organizations and researchers have found the youthful demographic profile of the interpreting profession. For example, the *2007 Transn China Interpreters' Living Conditions Report* indicated that interpreters aged 26 to 40 constituted 68% of survey respondents. This age distribution remained largely unchanged fifteen years later: the 2022 talent development report issued by the Translators Association of China (中国翻译协会, 2022) found that interpreters aged 40 or below accounted for 68.3% (p.10). Smaller-scale academic investigations corroborate these findings. For instance, in Han's (2016b) survey of English–Chinese interpreting practitioners in mainland China, 85.7% of participants were under 35 years old, implying that only around 14% were 35 or older. Setton and Guo's (2009) survey targeted the alumni of translation and interpreting postgraduate training programs, AIIC members in China, both in-house and freelance interpreters in Shanghai and the Taiwan Province of China, as well as official interpreters working for the Shanghai government, found a mean age of 35.5 and a median of 27.5, which corroborated the youthfulness of our profession.

Interestingly, Setton and Guo also reported that within their sample, AIIC members had an estimated mean age of 49. It is thus an older bracket compared to the pool's average. This pattern aligns with international data showing that AIIC members worldwide tend to be older. Nicodemus et al. (2021), for instance, observed that the largest age group among AIIC members is between 50 and 60 years old. While the interpreter workforce overall is relatively young, a notable segment of elder professionals is still actively working in the field. For example, the ten interviewees in the study of Liu et al. (2020) were in their 70s. In the survey by Pöchhacker and Zwischenberger (2010), the oldest participant was 87. These observations suggest a stratification trend: a considerable number of interpreters exit the profession within 15 years of entry, whereas more senior interpreters often continue practising well into their very later life.

2.4.3. Interpreters' socioeconomic status

The interpreting profession could be considered a respectable yet not especially prestigious profession (Giustini, 2022). In China, for example, a 2016 occupational prestige study (宗刚等, 2016) found translators and interpreters ranked twentieth out of thirty occupations, which is on par with middle-status occupations like network engineers and bank clerks. This assessment aligns with interpreters' own self-perceptions, as they often compare their socioeconomic status to that of compulsory-education teachers, nurses, and journalists (Gentile, 2013, 2016).

However, a noteworthy feature of the Chinese interpreting market (possibly worldwide as well) is the large proportion of part-time, or ad hoc interpreters, who typically hold other full-time positions and accept interpreting assignments on an occasional basis. In 2022, the ratio between full-time versus ad hoc interpreters is 98 : 440 (中国翻译协会, 2022, 第 7 页), which means that interpreters work only part-time quadruple than their full-time counterparts. Given the larger proportion of part-time interpreters among practitioners (中国翻译协会, 2019) in China, an ad hoc interpreter may also possess multiple occupational identities, such as a university professor (Pöchhacker, 1995), a TV compère (Chiaro, 2002) or a medical doctor (Krystallidou *et al.*, 2018). As we elaborated in subsection 2.2.5, the three above-mentioned professions confer a higher socioeconomic status than a full-time interpreter (Gentile, 2013), and such occupations come with a professional title respected by the social public.

The socioeconomic status of the above-mentioned occupations is generally perceived as higher than full-time interpreters (Chai & Li, 2024; Gentile, 2016). Due to the halo effect (Thorndike, 1920), a cognitive bias wherein the perception of a positive trait leads to the biased assumption of other positive qualities (Forgas & Laham, 2017). Individuals with perceived higher status are often positively stereotyped as being more competent, confident, and articulate (Durante & Fiske, 2017). Such positive stereotypes of speakers with higher status can also lead to a bias in language comprehension where the utterance of high-status individuals is given more credibility or is interpreted more favourably, irrespective of the actual quality or result of the communication (Bornkessel-Schlesewsky *et al.*, 2013; Kilpatrick *et al.*, 2007). In particular, community interpreters and liaison interpreters merely share a less charismatic professional prestige as social workers (Gentile, 2016).

2.5. Summary

Perceiving interpreting quality differs from assessing it. The assessment of interpreting quality

involves several criteria that jointly contribute to an interpreter's holistic performance. While expert assessment of interpreting quality complies with objective measures, such as fidelity and fluency as the two general categories, users expect an interpreter to provide quality renditions as well. However, the significant challenge in users' perceptions of interpreting quality is their inability to comprehend the content of the source speech. Users' emphasis on delivery aspects includes aspects like fluency, tone, and pace were found, which sometimes overshadow the accuracy of rendition. Moreover, the impact of an interpreter's social categorical information, such as gender, age, and socioeconomic status was also found on users' quality perceptions.

The notion of social categorization and its automatic effects on interpreting judgments is an important area of study. Often without consciousness, social categorization implicitly shapes how interpreters are perceived, particularly in terms of gender, age, race and socioeconomic status reflected through professional titles. These automatic categorizations can lead to stereotypical judgments. For instance, a non-local accent may trigger doubts about the speaker's education levels, and a darker skin colour tone may evoke associations with crimes. We found common ground between such stereotyped perceptions interpreting quality and experimental outcomes through controlled social psychological experiments. Driven by the automaticity of social categorization, stereotypes bias individuals' perception of impression formation based on characteristics like gender, age, and socioeconomic status. Traditional stereotypical beliefs, such as the notion that language-related professions are feminine, that older individuals must experience cognitive decline, and that professionals with titles are inherently competent and agentic, not only shape psychological impressions but also influence judgments. Likewise, an interpreter's perceivable social categorical information, which is unavoidable from being perceived and stereotyped, may act as triggers from such biased perceptions on whose performance.

Despite the insufficient survey outcomes for the general gender ratio, age groups and socioeconomic status of the interpreting profession in China, it may provide clues to the image of a stereotyped interpreter as perceived by the general social public. Since female interpreters, and young practitioners consist of the majority of this profession, a "typical interpreter" in China may appear as a young female, whose socioeconomic status is comparable to the middle-status professionals, but lower than those ad hoc colleagues with high-status professional roles. Based on the summary of the components of interpreting quality, the social categorical cues as an individual's salient personal attributes, and the review of the stereotypes induced by these cues in social psychology, it motivates the depiction of a typical interpreter's public stereotype in the next chapter, thereby providing a theoretical foundation for the controlled experiments in

Chapter 4 and the correction model in Chapter 5.

Chapter 3 Exploring the stereotype of an interpreter in China

In the last chapter, we reviewed the existing demographic information of the interpreters group in mainland China. It found a relative homogeneity within the interpreter population in terms of gender and age: the majority of interpreters in China are female and under 40 years old. In terms of the socioeconomic status of the interpreting profession, the latest available occupational prestige survey found full-time interpreters and translators ranking 20th among the 30 selected occupations (宗刚等, 2016). It indicates that the socioeconomic status of interpreters is modest, generally placing them within the middle stratification. These findings provide a foundational understanding of the group image of interpreting practitioners in China. However, general public stereotypes of some specific social groups often diverge significantly from their actual demographic characteristics. For instance, despite data showing that women are statistically safer drivers than men (廖熠佳, 2019), but societal stereotypes frequently depict female drivers as dangerous road threats (Li & Luo, 2020). Thus, this chapter aims to explore the public's general stereotypes of Chinese interpreters to address the first research question: "What is the stereotype of a Chinese interpreter held by the general Chinese public?" We begin by reviewing several mainstream methods used in social psychology research for collecting stereotype data, and then the approach that is mostly suited for this study. Subsequently, we designed a questionnaire and collected responses from the general public in mainland China. After filtering the data, removing noise, and conducting analysis, we present the major finding of the public's stereotypes about interpreters, together with some interesting minor findings. We conclude this chapter with a discussion of the findings.

3.1. Methods for stereotypes measurement

The content and intensity of stereotypes can be measured using a variety of empirical methods in social psychology, including but not limited to pictorial materials, free association tasks (Joffe & Elsey, 2014), typical trait assignments (Katz & Braly, 1933), assessments of shared attribute percentages (Brigham, 1971), homogeneity evaluation techniques (Linville *et al.*, 1989) etc. This section reviews three representative measurement approaches: open-ended or semi-open feature elicitation methods, dual-dimension measurement methods, and implicit–explicit measurement methods. The rationale for selecting the most suitable measurement method for this study was discussed thereafter.

3.1.1. Open-ended and semi-open feature collection methods

During the early stages of stereotype research approximately spanning from the early twentieth century to the 1980s, theoretical frameworks surrounding stereotypes were still in their infancy, leaving significant gaps in the field. One of the primary objectives during this period was to quickly identify culturally shared stereotypical traits to facilitate large-scale mapping of social stereotypes. Consequently, early stereotype researchers have introduced schema-oriented methods such as free association and typical trait assignment. The free association method in psychology originates from Freud's clinical practice, where patients were encouraged to verbalize any thoughts that spontaneously emerged in their minds. Freud posited that this technique could bypass conscious defence mechanisms and reveal repressed unconscious content, such as desires and conflicts (Jones *et al.*, 1963). Inspired by this approach, social psychologists adapted free association to investigate the subconscious cognitive structures underlying group stereotypes (Joffe & Elsey, 2014). The procedure for using free association in stereotype research is relatively straightforward: participants are presented with a target group and asked to generate the first word or phrase that comes to mind in relation to that group (Nelson *et al.*, 2000). When participants provide between three to ten adjectives, and with a sufficiently large sample, researchers can extract a high-frequency word pool that reflects the predominant stereotypes associated with that group.

Before the widespread availability of neuroimaging, eye-tracking, and reaction time-based methods, free association was one of the few techniques capable of indirectly accessing naturalistic thoughts (Joffe & Elsey, 2014). Compared to interviews and self-reports which rely entirely on introspective responses, free association has been considered a method with higher ecological validity, as it allows for the spontaneous elicitation of socially shared cognitive representations. Therefore, the free association method is particularly suitable for collecting explicit stereotypes about a given group, especially in cases where there is limited prior knowledge about that group. This approach continues to be widely used in psychological research today. For example, Lindqvist et al. (2017) employed free association to examine stereotypes held by Swedish citizens (n=102, M = 38 years, SD=12) toward those economically disadvantaged individuals receiving welfare benefits. The study aimed to determine whether stereotypes about the poor in Sweden were as negative as those found in other countries (Cozzarelli *et al.*, 2001; Durante & Fiske, 2017; Wang & Yang, 2020). A total of 578 words were collected and subsequently categorized into 20 thematic groups using Povee and Roberts' (2014) thematic analysis method. Their findings revealed that Swedish citizens associated the welfare-receiving poor with various dimensions of stereotype cues, including lazy, uneducated,

drug addict, and migrant background. Such cues also reflect citizens' underlying negative attitudes toward this group. Another example is the study of Lyu & Liu (2018) which applied the free association method to collect stereotypes about medical doctors from both the patients' and medical professionals' perspectives. Participants were asked to generate between three and six words that they associated with doctors. The study identified shared stereotypes from both patients and doctors, such as high medical ethics, serious work attitude, and a strong sense of social responsibility, which were generally positive. However (and interestingly), it also uncovered differences in perception: doctors themselves perceived their income as low, whereas patients stereotyped doctors as earning high salaries.

In addition to the fully open-ended free association method, a semi-open approach known as typical traits assignment was introduced in the mid-20th century. Typical traits assignment is similar to free association, but differs in its semi-open structure. In this method, researchers provide participants with a predefined list consisting of dozens of adjectives for participants to choose from. Participants were also allowed to add additional traits if they found the provided list insufficient. One of the very "classic" studies utilizing this approach (Fiske *et al.*, 2002, p. 879) is Katz and Braly's (1933) seminal research on racial stereotype representation and intensity. In their study, one hundred Princeton University students were asked to select the five most characteristic traits of ten different national or ethnic groups (such as Germans, African Americans, and Jewish people) from a list of 84 adjectives. If they felt the list was inadequate, they were permitted to supplement it with additional descriptors. The findings revealed highly consistent stereotypes among participants, such as "scientifically-minded" (78%) and "industrious" (65%) for Germans, "lazy" (75%) for the Blacks, and "shrewd" (79%) for Jewish individuals (Katz & Braly, 1933, p. 284-285). Gilbert (1951) later replicated Katz and Braly's (1933) study using the same list of 84 traits and the same ten ethnic groups. This longitudinal design ensured data consistency over time, which allows for direct comparisons of stereotype evolution through two decades. By analyzing the five most frequently chosen adjectives for each group and their selection frequencies, researchers were able to transform abstract social attitudes into quantifiable indicators. For example, the stereotype of the Blacks as "superstitious" was endorsed by 84% of students in 1932, but this figure dropped to 41% in 1950, reflecting a measurable weakening of the stereotype. Subsequent replications of this typical traits experiment continued at Princeton over approximately 20-year intervals (Bergsieker *et al.*, 2012; Devine & Elliot, 1995; Karlins *et al.*, 1969; Madon *et al.*, 2001).

Compared to fully open-ended free association tasks, the semi-open typical trait assignment method offers a balanced approach that integrates both data standardization and flexibility. Such

a semi-open format facilitates systematic and efficient cross-group comparisons. When participants evaluate multiple groups simultaneously, a unified adjective list ensures that all groups are assessed within the same evaluative framework and eliminates inconsistencies in wording that could arise in free association tasks. For instance, through an open-ended free association approach, one participant might describe Germans as "diligent" while referring to Japanese individuals as "devoted". Determining whether these two descriptions reflect the same underlying trait would require additional interpretative analysis. By contrast, in a trait assignment task, as participants assign "diligent" to both Germans and Japanese but associate "lazy" exclusively with the Blacks (Katz & Braly, 1933, p. 284-285), these data would provide clearer evidence of specific biases. Thus, while semi-open methods enhance cross-group comparability and analytical efficiency, they inherently limit the potential for uncovering unexpected, novel stereotype content that might emerge through fully open-ended methods.

3.1.2. Binary-structure measurement methods

While the typical trait assignment method provides a "consistently documented measure of stereotypes across groups" (Fiske *et al.*, 2002, p. 879), it received doubts from Fiske et al. (2002) and other researchers that lacked a solid theoretical basis. From the perspective of contemporary social psychological research, this method overlooks the influence of social power dynamics in shaping stereotypes. Traditional approaches merely describe what traits are assigned to a group without addressing why certain groups are associated with these traits. Moreover, as participants must select traits that are entirely positive or negative, such as lazy or intelligent, this method's reliance on unidimensional labels may fail to capture the coexistence of both positive and negative evaluations. For instance, Asian Americans, often regarded as a model minority, are perceived as highly competent but simultaneously stereotyped as lacking warmth, being socially reserved, and overly self-focused (Lin *et al.*, 2005). To address these limitations, Fiske et al. (2002) developed the Stereotype Content Model, which is a 2×2 framework based on the interaction between warmth and competence. This model is now considered one of the "most influential frameworks" for understanding group stereotypes (佐斌等, 2006, 第 138 页). According to this model, stereotype dimensions emerge from interpersonal and intergroup interactions. When individuals encounter others, either as individuals or as group members, they seek to understand their intentions and their ability to pursue those intentions effectively: whether they are positive or negative toward the self or in-group, and how is their ability to pursue those intentions effectively. These two key characteristics correspond to perceptions of warmth and competence. Thus, while stereotypes may be expressed using a wide variety of

descriptive labels, they ultimately converge onto these two fundamental dimensions. Thus, stereotypes can be categorised into four clustered patterns:

1. Admired (high warmth and high competence): groups that are both liked and respected, such as firefighters.
2. Paternalistic (high warmth but low competence): groups viewed as warm but incapable, such as the aged or individuals with disabilities.
3. Envious (high competence but low warmth): groups perceived as capable but cold or untrustworthy, such as wealthy individuals.
4. Contemptuous (low warmth and low competence): groups that are both disliked and seen as incompetent, such as the beggars.

In empirical applications of the Stereotype Content Model, researchers often employ structured questionnaires that require participants to evaluate target groups using standardized scales. Participants rate groups on competence-related traits (e.g., efficient, professional) and warmth-related traits (e.g., sincere, approachable). A representative example of its validation in in mainland China is the study by Ji et al. (2021) which examined the applicability of the model in a Chinese context. The researchers selected 41 representative target groups such as firefighters and migrant workers. 1,315 undergraduate participants completed paper-and-pencil questionnaires assessing perceptions of warmth and competence using 18 items rated on a 5-point Likert scale. Additionally, a separate sample of 268 undergraduates evaluated the perceived social status and competitive threat associated with these groups. Subsequent paired-sample t-tests and ANOVA analyses were conducted to examine both group-level and cluster-level differences. This analysis allowed for the classification of groups into the four prototypical quadrants as mentioned above. By calculating the mean scores for all groups on the warmth and competence dimensions, a two-dimensional stereotype perception map was constructed, which visually represented how different social groups are perceived.

This binary-structure measurement method based on competence and warmth has demonstrated its ability to move beyond the traditional unidimensional evaluation of stereotypes as simply positive or negative. Since previous approaches often treated stereotypes as isolated judgments, establishing meaningful relations between different categories is not a theoretically solid approach. While the Stereotype Content Model highlights that stereotypes can be mixed, and certain groups may simultaneously be attributed both positive and negative traits that appear contradictory. This framework helps explain why some social groups are

perceived with a mix of favourable and unfavourable characteristics.

3.1.3. Explicit-implicit measurement methods

Both open-ended methods such as free association and typical traits assignment, and binary-structured models like the Stereotype Content Model rely on questionnaire-based data collection to measure participants' stereotypes. According to the Dual Process Theory (Edwards, 1953), human cognition operates through two parallel systems: an explicit system which involves rational, conscious processing, and an implicit system which consists of automatic, unconscious attitudes. A major limitation of explicit measurement methods is their susceptibility to social desirability bias (Phillips & Clancy, 1972), that is the tendency for individuals to align their responses with socially accepted norms rather than revealing their authentic beliefs. Respondents may present politically correct views that do not necessarily reflect their internal attitudes. For example, heterosexual participants may outwardly express tolerance toward sexual minority groups but still hold subconscious transphobic biases against transgender individuals (Wang-Jones *et al.*, 2018). This attitude-behaviour dissociation (Swanson *et al.*, 2001) makes explicit measures inadequate for capturing the automatic and unconscious structures underlying social cognition.

To address this limitation, Greenwald & Banaji (1995) introduced the Implicit Association Test (IAT) which is designed to quantify the unconscious associations individuals hold between specific social categories (e.g., race, gender, age) and attribute concepts (e.g., positive or negative traits, occupational roles). By measuring differences in reaction times during categorization tasks, this test evades conscious control mechanisms and directly assesses the strength of automatic associations. In a standard IAT procedure, participants categorize target concepts (such as Black or White) to either the left or right side of the screen using a designated key. They then perform a similar categorization task for attribute words (such as intelligent or clumsy). The critical phase consists of two conditions: in the congruent task, participants categorize target concepts and attribute words using the same key. In the incongruent task, the pairing contradicts implicit cognitive expectations. By calculating the reaction time differences between these two conditions being measured in milliseconds by computers, researchers can infer the direction and strength of an individual's implicit bias: the shorter the response time, the stronger the psychological association between concepts.

This implicit-explicit measurement approach has been instrumental in shifting psychological research from a consciousness-dominated paradigm to a parallel-processing model, where automatic and controlled cognitive systems coexist. It provides a crucial

methodological tool for uncovering the mechanisms of implicit social cognition, offering deeper insights into biases that traditional explicit measures fail to detect (Schnabel & Asendorpf, 2013).

3.1.4. The rationale for adopting the free association

Based on the brief overview of the three stereotype measurement methods discussed above, there is a need to determine the most suitable approach for collecting public stereotypes of an interpreter in China. We selected the free association method as the approach in the current study, in which each participant is asked to provide up to five adjectives that best characterize their impression of the interpreter profession.

From the perspective of information richness, the free association method is better suited than the other two approaches for collecting data from a newly examined social group that starting from scratch. According to free association guidelines, each participant is required to generate between 3 to 10 adjectives based on their first impression of the target group. With a sufficiently large sample, this approach can yield a vast pool of descriptive terms. By manually organizing and categorizing these responses, it enables us to identify the specific traits that constitute the Chinese public's stereotypes of interpreters. For example, we anticipate that adjectives such as "linguistically skilled" may frequently emerge. Additionally, free association enables the collection of valuable qualitative data, including unexpected or novel perceptions that may not have been anticipated. In contrast, the typical traits assignment method where participants select from a predefined list of adjectives may restrict responses to the researchers' pre-established categories. Given that we, as professionals within the language service industry, may hold a different perspective on our profession compared to the general public, pre-selecting adjectives may introduce bias. This concern is supported by prior research on stereotypes of Chinese doctors, where patients reported diverged significantly stereotyped income levels from what practitioners themselves perceived (吕小康和刘颖, 2018). Such constraints may result in capturing only the stereotypes that we researchers presume to exist, rather than those held by the general public. On the other hand, the Stereotype Content Model is primarily designed for large-scale comparisons across multiple social groups, which aims to determine their relative placement along the warmth and competence dimensions. If applied to this study, this model would allow participants to rate an interpreter on warmth and competence, providing quadrant classification (which would possibly place interpreters in the high competence & high warmth category). However, beyond these quantifiable attributes, the method would yield little meaningful qualitative data about how interpreters are stereotyped. While employing an

explicit–implicit measurement approach would likely produce binary or internally contradictory results. For example, the general public may explicitly perceive interpreters as warm while implicitly associating them with coldness. Though such findings would reveal implicit biases, they would be limited in scope and depth, failing to provide a relatively comprehensive understanding of the specific content of interpreter stereotypes.

Although social desirability bias (Phillips & Clancy, 1972) suggests that socially accepted norms influence respondents to align their answers with prevailing societal expectations rather than their genuine beliefs, we state that this concern is mitigated in the present study. Since participants provide their responses through an online open questionnaire without direct face-to-face interaction or telephone communication with the experimenters, there is no opportunity for their impression formation (Brewer, 1988) on the experimenter, reducing the likelihood of response bias. In terms of affective attitude toward certain occupations that may affect decisions , interpreting is a relatively neutral profession, with interpreters expected to maintain impartiality in their work (Ren, 2020). As a result, public perceptions of this occupation are unlikely to be as heavily shaped by social values as those of more ideologically charged professions, such as military personnel or criminals. Given these factors, the use of an explicit stereotype measurement method in this study is expected not to be heavily affected by social desirability bias.

From an experimental feasibility perspective, the free association method may offer the most operational approach to data collection. While state-of-the-art measurements such as neuroimaging and reaction time-based methods can capture perceptual responses with greater precision, they require substantial time and resource investments for each participant, making them less viable for larger-scale stereotype research. For example, the number of participants in an Implicit Association Test study often ranges between 50 and 100 (Swanson *et al.*, 2001; 胡志海, 2005). Despite being a century-old traditional method in psychology, free association remains actively utilized in social psychology and is often integrated with contemporary models such as the Stereotype Content Model as part of hybrid measurement approaches. An example can be found in Ji et al. (2021), as it incorporated free association to collect typical stereotype descriptors alongside warmth and competence ratings. This approach identified characteristic traits for groups perceived as high warmth & high competence, such as “scholarly” and “knowledgeable” for university professors, and supported subsequent research on “stereotype activation, stereotype application, and impression management” (p.12). Through personal smartphones, participants can conveniently complete free association tasks via online survey forms, allowing for efficient data submission and improving response rates.

Certain challenges associated with the free association method should be carefully addressed before questionnaire distribution. First, this approach generates a substantial volume of qualitative data, which cannot be processed through conventional statistical techniques such as outlier detection or automated noise reduction. A meticulous manual review of each response is required for validation. Additionally, the inherently subjective nature of qualitative responses necessitates careful categorization. As elaborated in previous Princeton studies (Gilbert, 1951; Katz & Braly, 1933), distinctions between seemingly similar descriptors such as industrious for Germans and progressive for Japanese may not be entirely clear-cut. It requires analytical rigour to minimize interpretative ambiguity.

3.2. Exploring the public stereotype of an interpreter

We aim to measure the public stereotype of an interpreter's occupational image in China. Based on the previous literature review, we focus specifically on three key categories: gender, age, and social status.

3.2.1. Design of the questionnaire

The title of the questionnaire presented to participants is "A study on the public perception of the occupational image of interpreters in China". We deliberately omitted the term stereotype in the title and replaced it with occupational image, thus creating the impression that the study focuses on the general public perception of a professional group.

The full questionnaire is available in Appendix 1. It consists of six sections: (1) the respondents' demographic information, (2) the frequency of respondents' prior interactions with interpreters, (3) respondents' gender stereotypes of interpreters, (4) respondents' age-related stereotypes of interpreters, (5) respondents' perceptions of interpreters' social status, and (6) five open-ended descriptors eliciting spontaneous associations with interpreters.

In the demographic information section, we collected data on respondents' gender (male, female, non-binary), age, political affiliation, geographic region, occupation (including full-time student), household income (parent's income for students), professional ranking, highest level of education and field of study, acquired foreign language, and self-assessed foreign language proficiency.

The theoretical rationale for measuring respondents' frequency of contact with interpreters is based on the "cross-group contact hypothesis", which posits that increased interactions with outgroup members contribute to the reduction or modification of stereotypes (Wright, 2009). To examine this aspect, we included a multiple-choice question asking respondents about the

gender distribution of the interpreters they had contact with. Additionally, we designed another multiple-choice question to investigate the sources through which respondents became familiar with the interpreter community. This question aims to explore the origins of interpreter-related stereotypes, given that media exposure is a well-established factor in reinforcing stereotypes (Ash *et al.*, 2022; Seiter, 1986; Ward & Grower, 2020), and previous studies have identified that media consciously reinforce the feminine stereotype of interpreters (Du, 2021; Hong & Choi, 2020).

Regarding respondents' gendered perceptions of interpreters, we first inquired about the perceived gender ratio within the interpreting profession. To prevent response polarization, we adopted a five-point Likert scale, ranging from "predominantly male" and "mostly male, with some female interpreters" to "equally male and female," "mostly female, with some male interpreters," and "predominantly female." Subsequently, we included a multiple-choice question assessing whether respondents perceived the interpreting profession as masculine, feminine, or neutral. Furthermore, we examined the underlying reasons behind the gendered perception of interpreting. Specifically, we asked whether respondents attributed their perceptions to cross-group contact leading to stereotype modification, the numerical predominance of female practitioners reinforcing gender stereotypes, or whether they linked stereotypes to aspects of interpreting performance such as fidelity, fluency, tone (paralinguistic features), language aptitude, or appearance, with the latter potentially influenced by the halo effect (Fisicaro & Lance, 1990; Forgas & Laham, 2017; Thorndike, 1920).

Additionally, we guided respondents in forming a prototype (Rosch, 1973) of interpreters based on their stereotypes, which is the anchor for the group homogeneity. These questions consist of three multiple-choice questions on appearance, height, and body shape.

Stereotypes regarding interpreters' age were measured using a five-point scale, with age brackets ranging from under 25 to over 55 years old. Although the conventional definition of old age is 65 years and above (as elaborated in subsection 2.3.4), research suggests that women tend to be perceived as entering old age earlier than men (Hubbard & Rockwood, 2011; Shaw, 2006). In China, the official retirement age further reinforces this perception, as women typically retire at 55, while men retire at 60 (新华社, 2024). Given this context, our study set 55 and 60 separately as the threshold for categorizing an interpreter as "aged".

To assess respondents' perceptions of interpreters' social status, this study employs two types of questions designed to examine underlying stereotypes. As discussed in Subsection 2.3.5, professional titles serve as an indicator of socioeconomic status; however, such status is constructed through a combination of social capital, economic capital, and cultural capital

(Borthwick *et al.*, 2015; Bourdieu & Wacquant, 2013, 2013). In the research field of management science, occupational prestige is conceptualized as a composite of political status, social status, and economic status, with researchers evaluating these three dimensions before deriving a comprehensive prestige score for a given profession. The most recent study examining the social prestige of translators and interpreters is that of Zong, Li & Sun (2016), as they reported rankings of 18/30 for political status, 19/30 for economic status, and 22/30 for social status, culminating in an overall ranking of 20/30. It is worth noting that these findings indicate a decline in the perceived status of translators and interpreters compared to the results of Tian, Kuang & Liu (2013) who adopted the same methodology, ranked the profession 25/86 in political status, 19/86 in economic status, 32/86 in social status, and 28/86 in overall occupational prestige. However, the factors contributing to this downward trend fall beyond the scope of the present study and are thus not addressed further.

Given that this study specifically focuses on the status of interpreters other than multiple occupations, we categorized the 30 occupations from Zong, Li & Sun (2016) into ten groups, each consisting of three professions, from which we selected one representative occupation per category. Respondents were then asked to compare the political status, economic status, and social status of interpreters against these ten selected occupations using a five-point Likert-type scale, with response options ranging from “much lower” and “somewhat lower” to “comparable,” “somewhat higher,” and “much higher.” Thus, through statistical analysis, it allows for the identification of the profession most closely aligned with interpreters in relative prestige of the interpreting profession.

At the end of the questionnaire, we incorporated a free association task, in which respondents were asked to freely provide five descriptors that best reflect their impressions of interpreters. This approach was designed to collect qualitative data on the stereotypical content associated with interpreters.

3.2.2. Distribution and collection of questionnaires

Since the target respondents of this study were ordinary Chinese citizens, the questionnaire was distributed via Sojump (问卷星)[4], the earliest and largest online survey platform in mainland China (周晓杰, 2020). The questionnaire was finalized on May 7th, 2023, and the period open for responses starts from May 7th, 2023 to August 31st, 2023. To maximize outreach, the experimenter disseminated the survey through WeChat group chats, WeChat Moments community, and referrals from friends and family to reach broad public participation. To

[4]. https://www.wjx.cn/

enhance the response rate, an incentive mechanism was implemented: the questionnaire description page indicated that participants who completed the survey conscientiously would have a chance to receive a 10 Chinese Yuan (1.4 USD) cash reward (50 recipients in total) as a token of appreciation. Among the responses that passed the experimenter's manual inspection, 50 submissions were randomly selected for the reward. The selected participants were compensated via Alipay transfer in April 2024, using the payment details they had provided. Therefore, the sampling method employed in this study is a respondent-driven sampling method, which slightly differs from traditional snowball sampling (赵延东和 Pedersen J, 2007). Since the pool of respondents was expanded through chain referral through participants who have submitted questionnaires, and the experimenter actively interposed in the distribution of sampling and went through rigorous further data cleaning, it avoided such sampling bias and unknown sampling probability caused by snowball sampling in some degrees (赵延东和 Pedersen J, 2007; 风笑天, 1994).

3.2.3. Data cleaning

A total of 349 complete responses were collected for this survey. The raw dataset comprised 148 male respondents (42.41%) and 201 female respondents (57.59%), with no respondents self-identified as non-binary. The mean age of male respondents was 37.10 years (Median=31, *SD*=15.78), with the oldest being 89 years old and the youngest 8 years old. For female respondents, the mean age was 33.75 years (Median=29, *SD*=13.01), with the oldest being 85 years old and the youngest 18 years old. Respondents were long-term residents of various regions in mainland China, including Taizhou (Jiangsu), Shaoxing (Zhejiang), Mianyang (Sichuan), Weihai (Shandong), Aksu (Xinjiang), and Sanya (Hainan), among others. Their occupations ranged from positions inside the system (体制)[5] such as government officials, teachers, and non-institutional occupations such as baristas and freelancers. This diverse occupational and geographical distribution suggests that the survey achieved its goal of capturing a broad cross-section of the general public.

Different data cleaning strategies were employed for the multiple-choice questions and the free association descriptors, considering that respondents may have allocated differing levels of engagement to these sections. For instance, some respondents may have answered multiple-choice questions arbitrarily while showing more interest in the free association task.

For the multiple-choice responses, the duration was a key criterion for evaluating data

[5]. The system refers to institutional roles within China's state authority-managed system, which includes government agencies, public institutions, and state-owned enterprises.

reliability. The experimenter conducted two simulated completions of the questionnaire to establish a reference for reasonable response times. The first simulation, performed on a desktop computer, assumed a respondent without foreign language proficiency or prior contact with interpreters, requiring approximately 4.5 minutes (270 seconds) to complete. The second simulation, conducted on a smartphone while providing genuine responses, took approximately 8 minutes (480 seconds). The results indicate that computer-based responses tend to be more time-efficient than mobile-based responses. Among the 349 completed questionnaires, response times ranged from 126 seconds (minimum) to 7,538 seconds (maximum). The 25th percentile was 339 seconds, the median was 477 seconds, and the 75th percentile was 653 seconds. Given these benchmarks, responses with durations between 126 and 339 seconds were carefully inspected, revealing that most of them were noise. Conversely, excessively long response times could be attributed to factors such as aged respondents unfamiliar with smartphones or respondents being interrupted while completing the questionnaire. To assess the quality of prolonged responses, data between 1,163 and 7,538 seconds were also manually reviewed, with most responses demonstrating acceptable quality.

Subsequently, a focused inspection was conducted on responses in which respondents reported their long-term residence as Dongcheng District, Beijing. The rationale for this inspection is that Dongcheng District, Beijing was the first drop-down menu selection question encountered after respondents had completed the basic demographic questions on gender, age, and political affiliation. While it is possible that some respondents genuinely reside in Dongcheng District, a subset of respondents might have selected the first available option in this drop-down list to expedite survey completion, potentially with the intention of obtaining this 1.4 USD reward. After reviewing response patterns in other sections of these submissions, 10 invalid responses were identified and removed.

We then examined the remaining responses. First, responses exhibiting patterned answering (e.g., selecting options in sequential order from the first to the fifth or in a reversed sequence) were identified and excluded. Next, uniform responses were removed, where all answer choices were consistently assigned the same response category. Subsequently, logically inconsistent responses were eliminated. For instance, responses indicating that interpreters have a lower political status than housemaids but a higher political status than judges were considered illogical and removed from the dataset. Finally, data from two respondents who self-reported as translators or interpreters was excluded. Following the data-cleaning process, a total of 112 invalid responses were excluded, resulting in 237 valid responses. The valid response rate for the multiple-choice section was 67.91%.

For data cleaning of the free association task descriptors, they were inspected by the experimenter item by item. During this process, two respondents' all descriptors were removed due to meaningless input (e.g., "null"). This suggests that the majority of respondents who answered the multiple-choice section conscientiously also provided thoughtful responses in the free association task. Next, all descriptors from invalid responses were examined. Among these, ten respondents' descriptors were removed due to content irrelevant to the research context (e.g., "An interpreter consists of three parts: a transmitter, a receiver, and a charging case. It looks delicate and is portable for users"). These respondents exhibited lower response quality than those in the valid response group, though the level of engagement was higher than expected. This finding supports our earlier assumptions regarding response behaviour. Next, redundant expressions from the same respondent were removed. For instance, if a respondent provided both "attractive" and "nice-looking", the less representative term (e.g., "handsome") was deleted, while the retained term was recorded as a single deletion. In total, 163 descriptors were removed through this process. Following this, compound descriptors entered in the same response field were segmented into individual terms. For example, "gentle and beautiful" was split into "gentle" and "beautiful". However, if a newly segmented descriptor was identical to an existing response from the same respondent, it was not counted again. This process yielded 28 additional descriptors. Finally, spelling errors were corrected, filler words were removed, and English responses were translated into Chinese to ensure consistency in data formatting. As a result, a total of 1441 unique descriptors were obtained.

After completing the data-cleaning process, the reliability and validity of each question set were assessed. The most commonly employed reliability statistic is Cronbach's Alpha, where a threshold of 0.7 or above is generally considered indicative of acceptable reliability. The reliability coefficients for respondents' stereotyped interpreters' political status (α=0.65), economic status (α=0.50), social status (α=0.57), and overall status (α=0.58) suggest varying levels of internal consistency across the constructs. The validity analysis was performed using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's test of sphericity, both of which assess the suitability of the dataset for factor analysis. In confirmatory questionnaire designs, a KMO value greater than 0.7, coupled with a significant Bartlett's test result (p<0.05), is typically considered indicative of adequate structural validity. The KMO statistics and corresponding Bartlett's test results for respondents' stereotyped interpreters are reported: political status (KMO=0.71, $\chi^2(45)=260.18$, p<0.001), economic status (KMO=0.61, $\chi^2(45)=258.38$, p<0.001), social status (KMO=0.68, $\chi^2(45) = 352.50$, p<0.001), and overall status (KMO=0.65, $\chi^2(45) = 360.55$, p<0.001). Given that the present study employs an

exploratory rather than confirmatory research design, the reliability and validity statistics are reported for reference only, without further interpretation of the results.

3.2.4. Demographic statistics of respondents

After data cleaning, there were 237 valid respondents. Descriptive statistics were conducted on the demographic information collected in the first section of the questionnaire. The mean time spent completing the survey was 693.64 seconds (Median=508, *SD*=690.55), with the shortest time being 248 seconds and the longest time being 7,538 seconds. The average age of all respondents was 38 years (Median=32, *SD*=14.9). Among them, there were 91 male respondents, with an average age of 41 years (Median=35, *SD*=16.23), the oldest being 82 years and the youngest being 18 years. There were 146 female respondents, with an average age of 36 years (Median=31, *SD*=13.74), the oldest being 85 years and the youngest being 19 years. No respondents identified as non-binary.

As noted in the previous subsection, the respondents' occupations spanned a range of both institutional (within the system) and non-institutional roles. Among the respondents, 47 (19.83%) respondents were students, 64 (27.00%) clearly reported positions within the system (e.g., government officials, state-owned enterprise employees, teachers), and 36 (15.19%) respondents were retirees (defined as females aged 55 and above, males aged 60 and above, and/or those who identified retired as their occupation). There are 197 (83.12%) respondents who acquired one foreign language or above, mostly English. Other demographic information that can be summarized in a table is provided in Table 3.1.

Table 3.1 Other demographic information of respondents

Category	**Content**	***n***	**Percentage**
Gender	Male	91	38.40%
	Female	146	61.60%
Political affiliation	The mass	88	37.13%
	Communist Party of China	98	41.35%
	Communist Youth League	49	20.68%
	Democratic parties	2	0.84%
Province or provincial-level jurisdiction	Jiangsu	78	32.91%
	Hubei	31	13.08%
	Shandong	22	9.28%

	Guangdong	17	7.17%
	Sichuan	16	6.75%
	Yunnan	8	3.38%
	Zhejiang	7	2.95%
	Henan	7	2.95%
	Hunan	6	2.53%
	Beijing	6	2.53%
	Tianjin	6	2.53%
	Shanghai	5	2.11%
	Jiangxi	5	2.11%
	Qinghai	3	1.27%
	Chongqing	3	1.27%
	Anhui	3	1.27%
	Hebei	2	0.84%
	Ningxia	2	0.84%
	Xinjiang	2	0.84%
	Guangxi	1	0.42%
	Gansu	1	0.42%
	Guizhou	1	0.42%
	Hainan	1	0.42%
	Jilin	1	0.42%
	Shanxi	1	0.42%
	Inner Mongolia	1	0.42%
	Fujian	1	0.42%
Ranking	Not applicable	113	47.68%
	Junior	33	13.92%
	Intermediate	53	22.36%
	Associate senior	35	14.77%
	Senior	3	1.27%
Education	Primary school	1	0.42%
	Middle school	3	1.27%
	High school or vocational	22	9.28%
	Associate degree	32	13.50%
	Bachelor	106	44.73%

	Master	64	27.00%
	Doctor	9	3.80%
Major	Not applicable	26	10.97%
	Literature	80	33.76%
	History	4	1.69%
	Philosophy	2	0.84%
	Economics	18	7.59%
	Management	23	9.70%
	Law	7	2.95%
	Education	19	8.02%
	Arts	7	2.95%
	Science	14	5.91%
	Engineering	26	10.97%
	Medicine	10	4.22%
	Agricultural Science	1	0.42%
Foreign language acquired	Not applicable	40	16.88%
	English	194	81.86%
	Japanese	30	12.66%
	French	21	8.86%
	German	6	2.53%
	Korean	6	2.53%
	Russian	5	2.11%
	Arabic	2	0.84%
	Burmese	1	0.42%

3.2.5. Major findings

Given the role of media exposure in shaping group stereotypes (Seiter, 1986; Ward & Grower, 2020), we first examined the sources from which participants gained knowledge about the interpreting profession. Analysis of respondents' answers to the question "From which sources do you typically learn about interpreters?" revealed that the top three sources were official television media (114 respondents, 48.10%), official print media (90 respondents, 37.97%), and mobile short-video applications such as Douyin[6] (78 respondents, 32.91%). These findings suggest that official media outlets serve as the primary source of public knowledge about

[6]. Douyin is the Chinese version of TikTok.

interpreters, although grassroots social media platforms such as Douyin also exert a certain degree of influence. Additionally, 73 respondents (30.8%) reported that their perceptions of interpreters were reinforced through informal conversations with friends, which further contributed to the formation of stereotypical beliefs about the profession.

Among all respondents, 51 individuals (21.51%) reported having had direct contact with interpreters in work settings. Specifically, 21 respondents (8.86%) had attended online conferences featuring international speakers with interpreter support, 31 respondents (13.08%) had participated in offline conferences where interpreters were present, 19 respondents (8.02%) had attended international exhibitions with interpreter services, and 1 respondent (0.42%) had encountered an interpreter during an in-person volunteer training session. As this was a multiple-choice question with multiple answers, there was an overlap in responses, meaning that a single respondent could have encountered interpreters in both online and offline settings. Among these 51 respondents, 19 individuals reported having interacted with an interpreter only once or twice in their lifetime, 12 respondents encountered interpreters once or twice per year, 11 respondents had contact with interpreters three to ten times annually, and 7 respondents engaged with interpreters more than ten times per year.

In terms of gender-based stereotypes concerning interpreters, a majority of respondents perceived the interpreting profession as female-dominated. Specifically, 135 respondents (56.96%) indicated that interpreters were predominantly female, selecting either “overwhelmingly female” or “mostly female, with some male practitioners”. This proportion exceeded the number of respondents who perceived the profession as gender-balanced (*n*=87, 36.71%) or male-dominated (*n*=15, 6.32%, including those who selected “overwhelmingly male” or “mostly male, with some female practitioners”). These findings are consistent with the gender distribution of interpreters discussed in subsection 2.4.1, which highlights the numerical predominance of female practitioners within the profession.

With regard to age-related stereotypes, respondents demonstrated a high degree of consistency in their perceptions of the typical age range of interpreters. Specifically, 180 respondents (75.95%) identified interpreters as being primarily between 25 and 35 years old, while 48 respondents (20.25%) perceived them to be within the 35–45 age range, collectively accounting for 96.2% of the sample. Notably, no respondents (0%) associated interpreters with an age of 55 years or older. This perception stands in contrast to the survey data reported by the Translators Association of China (2022), which reported that 18.3% of full-time translators and interpreters are 45 years or older. Furthermore, the proportion of interpreters aged 25–35 years in the actual workforce (37.5%) is considerably lower than the respondents’ subjective

estimations, thus revealing such a discrepancy between public perception and the actual age demographics of the profession.

Given that social status is a composite construct influenced by multiple factors, this survey collected respondents' perceptions of interpreters' educational background, income, political status, economic status, and social status. In terms of education, the majority of respondents (*n*=157, 66.24%) considered a master's degree to be the typical level of educational attainment for interpreters, while 41 respondents (17.3%) believed that a doctoral degree would be appropriate for the profession. Notably, no respondents perceived individuals with associate degrees or lower as suitable for the role of interpreter. This perception also stands in contrast to the actual educational profile of translator and interpreter practitioners. According to the Translators Association of China (2022), 48.6% of practitioners hold a bachelor's degree, 35.7% possess a master's degree, and 8% have attained a doctoral degree. These discrepancies suggest that respondents tend to hold a positive stereotype regarding the educational qualifications of interpreters, perceiving the profession as more academically advanced than it is in practice.

Regarding the stereotyped annual income of interpreters, after removing 19 outliers using the interquartile range (IQR) method, the mean income was calculated to be ¥298,600 (Median = ¥300,000, *SD*=¥134,900). In comparison, the average annual disposable income for residents in China in 2024 was ¥41,314 (国家统计局, 2025), making this perceived figure approximately seven times the average income. Therefore, it placed this income within the high-income category. This figure also exceeds the actual income data for translators and interpreters, with 56.2% of full-time translators earning a monthly salary between ¥3,000 and ¥8,000 (中国翻译协会, 2019).

In comparing the political status, economic status, social status, and overall status of interpreters with other professions, we first conducted descriptive statistics. The responses for the scale ranging from "much higher" to "much lower" were converted into numerical values from +2 to -2, and the average scores were computed for each comparison. The profession with the smallest mean difference was then identified. The results indicate that, according to respondents, interpreters' political status is most similar to that of network engineers (*M*=-0.04, *SD*=0.85), their economic status is closest to that of network engineers (*M*=0.05, *SD*=0.94), their social status aligns most closely with that of network engineers (*M*=0.33, *SD*=0.89), and their overall status is also most similar to that of network engineers (*M*=0.03, *SD*=0.91).

Subsequently, we conducted significance testing. Since each status category listed 10 professions ranked from high to low, and each profession was assigned a score ranging from +2 to -2, a 10×5 numerical matrix was created. We encoded the ten professions in each category

based on their rank in terms of status (1–10) and fitted a linear mixed model, designating occupation status as the fixed factor and respondents as the random intercept. The model formula was: *lmer(scoring~position+(1|respondent ID))*. For political status, the standard deviation of the scores reflecting variation between respondents was *SD*=0.38. For each unit decrease in political status, the average score increased by 0.21 (*SE*=0.01, *df* = 2132, *t*=30.27, *p*<0.001). Using the fixed effects equation, we calculated the occupation closest to interpreters by determining the ordinal value at which the score is zero. The equation is as follows: *Status ranking=-intercept÷slope*. This yielded an ordinal value of 4.98, which approximates to 5. Thus, in terms of political status, interpreters are most similar to journalists. For economic status, the standard deviation of respondents' scores was *SD*=0.55. For each unit decrease in economic status, the average score increased by 0.29 (*SE*=0.01, *df*=2132, *t*=48.67, *p*<0.001). The ordinal value for this status category was 5.62, which approximates 6, corresponding to the profession of programmer. In terms of social status, the standard deviation of scores was *SD*=0.89, and for each unit decrease in social status, the average score increased by 0.27 (*SE* =0.05, *df*=1895, *t*=39.10, *p*<0.001). The ordinal value was 5.69, which approximates to 6, corresponding to the occupation of a musician. Finally, referencing the approach used by Zong, Li, & Sun (2016), who calculated overall status by averaging the scores for political, economic, and social status, we derived respondents' overall status score with *SD*=0.49. For each unit decrease in overall status, the average score increased by 0.26 (*SE* =0.01, *df*=6398, *t*=68.92, *p* <0.001). The ordinal value was 5.51, which approximates 6, corresponding to the occupation of a private entrepreneur. In Zong, Li, & Sun (2016)'s survey, the overall status ranking for private entrepreneurs was 17/30, placing them in the middle range, lower than doctors (15/30) and university professors (7/30).

We subsequently categorized and labelled all descriptors collected through the free association task. First, we clustered them into 20 categories based on the most prominent characteristics (e.g., "high professionalism"). Next, we incorporated relatively ambiguous descriptors into these 20 primary categories, while removing those that were vague in meaning or difficult to categorize. Finally, the 20 categories were consolidated into 10 major categories, as presented in Table 3.2:

Table 3.2 The ten most frequently assigned stereotypical traits of the interpreting profession by the Chinese public.

Ranking	**Traits**	**Five examples of descriptors**	***n***
1	High professional	Solid expertise, strong translation skills, strong	345

	ability	comprehension, erudite, professional literacy (专业知识牢固、翻译能力强、理解能力强、博学多识、专业素养好)	
2	Strong psychological resilience	Mature and composed, emotionally stable, calm under pressure, unflappable, confident and poised (成熟稳重、情绪稳定、临危不乱、遇事不惊、自信大气)	155
3	Good disposition	Gentle and elegant, mentally alert, civilized and polite, poised and graceful, elegant demeanour (温和儒雅、外表精神、文明礼貌、端庄大方、风姿绰约)	147
4	Fluent expression	Smooth articulation, eloquent, articulate, quick-witted, pleasant voice (表达流畅、口才好、口齿伶俐、伶牙俐齿、声音好听)	135
5	High IQ	Wise, clever, sharp-minded, quick-thinking, mentally agile (睿智、聪明伶俐、机制敏捷、才思敏捷、头脑灵活)	122
6	Attractive appearance	Good-looking, well-proportioned features, attractive, high-end appearance, sweet-looking (长得好看、五官端正、颜值在线、形象高级、长相甜美)	114
7	Well-educated	Highly educated, cultured, university graduate, high academic degree, knowledgeable (文化程度高、有文化修养、高校毕业、高学历、学识有为)	57
8	Strong stress tolerance	Able to handle pressure, high mental resilience, always on call, hard-working, busy (面临压力、心里素质高、随时待命、工作辛苦、忙碌)	51
9	Respected	Dignified, prestigious, high social status, role model, respectable (体面、高大上、社会层次高、吾辈楷模、令人	43

		敬佩)	
10	Well-paid	High salary, decent income, well-compensated, generous pay, high hourly wage (高薪、收入可以、薪水不低、大方有钱、时薪高)	38

Therefore, in response to the first research question, "What is the stereotype of a Chinese interpreter held by the general Chinese public?", it can be concluded that the general stereotype of an interpreter among the Chinese public is that they are a young (aged 25-35) female, with a moderate overall social status. The most notable professional attribute assigned to interpreters is high professional competence.

3.2.6. Minor findings

In addition to the major findings, several minor findings are also worth mentioning. First, we visualized the descriptors provided by respondents through free association. Using Xshell on a virtual private server running Ubuntu 20, we installed RStudio Server 4.4.2 and generated a word cloud using the ggwordcloud[7] package in R (R Core Team, 2020), as shown in Figure 3.1. The top 10 most frequently mentioned descriptors were: high education (*n*=24), good appearance (22), quick-witted (17), fast reaction (15), articulate (13), professional (13), knowledgeable (11), beautiful (11), intelligent (11), and equally ranked descriptors such as well-versed in knowledge (10), eloquent (10), strong stress resilience (10), meticulous (10), and quick to respond (10).

[7]. https://cran.r-project.org/web/packages/ggwordcloud/vignettes/ggwordcloud.html

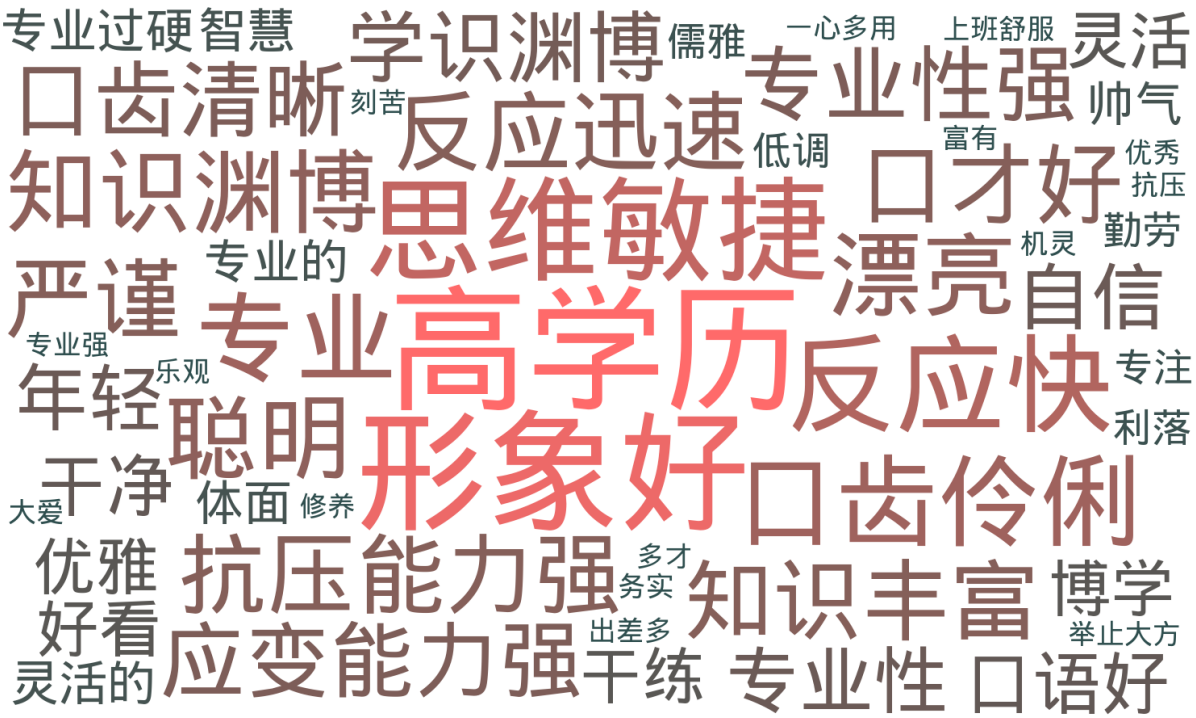


Figure 3.1 The word cloud of the most frequently mentioned descriptors for the stereotype of interpreters. The top ten descriptors are outlined in the first paragraph of this subsection.

Given that these descriptors inherently carry emotional connotations, we further examined the emotional content and intensity they project. In social psychology, emotions are known to influence cognitive judgments, with positive emotions typically reinforcing favourable stereotypes of a group, and conversely, negative emotions amplifying unfavourable perceptions (Bijlstra *et al.*, 2014). To assess the emotional dimensions of respondents' perceptions of the interpreter group, we utilized the Chinese sentiment lexicon developed by the Information Retrieval Research Laboratory of Dalian University of Technology (徐琳宏等, 2008). This lexicon is based on the six primary emotion categories proposed by Ekman (1973), but further refines the positive category into more detailed subcategories, resulting in a classification system encompassing seven broad emotional categories and twenty-one subcategories. We imported this lexicon in RStudio, then aligned all valid descriptors submitted by respondents using a tokenizer, and matched them with the lexicon to categorize the emotion. After that, we measured emotional intensity, and determined emotional polarity (using weighted intensity, with the mode of polarity extracted; empty values were returned when no match was found). After these steps, 653 empty values were discarded. The results indicated that the majority (n=665, 86.81%) of respondents expressed positive sentiments towards the interpreter group,

while only 6 (0.78%) descriptors reflected a negative sentiment, as illustrated in Figure 3.2.

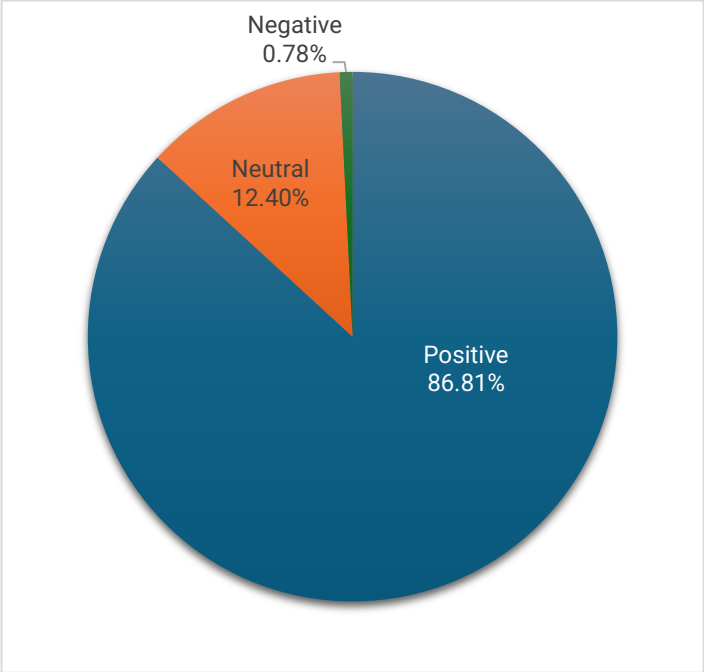


Figure 3.2 The sentiment analysis of the descriptors reveals that the majority of respondents hold positive emotional perceptions of the interpreter group.

We further calculated the frequency and polarity distribution of emotional categories, which revealed the predominant emotions associated with the interpreter group, as shown in Table 3.3. It is evident that praise (*n*=706, 92.17%), representing positive sentiments, constitutes the majority of the respondents' emotional content.

Table 3.3 Sentiment analysis results of the descriptors submitted by respondents through free association regarding the interpreter group.

Emotion	Frequency	Intensity (0-9)	Percentage
Praise	706	5.52	92.17%
Pleasant	18	4.67	2.35%
Relieved	9	4.11	1.17%
Trust	7	5.43	0.91%
Favour	6	3	0.78%
Wonder	6	5.67	0.78%
Respect	6	3.67	0.78%
Blame	5	3.8	0.65%
Fear	2	3	0.26%

Fretful	1	3	0.13%

As the Chinese Sentiment Lexicon (徐琳宏等, 2008) does not specifically establish a "Neutral" emotional category between Positive and Negative, it instead uses a polarity score of 0 to indicate neutrality. Therefore, the percentage values in this table are higher than those shown in Figure 3.2. The classification of descriptors reveals two categories of terms that are emotionally negative. The first category expresses concerns about the limited career development opportunities for interpreters (*n*=7), such as "restricted development opportunities" and "short future career cycle". The second category (*n*=13) reflects perceptions of the interpreter group as having low warmth, with descriptors such as "enigmatic", "somewhat aloof", and "Buddist-minded".

Additionally, a category of descriptors (*n*=24) praises the political consciousness of interpreters, with terms like "deeply rooted in the socialism values", "ambassadors of peace", and "representatives of China's national image".

Lastly, there is a category of descriptors that portrays the stereotypical physical appearance of interpreters (e.g., "height 165 cm", "slim build", "ponytail", and "wears glasses").

Finally, we employed a text-to-image generation model based on generative artificial intelligence to create prototype portraits of a Chinese interpreter, as conceptualized by the general public. After evaluating various platforms, including Midjourney, DALL·E 3, AWPortrait FL, and Stable Diffusion XL, the experimenter selected DreamStudio[8], based on Stable Diffusion XL 1.0 as the generation platform. Given the established public stereotypes regarding the gender and age of interpreters, two distinct prompts were formulated to reflect the specific demands of consecutive and simultaneous interpreting. The prompts were as follows: "*medium shot, young, female, simultaneous interpreter, Chinese, simultaneous interpreting, headphone, business suit*" and "*medium shot, young, female, consecutive interpreter, Chinese, consecutive interpreting, listening, taking notes, business suit*". The configuration of generation steps was set to 100. Approximately 250 images were generated for each prompt, which was 500 images in total. Subsequently, five images representing consecutive interpreters and five representing simultaneous interpreters were selected, and their backgrounds were removed using Adobe Express. A further small-scale survey was conducted with 20 Chinese citizens, who were asked to vote on which of the 10 images most accurately depicted a "Chinese interpreter". The five images with the highest number of votes are

[8]. https://dreamstudio.ai/generate

presented in Figure 3.3.

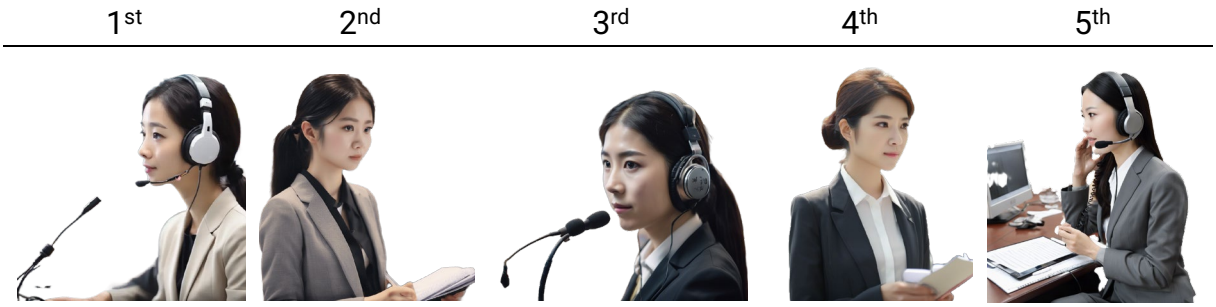


Figure 3.3 The top five voted images of an AI-generated prototype of a Chinese interpreter, generated by Stable Diffusion XL 1.0 and voted by 20 native Chinese citizens.

3.3. Discussion

Through the literature review presented in the previous chapter, we have established that interpreters' gender, race, accent, age, and professional title influence users' perceptions of interpreting quality. For instance, Asian female interpreters are more likely than their male counterparts to receive praise and sympathy for their interpreting performance (Yang, 2023). Similarly, aged interpreters may experience a loss of interpreting opportunities due to their advanced age, whereas interpreters with a professional title, even without formal interpreter training, tend to receive higher quality ratings and trust scores from users. These findings align with research in social psychology, which examines how stereotypes associated with gender, age, and professional titles reflect underlying socioeconomic status perceptions. As natural and social cues inherent to every individual, gender and age are universal natural cues possessed by all interpreters, while ad hoc interpreters additionally carry a professional title associated with their primary occupation alongside the role of interpreter. Given these factors, such cues inevitably introduce bias into users' perceptions of interpreting quality. However, due to the scarcity of existing literature, particularly the lack of controlled variable-based quantitative studies on gender and age, further experimental research on these aspects is warranted to generate empirical insights.

Before designing controlled variable experiments, it is necessary to establish baseline criteria for the control group, specifically defining the perceived gender and age of the reference interpreter as perceived by user participants. Additionally, the ranking of an interpreter's professional title within Chinese society should be considered as well. This ranking should ideally be at least one tier higher than that of interpreting as a profession to create a distinct contrast in socioeconomic status compared to interpreting practitioners, such as positioning the

upper-class relative to the middle class or the middle class relative to the lower class. It is also necessary to formulate hypotheses to address the research questions. For instance, regarding the second research question, to determine whether an interpreter's perceived gender, age, or professional status introduces a positive or negative bias. To provide empirical insights into these considerations, the present chapter examines the stereotypical perceptions of the interpreting profession among the Chinese public. Data collected using five-point Likert-type scales reveal that the dominant gender stereotype associated with interpreters is female. While respondents did not explicitly state that interpreting is a pro-feminine profession in terms of professional competence, the overall stereotypical perception of the profession is feminine. This is reflected not only in questionnaire responses but also in free association tasks, where respondents frequently employed feminine descriptors. Regarding the age of interpreters, respondents predominantly perceived the interpreter group to be within the 25-35 age range, with no respondents considering individuals aged 55 or older to be suitable for the profession. These findings will serve as a key starting point for the next chapter, where experimental designs will be developed to further investigate the impact of such stereotypes on quality perception.

In terms of the public perception of the socioeconomic status of the interpreter group, we combined the findings from the occupational prestige social survey by Zong, Li, & Sun (2016) and verified them through both descriptive statistics and a set of linear mixed models. The results revealed that, in terms of overall socioeconomic status, the interpreter group is most closely associated with the occupation of private entrepreneurs, placing them at a moderately lower level in the hierarchy. This finding aligns with the interpreter group's moderate ranking of occupational prestige in the study of Zong, Li, & Sun (2016). Thus, the interpreting profession, in terms of its overall professional status, cannot be equated with traditionally high-status professions such as judges or university professors. This conclusion provides a theoretical basis for assigning an elevated socioeconomic status professional title to interpreters in the next chapter. Furthermore, we observed that the descriptors derived from the free association exercise predominantly evoke positive emotions, with praise being the most frequent sentiment. Therefore, utilizing a professional title associated with a higher socioeconomic status in the experiment will not lead to a shift in participant' attitudes from negative to positive, effectively eliminating this as a confounding variable.

The classification of descriptors reveals that the public stereotype of interpreters largely focuses on attributes such as strong professional competence, psychological resilience, fluent delivery, and physical attractiveness, all of which align with traditional media narratives. Even

though the majority of respondents have not personally interacted with interpreters or listened to live interpreting sessions, it can be inferred that media exposure plays a significant role in reinforcing these stereotypes. Among the top ten descriptors, fluency (i.e. expressiveness) is frequently emphasized in relation to interpreting quality, while fidelity (i.e., accuracy relative to the source text) is seldom mentioned. This finding is consistent with the discussion in subsection 2.1.3, which highlighted the tendency of users to place greater emphasis on the expressive abilities of interpreters, while often overlooking the accuracy of their translations. Thus, stereotypes can significantly influence, and even interfere with, users' perception of interpreting quality. Hypothetically speaking, an interpreter who is physically attractive, demonstrates fluent delivery, and exudes psychological resilience might be perceived positively. However, if this interpreter has insufficient foreign language listening comprehension skills, leading to substantial loss or distortion of information in whose interpretation, their performance might still be well-received due to their alignment with respondents' stereotypes. This could lead some interpreters to prioritize enhancing traits that align with public stereotypes, such as psychological resilience (ranked third), physical attractiveness (ranked sixth), or educational qualifications (ranked seventh), rather than focusing on improving their foreign language listening comprehension skills. This would, in turn, boost their perceived professional competence, even if their actual interpreting abilities are compromised. In fact, such behaviours have been reported by Hong & Choi (2020), where it was noted that interpreters may activate and reinforce stereotypes to gain greater recognition. However, this approach is detrimental to both ensuring interpreting quality and enhancing the professional capabilities of interpreters. It also undermines the long-term health and sustainable development of the interpreting profession, as it prioritizes conforming to superficial perceptions over improving the core skills essential for high-quality interpreting.

In addition to the major findings, one noteworthy minor finding was that users, through free association, reported their mental prototype of an interpreter. This prototype represents a schema in psychology ', serving as a tangible manifestation of the stereotype associated with a group. Several descriptors consistently depicted interpreters as "slender", "around 165 cm tall", and "female with glasses". However, not all interpreters conform to this physical prototype. For instance, the professional interpreter Dr Lin Chaolun is a tall aged[9] male, which contrasts with the stereotypical image. Media exposure constitutes one of the primary factors facilitating the formation of this prototype. However, the emergence of this prototype also serves as a critical

[9]. Dr. Lin Chaolun is 67 years old in the year 2025.

reference for the experimental stimuli in the subsequent chapter. By utilizing contemporary generative artificial intelligence technology, it becomes possible to translate textual stereotypes of interpreters into visual representations, thereby concretizing the otherwise abstract prototype. This methodology not only enhances the clarity of the stereotype's physical manifestation but may also provide referential value for further empirical research.

3.4. Summary

The primary objective of this chapter is to address the first research question: "What is the stereotype of a Chinese interpreter held by the general Chinese public?" To this end, we employed free association tasks, a widely recognized method in social psychology for eliciting stereotypes, to categorize and analyse the descriptors provided by 239 participants from the general public. In addition, the survey gathered respondents' perceptions regarding the gender, age, and socioeconomic status of a stereotypical interpreter. The findings align closely with the phenomena discussed in the literature review. Specifically, the Chinese public predominantly associates the typical interpreter with a young female (aged between 25 and 35), with a socioeconomic status corresponding to the middle class, a result that is consistent with Zong, Li, and Sun's (2016) research on occupational prestige. Through categorizing the descriptors provided via free association, the five most salient stereotypical traits attributed to interpreters were identified: high professional ability, strong psychological resilience, good disposition, fluent expression, and high IQ. These descriptors were subjected to sentiment analysis using the Chinese Sentiment Lexicon (徐琳宏等, 2008), revealing that the overall emotional tone associated with these traits is overwhelmingly positive, predominantly characterized by praise. Consequently, the public's attitude toward the interpreting profession is largely favourable, with no indication that the professional title of an ad hoc interpreter would lead to a shift in emotional perception. Additionally, we utilized generative artificial intelligence to produce a prototypical image of an interpreter, based on the primary stereotypes identified. This image serves as a reference for the design of experimental stimuli in the subsequent chapter.

Chapter 4 Exploring users' stereotyped interpreting quality perception

In the preceding chapter, we examined the stereotypes held by the Chinese public regarding a Chinese interpreter's gender, age, and social status. The findings revealed that users predominantly perceive interpreters as young, female, and positioned within the middle stratification in terms of socioeconomic status. The current chapter seeks to explore how these three categories of social categorical information are extrinsic to interpreting quality but may still bias users' perceptions of interpreting quality. In specific, this chapter seeks to ascertain whether these biases manifest as positive or negative influences. Consequently, the primary objective of this chapter is to address the second research question: "Does an interpreter's perceived social categorical information activate users' stereotypes, thus biasing their perceptions of interpreting quality? If so, in which direction?" In practical contexts, an interpreter typically embodies multiple categories of social cues, such as variations in both gender and age. As a result, through multivariate analysis, we seek to further investigate the third research question: "How do multiple variates (gender, age and professional title) interact with each other, and does one category exert a major influence?" To this end, the chapter first outlines the study's variables, research framework, and participant profile. A pilot study was then conducted, with the professional title as the sole independent variable. Following this, a formal experiment was executed to collect data on gender, age, and professional title, which was subsequently subjected to cross-validation. The results of these analyses are discussed in the latter section of this chapter.

4.1. Experiment design

4.1.1. Experimental framework and variable control

In terms of empirical research methodology, this chapter generally adopts an experimental design that controls for variables. Specifically, it aims to maintain consistency in the experimental conditions between the experimental group and the control group, ensuring that the overall level of participants remains as uniform as possible. The only distinction lies in the experimental stimulus provided to each group, which differs for the experimental group while remaining consistent for the control group. In the field of interpreting studies, several experiment-based studies have already established complete research designs and achieved expected results in data collection and analysis. For instance, Cheung's (2013, 2020, 2021)

series of studies validated how factors such as the interpreter's perceived non-native accent, nationality, and the mode of interpreting (onsite versus remote) influence users' perceptions of interpreting quality. The experimental outcomes confirmed both the theoretical framework from the literature and the hypotheses. Therefore, compared to qualitative studies such as those by Hong & Choi (2020) and Liu et al., (2020) which provide in-depth interview data but lack statistical significance, the approach taken in this chapter better addresses the limitation of qualitative research failing to draw conclusions from a statistical perspective. Naturally, the data obtained from experimental research does not conflict with interview-based studies but rather provides a clear directional foundation for subsequent qualitative research or quasi-experimental studies, which involve handling the original groups in relatively natural settings (Creswell, 2012).

The experimental variables in this chapter follow the three main categories discussed in the previous chapters, namely the gender, age, and professional title of the interpreter as perceived by users. These categories correspond to the socio-economic status associated with the interpreter's role as an ad hoc professional. The rationale for not including race among the "big three" categories is provided in section 2.2. Furthermore, the set of experiments in this chapter aims to address a research gap, specifically the lack of intersectional studies involving multiple social categorical variables in prior interpreting research. In the field of interpreting studies, most research has failed to implement strict variable control. For example, Christeny Gracia's (2023) experiment-based study included a large number of interpreters, but the social categorical differences among them introduced confounding variables, thus weakening the study's persuasive power. Although Cheung's series of experiments has brought a new social-psychological perspective to interpreting studies, and the findings have significantly supplemented the conclusions of quantitative research, some studies, such as Cheung (2020), still suffer from such a confounding variable that mixed an interpreter's perceived nationality with his non-native accent. That is to say, the experiment manipulated an interpreter speaking Mandarin Chinese with an American accent, yet the main purpose of the study was to examine the impact of the interpreter's nationality on users' perceptions of interpreting quality. In contrast, the set of experiments in this chapter, due to its shorter time span, more homogeneous demographic information of participants, and consistent experimental stimuli and settings, may provide a more robust framework for analysing the effects of activating these three variables individually or in combination.

4.1.2. Participants

In terms of participant selection, Shao (2018, 第 44 页) noted that "research subjects in interpreting studies are often drawn from three categories: untrained bilingual, student interpreters, and professional interpreters." An important and often larger group in interpreting assignments, namely, interpreting users should also be considered. The importance of the user group has already been discussed in subsection 2.1.2. The user group is vast and heterogeneous, meaning it cannot be assumed that all users are monolingual individuals with no foreign language proficiency, nor that they are all untrained bilinguals. In China, where English is taught as a foreign language, students are required to undergo nine years of compulsory English education starting from elementary school. As a result, interpreting users in China generally possess basic but not necessarily advanced English proficiency. According to the EF English Proficiency Index[10], China ranks 91st out of 116 countries, below the global average. However, the group of interpreters also includes users with higher levels of English proficiency. Therefore, the research participants in this study are appropriately selected from a group of users with medium-level English proficiency based in mainland China.

4.1.3. Independent and dependent variables

In alignment with the research questions and the selected methodology of this chapter, the experimental stimulus consists of two main components: the audio recording of the rendition and the social categorical information of the interpreter as perceived by the users. The social categorical information includes three variables: gender, age, and professional title. Drawing on the experimental setup from Cheung's (2013, 2020, 2021) series of studies, the most suitable experimental stimulus in this study is a rendition produced by an interpreter but has been carefully manipulated. The rendition should remain identical in terms of content and paralinguistic features to minimize confounding variables. Regarding the social categorical information of gender, age, and professional title perceived by users, the approach follows Cheung (2020) which utilized an image of the interpreter to activate participants' stereotypes. As a visual representation of an individual's appearance, photos can effectively trigger stereotypes (Hansen *et al.*, 2016; Qi *et al.*, 2018). Thus, in this study, the interpreter's photo will be carefully manipulated and presented as a variable to participants. In Cheung (2020), ten photos from the open Internet were selected and two among them were chosen after a vote by 100 undergraduate students. In the preliminary phase of this study, we first invited two different interpreters and tried to request them to produce renditions that were mostly identical. However,

[10]. https://www.ef.com/wwen/epi/regions/asia/china/

despite providing rendition content that included filled pauses, speech rate annotations, and key timestamps, it was observed that aligning the renditions was challenging. Additionally, the renditions sounded unnatural due to the imposed reading requirements. Given the advancements in digital technology, an alternative and effective approach is employed: the use of face-swapping technology to alter the gender and age in the photographs. Moreover, since the gender and age in the photos are modified, adjustments to the gender or age in the audio rendition are necessary to ensure consistency between the social categories of the voice and the image. This can be achieved using professional audio editing software to modify the voice tone accordingly.

In accordance with the research questions of this dissertation, the independent variable in this study is the social categorical information of the interpreter as perceived by the users, while the dependent variable is the perceived interpreting quality. Interpreting quality is a broad and complex concept that, despite extensive research and discussion by interpreting scholars over the decades, still lacks a universally accepted, clearly defined set of criteria. This issue was also reviewed in Chapter 2. Additionally, since users have not undergone systematic interpreting training, overly detailed categorization of interpreting quality would not only increase the cost of user education (范志嘉, 2018), but also make it difficult for users to clearly differentiate between categories during the evaluation process. Therefore, this study primarily adopts Liu's (2015) two main categories of interpreting quality: *Accuracy* and *Delivery* (see Table 2.2 for full definitions). In this study, these are referred to as *Fidelity* and *Delivery*, respectively.

Furthermore, as this experiment generally introduced social psychological approaches to our interpreting research, two additional categories from the intersection of interpreting studies and social psychology are introduced: *Domain knowledge* and *Trust*. Domain knowledge in interpreting refers to the accurate translation of terminologies and the overall style of the translation conforming to the linguistic practices of industry practitioners. In the work of other interpreting scholars, domain knowledge is often cited as a subcategory of interpreting quality (see Table 2.1). Trust, on the other hand, plays a crucial role in the interpreter-user relationship and significantly influences users' acceptance of the rendition. For instance, users believe that personal character and trust are important in understanding what constitutes good interpreting (Edwards *et al.*, 2005), while lack of trust can lead users to question the quality of the translation, thereby affecting communication effectiveness between users and interpreters (Wilson & Walsh, 2019) and creating tension in the working relationship (Hsieh *et al.*, 2010). We state that users' recognition of the interpreter's competence in domain knowledge and their level of trust in the interpreter are key factors in shaping perceptions of interpreting quality. Therefore, in the

analysis of the experimental results in this chapter, we not only examine fidelity and fluency in traditional settings but also analyse their perceptions of domain knowledge and trust. A summary of the variables, their content, and the methods for measuring or controlling these variables can be found in Table 4.1.

Table 4.1 Variables, content, and methods of measurement or control in this study

Variable type	Category	Content	Control / Measurement
Independent variables	Gender	Male, Female	Face swapping driven by computer technology
	Age	Young (below 40), Aged (above 60)	Face swapping driven by computer technology
	Professional title	Job title, educational background, professional qualifications	Presented or implicit in the experimental stimulus
Dependent variables	Fidelity	Users' perception of the accuracy of the rendition relative to the original text	Users' ratings based on a Likert-type scale
	Delivery	Users' perception of whether the interpreter's delivery is sufficiently idiomatic	Users' ratings based on a Likert-type scale
	Domain knowledge	Users' perception of the interpreter's professionalism in the medical domain	Users' ratings based on a Likert-type scale
	Trust	Users' trust in the interpreter	Users' ratings based on a Likert-type scale

4.1.4. The experimental stimulus

In selecting the audio materials for interpreting, several considerations were taken into account to ensure the appropriateness of the experimental stimulus. According to Gile (2009), if the source speech contains excessive information density, is highly specialized, or is delivered at an excessively fast pace, it may pose comprehension and translation difficulties even for professional interpreters. Such conditions would be unfavourable for the preparation of the

experimental stimulus. Conversely, if the material is too simple, it may result in a ceiling effect, wherein participants rely less on the interpreter and affect the validity of the experiment. Therefore, the selected interpreting audio should strike a balance: the English content should be moderately challenging but not overly difficult, possess domain-specific terminology without being excessively specialized, maintain a moderate speech rate, and control for non-native accents to avoid accent-related biases (particularly those linked to racial perceptions). Additionally, the duration of the audio should not be excessively long. A length of approximately 10 minutes is considered optimal (e.g., Cheung 2013), as excessively long audio content may lead to participant fatigue, reducing engagement and motivation. Based on these considerations, the experimenter searched publicly available English speeches on the Internet targeted at native Chinese users and ultimately selected a popular medical science lecture on delaying ageing and preventing cancer. The speaker is a native Australian cardiologist, and the original video lasts approximately 12 minutes, with a speaking pace of 118 words per minute. To assess the readability of the content, the Flesch Reading Ease Score was calculated, yielding a score of 67.00, while the Flesch-Kincaid Grade Level was 8.17. These scores indicate that the material presents a moderate level of comprehension difficulty for ordinary Chinese users while remaining manageable for professional interpreters.

This study adopts consecutive interpreting as the primary interpreting mode. The rationale behind this choice is that compared to simultaneous interpreting, the consecutive way more prominently exposes the interpreter's social categorical information to users. While the codes of conduct of national and international professional organizations require interpreters to remain transparent, invisible, passive, neutral, and detached in their professional practice (Mason & Ren, 2012), the degree of visibility varies significantly between interpreting modes. In simultaneous interpreting, the interpreter operates from within a booth, with their only connection to users being the audio output through the interpreting receiver. This setting largely minimizes the visibility of the interpreter's social categorical attributes. Conversely, in consecutive interpreting, the interpreter is physically present in the same space as the users, making their appearance, professional title, and other social category-based characteristics readily identifiable. In some cases, this visibility may even become excessive (Zheng & Xiang, 2018). Therefore, choosing consecutive interpreting as the primary interpreting mode enables a more effective presentation of the experimental stimuli, thus improving the validity and impact of the study's findings.

Selecting an appropriate interpreter is a crucial consideration in this study. The foremost requirement is that the interpreter should be capable of accurately and comprehensively

rendering the source language into the target language in a single take. Choosing a qualified interpreter who can produce a natural and high-quality interpreting audio recording is required. This minimizes potential issues related to fidelity, such as omissions or mistranslations, as well as fluency-related disfluencies, such as hesitations or self-corrections, which could negatively influence users' quality perceptions and, consequently, interfere with the intended examination of stereotype activation. Additionally, given that the interpreter's professional title serves as a social cue in this study, the ideal scenario would involve selecting an interpreter who already holds a high professional title while also being sufficiently qualified in interpreting. This approach ensures the generation of high-quality interpreting audio while simultaneously leveraging the interpreter's professional title, thereby safeguarding the validity and naturalness of the experimental stimulus (邵张旻子, 2018). Since the stimulus revolves around medical interpreting, the selection criteria were further refined to prioritize an interpreter with a formal academic background in medicine and a high socioeconomic status (as reflected by whose educational credentials and professional titles in the medical field) but who has since transitioned to full-time interpreting practice.

4.1.5. Preliminary set-up

After finalizing the design framework for the experimental stimulus and confirming both the independent and dependent variables, the next step is the production of the experimental stimulus. Through the experimenter's networking within China's community of professional interpreters, a qualified interpreter was finally identified. The selected interpreter is a male, approximately 35 years old, with a combined bachelor's and master's degree in medical science from a top-tier university in China. Additionally, he has held a visiting scholar position at Harvard Medical School. He formerly served as a deputy chief physician before transitioning to a full-time career as a professional medical interpreter. This background provides an optimal foundation for the four versions of video stimuli prepared for user participants across different experimental groups.

This professional medical interpreter selected for this study was invited to record the experimental audio stimulus. He signed an informed consent form and received compensation for his participation. The consecutive interpretation was performed and recorded using Tencent VooV Meeting software. In the working flow of consecutive interpreting, speakers often pause every two to three minutes to allow the interpreter to render the translation. In this experiment, the interpreter was not instructed on a fixed time interval for interpretation but was instead advised to maintain consistent pauses. As per the interpreter's request for concealing his

personal information, an open-source resume photo licensed under Creative Commons 2.0 from Flickr was used in place of his actual resume photo. Subsequently, an AI-powered photo editor FaceLab[11] was used to create modified versions of the photo. The gender-swapping function was applied to generate a female version, while the age-progression function was used to produce an aged version (see Figure 4.1).

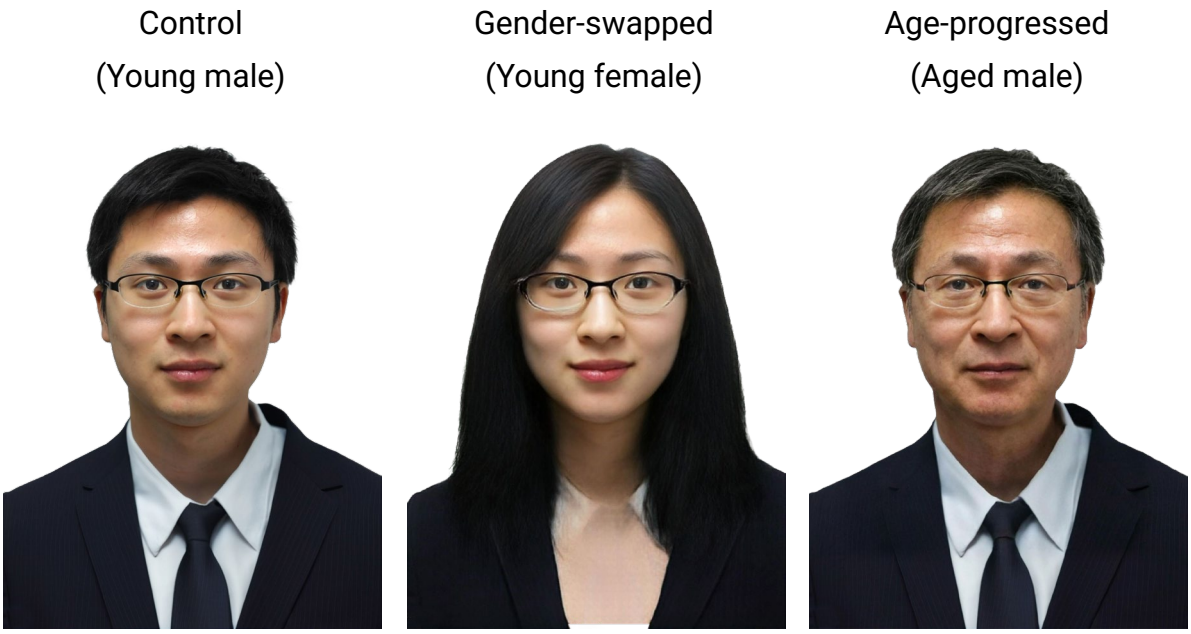


Figure 4.1 Photos presented in the control group, gender-swapped female group, and age-progressed group

To verify whether the three images accurately represented the young (under 35) and aged (over 60) categories, 10 online age detection tools including FacialAge[12] were used to assess the estimated ages of the three photos. A Kruskal-Wallis test was conducted to analyse potential intra-group differences in the age assessment results. No significant differences were found among the ratings provided by the age detectors, which demonstrated consistency in the automatic age ratings for the three images: the aged male had a mean estimated age of 62.5 years (SD=2.33, p=0.44), the young male had a mean estimated age of 36.5 years (SD=2.45, p =0.44), and the young female had a mean estimated age of 28.5 years (SD=2.50, p=0.44). To ensure that the voice characteristics corresponded to the altered images, an audio editing

[11]. https://apps.apple.com/us/app/facelab-face-age-aging-hair/id1530776865

[12]. https://www.facialage.com/

software AudioDirector[13] was used to modify the voice timbre via the Audio-Voice Changer function. After conducting individual auditory tests on different voice outputs, the experimenter selected "ZH-Guo Matilda" for the young female voice and "ZH-Richard Wei" for the aged male voice to best match the respective image transformations.

After collecting the rendition audio, the experimenter merged the speech video with the audio to create four video stimuli. Each video was then paired with a different photo according to the priming conditions. For stimuli that included a professional title, the title was explicitly displayed in the video (see the experimental design subsection for details). The study did not produce a stimulus for an aged female interpreter (both gender-swapped and age-progressed) due to potential distortions that could arise from two consecutive transformations applied to the image and voice. Our consideration is that such distortions could lead to an unnatural perception among participants, thereby compromising the quality of the collected data.

4.2. The pilot study

After formulating the hypotheses, we conducted a pilot experiment using professional title as the sole variable. The study yielded promising findings, and the relevant results were published as a preliminary outcome of the PhD stage in the *Journal of Foreign Languages* (李天韵和崔英, 2022)。

4.2.1. Participants

The pilot study was conducted in the year 2022. Due to COVID-19-related restrictions, the final sample consisted of 113 university participants from Shandong University at Weihai. Among the 113 participants, 60 of them were non-English major undergraduates and 53 were undergraduate students from the English translation major. For those non-English major students, who had not received any formal interpreting or translation training, were categorized as low-proficiency interpreting users. In contrast, English translation major students, all of whom had passed a selection test assessing their translation and interpreting competence, were classified as high-proficiency interpreting users. Additionally, four experts were invited to provide baseline reference ratings for the interpreter's performance.

[13]. https://www.cyberlink.com/blog/audio-editing/1435/how-to-use-voice-changer-online-windows-free

4.2.2. Experimental procedures

The pilot study employed video recordings of simulated real-life conference interpreting as the experimental material. Quantitative and qualitative data from participants were collected via questionnaires. The pre-recorded conference interpreting session featured a public lecture on delaying ageing and cancer prevention, a highly industry-specific topic, with no subtitles in the video. The content was unrelated to the participants' academic curriculum. The speaker was a native Australian English speaker with a professional identity as a preventive health specialist, cardiologist, and a Fellow of the Royal Australasian College of Physicians. The interpreter's native language was Chinese, with English as the first foreign language. His professional background includes roles as an associate chief orthopaedic physician, a holder of a master's degree in clinical medicine, and a visiting scholar at Harvard Medical School. As such, the interpreter possessed multiple professional titles, namely interpreter, physician, and visiting scholar.

Prior to the commencement of the experiment, all participants were randomly assigned to either the experimental group or the control group. The experimental group consisted of 60 participants, including 33 low-proficiency interpreting users and 27 high-proficiency interpreting users. The control group comprised 53 participants, including 27 low-proficiency interpreting users and 26 high-proficiency interpreting users. At the beginning of the experiment, participants were informed that they would be watching a medical-themed video in English, in which the speaker paused approximately every two minutes, allowing the interpreter to provide consecutive interpretations. Upon completion of the video, participants were required to complete a 10-item true/false comprehension test followed by a questionnaire. To motivate participants to engage seriously with the test and enhance data reliability, they were informed that the top 10 scorers on the true/false test would receive a monetary reward. Additionally, as mistranslations or omissions are inevitable in consecutive interpreting, participants were advised to listen carefully to the English content rather than relying solely on the Chinese interpretation.

The video recordings viewed by both the experimental group and the control group were identical, with the only stimulus variable being: before the video commenced, the experimental group was shown a 10-second introduction of the speaker's professional titles, followed by a 10-second introduction of the interpreter's professional titles, whereas the control group was shown only the 10-second introduction of the speaker's professional titles, with no introduction of the interpreter's professional titles. This stimulus variable functions as a priming mechanism, activating cognitive heuristics based on implicit stereotypes among participants. Accordingly,

the control group was primed with the identity of an "interpreter", whereas the experimental group was primed with the identity of an "ad hoc interpreter undertaken by a highly educated professional medical doctor".

The primary focus of this experiment is to highlight the ad hoc interpreter's professional title as a medical doctor. The general stereotype of medical doctors, as elaborated in Chapter 3, is generally positive among the social public. In contrast, the language service profession holds a relatively lower social status, with professional interpreters being perceived less favourably than physicians (中国翻译协会, 2019; 宗刚等, 2016). Based on these considerations, the research hypothesis for this experiment is that there will be a significant difference in the perceived interpreting quality among users in the experimental group, while no significant difference will be observed in the control group. Additionally, drawing on Chiaro's (2002) findings, this study further hypothesizes that the disparity in quality perception between the experimental and control groups will be smaller among high-proficiency foreign language users compared to low-proficiency users.

4.2.3. Quantitative data analysis

A total of 60 questionnaires were collected from the experimental group and 53 questionnaires from the control group, resulting in a 100% response rate. After excluding invalid responses due to missing answers, identical responses, or patterned answering, 9 invalid questionnaires were removed from the experimental group, yielding 51 valid responses, while 1 invalid questionnaire was removed from the control group, leaving 52 valid responses.

The quantitative data analysis focused on four categories of interpreting quality as elaborated in section 4.1: fidelity, delivery, domain knowledge and trust. Each dimension was assessed twice with positively framed items and once with a negatively framed item, which was later reverse-coded to obtain an average score from three measurements, ensuring data reliability. Consequently, the quantitative questionnaire consisted of 12 items, each rated on a five-point Likert-type scale, with response options ranging from "strongly disagree", "disagree", "neutral", "agree", to "strongly agree". Before conducting data analysis, the Kolmogorov-Smirnov test was applied to assess the normality of the dataset. The results confirmed a normal distribution across all samples (p=0.19), low-proficiency foreign language users (p=0.20), and translation major students (p=0.20).

To compare the experimental group and control group across the four dimensions, we conducted mean comparison analysis and independent-sample t-tests, as presented in Table 4.2. The results indicate significant differences in scores between the experimental group and the

control group across all four categories, despite both groups evaluating the same interpreting video. This finding suggests that when the ad hoc interpreter's professional titles were emphasized, it significantly affected users' perceptions of interpreting quality. Since the experimental group consistently exhibited higher mean scores than the control group, it can be inferred that highlighting the interpreter's professional titles introduced a positive bias in users' quality perception, aligning with the public's positive stereotypes toward physicians and highly educated individuals. Furthermore, the differences in mean scores between the experimental and control groups were particularly pronounced in industry competence and trust, with high t-values and a p-value close to 0. This suggests that emphasizing the interpreter's social identity significantly strengthened users' perceptions of the interpreter's domain competence and trustworthiness.

Table 4.2 Overall independent sample t-test analysis between the experimental group and control group across the four categories

	Experimental group mean	**Control group mean**	**Mean difference**	***t***	***p***
Fidelity	4.19	3.90	0.29	2.50	.01
Delivery	3.60	3.21	0.39	2.21	.03
Domain knowledge	4.16	3.34	0.82	5.22	<.001
Trust	4.19	3.56	0.63	3.69	<.001

Subsequently, an independent-sample t-test was conducted to compare the high-proficiency foreign language user groups in the experimental and control groups, examining whether there were significant differences in their ratings. The results are presented in Table 4.3：

Table 4.3 Independent-sample t-test analysis of high-proficiency foreign language users in the experimental and control groups

	Experimental group mean	**Control group mean**	**Mean difference**	**t**	***p***
Fidelity	4.09	3.69	0.40	2.31	.03
Delivery	3.27	2.85	0.42	1.83	.07
Domain	3.93	3.04	0.89	3.79	<.001

knowledge					
Trust	3.94	3.20	0.74	3.16	<.01

The results indicate that high-proficiency foreign language users in the experimental group exhibited significant differences in their ratings of fidelity, domain knowledge and trust compared to their counterparts in the control group. Additionally, the differences in linguistic expression were approaching significance. This suggests that even when users possess higher foreign language proficiency, their perception of interpreting quality remains influenced by the interpreter's professional titles. Similar to the overall sample, the mean differences in domain knowledge and trust between the experimental and control groups were particularly pronounced, with high t-values and statistical significance (p<0.001). This indicates that the interpreter's perceived professional titles also reinforced high-proficiency users' recognition of the interpreter's industry competence and trustworthiness. Compared to the overall dataset, the t-values in the high-proficiency user group were generally lower, though they remained statistically significant. To further examine the relationship between users' foreign language proficiency and quality perception bias, a two-way ANOVA was conducted to compare low-proficiency and high-proficiency users across the four dimensions. The analysis results showed no significant interaction effects within groups (p=0.42, 0.81, 0.70 and 0.46), demonstrating that users, regardless of their foreign language proficiency level, were subject to perception bias induced by the interpreter's perceived professional titles.

To verify whether expert ratings were stereotyped by the interpreter's professional titles, we invited professional interpreters as expert representatives to assign scores. All four expert evaluators held a master's degree in translation and interpreting studies, and had passed the national Translation Professional Qualification (CATTI) exam at Level 2 or Level 1 for interpreting, and had at least five years of professional interpreting experience. Two experts watched the video from the experimental group, while the other two watched the video from the control group. They then rated the performance based on the same four dimensions: fidelity, delivery, domain knowledge and trust. A non-parametric independent-sample Wilcoxon ranked test revealed no significant differences between the two groups of expert ratings (p=1.00, 1.00, 0.33 and 0.67), indicating that expert assessments were largely unaffected by the emphasis on the interpreter's professional title. This suggests that the experts adhered to the objective and faithful evaluation principles outlined in interpreting assessment frameworks and upheld their professional responsibilities as evaluators (Lee, 2008). Another explanation is that before conducting the quality evaluation, the experts were explicitly informed that their role was that

of an evaluator rather than an interpreting user. This distinction may have enabled them to step outside the speaker-interpreter-user interaction system, making them less susceptible to the social attributes of the speaker or the interpreter. Nevertheless, while the differences in industry competence ratings between the experimental and control group experts did not reach statistical significance, they showed a more noticeable trend compared to the other three dimensions. This suggests that expert ratings may have been somewhat influenced by the interpreter's professional identity. One possible explanation is that, given that all four evaluators had backgrounds in translation rather than medicine, they relied more on the interpreter's professional identity when assessing industry competence. This could be interpreted as a reflection of respect for the interpreter's expertise within their designated professional field.

Subsequently, we conducted a non-parametric independent-sample Kruskal-Wallis test, comparing the mean expert ratings with the mean scores of the experimental and control groups, and visualized the results using a box plot. The results are presented as follows:

Table 4.4 Comparison of mean expert ratings with experimental and control group score differences

	Mean expert ratings	**Difference from the experimental group**	**Difference from the control group**
Fidelity	4.06	0.03	0.37
Delivery	3.50	0.23	0.65
Domain knowledge	4.06	0.13	1.02
Trust	3.68	0.26	0.48

As shown in Figure 4.2, the ratings from the experimental group were closest to the expert scores, whereas the ratings from the control group deviated more significantly from the expert assessments. This indicates that, within the experiment, when users were not exposed to the interpreter's activated professional titles, their overall rating levels tended to be lower. In contrast, awareness of the interpreter's professional titles led to ratings that aligned more closely with expert evaluations.

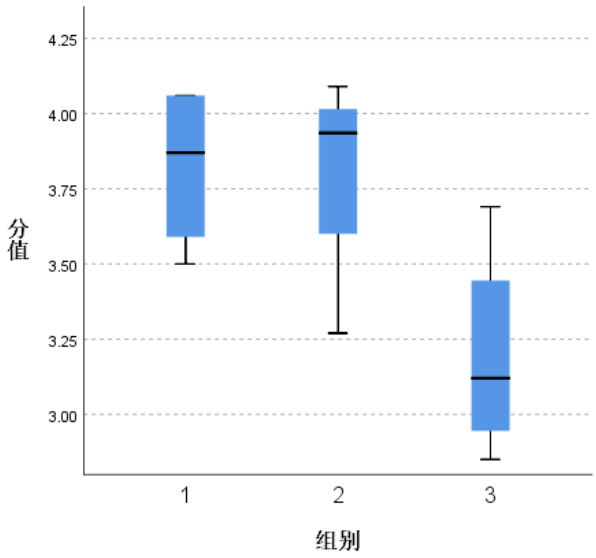


Figure 4.2 Independent-sample Kruskal-Wallis test results of the three group means, group 1 represents expert ratings, group 2 represents the experimental group, and group 3 represents the control group.

4.2.4. Qualitative data analysis

The qualitative data was collected through an open-ended response question, where participants provided descriptors of their overall perception of the interpreter's performance. After data collection, subjective evaluations were first manually categorized into keyword labels, with each participant potentially generating two or more keywords. Subsequently, the keywords from the experimental group and control group were separately imported into the qualitative analysis software Atlas.ti, where word lists were generated. The vocabulary was then categorized and summarized based on fidelity, delivery, domain knowledge and trust. The results are presented in Table 4.5:

Table 4.5 Sorted results of qualitative data

Category	**Experimental group (*n*=55)**	**Control group (*n*=81)**
Praise for comprehensive information	18.2% (10)	12.3% (10)
Doubt about information accuracy	1.8% (1)	30.9% (25)
Praise for fluent expression	7.3% (4)	19.8% (16)
Perceived issues in expression	21.8% (12)	22.2% (18)
Acknowledgement of industry	43.6% (24)	4.9% (4)

competence		
Doubt about industry competence	0% (0)	2.5% (2)
Expressed trust	3.6% (2)	6.2% (5)
Expressed distrust	3.6% (2)	1.2% (1)

As shown in Table 4, in the control group, where professional titles were not perceived as an intervention, a considerable number of participants expressed doubts regarding the interpreter's information fidelity and found the linguistic expression to be suboptimal. However, some participants also praised the interpreter for fluent delivery and comprehensive information fidelity. Notably, only a few users in the control group acknowledged the interpreter's industry competence. In contrast, after perceiving the intervention of the professional titles, most users in the experimental group recognized the interpreter's industry competence, and doubts about information fidelity were significantly reduced. Furthermore, no participants in the experimental group questioned the interpreter's industry competence, aligning with the conclusions drawn from the quantitative analysis. Although the number of participants in the experimental group who praised the fluency of expression and expressed trust in the interpreter was slightly lower than in the control group, we attribute this to a concentration of responses shifting toward the category of acknowledging industry competence. Overall, emphasizing the interpreter's professional identity substantially reduced scepticism while significantly increasing positive evaluations, thereby confirming the positive impact of implicit stereotypes.

4.2.5. Summary

The quantitative data analysis indicates that before recognizing the interpreter's highlighted professional titles, the overall interpreting users in this study rated interpreting quality lower than the expert evaluations. However, when the interpreter's professional titles were made salient, users' perceptions of interpreting quality exhibited a positively biased shift. This not only led to higher ratings in terms of information fidelity and linguistic expression but also, under the influence of stereotypes, significantly enhanced users' recognition of the interpreter's competence within the industry and increased their trust in the interpreter. Compared to low-proficiency interpreting users, high-proficiency users were also influenced by the professional titles in their quality perception bias, though their bias was less pronounced in the dimensions of information fidelity and linguistic expression. Thus, regardless of foreign language proficiency levels, users were consistently subject to perception biases induced by professional titles.

When the interpreter's professional titles were made salient, users' perceived interpreting quality levels became more aligned with expert ratings. Although this phenomenon lacks a definitive explanation from social psychological theory, it may be related to the level of societal respect for professional interpreters (宗刚等, 2016). The four expert evaluators, as fellow interpreters, demonstrated a high degree of respect for the professional status of the interpreter in the experiment, resulting in higher scores. In contrast, the user group's ratings may have been influenced by the socioeconomic status of the interpreting profession. Additionally, the experts' rating range was relatively narrow, which could be attributed to the small sample size but also suggests that expert ratings were more stable and less variable than those of interpreting users, indirectly reflecting the relative consistency in expert evaluations.

The qualitative data analysis further supports these findings. Users who did not recognize the interpreter's professional titles generally expressed scepticism about the interpreter's competence, doubted the interpreter's ability to accurately render the source information, and rarely acknowledged industry expertise. However, when users perceived the interpreter's professional titles, their evaluation of the interpreter's competence shifted significantly. Doubts about the accuracy of the interpretation decreased substantially, while nearly half of the responses focused on affirming the interpreter's domain competence. Thus, the perception bias introduced by professional titles was also evident in the qualitative data. The respect for the interpreter's overall competence in this experiment was driven more by their professional background than by the actual quality of the interpretation itself. This finding supports the concept of implicit evaluation, where the explicit assessment reflects a user's perceived experience of interpreting quality, yet the implicit evaluation is influenced by the interpreter's professional background. Therefore, the tendency and independence of interpreting users' quality perceptions warrant further investigation and validation.

4.3. The formal study

4.3.1. Hypotheses

Before initiating the experiment, we formulated hypotheses for Research Questions 2 and 3. For Research Question 2, "Does an interpreter's perceived social categorical information activates users' stereotypes, thus biasing their perceptions of interpreting quality? If so, in which direction?", this inquiry aims to explore the influence and extent of individual stimuli on quality perception. Since the literature review has already examined the role of gender, age, and professional title-related stereotypes, hypotheses can be proposed for these three variables as follows:

1. Gender: Given the general stereotype that "women have a greater linguistic aptitude" and that the "language service industry is predominantly female," we hypothesize that a perceived female interpreter will receive higher scores than a male interpreter in fidelity and delivery, while no significant difference will be observed in domain knowledge and trust.
2. Age: Due to stereotypes linking aged individuals with cognitive decline and memory deterioration, we hypothesize that a perceived aged interpreter will receive lower ratings across fidelity, delivery, domain knowledge, and trust compared to a younger counterpart.
3. Professional Title: Since the halo effect induced by professional titles is a form of positive stereotype, where a positive impression in one area generalizes to an overall favourable perception, we hypothesize that when users perceive an interpreter's professional title, their ratings across fidelity, delivery, domain knowledge, and trust will be significantly higher.

For Research Question 3, "How do multiple variates (gender, age and professional title) interact with each other, and does one category exert a major influence?", the lack of prior research makes it challenging to formulate theoretically grounded hypotheses. However, given that gender stereotypes and professional titles generally induce positive biases, whereas aged stereotypes are associated with negative biases, we establish a young male interpreter without professional titles as the control group. Based on this, we hypothesize that a female interpreter, regardless of whether she holds a perceived professional title, will be rated more favourably than her male counterpart. Since socioeconomic status has been shown to partially counteract the dominance of race as a single categorical dimension through cognitive reframing (Weeks & Lupfer, 2004), and existing research on socioeconomic status suggests that it exerts a stronger

influence than natural cues such as race (Mattan *et al.*, 2019), we hypothesize that the positive stereotype associated with professional titles will outweigh the negative bias linked to aged interpreters. Consequently, we predict that an aged interpreter without a perceived professional title will be rated lower in rendition quality perception than a younger counterpart without a perceived title. However, once the professional title is perceived, the quality perception of the aged interpreter's rendition will be higher than that of a young male interpreter without a perceived professional title. Finally, in scenarios where both interpreters' professional titles are salient, we hypothesize that the young male interpreter will receive higher ratings across all dimensions of interpreting quality than the aged interpreter.

We designated the young male interpreter without perceivable professional titles as the control group, and the ratings submitted by participants who perceived this social categorical information were encoded as Y-M-NS. Accordingly, the five experimental groups were categorized as Y-M-NS, Y-F-NS, Y-F-WS, O-M-NS, and O-M-WS. Based on the hypotheses above, the primary group relationships and influencing factors of the control group were organized into a conceptual framework, as illustrated in Figure 4.3:

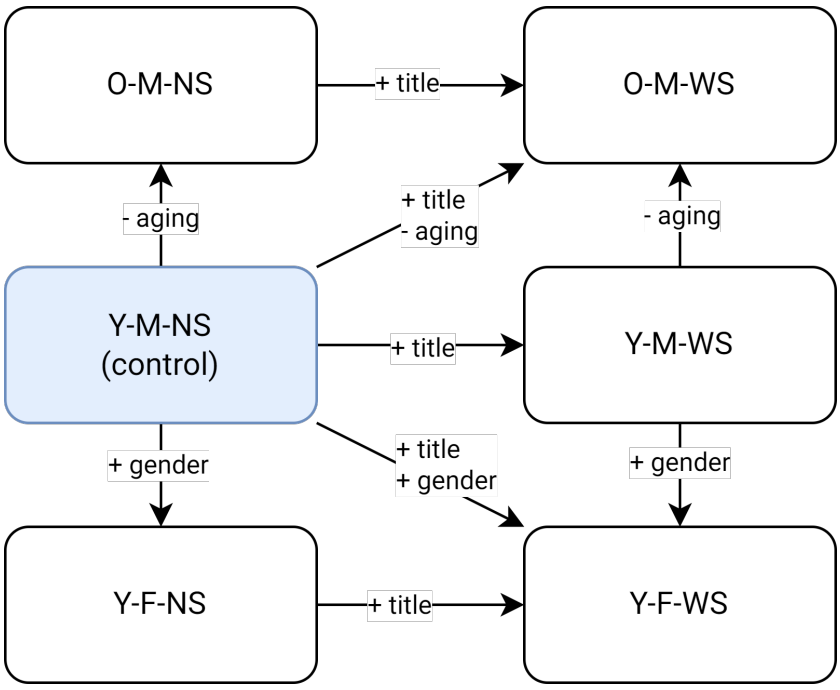


Figure 4.3 The group relations between Y-M-NS as the control group and the other five experimental groups.

4.3.2. Participants

The study recruited a total of 452 participants from a university in Jiangxi Province, Mainland

China. Participation in the experiment was voluntary and facilitated by an administrative staff member. All participants were native Mandarin speakers, with Chinese as their native language and English as their first foreign language. They originated from various provinces and municipalities across China, with a large portion of them were from Jiangxi province. Among the participants who returned valid questionnaires and passed the experimenter's data inspection, 177 were female and 155 were male. No participants self-identified as non-binary. The mean age of the participants was 19.53 years (*SD*=2.52). Except for one group (Y-F-NS) that has not been organized to attend the College English Test Band Four (CET-4) which covers a vocabulary range of 4200 English words, the CET-4 pass rate among the remaining participants was 37.65%. Consequently, the overall English proficiency of the participants can be categorized as moderate. None of the participants self-reported having received formal interpreting training. Detailed information on the effective sample size for each participant group is presented in Table 4.6:

Table 4.6 Effective sample size, mean age, and CET-4 pass rate for each participant group

Group	Effective sample size (total sample size)	Mean age (*SD*)	CET-4 pass rate
Y-M-WS	40 (51)	20.13 (.79)	67.50%
Y-M-NS	54 (95)	20.61 (4.17)	75.93%
Y-F-WS	54 (95)	18.98 (.79)	46.60%
Y-F-NS	59 (66)	18.95 (4.77)	not applicable
O-M-NS	54 (62)	19.91 (1.25)	46.30%
O-M-WS	71 (83)	19.92 (1.09)	42.25%

4.3.3. Procedure

The experiments were organized by the experimenter, with participants assigned to different groups to take part in each session. All groups went through identical experimental procedures, with the only difference being the video stimuli presented to them. The experiments were conducted in the same multimedia classroom, equipped with displays and headsets for each participant.

Before the experiment began, participants were informed that they would be watching a popular medical science video delivered in English but provided with Chinese consecutive interpreting. After viewing the video, they were to take a short comprehension test and a questionnaire. To encourage their attentive participation, they were also told that a cash reward may be granted to those who achieved a certain score on the test and completed a valid

questionnaire. The test consists of ten multiple-choice questions based on the video, all of which were correctly rendered by the interpreter. Each question was measured three times, once directly, once in reverse format and once repeatedly. These questions assessed participants' perceptions of the interpreter's fidelity, delivery, domain knowledge, and trustworthiness. Participants completed both the test and the questionnaire by scanning two QR codes displayed on-screen using their smartphones. All participants returned their questionnaires. After manually cleaning the data to remove biases (such as uniform or patterned responses), 256 valid samples were retained.

4.3.4. Analysis of expert assessment

The interpreter's rendition was first assessed by twelve peer experts. All expert interpreters had been working in the profession for five years or above, and held level-two certification from the China accreditation test for translators and interpreters (CATTI-2). CATTI-2 certification requires interpreters to possess "a relatively comprehensive foundation in foreign language knowledge and interpreting theory, ensuring fluent and accurate translations" (CICG Academy of Translation and Interpretation, 2023). The experts were divided into six groups of two, with each group watching one of the video conditions separately. They then assessed the interpreter's performance in terms of fidelity, delivery, domain knowledge, and their level of trust in the interpreter. The scoring range for the experts (from 0 to 100) was mirrored in the Likert scale used for the user participants.

We then measured if the experts exhibited rating differences in their perceived gender, professional titles and age of the interpreter. For gender, a Mann-Whitney U test indicated experts' ratings on fidelity (U=5.50, p=0.41), delivery (U=2.50, p=0.10), domain knowledge (U=5.00, p=0.34) and trust (U=6.50, p=0.62) not being significantly influenced by perceived gender. For the professional title, apart from a significant enhancement in domain knowledge (U=30.00, p=0.04), experts' assessments were not significantly affected by the perceived title on fidelity (U=16.50, p=0.79), delivery (U=18.00, p=0.77) and trust (U=25.50, p=0.17). While for age, no statistically significant differences were found in either fidelity (U=3.00, p=0.12), delivery (U=3.00, p=0.12), domain knowledge (U=3.00, p=0.13) and trust scores (U=10.00, p=0.13). We present the mean assessment scores of experts are presented in the table below:

Table 4.7 Assessment scores of experts. The experts used a 100-point scoring system, which was subsequently converted by the experimenter to align with the 5-point Likert-type scale used

by the participants.

Categories	Minimum	Maximum	*M*	*SD*
Fidelity	4.00	4.50	4.19	.22
Delivery	3.25	4.00	3.67	.25
Domain knowledge	4.00	4.50	4.25	.21
Trust	3.50	4.00	3.69	.16

4.3.5. Descriptive analysis of users' perception results

To examine the effects of the interpreter's perceived social categorical information (gender, age and professional title) on users, we first employed an interaction plot to present the lifting effect of perceived socioeconomic status across genders, as shown in Figure 4.4:

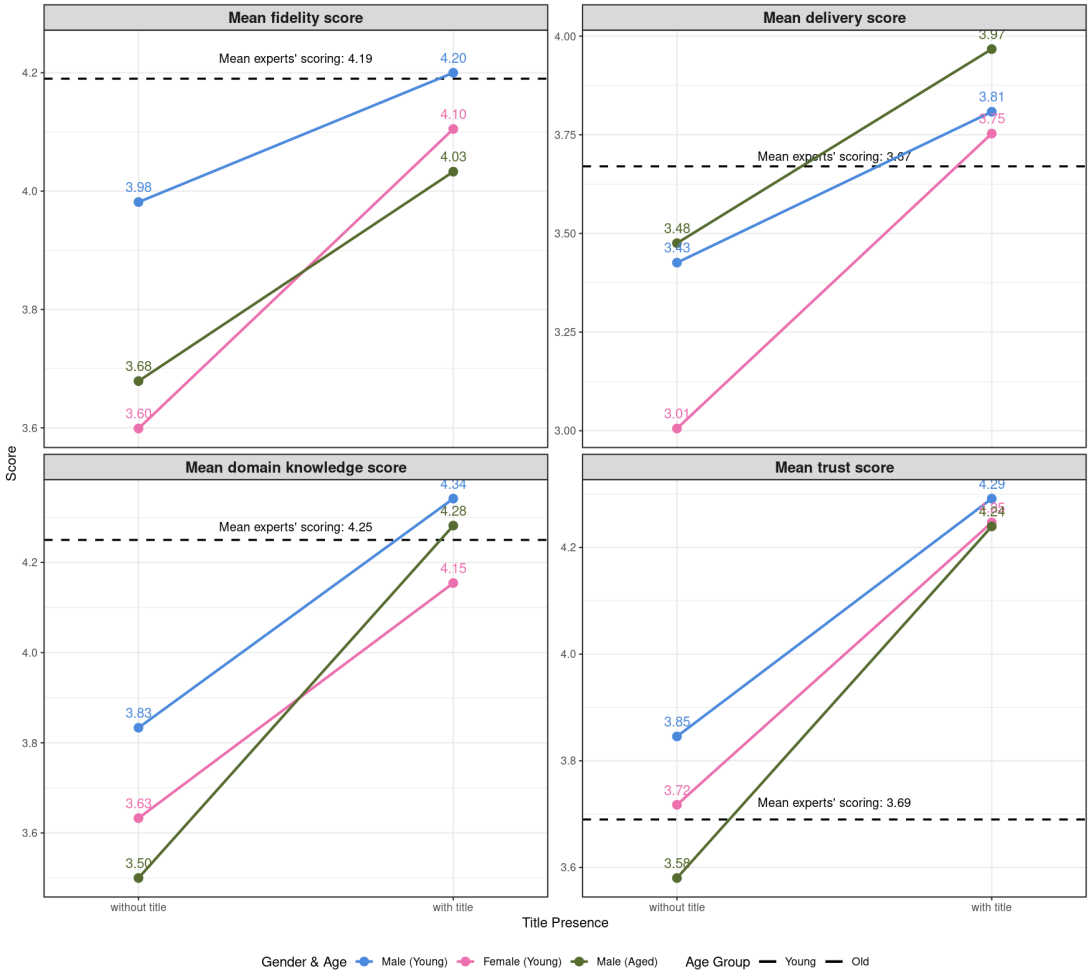


Figure 4.4 An interaction plot of mean scores by conditions.

(The solid lines represent the changes in users' quality perception ratings, from without a perceived professional title to before and after the presentation of the professional title. The blue solid line represents the young male interpreter, the pink solid line represents the young

female interpreter, the green solid line represents the aged male interpreter, and the black horizontal dashed line represents the average experts' quality assessment scores.)

Although the results reflected in the interaction plots do not show statistical significance, this interaction plot has indicated that the presentation of social categorical information has a significant impact on users' quality perceptions across all four dimensions. For fidelity, users' ratings for the interpreter without perceiving professional titles were generally lower than the experts' scores. Among the users' ratings, the young female interpreter received slightly lower scores than the aged male interpreter, and significantly lower scores than the young male interpreter. After the activation of professional titles, the ratings for all three interpreters improved. The young male interpreter reached the level of the expert scores, while the young female interpreter and the aged male interpreter were slightly below the expert ratings.

In terms of delivery, users' ratings for the interpreter without perceiving professional titles were generally lower than the experts' scores. Among the ratings, the aged male interpreter scored slightly higher than the young male interpreter, but both were considerably higher than the young female interpreter. Similarly, after the activation of professional titles, the ratings for all three interpreters improved and exceeded the expert ratings. The aged male interpreter received the highest score, while the young male interpreter and the young female interpreter had relatively closer scores.

In terms of domain knowledge, users' ratings for the interpreter without perceiving professional titles were generally lower than the experts' scores. Among the ratings, the elderly male interpreter received the lowest score, while the young male interpreter had the relatively highest score. After the activation of professional titles, the scores for all three interpreters improved. Notably, both perceived male interpreters surpassed the expert assessment scores, while the perceived female interpreter remained slightly below the expert ratings.

Finally, in terms of trust, users' ratings for the interpreter without perceiving professional titles were generally lower than expert scores, except for the young interpreters, whose ratings were higher than those of the experts. However, after the activation of professional titles, users' ratings for all categories significantly increased and were far higher than the expert assessments.

4.3.6. Significance analysis of users' perception results

Before the analysis, we first assessed the reliability and validity of users' responses. In terms of reliability, we conducted an overall measurement of the 12 quality perception ratings (i.e., three ratings for each of the four categories) and found that the reliability was good (Cronbach's α

=0.90). In terms of validity, we conducted KMO and Bartlett's test, with a KMO value of 0.90, and Bartlett's sphericity test indicating good validity (KMO=0.71, $\chi^2(66)=1992.16$, $p<0.001$).

Since the data in this study is a 4×3 matrix (i.e., four quality categories and three independent variables), we used generalized linear models in RStudio to test the significance of the data. Although a linear mixed model could measure random effects more accurately, it may be a highly uncertain model, and improper use may lead to an inflated risk of false positives. Psychologists have expressed concerns about the excessive use of linear mixed models and how models are reported (Meteyard & Davies, 2020). Additionally, since this study was not intended to explore too many random effects, we opted for the relatively more robust generalized linear model. We set the fixed effects to be groups and participants' genders. The formula is: *outcome ~ group + participant_gender*. The inclusion of participants' gender was motivated by the findings of Iglesias (2013), who found a gender bias in quality perception: female users tend to have higher expectations for delivery, whereas male users show greater sensitivity to domain knowledge. The results are summarized in Table 4.8：

Table 4.8 Summary of users' stereotypical perceptions of interpreting quality across conditions

Stimulus	**Pairs**	**Category**	***B***	***SE***	***t***	***p***
Gender	Y-M-NS vs Y-F-NS	fidelity	0.38	0.11	3.35	0.001
		delivery	0.43	0.16	2.76	0.007
		domain knowledge	0.22	0.13	1.71	0.09
		trust	0.14	0.14	0.97	0.33
	Y-M-WS vs Y-F-WS	fidelity	0.05	0.13	0.40	0.69
		delivery	-0.01	0.15	-0.03	0.97
		domain knowledge	0.19	0.14	1.41	0.16
		trust	0.03	0.13	0.20	0.84
Aging	Y-M-WS vs O-M-WS	fidelity	0.00	0.23	-0.02	0.98
		delivery	-0.20	0.26	-0.77	0.44
		domain knowledge	-0.09	0.24	-0.38	0.70
		trust	0.08	0.24	0.34	0.73
	Y-M-NS vs O-M-NS	fidelity	0.38	0.17	2.23	0.03
		delivery	0.14	0.19	0.74	0.46
		domain knowledge	0.41	0.19	2.14	0.03
		trust	0.30	0.19	1.57	0.12

Title	Y-M-NS vs Y-M-WS	fidelity	-0.22	0.13	-1.67	0.10
		delivery	-0.43	0.16	-2.71	0.008
		domain knowledge	-0.56	0.15	-3.69	0.001
		trust	-0.47	0.15	-3.16	0.002
	O-M-NS vs O-M-WS	fidelity	-0.35	0.13	-2.74	0.007
		delivery	-0.49	0.15	-3.40	0.001
		domain knowledge	-0.78	0.14	-5.76	0.001
		trust	-0.66	0.14	-4.68	0.001
	Y-F-WS vs Y-F-NS	fidelity	0.51	0.10	4.91	0.001
		delivery	0.75	0.14	5.18	0.001
		domain knowledge	0.52	0.11	4.69	0.001
		trust	0.53	0.13	4.06	0.001
Aging & gender	Y-F-WS vs O-M-WS	fidelity	0.05	0.13	0.43	0.67
		delivery	-0.24	0.15	-1.55	0.12
		domain knowledge	-0.14	0.14	-1.01	0.31
		trust	0.04	0.14	0.31	0.76
	Y-F-NS vs O-M-NS	fidelity	-0.11	0.15	-0.73	0.47
		delivery	-0.44	0.18	-2.43	0.02
		domain knowledge	0.14	0.15	0.94	0.35
		trust	0.13	0.17	0.73	0.47
Aging & title	Y-M-WS vs O-M-NS	fidelity	0.46	0.28	1.63	0.11
		delivery	0.57	0.30	1.89	0.06
		domain knowledge	0.70	0.28	2.52	0.01
		trust	0.46	0.29	1.58	0.12
	Y-M-NS vs O-M-WS	fidelity	-0.02	0.15	-0.16	0.88
		delivery	-0.44	0.18	-2.47	0.01
		domain knowledge	-0.40	0.17	-2.34	0.02
		trust	-0.30	0.17	-1.79	0.08
Gender & title	Y-M-WS vs Y-F-NS	fidelity	0.52	0.13	3.92	0.001
		delivery	0.75	0.18	4.28	0.001
		domain knowledge	0.72	0.13	5.59	0.001
		trust	0.56	0.16	3.47	0.001
	Y-M-NS vs Y-F-WS	fidelity	-0.12	0.11	-1.09	0.28

		delivery	-0.32	0.14	-2.28	0.03
		domain knowledge	-0.30	0.14	-2.26	0.03
		trust	-0.39	0.13	-3.04	0.003
Aging, gender & title	Y-F-WS vs O-M-NS	fidelity	0.43	0.15	3.00	0.003
		delivery	0.31	0.17	1.85	0.07
		domain knowledge	0.65	0.16	4.18	0.001
		trust	0.65	0.16	4.08	0.001
	Y-F-NS vs O-M-WS	fidelity	-0.48	0.13	-3.84	0.001
		delivery	-0.98	0.16	-5.98	0.001
		domain knowledge	-0.65	0.13	-4.92	0.001
		trust	-0.49	0.15	-3.24	0.002

We first conducted analyses for each of the three individual factors: gender, ageing, and professional title. In terms of gender, male interpreters without professional titles scored higher than their female counterparts in fidelity, delivery, and trust, although no significant difference was found in domain knowledge. This suggests that, at least in this study, users perceive male interpreters as more competent than female interpreters. However, when professional titles were presented, there were no significant differences between male and female interpreters across all categories, indicating that the stereotype associated with professional titles had a greater influence than gender-based stereotypes.

In terms of ageing, the aged male interpreter without professional titles was perceived to have lower fidelity than younger interpreters, although no difference was found in delivery. This suggests that the stereotype related to cognitive decline and reduced memory capabilities associated with ageing was reflected in users' perceptions of fidelity, but not in delivery. Similarly, in domain knowledge, aged interpreters scored lower than both male and female younger interpreters, indicating that users perceived aged interpreters as less capable of handling industry-specific terminology, implying that they were seen as unsuitable for technical interpretation. However, once professional titles were activated, the stereotype associated with ageing was mitigated, and there were no statistically significant differences in users' perceptions of interpreters across all categories.

In terms of professional titles, male interpreters who were presented with titles saw significant improvements in delivery, domain knowledge, and trust, with fidelity also showing a marginal statistical significance. This marginal effect in fidelity can be attributed to the fact that male interpreters, even without titles, received the highest fidelity scores. Similarly, both

female interpreters and aged interpreters also experienced significant improvements across all dimensions when professional titles were presented. This underscores the substantial positive impact of professional titles in enhancing users' perceptions of interpreters across all perceived categories.

The subsequent two-way analysis allows us to explore the relative dominance of each factor in combination. In the case of ageing and titles, aged male interpreters with professional titles received more positive quality perceptions in delivery and domain knowledge compared to young interpreters without titles, as well as stronger trust ratings from users. This suggests that professional titles have a greater influence than age. Similarly, when considering the interaction of titles and gender, young female interpreters with titles showed significant differences in delivery, domain knowledge, and trust compared to their male counterparts without titles, also closing the gap in fidelity. This indicates that professional titles have a stronger impact than gender. Given that the influence of professional titles outweighs that of gender and ageing, when both ageing and gender are considered as interacting factors, the perceived aged male interpreter receives significantly higher user ratings in delivery compared to his young female counterpart, with no significant differences across the other dimensions. This suggests that the influence of gender is slightly stronger than age. Thus, the overall ranking of influence is as follows: professional titles > gender > age.

Therefore, the results of the three-way interaction analysis, which combined all categorical information, align with expectations. The aged male interpreter with a professional title received significantly higher quality perceptions across all categories than a perceived young female interpreter. However, once the professional title is attributed to the young female interpreter, users' quality perception of her significantly surpasses that of the aged male interpreter.

4.4. Discussion of the formal study

This study was designed to investigate how an interpreter's perceived social categorical cues, such as age, gender and socioeconomic status reflected through professional titles subtly affect users' perceptions of interpreting quality. We carefully manipulated the interpreter's perceived gender and socioeconomic status in the experimental stimuli, while keeping all other linguistic and paralinguistic features consistent. We hypothesized that users' quality ratings would be skewed by these perceptions and found that their rating patterns were indeed influenced by these identities, independent of linguistic factors.

4.4.1. Severity or leniency in expert assessors' ratings of interpreter performance

The interpreter's performance was first evaluated by twelve peer experts, as we aimed to establish a baseline reference for the interpreter's performance from an expert perspective. This also allows us to examine whether expert assessments are influenced by social categorical indicators. Notably, although the expert assessors had undergone systematic interpreter training and were well-versed in the interpretation scoring guidelines, they were also bound by the interpreters' code of ethics (see Chapter 2 for detailed elaborations). However, this does not guarantee that all expert assessors would necessarily maintain a neutral and objective stance in their ratings. In fact, raters can sometimes exhibit a consistently biased attitude, either being excessively harsh or lenient toward particular groups of examinees. In turn, it may result in a higher degree of severity for some responses and leniency for others (Han, 2015). Fortunately, the expert raters in this study adhered to the code of ethics, which emphasizes that interpreters should remain impartial and neutral (Setton & Dawrant, 2016, p. 152), no significant differences in expert ratings were found across the three experimental groups and the control group in fidelity and delivery. This suggests that the experts adhered to the principles of objective assessment and upheld the professional norms expected of assessors (AIIC, 1999), at least in the present study. Their trust scores also did not increase when assessing M-WS or F-WS, remaining at a consistent level. This indicates that experts were not easily swayed by the positive bias often associated with high socioeconomic status (Li *et al.*, 2023). The trust-related decisions of experts remained unchanged, markedly contrasting with the behaviour of user participants. Unlike the user participants, the experts' trust-related decisions remained unaffected. However, their scores for domain knowledge were influenced by the perceived status of the interpreter, which may be explained by the fact that the experts came from a language background rather than a medical one, leading them to place greater respect on the medical interpreter's expertise in his specialized field.

4.4.2. Gender stereotype on quality perception

We hypothesized that the rendition of the perceived female interpreter would be rated more favourably than that of her male counterpart. Contrary to our initial expectations, presenting a female face led to a reduction, rather than enhancement in users' ratings of fidelity and delivery. This suggests a user bias favouring male over female interpreters in terms of perceived competence. This finding is somewhat counterintuitive, given the common stereotype that associates women with greater linguistic competence. We offer two possible explanations: first, the speech video used in this experiment had a strong disciplinary focus on medical science, a

field stereotypically viewed as masculine (Wilbourn & Kee, 2010). The tendency for a "masculine default" in gender role perceptions may have led users to perceive men as the normative standard, a bias further reinforced by the original speaker's identity as a male cardiologist. Consequently, the perceived female interpreter may have suffered from a lack of gender congruence with the male speaker, compounded by the "technical" stereotypes typically associated with men (Rice & Barth, 2017), which could have negatively influenced users' quality ratings. Second, while traditional gender roles often deem females more suitable for language service roles, these stereotypes are largely based on superficial perceptions such as pleasant vocal tone (Iglesias, 2013) and beautiful professional image (Du, 2021; Hong & Choi, 2020). According to the stereotype fidelity model (Fiske *et al.*, 2002), stereotypes operate on two dimensions: competence (perceived capabilities and skills) and warmth (perceived friendliness and kindness). Since the traditional gender-congruent stereotype assumes females are "warmer" but less "competent" than males (de Lemus *et al.*, 2014), such traits of a typical interpreter are more related to warmth instead of competence. Since most participants in this study obtained insufficient English proficiency to comprehend the original English speech, they relied heavily on the interpreter. As they sought to complete a comprehension test with the assistance of the interpreter's knowledge, it may result in their greater need for the interpreter's competence (domain knowledge) over warmth (communication).

4.4.3. Professional titles mitigate the negative stereotypes of natural cues

In line with our expectations, an interpreter's perceived titles enhanced users' quality perceptions. When comparing Y-M-WS with Y-M-NS and Y-F-WS with Y-F-NS, such enhancement was evident not only in domain knowledge and trust, but also in fidelity and delivery. The facilitative effect of positive stereotypes, or the halo effect, was thus consistent across rating indicators. With the boost from perceived titles, users' ratings are closely aligned with the experts' assessments of fidelity, delivery, and domain knowledge. However, in the absence of title priming, the interpreter's performance tended to be undervalued, revealing a tendency among users to underestimate the work of full-time interpreters. This reveals an underlying tendency among users to undervalue the effort of full-time interpreters who work without the protection of titles. Negative judgments from users about interpreting quality could reinforce conference organizers' biases (Cheung, 2020), potentially resulting in missed opportunities for interpreters in domain-specific assignments due to status-driven stereotypes.

While experts did not exhibit a significant increase in trust under the Y-M-WS and Y-F-WS conditions, users demonstrated a considerable rise in trust scores. This indicates that an

interpreter's perceived titles, reflecting high socioeconomic status, effectively enhanced users' trust. Trust plays a crucial role in the perception of interpreting quality, as it ensures a steady comprehension of the interpretation (Hsieh *et al.*, 2010), while distrust may generate negative attitudes towards the interpreter, potentially undermining the acceptance of renditions (Wilson & Walsh, 2019). In this study, we may explain the interpreter's perceived socioeconomic status as a facilitator of trust, as supported by past trust-game studies (Blue *et al.*, 2020; Cheng X *et al.*, 2022; Qi *et al.*, 2018). Although perceived gender differences led to stereotyped ratings in fidelity and fluency, these gender-driven stereotypes were effectively mitigated by the perceived socioeconomic status. No statistical significances were found between M-WS and F-WS in domain knowledge and trust, suggesting that title may have a mitigating effect on gender-related biases. Given the difficulty of listening comprehension, users were largely deprived of the ability to comprehend the original speech. This dynamic creates what is known as "coercive trust" (Robb & Greenhalgh, 2006) where users, having no alternative, are forced to place unwarranted trust in the interpreter. In such a passive trust relationship, users are prone to challenge the interpreter's abilities, especially if they feel potential discrepancies between the rendition and the original speech (Robb & Greenhalgh, 2006; Wilson & Walsh, 2019). This might explain why users assigned significantly lower scores to the interpreter without titles compared to experts' assessments across the dimensions of fidelity, delivery, domain knowledge, and trust. However, when an interpreter is granted notable titles, users may perceive them as endorsements of their professional capacities by the institution they present, transforming their trust from coercive to voluntary (Robb & Greenhalgh, 2006). This may also explain why users' trust scores for an ad hoc interpreter were significantly higher than those of experts. In such settings, an interpreter's perceived title can be an effective tool for building trust. If this positive stereotype leads to enhanced perceptions of interpreting quality, it might not necessarily be detrimental – unlike innate attributes, title can be acquired. Similar to high socioeconomic status or company endorsements helping black clients reduce bias and increase loan approval rates (Scott *et al.*, 2023), interpreters specializing in domain-specific topics may gain user and client trust by obtaining higher academic degrees or industry accreditations, thereby alleviating the negative effects of coercive trust.

4.4.4. Ageing stereotype on quality perception

Our hypothesis was that when users are presented with an aged male interpreter without professional titles, the ratings in the four quality-related categories would be significantly lower compared to those for a young male interpreter. Even with the activation of professional titles,

the aged interpreter would still not have an advantage in any of the categories. Some of our hypotheses were validated, while other research outcomes did not align with our expectations.

First, according to the results from the generalized linear model, aged interpreters without professional titles had significantly lower scores than their younger counterparts in fidelity and domain knowledge, but there was no significant difference in delivery. Although there was a mean difference in trust scores, the significance was marginal and not statistically significant. Given that there were no differences in content and paralinguistic features across all versions of the renditions, the users' stereotypes about aged interpreters seem to focus primarily on their perceived deficiencies in content-related information completeness and accuracy, as well as their lesser command of domain-specific terminology compared to younger interpreters. This stereotype may stem from the belief that older individuals are not ideal candidates for interpreting roles. As mentioned in subsection 2.3.4, ageing stereotypes are often associated with ageing, sluggishness, and cognitive decline (佐斌, 2015; 佐斌等, 2007). However, interpreting is a fast-paced and demanding profession, and this intense, rapid pace contrasts with the traditional characteristics associated with ageing stereotypes. As a result, users may subconsciously harbour biases that aged interpreters are prone to forgetfulness, thereby lowering the ratings in fidelity. Additionally, the slower adaptation of the elderly to technological advancements, a stereotype reinforced through media exposure, may further contribute to the lower ratings in domain knowledge, with users perceiving aged interpreters as less capable of accurately rendering medical terminology in the video.

This age-related stereotype in interpreting does not align with reality. For example, due to their cognitive training, interpreters generally possess stronger cognitive functions than non-trained bilinguals (Boos *et al.*, 2022), including better abilities to manipulate information in working memory and process or store sub-lexical phonological representations (Signorelli *et al.*, 2011). This cognitive training can effectively delay the ageing of the brain. Research has shown that the number of years of simultaneous interpreting experience enhances the cognitive function of selective attention during interpretation (Yagura *et al.*, 2021), and experience also affects the quality of the work produced by experienced interpreters (Tiselius & Jenset, 2011). Although ageing inevitably leads to cognitive decline (Liu *et al.*, 2020), the cognitive functions of older interpreters do not experience a sharp decline compared to younger interpreters; rather, they are only slightly less efficient, as younger interpreters were found to be marginally better than older interpreters in nonword repetition and cued recall (Signorelli *et al.*, 2011). On the other hand, in terms of terminology and other domain knowledge, the stereotype is contrary to the reality. Zdravkovska (2019) found that the largest percentage of interpreters aged 26 to 30

indicated that nomenclature was their biggest challenge, while respondents from other age groups reported pre-existing knowledge as their main challenge. Therefore, older interpreters, with their years of experience, tend to have a superior command of terminology compared to their younger counterparts.

4.5. Summary

Genuine conference users are the most direct recipients of interpreting services, they consist of an important component of interpreting quality perception. Unlike expert assessors who adhere to established criteria in quality assessment, users can be influenced by an interpreter's perceivable paralinguistic and non-linguistic features, as demonstrated by past empirical studies. Based on the previous experimental outcomes of Li & Cui (2022) and Cheung (2013, 2020), this study further explored the interaction between an interpreter's perceived title and gender. We first examined expert assessors' ratings of interpreting quality when they perceive the interpreter of different genders and possession of titles. We found that the experts maintained appropriate scoring standards, with no significant influence from these factors. However, users showed a reduction in quality ratings for fidelity and delivery when the interpreter was perceived to be female, while their ratings in domain knowledge and trust remained relatively stable. This suggests that female interpreters in technical conference settings are stereotyped as less competent than their male counterparts. Furthermore, we observed a significant improvement in perceived quality due to the interpreter's title, which not only enhanced users' overall quality ratings but also helped to moderate the discrepancies caused by gender. In particular, the title reduced rating gaps in domain knowledge and trust. We interpret this as a halo effect, where the interpreter's title establishes greater trust among users, shifting their acceptance of the renditions from coercive to voluntary.

As it is often less practical to invite experts to assess interpreting quality in conference settings, proposing a model for rectifying users' quality ratings may serve as a viable method to restore an interpreter's actual performance. Future research could continue to adopt similar experimental settings to investigate an interpreter's other perceivable traits, such as age, comparing older versus younger interpreters. Cross-analysis of multiple variables could then be conducted to identify the non-linguistic features that most significantly influence users' perceptions of quality, enabling the development of a more refined model for interpreting quality assessment.

Chapter 5 A correction model for users' stereotyped interpreting quality perceptions

Drawing on social psychologists Freeman and Ambady's (2011) dynamic interactive model of person construal, we have identified a theoretical framework that elucidates why users tend to assign lower fidelity and domain knowledge ratings to an aged male interpreter. This chapter aims to answer the last research question, i.e., "Is it feasible to develop a correction model to modify the average score of users' quality ratings?" Building on this perspective, we propose an interactive model of interpreting quality perception (iMQP) to illustrate how social and cognitive processes shape users' evaluative judgments. In particular, the author suggests two possible methods for mitigating biased quality perceptions: (1) pre-rating calibration through introducing higher-level cues (e.g., professional titles, contextual reminders) before the rating process, and (2) post-rating correction informed by the interpreter's perceived social categorical attributes once the ratings have been collected.

5.1. The dynamic interactive model of person construal explained

The dynamic interactive theory of person construal proposed by Freeman & Ambady (2011) posits that person construal is a process driven by a multi-level dynamical system. This system integrates low-level sensory cues (e.g., facial features, vocal characteristics, bodily attributes) with high-level social cognition (e.g., stereotypes, motivations, task demands) through continuous interactions and dynamic competition to generate stable person perceptions. It emphasizes that person perception is not a static categorization process but rather a temporally evolving phenomenon characterized by probabilistic activation of parallel representations. See Figure 5.1:

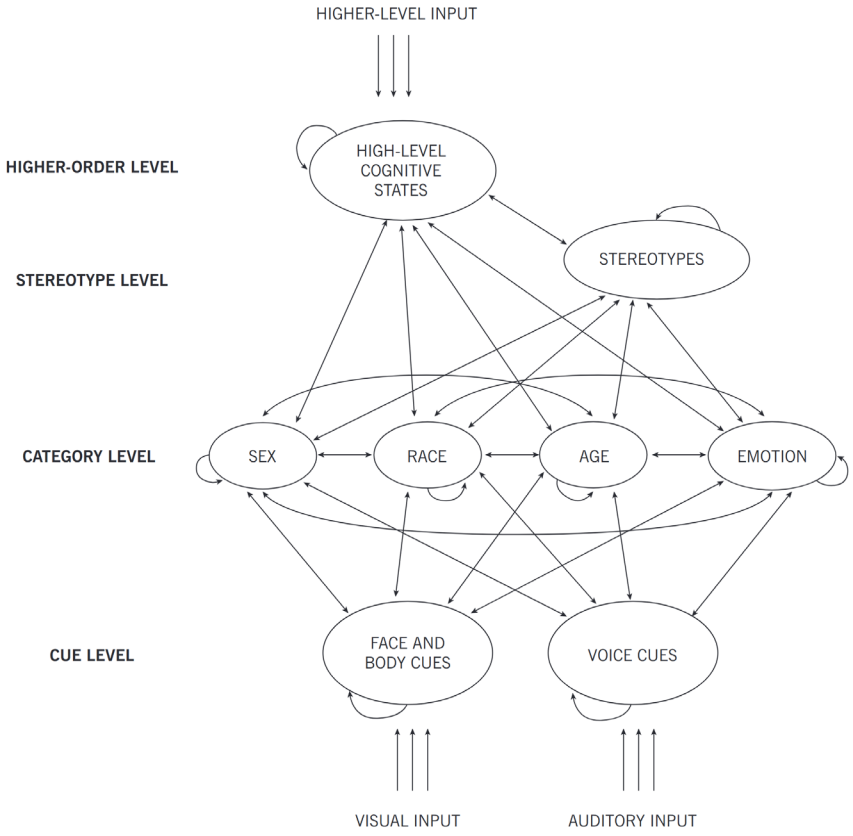


Figure 5.1 A general diagram of the dynamic interactive model of person construal from Freeman and Ambady (2011).

The dynamic interactive theory of person construal states that person perception is not a discrete categorization process, but a rather stable state formed gradually through the continuous interaction and competition of multi-level information. This model consists of four major levels: the cue level, the category level, the stereotype level, and the higher-order level. These levels are interconnected bidirectionally, and thus form a dynamic feedback loop that enables the real-time integration of perceptual cues, social categories, cultural knowledge, and individual motivations.

The cue level is the most basic level, it is responsible for receiving and processing raw input information from the sensory system, it consists of two independent processing pools: the face/body cues pool for visual cues and the voice cues pool (auditory cues). The visual system detects physical traits such as gender and race (such as the darker skin tones for the Blacks) through the face/body cues pool, while the auditory system analyses acoustic parameters like pitch and formants through the voice cues pool. Each pool contains nodes that represent specific

sensory feature detectors, such as masculine facial features from perceived square chin and thick eyebrows, or low-pitched voice tone as detected by the hearing system. These nodes interact bidirectionally with the category level, as they send signals to compatible category nodes. Take an example, masculine cues activate the male category. Nodes within the same dimension, such as masculine cues and feminine cues compete through mutual inhibition to ensure that the system does not accept contradictory interpretations of ambiguous inputs. Freeman & Ambady (2011) used an example, both the masculine cues and feminine cues nodes may briefly coexist and activate when processing a gender-ambiguous face, but they will ultimately converge on a single dominant interpretation through competition. In addition, the design of the cue level allows for cross-modal integration. For example, visual cues such as masculine facial features combined with auditory cues like a deep voice synergistically enhance the activation of the masculine category, and reflect the combined effect of multi-sensory cues through social interactions.

At the second layer of the interactive model, the category level roles in transforming lower-level sensory cues into social category representations, such as gender, race and age. This level serves as the core of the dynamic interactive theory of person construal, and transforms the crucial function of integrating sensory cues into social category representations. It consists of multiple competitive pools, each corresponding to an independent social categorization dimension. Nodes within each pool represent mutually exclusive category options, such as male instead of female, Mongoloid instead of Caucasian. These nodes are connected by mutual inhibition, which forms a dynamic competition mechanism that ensures only one category within a dimension dominates during a given period. For example, in a gender categorization task, the “male” and “female” nodes overlap with each other, causing a gender-neutral face to converge to a stable state. The dynamic operation of the category level also relies on bidirectional regulation from both top-down and bottom-up, where higher-order cognitive states enhance the excitability of target dimension nodes, and the continuous input from lower-level cues adjusts the intensity of category activation. As noted by Freeman & Ambady (2011, p. 256), this serves as the initial activation entry point for social stereotypes, and shapes the cognitive groundwork from local to global representations.

Through the literature reviews in Chapters 2 and 3 of this dissertation, we should already be familiar with the stereotype level. This level is the core layer in the dynamic interactive theory of person construal responsible for integrating social knowledge. Its function is to dynamically associate social categories activated at the category level with stereotype content. For example, the stereotype of men being more capable at mathematics is represented at this

level as the understanding that men are more born suited for mathematics subjects other than females. The level is composed of a global pool of stereotype nodes, with each node representing a stereotype associated with a specific social group or trait. Also, the connections between these nodes reflect cultural experiences of stereotypes. For instance, the activation of the "good at mathematics" stereotype would activate descriptors such as "logical" and "rational", while excluding descriptors like "free-minded" and "sentimental". Furthermore, this stereotype level supports cross-category semantic integration: when categories from different dimensions share stereotypes, their co-activation creates semantic resonance across dimensions, which leads to indirect reinforcement between categories. An example of this is the stereotype of the "geeky man" (理工男). Thus, it not only explains the automatic activation of stereotypes (Devine, 1989) but also reveals their proactive construction in real-time perception. It reveals the active constructive role of categories in real-time perception, and reflects that stereotypes are not static knowledge repositories. Instead, they are formed through the integration of multiple sources of input, including categories, sensory cues, and higher-order cognitive states, to generate social judgments. Thus, this highlights the cultural embeddedness and cognitive efficiency of the person construal system.

The final and highest level of the dynamic interactive theory of person construal is the higher-order level which serves as the central regulatory hub of the model. This level is responsible for integrating high-level cognitive states such as task demands, motivation, biases, cultural knowledge, and situational goals. After stereotypes are formed, the higher-order cognitive level plays a crucial role in selectively activating and regulating stereotypes based on the context. Through its nodes such as gender stereotypes and task demands, the higher-order cognitive layer dynamically adjusts the applicability and intensity of stereotypes, enabling them to manifest flexibly in different situations. Take an example, when the "hostility" node in the stereotype layer is activated by the category layer, it facilitates stereotypical nodes in the higher-order cognitive layer being activated through increased weighting, thus they jointly formulated a positive feedback loop between bias and stereotypes. Therefore, when processing ambiguous angry faces related to race, individuals with higher levels of stereotypes will find that their node of racial stereotypes not only directly strengthens the activation of the Blacks, but also amplifies their perceptions of hostility on the stereotypical layer. In turn, even without any direct evidence in the image itself, the bias toward categorizing the face as Black will be activated.

In contrast, individuals with lower levels of stereotyping also experience stereotype-driven bias and can actively suppress the extent of stereotype activation through intrinsic reduction and task-demand-driven dual mechanisms. As for the intrinsic reduction, such nodes can be

regulated by societal values and personal long-term social experiences, which significantly lower its baseline activation level compared to high-bias individuals. This leads to limited bottom-up activation of related nodes in the stereotype layer. For example, when individuals are asked to assess someone's competence for a particular position, it may directly suppress task-irrelevant nodes (such as racial stereotypes) in the stereotype layer while enhancing task-related semantic nodes (such as task competence). It thus helps to redirect cognitive resources to the goal-relevant dimensions (Pauker *et al.*, 2009). This active suppression mechanism is more effective in low-bias groups than in high-bias groups, as the baseline activation of stereotypes is already weaker in the former, and task demands can more thoroughly block residual bias associations. Furthermore, the regulatory mechanism is not limited to inhibition but may also selectively enhance specific stereotypes. For example, in cooperative contexts, the node of trust motivation may activate the "Asian-diligent" stereotype, thereby promoting positive interaction judgments. Freeman and Ambady's experimental simulations demonstrated that the higher-order cognitive layer can dynamically adjust the connection weights according to task instructions, so that the manifestation of stereotypes is consistent with the long-term cultural shaping of stereotypes while also adapting to short-term goal demands.

Finally, we should clarify the distinction between higher-level input and the visual and auditory cues at the cue level. Sensory inputs which are directly derived from external physical stimuli, they are bottom-up in nature to undergo primary processing. Visual input such as facial and body features activates the face/body cues node at the cue level, while auditory input, such as pitch, and resonance activates the voice cues node. For instance, masculine facial features (visual) or a low-pitched tone (auditory) each trigger the corresponding cue nodes. In contrast, higher-level input is top-down cognitive input based on endogenous goals, long-term knowledge, or situational demands. In Freeman & Ambady's original work, the concept of high-status (p. 271) is mentioned, which is not derived from natural cues observed through sight or sound but is stored as a social role node within the higher-order cognitive layer. This type of input influences the category layer (e.g., by enhancing the professionalism stereotype) or modulates the stereotype layer (e.g., by linking medical doctors with authority). Hence, professional titles do not rely on immediate sensory input but function as intrinsic knowledge participating in dynamic constraint satisfaction.

5.2. Towards an interactive quality perception model in interpreting

Based on the dynamic interactive theory of person construal by Freeman & Ambady (2011), it is thus possible to interpret the experimental outcomes presented in Chapter 4 from a rather

theoretical construct. We first utilize Freeman and Ambady's model to explain the differentiated ratings of interpreting quality perceptions for the same rendition. After that, we propose a dynamic interactive model that is applicable to interpreting quality perception.

5.2.1. Explaining users' stereotypical quality ratings through the dynamic interactive model

To explain the effect of stereotypes on users' interpreting quality perception, we refer to the case of an aged male interpreter without a disclosed professional title, as one of the experimental stimuli in Chapter 4. By manipulating visual cues through the presentation of different photographs to user participants (see Figure 4.1) and introducing auditory cues via rendition audio which remained consistent across all conditions except for voice tone, it was observed that the perceived aged interpreter received significantly lower ratings in fidelity and domain knowledge compared to a young male interpreter without an activated professional title. This subsection employs Freeman & Ambady's (2011) theoretical framework to elucidate the underlying cognitive mechanisms driving these differential rating patterns.

At the cue level, the detection and integration of sensory input provide the foundation for the activation of social categories. The distinction between aged and young male interpreters at this level primarily manifests in the physical properties of sensory cues. In terms of visual input, the aged interpreter's photograph underwent age progression, as forehead wrinkles, grey hair, facial spots, and nasolabial folds cues characteristic of an elderly individual. Simultaneously, the rendition audio had a deeper and more resonant tone. These features collectively activated the "ageing-related features" node within the cue level. In contrast, user participants in the control group were exposed to cues associated with youth, such as smooth skin, dark hair, and a bright, clear vocal tone, thereby activating the "youth-related features" node. Given that cue nodes within the same social dimension are connected through mutual inhibition, the activation of the ageing-related features node in the aged interpreter's profile directly enhanced the activation of the elderly category node at the category level, while simultaneously suppressing the young category node.

Once these cues are transmitted to the category level, they undergo multidimensional competition and integration, transforming sensory input into stable social category representations. For the aged male interpreter, the core process at this level involves the co-activation of age and gender categories and their cross-dimensional interaction. In this case, the aged or young nodes within the age dimension pool engage in mutual inhibition, while the male or female nodes within the gender dimension pool compete for dominance. When sensory cues (e.g., aged facial features and a masculine vocal tone) are transmitted to the category level, the

elderly node in the age dimension and the male node in the gender dimension are parallelly activated. On the contrary, the young and female nodes are being suppressed. Through dynamic constraint satisfaction, the category level converges to a stable state, wherein the salient nodes remain persistently activated, leading users to perceive the interpreter as an elderly male.

At the stereotype level, the social categories activated at the category level are dynamically associated with the semantic network of stereotypes, shaping users' evaluations of the professional competence of the aged male interpreter. It is important to note that this evaluation is stereotype-driven, and being reflected in a conceptualized perception of quality rather than an assessment of the actual rendition quality perceived by users. For the aged male interpreter, the elderly category node is linked to negative ageing-related stereotypes, such as "cognitive decline" and "slower knowledge acquisition", in contrast to younger individuals. Meanwhile, the male category node may activate associations with technical expertise, positioning males as more technically driven compared to females. The resulting overall stereotype emerges through the reinforcement of dominant impressions, while weaker impressions are suppressed, forming a globalized representation of the interpreter's professional competence. As reflected in the experimental outcomes in Chapter 4, this composite stereotype manifests as lower ratings of professional competence for elderly male interpreters relative to their younger counterparts, despite no significant disadvantage in expressive features. It ultimately results in a systematic bias toward more negative evaluations. Through hierarchical recursive embedding, the stereotype level integrates into the global dynamics of person construal, and finally shapes users' cognitive biases toward aged male interpreters.

Finally, the higher-order level. When the experimenter introduced professional titles to intervene in age-related stereotypes, information reflecting higher socioeconomic status like "professional medical doctor" and "highly educated" was externally input into the higher-order cognitive layer. It eliminated the perceived differences in users' quality perception between aged and young male interpreters. First, the higher-order cognitive layer dynamically adjusts the activation weights of lower-level nodes through task demand nodes (such as "professionalism assessment") and long-term knowledge background (such as "occupational prestige"), and in turn, influences the pathway of stereotype formation. The introduction of professional titles activates the "professionalism" and "authority" nodes within the higher-order cognitive layer, which then suppresses the ageing-related "cognitive decline" stereotype node through inhibitory connections. Second, the activation of professional titles reduces the processing weight of the non-targeted dimension of ageing, and has shifted users' attentional resources during the interpreting task toward professional competence rather than age-related

attributes. This results in the diminished significance of age-related cues in the category level's competitive process, thereby reducing the infiltration of age bias within the stereotype level. Consequently, professional titles regulate stereotype activation pathways through the higher-order cognitive layer, inhibiting the negative impact of age-related stereotypes at the stereotype level while simultaneously enhancing profession-related stereotypes through the halo effect. This, in turn, balances the perceived quality ratings between aged and young interpreters. As we found in the experimental results in Chapter 4, despite the presence of age-related stereotypes, strong contextual cues introduced via higher-order cognition provided a feasible intervention pathway for mitigating social biases.

In summary, when an interpreter's perceived age or gender is altered, the change originates from the cue level, propagating upward to the category level, and ultimately manifesting at the stereotype level. However, when natural cues remain unchanged and professional titles are introduced, the change originates from the higher-order level, subsequently influencing the stereotype level in a top-down manner. Thus, modifications in natural cues (e.g., age and gender) shape stereotypes and user behaviour in a bottom-up process, whereas social cues such as professional titles intervene in stereotype processing in a top-down manner.

5.2.2. Constructing the interactive model of interpreting quality perception

Based on the dynamic interactive theory by Freeman & Ambady (2011), we propose a model that integrates social psychological theories with the assessment of interpreting quality in interpreting studies. This model aims to explain not only users' perceptions of interpreting quality but also expert assessors' quality evaluations. I refer to this model as the Interactive Model of Quality Perception (iMQP). The main procedural flow of the model is illustrated in Figure 5.2, and my interpretation of its core structure is as follows:

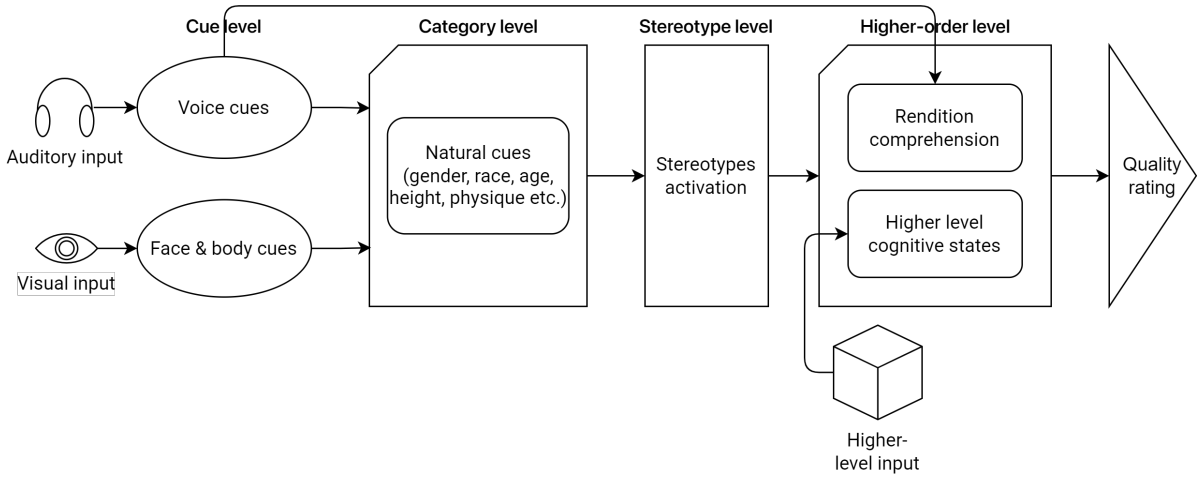


Figure 5.2 Construct of the interactive model of quality perception (iMQP).

Similar to the dynamic interactive model of person construal, iMQP consists of four hierarchical levels: the cue level, category level, stereotype level, and higher-order level. I propose that both users and experts process an interpreter's rendition through these sequential levels of information processing. The primary difference between experts and users lies in their higher-level input, which we will elaborate on this distinction at the later part of this subsection. We first attempt to explain the quality perception process from the perspective of general users.

Throughout an interpreting assignment, users receive more than just the content of an interpreter's rendition; they also process the rendition's paralinguistic features (i.e., voice cues) as well as the interpreter's facial and bodily cues, as this has been reviewed in Chapter 2. In simultaneous interpreting assignments, interpreters work inside a booth, separate from the event setting. Consequently, unless users make a deliberate effort to observe the interpreter in the booth (such as out of curiosity or due to a complaint), thus the visual input of the interpreter is effectively absent in simultaneous interpreting. On the contrary, both visual and auditory inputs contribute to voice cues and face & body cues respectively in a consecutive interpreting assignment. The paralinguistic features of voice cues, along with face & body cues, are processed at the categorical level, while the content of the rendition is transmitted to the higher-order level for rendition comprehension. Thus, these natural cues reach users' and experts' information processing systems prior to the semantic information of the rendition's content, providing categorical information with fundamental cue-based support at the cue level.

At the categorical level, both users and experts detect the natural cues associated with an interpreter's gender, age, race, and other social categorical information, thereby positioning the interpreter within a specific social category. Although simultaneous interpreting does not provide visual input for identifying an interpreter's race or other visual categories, as discussed in Chapter 2, existing literature suggests that users can infer race based on an interpreter's accent and certain acoustic characteristics. Following this categorization process, categorical information establishes weighted associative connections, laying the foundation for the stereotype level where a global stereotype representation is formed.

I state that both users and experts are influenced by stereotypes to some extent, and the primary distinction between them differentiates at the degree of higher-level cognitive states, see Figure 5.3 below:

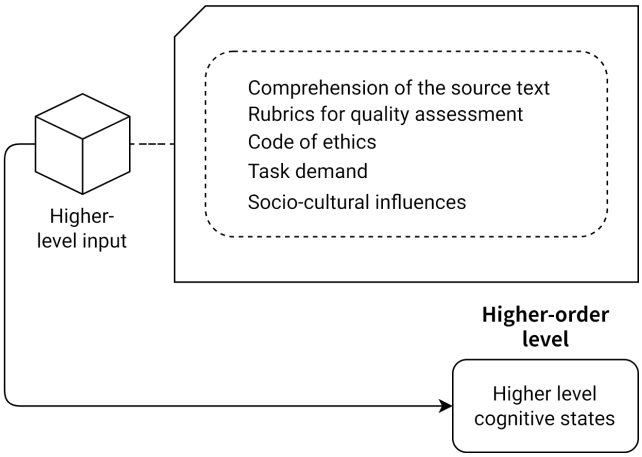


Figure 5.3 An example of the content within the higher-level cognitive states

First, experts are likely to hold weaker stereotypes compared to users. Experts are bilingual or multilingual individuals with strong proficiency in both the source language and the target language. This enables them to comprehend the source language more effectively than users. In contrast, a large proportion of interpreting users lack source language comprehension abilities, resulting in limited access to rendition comprehension at the higher-order level. Consequently, users rely more heavily on cue-level information, which is more salient and easier to process. Experts, on the other hand, possess stronger comprehension abilities at the higher-order level, reducing their dependence on cue-level information and leading to weaker stereotype activation at the stereotype level. Second, the higher-level cognitive states of experts receive different higher-level inputs than those of users. Since higher-level input also includes sociocultural influences, experts who have undergone systematic interpreter training and acquired a thorough understanding of professional codes of ethics tend to exhibit greater tolerance toward gender, race, and professional titles in cultural scripts. During interpreting quality assessment, experts are more likely to self-regulate their rating behaviour, preventing stereotypical biases from spreading unchecked. This explains why, as shown in Chapter 4's experiments, experts did not significantly increase their trust ratings in response to the salience of professional titles, whereas users exhibited a substantial increase. Finally, because experts receive sufficient endogenous input at the higher-order level, they do not require external reinforcement to influence their higher-level cognitive states in the same way that users do. As

a result, experts' quality ratings are more stable and reliable compared to those of users.

Of course, according to iMQP, expert assessors are not entirely exempt from the influence of stereotypes. If their higher-order cognitive regulation is insufficient, they may exhibit severity or leniency biases, as documented in Han (2015). Consequently, to achieve a more objective evaluation of an interpreter's performance, it is advisable to engage multiple raters in assessing the same rendition under standardized conditions. This allows for the identification and subsequent exclusion of outlier ratings that significantly deviate from the aggregate assessments. Nonetheless, from a theoretical standpoint, expert assessors are less prone than user groups to manifest systematic, large-scale stereotypical biases in their quality assessments.

5.3. Correction of users' biased quality perception ratings

While users' quality perception ratings may be subject to large-scale stereotype-driven biases, such biases can be correctified. I propose two possible methods for calibration: pre-rating calibrations through the infusion of higher-level cues before obtaining ratings, or post-rating corrections based on the interpreter's perceived social categorical information after the ratings have been collected.

5.3.1. Calibration of users' quality perceptions through higher-level input

As discussed in Section 5.2, once an interpreter's perceived natural cues such as gender, race, age, height, or physique, which are fixed and immutable, the information transmitted from the cue level to the category level remains constant. Therefore, the nodes formed by these factors propagate upwards to the stereotype level in a fixed and unmodifiable manner. However, by introducing higher-level input into users' cognitive processing, it is possible to alter their higher-order cognitive states, thereby attenuating the negative effects of stereotype activation (although this only mitigates its impact rather than eliminating stereotypes altogether). A long-term and effortful strategy for this correction is to implement user education (范志嘉, 2018). While for short-term intervention, higher-level input can be introduced before users begin evaluating an interpreter's performance to pre-emptively counteract biased ratings influenced by stereotypes. For example, prior to an interpreting assignment at a tech product launch conference, users could be explicitly reminded that interpreters, regardless of gender, age, or other social attributes, have already undergone a rigorous accreditation process during pre-assignment screening.

Another approach is to emphasize the interpreter's professional titles before the interpreting assignment begins. For instance, in court interpreting, users could be explicitly informed of an

ad hoc interpreter's legal credentials, including their law degree, professional certification, and legal titles. As demonstrated by the experimental outcomes in Chapter 4, professional titles enhance users' perceptions of fidelity, delivery, and domain knowledge, aligning them more closely with expert assessments. Thus, this serves as an effective corrective mechanism. While trust scores may still exhibit a certain degree of overestimation compared to experts' ratings, this is arguably preferable to the trust barriers that arise from users' lack of confidence in interpreters (Hsieh *et al.*, 2010; Wilson & Walsh, 2019). Although emphasizing professional titles may seem not fair enough to interpreters who are competent but lack formal titles, professional titles, as social cues other than natural cues, are more likely to be acquired than born. After all, obtaining a CATTI Level 2 certificate[14] is considerably more possible than increasing one's height by ten centimetres.

5.3.2. Manual correction of ratings based on an interpreter's perceived social categorical information

If higher-level input was not introduced before the interpreting assignment, and as a result, users provided biased quality perception ratings, these ratings can still be adjusted post hoc based on the interpreter's social categorical information. This allows for the correction of ratings to approximate the benchmark set by expert assessments.

Since this dissertation conducted controlled experiments and data collection on three social categorical variables (age, gender and professional title) among undergraduate interpreting users, the sample size is insufficient to develop a fully quantitative correction model (e.g., a linear regression model). However, existing research outcomes provide a general direction for correction. Thus, approximate corrections to users' stereotyped ratings can be made based on the interpreter's natural cues, with scores calculated on a 100-point scale, as follows:

- For an ad hoc interpreter with a professional title that confers higher socioeconomic status than a full-time interpreter, who is under 60 years old, and whose title has been activated, the existing score remains unchanged.
- If the above conditions apply, but the interpreter is a male over 60 years old, then fidelity increases by 4 points, while delivery decreases by 6 points.
- If the interpreter is a full-time young male without an activated professional title, then fidelity increases by 4 points, fluency increases by 5 points, and domain knowledge increases by 8 points.

[14]. CATTI refers to the China Accreditation Test for Translators and Interpreters.

- If the interpreter is a full-time young female, then fidelity increases by 12 points, delivery increases by 13 points, and domain knowledge increases by 15 points.
- If the interpreter is a full-time elderly male, then fidelity increases by 10 points, delivery increases by 4 points, and domain knowledge increases by 15 points.

5.4. Summary

Through bottom-up inputs (e.g., facial, vocal, and bodily cues) and top-down factors (e.g., stereotypes, motivations, and tasks), the iMQP model underscores how perceptions evolve probabilistically rather than forming instantaneously. This framework helps explain how age, gender, and professional cues can bias interpreting quality assessments. Yet, by leveraging higher-order regulatory mechanisms or emphasizing professional titles, such distortions can be mitigated and thus foster more accurate and equitable results of quality perceptions.

Chapter 6 Conclusion

This chapter provides a summative discussion of this dissertation, which synthesises the findings of this study's four research questions. It further analyses the implications of these conclusions and elaborates on the study's limitations, while outlining potential avenues for future research.

"There is nothing so obdurate to education or criticism as the stereotype" (Lippmann, 1922, p. 99). The aim of this dissertation is to examine whether an interpreter's natural cues, such as gender and age and social cues, such as an ad hoc interpreter's professional titles, are perceptible to users during the interpreting process. Furthermore, it seeks to determine whether users' perceptions of interpreting quality are influenced by these social categorical factors. If these factors exert negative effects on users' quality perception, the study aims to develop a correction model that can adjust users' ratings to approximate a more objective evaluation, aligning them with expert assessment scores. To this end, the study integrates theoretical constructs with experimental research methodologies.

From a theoretical perspective, it is first necessary to determine which disciplinary frameworks can be employed to deconstruct social categorical information and, secondly, to investigate why and how these natural or social cues exert either positive or negative influences. Additionally, a conceptual model should be developed to analyse the commonalities and divergences between interpreting quality assessment and quality perception within interpreting studies. This model should also incorporate the effects of social categorical information, bridging disciplinary gaps and extending the study of interpreting quality from controlled laboratory settings to practical scenarios characterized by greater variability and complexity. Given the scarcity of research on quality perception and users' biased treatment of interpreters, which is currently limited to a few interview-based studies and a small number of controlled experiments, this dissertation prioritizes the collection of extensive empirical data to validate its hypotheses. While from an experimental research perspective, this study first employed a combination of questionnaires and free association techniques to gather stereotypical perceptions of a Chinese interpreter from the general Chinese public. The findings from this initial phase provided a foundational reference for the control group in the controlled experiments. Subsequently, a series of controlled experiments were designed to examine how users' stereotypes regarding gender, age, and professional titles influence their perceptions of fidelity, delivery, and domain knowledge, as well as their trust in the interpreter.

Following the analysis of experimental outcomes, the study applied Freeman & Ambady's

(2011) dynamic interactive model of person construal to explain the underlying mechanisms of these perceptual biases. Based on these insights, the dissertation developed the interactive model of quality perception (iMQP), which serves as the basis for two corrective approaches: (1) mitigating users' biased quality perception ratings through higher-level input and (2) manual ad hoc correction of ratings. As stereotype-based information processing is not only a form of shallow cognitive processing but also an abstraction based on social group characteristics, this dissertation contributes to the understanding of biased quality perception from a social psychological perspective. By offering a theoretical interpretation of users' biased rating behaviours and proposing preliminary corrective measures, this study provides a foundation for addressing bias in interpreting quality perception at the group psychological level.

Certainly, research on language-related activities is inherently complex, influenced by a multitude of uncontrollable variables, and subject to individual differences among participants. Establishing causal relationships between two variables is therefore challenging, as statistical and correlational analyses cannot fully capture the intricate dynamics of variable interactions in real-world scenarios. Additionally, given the limitations of experimental conditions and sample size, this study has several notable constraints:

First, the generation of elderly female interpreter images through technological means required a two-step transformation process—first from male to female, and then from young to elderly. This process resulted in perceptual distortions in the generated images, which raised concerns about the naturalness of the stimuli. We were concerned that such distortions might lead user participants to question the experiment's authenticity, thereby compromising data quality. As a result, the formal study in Chapter 4 adopted a 2×3 matrix design, rather than the experimenter's originally planned 3×3 design. Consequently, the experimental outcomes lack data on users' perceptions of elderly female interpreters, as well as data from two dimensions—professional title activation and implicit perception. This structural limitation in the dataset results in incomplete coverage of relevant variables, thereby restricting the study's ability to examine how users' stereotypes influence perceptions of elderly female interpreters.

Second, the experimental setting was conducted within a university environment, with undergraduate participants serving as the sample population. Although the experimental outcomes demonstrated statistically significant results, undergraduate participants differ from genuine conference users, just as a university setting differs from authentic conference environments. As Liu (2023) states, we cannot assume that modifying context-independent research will produce reliable results for this emerging form of interpreting service. Therefore, future research should strive to conduct experiments in real conference settings, involving

genuine conference participants to enhance the ecological validity of the findings.

Third, due to insufficient data volume, the manual correction of ratings based on an interpreter's perceived social categorical information could only be approximated using a basic additive or subtractive adjustment, rather than the originally intended linear regression model or machine learning-based simulation. As a result, the corrected scores remain an estimated approximation, lacking strong statistical validity. Future studies should incorporate a larger and more diverse sample size, covering a broader range of participant conditions, to develop a fully functional mathematical model capable of direct input-output prediction.

Finally, while the survey on public stereotypes toward interpreters collected over 300 responses, the sample size remains relatively small compared to large-scale social surveys. Additionally, the majority of respondents were concentrated in coastal regions such as Jiangsu and Zhejiang, northern provinces like Shandong, and central regions like Hubei and Sichuan. This geographic concentration limits the representativeness of the findings, making them only generalizable to these specific regions rather than the entire country. Furthermore, the age distribution of respondents was skewed, with fewer middle-aged participants (40-60 years old) compared to younger respondents (20-30 years old). The non-normative age distribution necessitates further data collection to obtain a more balanced sample, thereby improving the robustness of future findings.

References

[1] Abbott A. The system of professions: An essay on the division of expert labor[M]. Chicago, IL, US: The University of Chicago Press, 1988: xvi, 435.

[2] Abele A E & Wojciszke B. Communal and agentic content in social cognition: A dual perspective model[M]//Olson J M, Zanna M P. Advances in Experimental Social Psychology: Vol. 50. Academic Press, 2014: 195-255.

[3] Acosta Vicente C. "You are a woman, so you understand what I mean": Public service interpreters' self-perceptions of the (in)significance of their gender[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2024, 26(1): 114-143.

[4] AIIC. Practical guide for professional conference interpreters[EB/OL]. (1999-12-01)[2023-11-03]. https://aiic.org/document/547/AIICWebzine_Apr2004_2_Practical_guide_for_professional_conference_interpreters_EN.pdf.

[5] Albl-Mikasa M. Global English and English as a lingua franca (ELF) : Implications for the interpreting profession[J]. Trans-kom, 2010, 3(2): 126-148.

[6] Allport G W. The nature of prejudice[M]. Oxford: Addison-Wesley, 1954: xviii, 537.

[7] American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing[M]. Washington, DC: American Educational Research Association, 2008.

[8] Anderson B. Perspectives on the role of interpreter[M]//Brislin R. Translation: Applications and Research. New York: Gardner Press, 1976: 209-228.

[9] Anderson K J. Students' stereotypes of professors: an exploration of the double violations of ethnicity and gender[J]. Social Psychology of Education, 2010, 13(4): 459-472.

[10] Angouri J. "We are in a masculine profession…": Constructing gender identities in a consortium of two multinational engineering companies[J]. Gender and Language, 2011, 5(2): 373-403.

[11] Artl G B. Act like a lady: The impact of gender identity on American sign language - English interpreters[D]. Monmouth: Western Oregon University, 2015.

[12] Arvey R D. Unfair discrimination in the employment interview: Legal and psychological aspects[J]. Psychological Bulletin, 1979, 86(4): 736-765.

[13] Ash E, Durante R, Grebenshchikova M, *et al.* Visual representation and stereotypes in news media[J]. SSRN Electronic Journal, 2022.

[14] Auspurg K & Düval S. Housework as a woman's job? What looks like gender ideologies could also be stereotypes[J]. Sociological Science, 2024, 11: 789-814.

[15] Barrett L F, Robin L, Pietromonaco P R, *et al.* Are women the "more emotional" sex? Evidence from emotional experiences in social context[J]. Cognition & Emotion, 1998, 12(4): 555-578.

[16] Bartłomiejczyk M. Interpreting quality as perceived by trainee interpreters: Self-evaluation[J]. The Interpreter and Translator Trainer, 2007, 1(2): 247-267.

[17] Bartłomiejczyk M. Parliamentary impoliteness and the interpreter's gender[J]. Pragmatics / Quarterly Publication of the International Pragmatics Association (IPrA), 2019, 30(4): 459-484.

[18] Bass B A. Effects of professional title and degree upon the perceived credibility of a sex therapist[J]. Journal of Sex Education and Therapy, 1986, 12(1): 47-50.

[19] Baysu G, Phalet K & Brown R. Dual identity as a two-edged sword: Identity threat and minority school performance[J]. Social Psychology Quarterly, 2011, 74(2): 121-143.

[20] Becerra O G. Do first impressions matter? The effect of first impressions on the assessment of the quality of simultaneous interpreting[J]. Across Languages and Cultures, 2016, 17(1): 77-98.

[21] Beeghly E. What is a stereotype? What is stereotyping?[J]. Hypatia, 2015, 30(4): 675-691.

[22] Beike D R & Sherman S J. Social inference: Inductions, deductions, and analogies.[M]//Handbook of social cognition: Basic processes; Applications, Vols. 1-2, 2nd ed. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, 1994: 209-285.

[23] Bergsieker H B, Leslie L M, Constantine V S, *et al.* Stereotyping by omission: Eliminate the negative, accentuate the positive[J]. Journal of Personality and Social Psychology, 2012, 102(6): 1214-1238.

[24] Berk-Seligson S. The impact of politeness in witness testimony: the influence of the court Interpreter[J]. mult, 1988, 7(4): 411-440.

[25] Berzonsky M. Identity formation: The role of identity processing style and cognitive processes[J]. Personality and Individual Differences, 2008, 44(3): 645-655.

[26] Bestelmeyer P E G. Regional accents: Spontaneous biases toward speakers who sound like us[J]. Journal of Language and Social Psychology, 2024: 0261927X241287189.

[27] Biernat M. Gender stereotypes and the relationship between masculinity and femininity: A developmental analysis[J]. Journal of Personality and Social Psychology, 1991, 61(3): 351-365.

[28] Bijlstra G, Holland R W, Dotsch R, *et al.* Stereotype associations and emotion recognition[J]. Personality and Social Psychology Bulletin, 2014, 40(5): 567-577.

[29] Blue P R, Hu J, Peng L, *et al.* Whose promises are worth more? How social status affects trust in promises[J]. European Journal of Social Psychology, 2020, 50(1): 189-206.

[30] Boas T C. Pastor Paulo vs doctor Carlos: Professional titles as voting heuristics in Brazil[J]. Journal of Politics in Latin America, 2014, 6(2): 39-72.

[31] Bodenhausen G V, Kang S K & Peery D. Social categorization and the perception of social groups[M]//Fiske S, Macrae C. The SAGE Handbook of Social Cognition. 1 Oliver's Yard, 55 City Road, London EC1Y 1SP United Kingdom: SAGE Publications Ltd, 2012: 311-329.

[32] Boos M, Kobi M, Elmer S, *et al.* The influence of experience on cognitive load during simultaneous interpretation[J]. Brain and Language, 2022, 234: 105185.

[33] Bornkessel-Schlesewsky I, Krauspenhaar S & Schlesewsky M. Yes, you can? A speaker's potency to act upon his words orchestrates early neural responses to message-level meaning[J]. PLoS ONE, 2013, 8(7): e69173.

[34] Borthwick A M, Boyce R A & Nancarrow S A. Symbolic power and professional titles: The case of "podiatric surgeon"[J]. Health Sociology Review, 2015, 24(3): 310-322.

[35] Bourdieu P, Nice R, Bourdieu P, *et al.* Distinction: a social critique of the judgement of taste[M]. Repr. Cambridge, Mass: Harvard Univ. Pr, 1984.

[36] Bourdieu P & Wacquant L. Symbolic capital and social classes[J]. Journal of Classical Sociology, 2013, 13(2): 292-302.

[37] Brescoll V L & Uhlmann E L. Attitudes toward traditional and nontraditional parents[J]. Psychology of Women Quarterly, 2005, 29(4): 436-445.

[38] Brewer M B. A dual process model of impression formation[M]//A dual process model of impression formation. Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, 1988: 1-36.

[39] Brey E & Pauker K. Teachers' nonverbal behaviors influence children's stereotypic beliefs.[J]. Journal of Experimental Child Psychology, 2019, 188: 104671.

[40] Brigham J C. Ethnic stereotypes[J]. Psychological Bulletin, 1971, 76(1): 15-38.

[41] Bühler H. Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters[J]. Multilingua, 1986, 5(4): 231-235.

[42] Cale E M. A quantitative review of the relations between the "Big 3" higher order personality dimensions and antisocial behavior[J]. Journal of Research in Personality,

2006, 40(3): 250-284.

[43] Calvo-Calvo M Á. Sexist stereotypes and biases associated with the female nursing model in advertising communication[J]. Texto & Contexto - Enfermagem, 2014, 23(3): 530-537.

[44] Carli L L, Alawa L, Lee Y, *et al.* Stereotypes about gender and science: Women ≠ scientists[J]. Psychology of Women Quarterly, 2016, 40(2): 244-260.

[45] Carlston D E & Skowronski J J. Savings in the Relearning of Trait Information as Evidence for Spontaneous Inference Generation[J]. Journal of Personality and Social Psychology, 1994, 66(5): 840-856.

[46] Carrington B. A quintessentially feminine domain? Student teachers' constructions of primary teaching as a career[J]. Educational Studies, 2002, 28(3): 287-303.

[47] Chaffee K E, Lou N M, Noels K A, *et al.* Why don't "real men" learn languages? Masculinity threat and gender ideology suppress men's language learning motivation[J]. Group Processes & Intergroup Relations, 2020, 23(2): 301-318.

[48] Chai Y & Li X. Interpreting students' perceptions of interpreter status, market order, and educational support and their commitment to the profession: a survey in China[J]. Onomázein, 2024(64): 01-40.

[49] Charlesworth T E S & Banaji M R. Gender in science, technology, engineering, and mathematics: Issues, causes, solutions[J]. The Journal of Neuroscience, 2019, 39(37): 7228-7243.

[50] Chen A. Perception of paralinguistic intonational meaning in a second language[J]. Language Learning, 2009, 59(2): 367-409.

[51] Chen J, Yang H & Han C. Holistic versus analytic scoring of spoken-language interpreting: a multi-perspectival comparative analysis[J]. The Interpreter and Translator Trainer, 2022, 16(4): 558-576.

[52] Chen S J. Linguistic dimensions of subtitling: Perspectives from taiwan[J]. Meta, 2004, 49(1): 115-124.

[53] Cheng X, Zhu Y, Hu Yinying, *et al.* Integration of social status and trust through interpersonal brain synchronization[J]. NeuroImage, 2022, 246: 118777.

[54] Cheng Y, Cui Q, Lin Y, *et al.* The reason behind female outnumbering male students in foreign language high schools in China[C]//2021 International Conference on Education, Language and Art (ICELA 2021). Sanya, China, 2022.

[55] Cheung A K F. Non-native accents and simultaneous interpreting quality perceptions[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2013, 15(1):

25-47.

[56] Cheung A K F. Scapegoating the interpreter for listeners' dissatisfaction with their level of understanding: An experimental study[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2015, 17(1): 46-63.

[57] Cheung A K F. Interpreters' perceived characteristics and perception of quality in interpreting[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2020, 22(1): 35-55.

[58] Cheung A K F. Listeners' perception of the quality of simultaneous interpreting and perceived dependence on simultaneous interpreting[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2021.

[59] Chiaro D. Linguistic mediation on Italian television: When the interpreter is not an interpreter: a case study[M]//Garzone G, Viezzi M. Interpreting in the 21st Century: Challenges and oppurtunities: Vol. 43. Amsterdam: John Benjamins Publishing Company, 2002: 215-225.

[60] Chiaro D & Nocella G. Interpreters' perception of linguistic and non-linguistic factors affecting quality: A survey through the world wide web[J]. Meta, 2004, 49(2): 278-293.

[61] Cho J. Why do interpreters need to be beautiful?: Aesthetic labour of language workers[J]. Gender and Language, 2017, 11(4): 482-506.

[62] Choi J. Assessing the impact of text length on consecutive interpreting[M]//Tsagari D. Assessment issues in language translation and interpreting. Frankfurt am Main: Peter Lang, 2013: 85-96.

[63] Christen y Gracia L. Communication mediated by conference interpreters: Age and sex stereotypes[J]. Obra digital, 2023(24): 125-146.

[64] Christensen T P. User expectations and evaluation: a case study of a court interpreting event[J]. Perspectives, 2011, 19(1): 1-24.

[65] Christoph D & Li A K F. Cognitive versus social rigidity in old age: implications for therapy[J]. Canadian Journal on Aging, 1985, 4(2): 59-65.

[66] CICG Academy of Translation and Interpretation. Introduction to the China accreditation test for translators and interpreters[EB/OL]. (2023-12-19)[2024-08-15]. http://www.aticicg.org.cn/2023-12/19/content_42612741.html.

[67] Clow K A, Ricciardelli R & Bartfay W J. Attitudes and stereotypes of male and female nurses: The influence of social roles and ambivalent sexism.[J]. Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement, 2014, 46(3): 446-455.

[68] Cohen H & Lefebvre C. Handbook of categorization in cognitive science[M]. 1st ed. Amsterdam Boston: Elsevier, 2005.

[69] Collados Aís Á. Quality assessment in simultaneous interpreting: The importance of nonverbal communication[M]//Pöchhacker F, Shlesinger M. The interpreting studies reader. London ; New York: Routledge, 2002: 327-336.

[70] Collados Aís Á. Quality assessment and intonation in simultaneous interpreting : evaluation patterns[J]. MonTI. Monografías de Traducción e Interpretación, 2016, 0(ne3): 213-238.

[71] Collados Aís Á & Becerra O G. Quality criteria[M]//Pöchhacker F, Grbić N, Mead P, et al. Routledge encyclopedia of interpreting studies. London ; New York: Routledge, Taylor & Francis Group, 2015: 337-338.

[72] Corcoran K & Mussweiler T. The cognitive miser's perspective: Social comparison as a heuristic in self-judgements[J]. European Review of Social Psychology, 2010, 21(1): 78-113.

[73] Cozzarelli C, Wilkinson A V & Tagler M J. Attitudes toward the poor and attributions for poverty[J]. Journal of Social Issues, 2001, 57(2): 207-227.

[74] Creswell J W. Educational research: planning, conducting, and evaluating quantitative and qualitative research[M]. 4th ed. Boston: Pearson, 2012.

[75] Crisp R J & Hewstone M. Multiple social categorization[M]//Advances in Experimental Social Psychology Volume 39: Vol. 39. Elsevier, 2007: 163-254.

[76] Cui Y & Liu X. How condensation and non-condensation impact viewers' processing effort and comprehension – an eye-tracking study on Chinese subtitling of English documentaries[J]. Perspectives, 2024: 1-19.

[77] Cvencek D, Nasir N S, O'Connor K, *et al.* The development of math–race stereotypes: "They say Chinese people are the best at math"[J]. Journal of Research on Adolescence, 2015, 25(4): 630-637.

[78] Czopp A M, Kay A C & Cheryan S. Positive stereotypes are pervasive and powerful[J]. Perspectives on Psychological Science, 2015, 10(4): 451-463.

[79] Dannenmaier W D & Thumin F J. Authority status as a factor in perceptual distortion of size[J]. The Journal of Social Psychology, 1964, 63(2): 361-365.

[80] de Lemus S, Moya M, Lupiáñez J, *et al.* Men in the office, women in the kitchen? Contextual dependency of gender stereotype activation in Spanish women[J]. Sex Roles, 2014, 70(11-12): 468-478.

[81] Deaux K & Lafrance M. Gender[M]//The handbook of social psychology: Vols. 1-2. 4th

ed. New York, NY, US: McGraw-Hill, 1998: 788-827.

[82] Deaux K & Lewis L L. Components of gender stereotypes[J]. Psychological documents, 1983, 13(2): 25.

[83] Defrancq B, Collard C, Magnifico C, *et al.* Sex and gender in conference interpreting[M]//Albl-Mikasa M, Tiselius E. The Routledge Handbook of Conference Interpreting. 1st ed. London: Routledge, 2021: 414-427.

[84] Déjean Le Féal K. Some thoughts on the evaluation of simultaneous interpretation[M]//Bowen M, Bowen D. Interpreting, yesterday, today, and tomorrow. Netherlands Philadelphia, PA: John Benjamins Pub. Co, 2008: 154-160.

[85] Delgado Luchner C. "A beautiful woman sitting in the dark": Three narratives of interpreter training at the University of Nairobi[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2019, 21(1): 91-114.

[86] Denton F T & Spencer B G. How old is old? Revising the definition based on life table criteria[J]. Mathematical Population Studies, 1999, 7(2): 147-159.

[87] Devine P G. Stereotypes and prejudice: Their automatic and controlled components.[J]. Journal of Personality and Social Psychology, 1989, 56(1): 5-18.

[88] Devine P G & Elliot A J. Are racial stereotypes really fading? The Princeton trilogy revisited[J]. Personality and Social Psychology Bulletin, 1995, 21(11): 1139-1150.

[89] Diriker E. De-/re-contextualizing conference interpreting: Interpreters in the ivory tower? Vol. 53[M]. Amsterdam: John Benjamins Publishing Company, 2004.

[90] Dobs K, Isik L, Pantazis D, *et al.* How face perception unfolds over time[J]. Nature Communications, 2019, 10(1): 1258.

[91] Dong Y & Li P. Attentional control in interpreting: A model of language control and processing control[J]. Bilingualism: Language and Cognition, 2020, 23(4): 716-728.

[92] Downie J. Interpreters vs machines: Can interpreters survive in an AI-dominated world?[M]. London ; New York: Routledge, Taylor & Francis Group, 2020.

[93] Du Y. "Beauties by the political figures" - professional female interpreters in Chinese media[J]. Feminist Media Studies, 2021, 23(2): 1-17.

[94] Durante F & Fiske S T. How social-class stereotypes maintain inequality[J]. Current Opinion in Psychology, 2017, 18: 43-48.

[95] Durante F, Fiske S T, Kervyn N, *et al.* Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap[J]. British Journal of Social Psychology, 2012, 52(4): 726-746.

[96] Durante F, Tablante C B & Fiske S T. Poor but warm, rich but cold (and competent):

Social classes in the stereotype content model[J]. Journal of Social Issues, 2017, 73(1): 138-157.

[97] Eagly A H & Koenig A M. The vicious cycle linking stereotypes and social roles[J]. Current Directions in Psychological Science, 2021, 30(4): 343-350.

[98] Eagly A H & Steffen V J. Gender stereotypes stem from the distribution of women and men into social roles[J]. Journal of Personality and Social Psychology, 1984, 46(4): 735-754.

[99] Edwards A L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed[Z]//Journal of Applied Psychology: Vol. 37. American Psychological Association (APA), 1953: 90-93.

[100] Edwards R, Temple B & Alexander C. Users' experiences of interpreters: The critical role of trust[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2005, 7(1): 77-95.

[101] Ekman P. Universal facial expressions in emotion[J]. Studia Psychologica, 1973, 15(2): 140-147.

[102] Ellemers N. Gender stereotypes[J]. Annual Review of Psychology, 2018, 69(1): 275-298.

[103] Ensmenger N. "Beards, sandals, and other signs of rugged individualism": Masculine culture within the computing professions[J]. Osiris, 2015, 30(1): 38-65.

[104] Eysenck H J. Crime and personality[M]. [Repr. of the 3., rev. ed., 1977], 1. publ. London: Routledge & Kegan Paul, 2013.

[105] FeldmanHall O & Shenhav A. Resolving uncertainty in a social world[J]. Nature Human Behaviour, 2019, 3(5): 426-435.

[106] Feng D, Zhao W, Shen S, *et al.* The influence of perceived prejudice on willingness to be a nurse via the mediating effect of satisfaction with major: A cross-sectional study among chinese male nursing students[J]. Nurse Education Today, 2016, 42: 69-72.

[107] Fernandes C, Macedo I, Gonçalves A R, *et al.* Effects of aging on face processing: An ERP study of the own-age bias with neutral and emotional faces[J]. Cortex, 2023, 161: 13-25.

[108] Fisicaro S A & Lance C E. Implications of three causal models for the measurement of halo error[J]. Applied Psychological Measurement, 1990, 14(4): 419-429.

[109] Fiske S T. Prejudices in cultural contexts: Shared stereotypes (gender, age) versus variable stereotypes (race, ethnicity, religion)[J]. Perspectives on Psychological Science, 2017, 12(5): 791-799.

[110] Fiske S T, Cuddy A J C, Glick P, *et al.* A model of (often mixed) stereotype content:

Competence and warmth respectively follow from perceived status and competition.[J]. Journal of Personality and Social Psychology, 2002, 82(6): 878-902.

[111] Fiske S T & Neuberg S L. Category-based and individuating processes as a function of information and motivation: Evidence from our laboratory[M]//Bar-Tal D, Graumann C F, Kruglanski A W, et al. Stereotyping and Prejudice. New York, NY: Springer New York, 1989: 83-103.

[112] Fiske S T & Neuberg S L. A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation[M]//Advances in Experimental Social Psychology: Vol. 23. Elsevier, 1990: 1-74.

[113] Fiske S & Taylor S. Social cognition evolves: Illustrations from our work on intergroup bias and on healthy adaptation[J]. Psicothema, 2020, 3(32): 291-297.

[114] Ford T E. Preconscious and Postconscious Processes Underlying Construct Accessibility Effects: An Extended Search Model[J]. Personality and Social Psychology Review, 2000, 4(4): 317-336.

[115] Forgas J P & Laham S M. Halo effects[M]//Pohl R. Cognitive illusions: Intriguing phenomena in thinking, judgment and memory, 2nd ed. New York, NY, US: Routledge/Taylor & Francis Group, 2017: 276-290.

[116] Freedman D J, Riesenhuber M, Poggio T, *et al.* Categorical representation of visual stimuli in the primate prefrontal cortex[J]. Science, 2001, 291(5502): 312-316.

[117] Freeman J B & Ambady N. A dynamic interactive theory of person construal[J]. Psychological Review, 2011, 118(2): 247-279.

[118] Freeman J B & Ambady N. The dynamic interactive model of person construal: Coordinating sensory and social processes.[M]//Dual-process theories of the social mind. New York, NY, US: The Guilford Press, 2014: 235-248.

[119] Freeman N J. Socioeconomic status and belief in a just world: Sentencing of criminal defendants[J]. Journal of Applied Social Psychology, 2006, 36(10): 2379-2394.

[120] Gentile P. The status of conference interpreters: A global survey into the profession[J]. Rivista internazionale di tecnica della traduzione, 2013, 15: 63-82.

[121] Gentile P. The professional status of public service interpreters: A comparison with nurses[J]. FITISPos International Journal, 2016, 3: 174-183.

[122] Gentile P. Through women's eyes. Conference interpreters' self-perceived status in a gendered perspective[J]. HERMES - Journal of Language and Communication in Business, 2018(58): 19-42.

[123] Gieshoff A C & Albl-Mikasa M. Interpreting accuracy revisited: a refined approach to interpreting performance analysis[J]. Perspectives, 2024, 32(2): 210-228.

[124] Gilbert D T, Fiske S T & Lindzey G. The handbook of social psychology: Vol. 18[M]. Boston, MA: McGrawHill, 1969.

[125] Gilbert G M. Stereotype persistence and change among college students[J]. The Journal of Abnormal and Social Psychology, 1951, 46(2): 245-254.

[126] Gile D. Basic concepts and models for interpreter and translator training: Revised edition: Vol. 8[M]. 2nd ed. Amsterdam: John Benjamins Publishing Company, 2009.

[127] Giustini D. 'I thought I could get away from gender discrimination': linguistic instrumentalism and self-actualisation of female interpreters in temporary employment in Japan[J]. Japan Forum, 2021, 33(4): 522-552.

[128] Giustini D. Haken conference interpreters in Japan: Exploring status through the sociology of work and of professions[J]. Interpreting and Society, 2022, 2(1): 3-31.

[129] Goffman E. Stigma: notes on the management of spoiled identity: Vol. 45[M]. London: Penguin, 1969.

[130] Gong Y, Zhao X, Ma Q, *et al.* Social identity threat attenuates own-race bias in face recognition under the "Asian-Caucasian" context[J]. Journal of Experimental Social Psychology, 2024, 113: 104624.

[131] Goodman M D & Gareis K C. The influence of status on decisions to help[J]. The Journal of Social Psychology, 1993, 133(1): 23-31.

[132] Goslin J, Duffy H & Floccia C. An ERP investigation of regional and foreign accent processing[J]. Brain and Language, 2012, 122(2): 92-102.

[133] Goy H, Kathleen Pichora-Fuller M & Van Lieshout P. Effects of age on speech and voice quality ratings[J]. The Journal of the Acoustical Society of America, 2016, 139(4): 1648-1659.

[134] Grbić N. Constructing interpreting quality[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2008, 10(2): 232-257.

[135] Grbić N. Quality[M]//Mikkelson H, Jourdenais R. The Routledge handbook of interpreting. London New York: Routledge, 2015: 333-336.

[136] Greenwald A G & Banaji M R. Implicit social cognition: Attitudes, self-esteem, and stereotypes[J]. Psychological Review, 1995, 102(1): 4-27.

[137] Hale S, Bond N & Sutton J. Interpreting accent in the courtroom[J]. Target. International Journal of Translation Studies, 2011, 23(1): 48-61.

[138] Han C. Investigating rater severity/leniency in interpreter performance testing: A

multifaceted Rasch measurement approach[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2015, 17(2): 255-283.

[139] Han C. Investigating score dependability in English/Chinese interpreter certification performance testing: A generalizability theory approach[J]. Language Assessment Quarterly, 2016a, 13(3): 186-201.

[140] Han C. A survey to profile conference interpreting practice in China[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2016b, 18(2): 259-272.

[141] Han C. Interpreting testing and assessment: A state-of-the-art review[J]. Language Testing, 2022, 39(1): 30-55.

[142] Han C & Lu X. Interpreting quality assessment re-imagined: The synergy between human and machine scoring[J]. Interpreting and Society, 2021, 1(1): 70-90.

[143] Han C & Lu X. Can automated machine translation evaluation metrics be used to assess students' interpretation in the language learning classroom?[J]. Computer Assisted Language Learning, 2023, 36(5-6): 1064-1087.

[144] Han C & Zhao X. Accuracy of peer ratings on the quality of spoken-language interpreting[J]. Assessment & Evaluation in Higher Education, 2021, 46(8): 1299-1313.

[145] Hansen K, Steffens M C, Rakić T, *et al.* When appearance does not match accent: neural correlates of ethnicity-related expectancy violations[J]. Social Cognitive and Affective Neuroscience, 2016, 12(3): 507-515.

[146] Hlavac J. A cross-national overview of translator and interpreter certification procedures[J]. The International Journal of Translation and Interpreting Research, 2013, 5(1).

[147] Hong L. A note on my experiences as a student, a teacher and an interpreter of English in China[M]//Burton P, Dyson K K, Ardener S. Bilingual Women. 1st ed. Routledge, 2020: 167-169.

[148] Hong S & Choi E. Struggling for professional identity: A narrative inquiry of Korean freelance male interpreters[J]. Babel. Revue internationale de la traduction / International Journal of Translation, 2020, 66(4-5): 674-688.

[149] House J. Translation, interpreting, and intercultural communication[M]//Jackson J. The Routledge handbook of language and intercultural communication: Vol. 13. Second edition. London: Routledge, 2013: 156-172.

[150] Hoyle D. ISO 9000 quality systems handbook-updated for the ISO 9001: 2015 standard: Increasing the quality of an organization's outputs[M]. Routledge, 2017.

[151] Hoza J. Team interpreting[M]//Stone C, Adam R, Müller De Quadros R, et al. The

Routledge Handbook of Sign Language Translation and Interpreting. 1st ed. London: Routledge, 2022: 162-178.

[152] Hsieh E, Ju H & Kong H. Dimensions of trust: The tensions and challenges in orovider-interpreter trust[J]. Qualitative Health Research, 2010, 20(2): 170-181.

[153] Hubbard R E & Rockwood K. Frailty in older women[J]. Maturitas, 2011, 69(3): 203-207.

[154] Hummert M L. A social cognitive perspective on age stereotypes[M]//Social Cognition and Aging. Elsevier, 1999: 175-196.

[155] Iglesias E. Understanding variability in interpreting quality assessment: User's sex and judgments for pleasant voice[M]//Way C, Vandepitte S, Meylaerts R, et al. Benjamins Translation Library: Vol. 108. Amsterdam: John Benjamins Publishing Company, 2013: 103-125.

[156] ISO/TC 37/SC 5. Interpreting services. General requirements and recommendations[S]. BSI British Standards, 2018.

[157] Ito T A & Urland G R. Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals.[J]. Journal of Personality and Social Psychology, 2003, 85(4): 616-626.

[158] Jaffe A. Differentiated eliteness: Socialization for academic leadership[J]. Social Semiotics, 2017, 27(3): 370-381.

[159] Jelsma E, Chen S & Varner F. Working harder than others to prove yourself: High-effort coping as a buffer between teacher-perpetrated racial discrimination and mental health among black american adolescents[J]. Journal of Youth and Adolescence, 2022, 51(4): 694-707.

[160] Jenkins R. Categorization: Identity, social process and epistemology[J]. Current Sociology, 2000, 48(3): 7-25.

[161] Jeon H A & Friederici A D. Degree of automaticity and the prefrontal cortex[J]. Trends in Cognitive Sciences, 2015, 19(5): 244-250.

[162] Ji Z, Yang Y, Fan X, *et al.* Stereotypes of social groups in mainland China in terms of warmth and competence: Evidence from a large undergraduate sample[J]. International Journal of Environmental Research and Public Health, 2021, 18(7): 1-21.

[163] Joffe H & Elsey J W B. Free association in psychology and the grid elaboration method[J]. Review of General Psychology, 2014, 18(3): 173-185.

[164] Jones E, Trilling L & Marcus S. The life and work of Sigmund Freud: Vol. 119[M]. [One vol. ed]. New York: Basic Books, 1963.

[165] Jones S. Professional titles matter[J]. Nature Microbiology, 2018, 3(12): 1329-1329.

[166] Kadushin A. Men in a woman's profession[J]. Social Work, 1976, 21: 440-447.

[167] Kaiser A. Re-conceptualizing "sex" and "gender" in the human brain[J]. Zeitschrift für Psychologie, 2012, 220(2): 130-136.

[168] Kalina S. Quality in interpreting and its prerequisites: A framework for a comprehensive view[M]//Garzone G, Viezzi M. Interpreting in the 21st Century: Vol. 43. Amsterdam: John Benjamins Publishing Company, 2002: 121-130.

[169] Kamushadze T. Playing with title: How to be the elite in a socialist way[M]//Antonyan Y. Elites and "elites" : Transformations of social structures in post-Soviet Armenia and Georgia. Yerevan: Academic Swiss Caucasus Net, 2016: 245-273.

[170] Kanahara S. A review of the definitions of stereotype and a proposal for a progressional model[J]. Individual Differences Research, 2006, 4(5): 306-321.

[171] Kang O & Rubin D L. Reverse Linguistic Stereotyping: Measuring the Effect of Listener Expectations on Speech Evaluation[J]. Journal of Language and Social Psychology, 2009, 28(4): 441-456.

[172] Karlins M, Coffman T L & Walters G. On the fading of social stereotypes: Studies in three generations of college students[J]. Journal of Personality and Social Psychology, 1969, 13(1): 1-16.

[173] Katan D. Occupation or profession: A survey of the translators' world[J]. Translation and Interpreting Studies, 2009, 4(2): 187-209.

[174] Katz D & Braly K. Racial stereotypes of one hundred college students[J]. The Journal of Abnormal and Social Psychology, 1933, 28(3): 280-290.

[175] Kilpatrick S G, Manktelow K I & Over D E. Power of source as a factor in deontic inference[J]. Thinking & Reasoning, 2007, 13(3): 295-317.

[176] Kirby N H, Nettelbeck T & Goodenough S. Cognitive rigidity in the aged and the mentally retarded[J]. The International Journal of Aging and Human Development, 1979, 9(3): 263-272.

[177] Klapper A, Dotsch R, Van Rooij I, *et al.* Four meanings of "categorization": A conceptual analysis of research on person perception[J]. Social and Personality Psychology Compass, 2017, 11(8): e12336.

[178] Kopczyński A. Quality in conference interpreting: Some pragmatic problems[M]//Snell-Hornby M, Pöchhacker F, Kaindl K. Translation Studies: An Interdiscipline: Vol. 2. Amsterdam: John Benjamins Publishing Company, 1994: 189.

[179] Kotler P & Armstrong G. Principles of marketing[M]. 8. ed. Upper Saddle River, NJ:

Prentice Hall, 1999.

[180] Krueger J. Social categorization, psychology of[M]//Smelser N J, Baltes P B. International Encyclopedia of the Social & Behavioral Sciences. Oxford: Pergamon, 2001: 14219-14223.

[181] Krystallidou D, De Walle C V, Deveugele M, *et al.* Training "doctor-minded" interpreters and "interpreter-minded" doctors: The benefits of collaborative practice in interpreter training[J]. Interpreting / International Journal of Research and Practice in Interpreting, 2018, 20(1): 132-150.

[182] Kuefner D. Early visually evoked electrophysiological responses over the human brain (P1, N170) show stable patterns of face-sensitivity from 4 years to adulthood[J]. Frontiers in Human Neuroscience, 2010, 3.

[183] Kunda Z & Spencer S J. When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application.[J]. Psychological Bulletin, 2003, 129(4): 522-544.

[184] Kurz I. Conference interpretation: Expectations of different user groups[J]. The Interpreters' Newsletter, 1993(5): 13-21.

[185] Kurz I. Conference interpreting: Quality in the ears of the user[J]. Meta, 2002, 46(2): 394-409.

[186] Lamont R A, Swift H J & Abrams D. A review and meta-analysis of age-based stereotype threat: Negative stereotypes, not facts, do the damage.[J]. Psychology and Aging, 2015, 30(1): 180-193.

[187] Lee J. Rating scales for interpreting performance assessment[J]. The Interpreter and Translator Trainer, 2008, 2(2): 165-184.

[188] Lee S B. Developing an analytic scale for assessing undergraduate students' consecutive interpreting performances[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2015, 17(2): 226-254.

[189] Lee S B. Holistic assessment of consecutive interpretation: How interpreter trainers rate student performances[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2019, 21(2): 245-269.

[190] Lee S Y, Pitesa M, Thau S, *et al.* Discrimination in selection decisions: integrating stereotype fit and interdependence theories[J]. Academy of Management Journal, 2015, 58(3): 789-812.

[191] Lerner J & Lomi A. Knowledge categorization affects popularity and quality of Wikipedia articles[J]. PLOS ONE, 2018, 13(1): e0190674.

[192] Levy B, Ashman O & Dror I. To be or not to be: The effects of aging stereotypes on the will to live[J]. OMEGA - Journal of Death and Dying, 2000, 40(3): 409-420.

[193] Li J, Faisal E & Al Hariri A. Numbers for boys and words for girls? Academic gender stereotypes among chinese parents[J]. Sex Roles, 2022, 87(5-6): 306-326.

[194] Li J & McLellan R. Is language learning a feminine domain? Examining the content and stereotype threat effect of female-language stereotypes among EFL learners in China[J]. Contemporary Educational Psychology, 2021, 66: 101991.

[195] Li M & Luo Z. The 'bad women drivers' myth: the overrepresentation of female drivers and gender bias in China's media[J]. Information, Communication & Society, 2020, 23(5): 776-793.

[196] Li S, Huang G, Ma Z, *et al.* Superior bias in trust-related decisions[J]. Current Psychology, 2023, 42(28): 24822-24836.

[197] Li X. The costs of workplace favoritism: Evidence from promotions in chinese high schools[C]//Chinese Economy Working Group Meeting, Spring 2020. Zoom, 2018: 1-77.

[198] Liang K. A descriptive study of age identity among older adults in China[J]. China Journal of Social Work, 2014, 7(3): 305-317.

[199] Lim B Y & Fan C. Simultaneous interpreting with accented mandarin: A study on the perception of quality in Taiwan[J]. INContext: Studies in Translation and Interculturalism, 2023, 3(2): 109-138.

[200] Lin M H, Kwan V S Y, Cheung A, *et al.* Stereotype content model explains prejudice for an envied outgroup: Scale of anti-Asian american stereotypes[J]. Personality and Social Psychology Bulletin, 2005, 31(1): 34-47.

[201] Lindqvist A, Björklund F & Bäckström M. The perception of the poor: Capturing stereotype content with different measures[J]. Nordic Psychology, 2017, 69(4): 231-247.

[202] Linton R. The study of man: an introduction[M]. New York: Appleton Century Crofts, 1936.

[203] Linville P W, Fischer G W & Salovey P. Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation[J]. Journal of Personality and Social Psychology, 1989, 57(2): 165-188.

[204] Lippmann W. Public opinion[M]. 1st Free Press pbks. ed. New York: Free Press Paperbacks, 1922.

[205] Liu M. Design and analysis of Taiwan's interpretation certification examination[M]//Tsagari D, van Deemter R. Assessment issues in language translation and interpreting: Vol. 23. Frankfurt: Peter Lang GmbH, Internationaler Verlag der

Wissenschaften, 2015.

[206] Liu M. User expectations research revisited: Methodological considerations[M]//Zwischenberger C, Reithofer K, Rennert S. Benjamins Translation Library: Vol. 160. Amsterdam: John Benjamins Publishing Company, 2023: 44-66.

[207] Liu M, Kurz I, Moser-Mercer B, *et al.* The interpreter's aging: A unique story of multilingual cognitive decline?[J]. Translation, Cognition & Behavior, 2020, 3(2): 287-310.

[208] Logan G D. Automaticity, resources, and memory: Theoretical controversies and practical implications[J]. Human Factors: The Journal of the Human Factors and Ergonomics Society, 1988, 30(5): 583-598.

[209] López-Sáez M & Lisbona A. Descriptive and prescriptive features of gender stereotyping. Relationships among its components[J]. Revista de Psicología Social, 2009, 24(3): 363-379.

[210] Louw P & de Wet F. The perception and identification of accent in spoken Black South African English[J]. Southern African Linguistics and Applied Language Studies, 2007, 25(1): 91-105.

[211] Luo R, Teng K, Xu L, *et al.* A decomposition of student achievement gap by gender in China: Evidence from random class assignment[J]. International Journal of Educational Research, 2021, 106: 101721.

[212] Macías M. Probing quality criteria in simultaneous interpreting: The role of silent pauses in fluency[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2006, 8(1): 25-43.

[213] Macías M. Fluency[M]//Pöchhacker F, Grbić N, Mead P, et al. Routledge encyclopedia of interpreting studies. London ; New York: Routledge, Taylor & Francis Group, 2015: 165-166.

[214] Macrae C N & Bodenhausen G V. Social cognition: Thinking categorically about others[J]. Annual Review of Psychology, 2000, 51(1): 93-120.

[215] Macrae C N & Bodenhausen G V. Social cognition: Categorical person perception[J]. British Journal of Psychology, 2001, 92(1): 239-255.

[216] Macrae C Neil, Milne A B & Bodenhausen G V. Stereotypes as Energy-Saving Devices: A Peek Inside the Cognitive Toolbox[J]. Journal of Personality and Social Psychology, 1994, 66(1): 37-47.

[217] Macrae C. Neil, Milne A B & Bodenhausen G V. Stereotypes as energy-saving devices: A peek inside the cognitive toolbox.[J]. Journal of Personality and Social Psychology,

1994, 66(1): 37-47.

[218] Madon S, Guyll M, Aboufadel K, *et al.* Ethnic and national stereotypes: The Princeton trilogy revisited and revised[J]. Personality and Social Psychology Bulletin, 2001, 27(8): 996-1010.

[219] Makarova E, Aeschlimann B & Herzog W. The gender gap in STEM fields: The impact of the gender stereotype of math and science on secondary students' career aspirations[J]. Frontiers in Education, 2019, 4: 60.

[220] Marrone S. Quality: A shared objective[J]. The Interpreters' Newsletter, 1993, 5: 35-41.

[221] Martinez A D, Laird M D, Martin J A, *et al.* Job title inflation[J]. Human Resource Management Review, 2008, 18(1): 19-27.

[222] Maryns K. Disclosure and (re)performance of gender-based evidence in an interpreter-mediated asylum interview[J]. Journal of Sociolinguistics, 2013, 17(5): 661-686.

[223] Mason I & Ren W. Power in face-to-face interpreting events[J]. Translation and Interpreting Studies, 2012, 7(2): 234-253.

[224] Mason M F, Cloutier J & Macrae C N. On construing others: Category and stereotype activation from facial cues[J]. Social Cognition, 2006, 24(5): 540-562.

[225] Mattan B D, Kubota J T, Li T, *et al.* Implicit evaluative biases toward targets varying in race and socioeconomic status[J]. Personality and Social Psychology Bulletin, 2019, 45(10): 1512-1527.

[226] Mervis C B & Rosch E. Categorization of natural objects[J]. Annual Review of Psychology, 1981, 32(1): 89-115.

[227] Meteyard L & Davies R A I. Best practice guidance for linear mixed-effects models in psychological science[J]. Journal of Memory and Language, 2020, 112: 104092.

[228] Moors A & Houwer J de. Automaticity: a theoretical and conceptual analysis[J]. Psychological Bulletin, 2006, 132(2): 297-326.

[229] Morrison J H & Baxter M G. The ageing cortical synapse: Hallmarks and implications for cognitive decline[J]. Nature Reviews Neuroscience, 2012, 13(4): 240-250.

[230] Moser P. Expectations of users of conference interpretation[J]. Interpreting. International Journal of Research and Practice in Interpreting, 1996, 1(2): 145-178.

[231] Moser-Mercer B. Quality in interpreting: Some methodological issues[J]. The Interpreters' Newsletter, 1996(7): 43-55.

[232] Moss-Racusin C A, Phelan J E & Rudman L A. When men break the gender rules: Status incongruity and backlash against modest men[J]. Psychology of Men & Masculinity, 2010, 11(2): 140-151.

[233] Mouzourakis P. Remote interpreting: A technical perspective on recent experiments[J]. Interpreting / International Journal of Research and Practice in Interpreting, 2006, 8(1): 45-66.

[234] Mouzourakis T. That feeling of being there: Vision and presence in remote interpreting[EB/OL]. (2003-06-19). http://www.aiic.net/community/print/default.cfm/page1173.

[235] Moya M & Fiske S T. The social psychology of the great recession and social class divides[J]. Journal of Social Issues, 2017, 73(1): 8-22.

[236] Mullen B, Migdal M J & Hewstone M. Crossed categorization versus simple categorization and intergroup evaluations: a meta-analysis[J]. European Journal of Social Psychology, 2001, 31(6): 721-736.

[237] Nakahara C. Expanding the collective narrative: exploring the experiences of American sign language/English interpreters of Asian heritage[D]. Monmouth: Western Oregon University, 2016.

[238] Nelson D L, Mcevoy C L & Dennis S. What is free association and what does it measure?[J]. Memory & Cognition, 2000, 28(6): 887-899.

[239] Nelson T D. Handbook of prejudice, stereotyping, and discrimination[M]. Second edition. New York London: Psychology Press, Taylor & Francis Group, 2015.

[240] Nicodemus B, Liu M & McClure S. The reading habits of professional signed and spoken language interpreters[J]. Translation and Interpreting Studies, 2021, 17(2): 287-312.

[241] Nienstedt B C, Zatz M S & Epperlein T. Court processing and sentencing of drinking drivers: Using new methodologies[J]. Journal of Quantitative Criminology, 1988, 4(1): 39-59.

[242] Nier J A, Bajaj P, McLean M C, *et al.* Group status, perceptions of agency, and the correspondence bias: Attributional processes in the formation of stereotypes about high and low status groups[J]. Group Processes & Intergroup Relations, 2012, 16(4): 476-487.

[243] North M S & Fiske S T. An inconvenienced youth? Ageism and its potential intergenerational roots.[J]. Psychological Bulletin, 2012, 138(5): 982-997.

[244] Obasi C. Race and ethnicity in sign language interpreter education, training and practice[J]. Race Ethnicity and Education, 2013, 16(1): 103-120.

[245] Otten S & Wentura D. About the impact of automaticity in the minimal group paradigm: evidence from affective priming tasks[J]. European Journal of Social Psychology, 1999, 29: 1049-1071.

[246] Paladino M P & Mazzurega M. One of us: On the role of accent and race in real-time in-

group categorization[J]. Journal of Language and Social Psychology, 2019, 39(1): 22-39.

[247] Park K. The interpreter's delivery: A spotlight on user perceptions[J]. Lebende Sprachen, 2014, 59(2).

[248] Pattyn S, Rosseel Y, Overwalle F, *et al.* Social classification occurs at the subgroup level[J]. Social Psychology, 2015, 46(1): 13-23.

[249] Pauker K, Weisbuch M, Ambady N, *et al.* Not so black and white: Memory for ambiguous group members[J]. Journal of Personality and Social Psychology, 2009, 96(4): 795-810.

[250] Peng K & Dai H. A model of risk factors in remote interpreting[M]//Kubincová Z, Hao T, Capuano N, et al. Emerging Technologies for Education: Vol. 14607. Singapore: Springer Nature Singapore, 2024: 3-15.

[251] Petrochuk N O. Notion of an accent in linguistic studies[J]. Science and Education a New Dimension, 2021, IX(257)(75): 54-56.

[252] Phillips D L & Clancy K J. Some effects of "social desirability" in survey studies[J]. American Journal of Sociology, 1972, 77(5): 921-940.

[253] Pietraszewski D & Schwartz A. Evidence that accent is a dedicated dimension of social categorization, not a byproduct of coalitional categorization[J]. Evolution and Human Behavior, 2014, 35(1): 51-57.

[254] Plaks J E, Stroessner S J, Dweck C S, *et al.* Person theories and attention allocation: Preferences for stereotypic versus counterstereotypic information[J]. Journal of Personality and Social Psychology, 2001, 80(6): 876-893.

[255] Pöchhacker F. "Those who do...": A profile of research(ers) in interpreting[J]. Target / International Journal of Translation Studies, 1995, 7(1): 47-64.

[256] Pöchhacker F. Quality assessment in conference and community interpreting[J]. Meta, 2002, 46(2): 410-425.

[257] Pöchhacker F. Introducing interpreting studies[M]. 1st edition. London and New York: Routledge, 2004.

[258] Pöchhacker F. User expectations[M]//Grbić N, Mead P, Setton R, et al. Routledge encyclopedia of interpreting studies. First issued in paperback. London New York: Routledge, Taylor & Francis Group, 2019: 430-432.

[259] Pöchhacker F & Liu M. Interpreting technologized: Distance and assistance[J]. Interpreting / International Journal of Research and Practice in Interpreting, 2024, 26(2): 157-177.

[260] Povee K & Roberts L D. Attitudes toward mixed methods research in psychology: the

best of both worlds?[J]. International Journal of Social Research Methodology, 2014, 18(1): 41-57.

[261] Qi Y, Li Q & Du F. Are rich people perceived as more trustworthy? Perceived socioeconomic status modulates judgments of trustworthiness and trust behavior based on facial appearance[J]. Frontiers in Psychology, 2018, 9: 512.

[262] R Core Team. R: A language and environment for statistical computing[EB/OL]. 2020. https://www.R-project.org/.

[263] Ren W. The evolution of interpreters' perception and application of (codes of) ethics in China since 1949: a sociological and historical perspective[J]. The Translator, 2020, 26(3): 274-296.

[264] Renner A M & Masch L. Emotional woman – rational man? Gender stereotypical emotional expressivity of German politicians in news broadcasts[J]. Communications, 2019, 44(1): 81-103.

[265] Rennert S. The impact of fluency on the subjective assessment of interpreting quality[J]. The Interpreters' Newsletter, 2010(15): 101-115.

[266] Reyna C. Lazy, dumb, or industrious: When stereotypes convey attribution information in the classroom[J]. Educational Psychology Review, 2000, 12(1): 85-110.

[267] Riach P A & Rich J. An experimental investigation of sexual discrimination in hiring in the English labor market[J]. The B.E. Journal of Economic Analysis & Policy, 2006, 6(2): 0000102202153806371416.

[268] Riccardi A. Evaluation in interpretation: Macrocriteria and microcriteria[M]//Hung E. Teaching Translation and Interpreting 4: Vol. 42. Amsterdam: John Benjamins Publishing Company, 2002: 115-126.

[269] Rice L & Barth J M. A tale of two gender roles: The effects of implicit and explicit gender role traditionalism and occupational stereotype on hiring decisions[J]. Gender Issues, 2017, 34(1): 86-102.

[270] Robb N & Greenhalgh T. "You have to cover up the words of the doctor": The mediation of trust in interpreted consultations in primary care[J]. Journal of Health Organization and Management, 2006, 20(5): 434-455.

[271] Roccas S & Brewer M B. Social identity complexity[J]. Personality and Social Psychology Review, 2002, 6(2): 88-106.

[272] Roets A, Kruglanski A W, Kossowska M, *et al.* The motivated gatekeeper of our minds: New directions in need for closure theory and research[M]//Olson J M, Zanna M P. Advances in Experimental Social Psychology: Vol. 52. Academic Press, 2015: 221-283.

[273] Ron Y, Dafni-Merom A, Saadon-Grosman N, *et al.* Brain system for social categorization by narrative roles[J]. The Journal of Neuroscience, 2022, 42(26): 5246-5253.

[274] Rosch E H. Natural categories[J]. Cognitive Psychology, 1973, 4(3): 328-350.

[275] Rossion B & Jacques C. The N170: Understanding the time course of face perception in the human brain[M]//The Oxford handbook of event-related potential components. New York, NY, US: Oxford University Press, 2012: 115-141.

[276] Rothermund K & De Paula Couto M C P. Age stereotypes: Dimensions, origins, and consequences[J]. Current Opinion in Psychology, 2024, 55: 101747.

[277] Rudman L A & Fairchild K. Reactions to counterstereotypic behavior: The role of backlash in cultural stereotype maintenance[J]. Journal of Personality and Social Psychology, 2004, 87(2): 157-176.

[278] Rudman L A & Glick P. Prescriptive gender stereotypes and backlash toward agentic women[J]. Journal of Social Issues, 2001, 57(4): 743-762.

[279] Rudman L A & Phelan J. Backlash effects for disconfirming gender stereotypes in organizations[J]. Research in Organizational Behavior, 2008, 28: 61-79.

[280] Rule N O & Sutherland S L. Social categorization from faces: Evidence from obvious and ambiguous groups[J]. Current Directions in Psychological Science, 2017, 26(3): 231-236.

[281] Russo M. Simultaneous film interpreting and users' feedback[J]. Interpreting / International Journal of Research and Practice in Interpreting, 2005, 7(1): 1-26.

[282] Ryan E B & Capadano H L. Age perceptions and evaluative reactions toward adult speakers[J]. Journal of Gerontology, 1978, 33(1): 98-102.

[283] Saeed M A, González E R, Korybski T, *et al.* Connected yet distant: An experimental study into the visual needs of the interpreter in remote simultaneous interpreting[M]//Kurosu M. Human-Computer Interaction. User Experience and Behavior: Vol. 13304. Cham: Springer International Publishing, 2022: 214-232.

[284] Schaller M. Social categorization and the formation of group stereotypes: Further evidence for biased information processing in the perception of group-behavior correlations[J]. European Journal of Social Psychology, 1991, 21(1): 25-35.

[285] Scheepers D, Spears R, Doosje B, *et al.* Diversity in in-group bias: structural factors, situational features, and social functions.[J]. Journal of Personality and Social Psychology, 2006, 90(6): 944-960.

[286] Schjoldager A. Assessment of simultaneous interpreting[M]//Dollerup C, Appel V. Teaching Translation and Interpreting 3: Vol. 16. Amsterdam: John Benjamins Publishing

Company, 1996: 187.

[287] Schmidt D F & Boland S M. Structure of perceptions of older adults: Evidence for multiple stereotypes.[J]. Psychology and Aging, 1986, 1(3): 255-260.

[288] Schnabel K & Asendorpf J B. Free associations as a measure of stable implicit attitudes[J]. European Journal of Personality, 2013, 27(1): 39-50.

[289] Scott M L, Bone S A, Christensen G L, *et al.* Revealing and mitigating racial bias and discrimination in financial services[J]. Journal of Marketing Research, 2023: 00222437231176470.

[290] Sczesny S, Nater C & Eagly A H. Agency and Communion[M]//Abele A E, Wojciszke B. Agency and Communion in Social Psychology. 1st ed. Routledge, 2018: 103-116.

[291] Sebastian M, Banate R M & Saquin M I. Gender roles among public elementary teachers: Basis for gender-responsive intervention activities[J]. International Online Journal of Primary Education, 2022, 11(2): 401-411.

[292] Seeber K G. Cognitive load in simultaneous interpreting: Existing theories — new models[J]. Interpreting. International Journal of Research and Practice in Interpreting, 2011, 13(2): 176-204.

[293] Seeber K G. Multimodal processing in simultaneous interpreting[M]//Schwieter J W, Ferreira A. The Handbook of Translation and Cognition. 1st ed. Wiley, 2017: 461-475.

[294] Seeber K G & Fox A B. Distance conference interpreting[M]//Albl-Mikasa M, Tiselius E. The Routledge Handbook of Conference Interpreting. 1st ed. London: Routledge, 2021: 491-507.

[295] Sehnert S, Franks B, Yap A J, *et al.* Scarcity, engagement, and value[J]. Motivation and Emotion, 2014, 38(6): 823-831.

[296] Seiter E. Stereotypes and the media: A re-evaluation[J]. Journal of Communication, 1986, 36(2): 14-26.

[297] Settles I H, Jellison W A & Pratt-Hyatt J S. Identification with multiple social groups: The moderating role of identity change over time among women-scientists[J]. Journal of Research in Personality, 2009, 43(5): 856-867.

[298] Setton R & Dawrant A. Conference interpreting – A trainer's guide: Vol. 121[M]. Amsterdam: John Benjamins Publishing Company, 2016.

[299] Setton R & Guo L. Attitudes to role, status and professional identity in interpreters and translators with Chinese in Shanghai and Taipei[J]. Translation and Interpreting Studies, 2009, 4(2): 210-238.

[300] Shannon C E & Weaver W. The mathematical theory of communication[M]. Urbana:

Univ. of Illinois Press, 1998.

[301] Shaw L B. Differing prospects for women and men: young old-age, old old-age, and elder[J]. SSRN Electronic Journal, 2006.

[302] Shlesinger M. Quality in simultaneous interpreting[M]//Gambier Y, Gile D, Taylor C. Benjamins Translation Library: Vol. 23. Amsterdam: John Benjamins Publishing Company, 1997: 123.

[303] Sigelman L & Tuch S A. Metastereotypes: Blacks' perceptions of Whites' stereotypes of Blacks[J]. Public Opinion Quarterly, 1997, 61(1, Special Issue on Race): 87.

[304] Signorelli T M, Haarmann H J & Obler L K. Working memory in simultaneous interpreters: Effects of task and age[J]. International Journal of Bilingualism, 2011, 16(2): 198-212.

[305] Silva S & Couto M J. Gender bias in the hospitality sector: Female and male jobs[C]//International Conference on Tourism Research: Vol. 6. Pafos, Cyprus, 2023: 292-297.

[306] Smeding A. Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance[J]. Sex Roles, 2012, 67(11-12): 617-629.

[307] Snytte J, Setton R, Mwilambwe-Tshilobo L, *et al.* Structure-function interactions in the hippocampus and prefrontal cortex are associated with episodic memory in healthy aging[J]. eneuro, 2024, 11(3): ENEURO.0418-23.2023.

[308] Spears R, Gordijn E, Dijksterhuis A, *et al.* Reaction in action: Intergroup contrast in automatic behavior[J]. Personality and Social Psychology Bulletin, 2004, 30(5): 605-616.

[309] Spencer S J, Fein S, Wolfe C T, *et al.* Automatic activation of stereotypes: The role of self-image threat[J]. Personality and Social Psychology Bulletin, 1998, 24(11): 1139-1152.

[310] Stangor C. The study of stereotyping, prejudice, and discrimination within social psychology: A quick history of theory and research[M]//Handbook of prejudice, stereotyping, and discrimination. 2nd Edition. New York: Psychology Press, 2009: 3-27.

[311] Stevenson A. Oxford dictionary of English[M]. 3rd ed. New York, NY: Oxford University Press, 2010.

[312] Stewart M A, Ryan E B & Giles H. Accent and social class effects on status and solidarity evaluations[J]. Personality and Social Psychology Bulletin, 1985, 11(1): 98-105.

[313] Stolier Ryan M & Freeman J B. Neural pattern similarity reveals the inherent intersection of social categories[J]. Nature Neuroscience, 2016, 19(6): 795-797.

[314] Stolier Ryan M. & Freeman J B. The neuroscience of social vision[M]//Absher J R, Cloutier J. Neuroimaging Personality, Social Cognition, and Character. San Diego: Academic Press, 2016: 139-157.

[315] Strong M & Rudser S F. The subjective assessment of sign language interpreters[J]. Sign Language Studies, 1986, 53(1): 299-314.

[316] Swanson J E, Swanson E & Greenwald A G. Using the Implicit Association Test to investigate attitude-behaviour consistency for stigmatised behaviour[J]. Cognition & Emotion, 2001, 15(2): 207-230.

[317] Taylor S E. The interface of cognitive and social psychology[M]//Harvey J H. Cognition, social behavior, and the environment. Hillsdale, N.J: L. Erlbaum, 1981: 189-211.

[318] Teas R K. Expectations, performance evaluation, and consumers' perceptions of quality[J]. Journal of Marketing, 1993, 57(4): 18.

[319] The ChinaMAP Consortium, Cao Y, Li L, *et al.* The ChinaMAP analytics of deep whole genome sequences in 10,588 individuals[J]. Cell Research, 2020, 30(9): 717-731.

[320] Thorndike E. A constant error in psychological ratings[J]. Journal of Applied Psychology, 1920, 4(1): 25-29.

[321] Timmermans S & Epstein S. A world of standards but not a standard world: Toward a sociology of standards and standardization[J]. Annual Review of Sociology, 2010, 36(1): 69-89.

[322] Tiselius E & Jenset G B. Process and product in simultaneous interpreting: What they tell us about experience and expertise[M]//Alvstad C, Hild A, Tiselius E. Methods and Strategies of Process Research: Vol. 94. Amsterdam: John Benjamins Publishing Company, 2011: 269-300.

[323] Torgrimson B N & Minson C T. Sex and gender: what is the difference?[J]. Journal of Applied Physiology, 2005, 99(3): 785-787.

[324] Tsang S. Cognitive discrepancy, dissonance, and selective exposure[J]. Media Psychology, 2017, 22(3): 394-417.

[325] Wadensjö C. Interpreting as interaction[M]. Transferred to digital print. London: Longman, 2005.

[326] Wang B. Evolution of interpreting as a social practice in China during the past four decades. An analysis of Chinese discourse on interpreting through the lens of social practice theory[J]. Interpreting and Society, 2023, 3(1): 24-45.

[327] Wang S & Yang D. The effects of poverty stereotype threat on inhibition ability in individuals from different income-level families[J]. Brain and Behavior, 2020, 10(12):

e01770.

[328] Wang-Jones T "Tie" S, Hauson A O, Ferdman B M, *et al.* Comparing implicit and explicit attitudes of gay, straight, and non-monosexual groups toward transmen and transwomen[J]. International Journal of Transgenderism, 2018, 19(1): 95-106.

[329] Ward L M & Grower P. Media and the development of gender role stereotypes[J]. Annual Review of Developmental Psychology, 2020, 2(1): 177-199.

[330] Wattenmaker W D, Mcquaid H L & Schwertz S J. Analogical versus rule-based classification[J]. Memory & Cognition, 1995, 23(4): 495-509.

[331] Weeks M & Leavitt P A. Using occupational titles to convey an individual's location in social stratification dimensions[J]. Basic and Applied Social Psychology, 2017, 39(6): 342-357.

[332] Weeks M & Lupfer M B. Complicating race: The relationship between prejudice, race, and social class categorizations[J]. Personality and Social Psychology Bulletin, 2004, 30(8): 972-984.

[333] Weisman K, Johnson M V & Shutts K. Young children's automatic encoding of social categories[J]. Developmental Science, 2015, 18(6): 1036-1043.

[334] Weiss I. Interest in working with the elderly: A cross-national study of graduating social work students[J]. Journal of Social Work Education, 2005, 41(3): 379-391.

[335] Wen F, Zuo B, Wang Y, *et al.* The (continuous) nature of perceived gender counter-stereotype: A threshold model of gender stereotype maintenance[J]. Archives of Sexual Behavior, 2020, 49(7): 2511-2530.

[336] Wiese H, Schweinberger S R & Hansen K. The age of the beholder: ERP evidence of an own-age bias in face memory[J]. Neuropsychologia, 2008, 46(12): 2973-2985.

[337] Wilbourn M P & Kee D W. Henry the nurse is a doctor too: Implicitly examining children's gender stereotypes for male and female occupational roles[J]. Sex Roles, 2010, 62(9-10): 670-683.

[338] Williams A & Giles H. Sociopsychological perspectives on older people's language and communication[J]. Ageing and Society, 1991, 11(2): 103-126.

[339] Wilson L & Walsh D. Striving for impartiality: Conflicts of role, trust and emotion in interpreter-assisted police interviews[J]. Pragmatics and Society, 2019, 10(1): 122-151.

[340] Wilson P R. Perceptual distortion of height as a function of ascribed academic status[J]. The Journal of Social Psychology, 1968, 74(1): 97-102.

[341] Wright S C. Cross-group contact effects[M]//Otten S, Sassenberg K, Kessler T. Intergroup Relations: The Role of Motivation and Emotion (A Festschrift for Amélie

Mummendey). Psychology Press, 2009: 293-315.

[342] Xu R. “When boys become the second sex”: The new gender gap among chinese middle school students[J]. The Yale Undergraduate Research Journal, 2020, 1(1): 6.

[343] Yagura H, Tanaka H, Kinoshita T, *et al.* Selective attention measurement of experienced simultaneous interpreters using EEG phase-locked response[J]. Frontiers in Human Neuroscience, 2021, 15: 581525.

[344] Yang Y. “Feel sorry for Miss translator!!!”: A danmu-based case study of Bilibili users’ parasocial interactions about the ad hoc interpreter on screen[J]. Interpreting, 2023, 25(1): 61-86.

[345] Zdravkovska E. Most common problems the interpreters face while working with medical terminology[J]. Knowledge International Journal, 2019, 31(6): 1777-1781.

[346] Zhang X, Li Q, Sun S, *et al.* The time course from gender categorization to gender-stereotype activation[J]. Social Neuroscience, 2016, 13(1): 52-60.

[347] Zhang Y & Tsang M. Gender gap in the National College Entrance Exam performance in China: A case study of a typical Chinese municipality[J]. Asia Pacific Education Review, 2014, 16(1): 27-36.

[348] Zheng B & Xiang X. Between invisibility and over-visibility: Self-perception and user expectations of liaison interpreters in business settings[J]. Babel / Revue internationale de la traduction / International Journal of Translation, 2018, 64(1): 1-32.

[349] Zuo B, Wen F, Wang M, *et al.* The mediating role of cognitive flexibility in the influence of counter-stereotypes on creativity[J]. Frontiers in Psychology, 2019, 10: 105.

[350] Zwischenberger C. Quality criteria in simultaneous interpreting: An international vs. a national view[J]. The Interpreters’ Newsletter, 2010(15): 127-142.

[351] Zwischenberger C. Professional self-perception of the social role of conference interpreters[M]//Biagini M, Boyd M S, Monacelli C. The changing role of the interpreter: Contextualising norms, ethics and quality standards. 1st ed. Routledge, 2017.

[352] 程梦瑶，黄娇娇和张雨．男女比例不均对外语类专业学生的影响——以华中师大外院为例[J]．英语广场(学术研究)，2014（7）：120-123．

[353] 崔琳杰.高校外语专业“女强男弱”现象的原因及对策分析[J].长春教育学院学报，2015，31（5）：87-88．

[354] 范志嘉．用户教育:口译质量评估的隐性参数[J]．解放军外国语学院学报，2018，41（1）：150-158．

[355] 方彬楠和冉黎黎．在华常住外国人恢复至 2019 年 85%[N]．北京商报，2024-01-19（002）．

[356] 风笑天．方法论背景中的问卷调查法[J]．社会学研究，1994（3）：13-18．
[357] 国家统计局．2024 年居民收入和消费支出情况[EB/OL]．（2025-01-17）[2025-02-15]．https://www.stats.gov.cn/sj/zxfb/202501/t20250117_1958325.html．
[358] 国务院第七次全国人口普查领导小组办公室．中国人口普查年鉴 2020[M]．Di 1 ban．北京：中国统计出版社，2022．
[359] 胡志海．大学生职业性别刻板印象的内隐研究[J]．心理科学，2005（5）：1122-1125．
[360] 黄敏和王斌华．口译资格考试的评分员培训机制研究——CATTI 考试二十年回顾[J]．外语电化教学，2024（3）：56-61+116．
[361] 纪莉和吴逸悠．口音歧视与社会群体的文化规训——以 30 年春晚小品的口音研究为例[J]．现代传播(中国传媒大学学报)，2015，37（07）：81-85+146．
[362] 李莲莲．语言标记与性别刻板印象的发展演变——以新闻报道中若干女性群体标记为例[Z]//青年记者．2016：41-42．
[363] 李强．转型时期冲突性的职业声望评价[J]．中国社会科学，2000（4）：100-111+206．
[364] 李强和刘海洋．变迁中的职业声望——2009 年北京职业声望调查浅析[J]．学术研究，2009（12）：34-42．
[365] 李天韵和崔英．内隐社会认知路径的口译质量感知偏差研究[J]．外国语(上海外国语大学学报)，2022，45（5）：100-109．
[366] 李廷茹．行走在男性和女性边缘——女博士“第三性”现象文化解读[J]．兰州学刊，2008（01）：77-80．
[367] 李欣璐，段夏媛和张晓斌．他人知觉过程中群体刻板印象激活的机制[J]．中国社会心理学评论，2023（1）：174-188+246．
[368] 李志民．中国院士多了还是少了[N]．中国青年报，2015-11-10（2）．
[369] 廖熠佳．试论女司机刻板印象的形成及应对措施[J]．新闻研究导刊，2019，10（02）：90．
[370] 刘劼成和赵西亮．家长的数学-性别刻板印象与子女教育投入[J]．教育经济评论，2024，9（4）：49-70．
[371] 刘晅和佐斌．性别刻板印象维护的心理机制[J]．心理科学进展，2006（3）：456-461．
[372] 刘益光．人类智能优势在口译中的体现[D]．浙江大学，2021．
[373] 吕小康和刘颖．医生角色的刻板印象及其在医患群体间的差异[J]．南京师大学报(社会科学版)，2018（1）：85-93．
[374] 马莉萍，由由，熊煜，等．大学生专业选择的性别差异——基于全国 85 所高校的调查研究[J]．高等教育研究，2016，37（5）：36-42．
[375] 秦启文和余华．性别角色刻板印象的调查[J]．心理科学，2001（5）：593-594．

[376] 全国翻译专业资格（水平）考试办．英语二级翻译口笔译考试大纲[M]．北京：外文出版社，2004．

[377] 任文和伊安·梅森．对话口译中的权力——口译社会学解读[J]．四川大学学报(哲学社会科学版)，2011（6）：61-69．

[378] 邵张旻子．同声传译局部认知负荷研究[D]．上海：上海外国语大学，2018．

[379] 宋强和林新奇．浅议“头衔通胀”[J]．企业管理，2013（4）：34-35．

[380] 宋全成．外国人及港澳台居民在中国大陆的人口社会学分析[J]．山东大学学报(哲学社会科学版)，2013（2）：89-99．

[381] 宋全成和熊方洲.外国人及港澳台居民在中国内地(大陆)的状况变动研究——基于2010—2020 年全国人口普查数据的人口社会学比较分析[J].厦门大学学报(哲学社会科学版)，2024，74（3）：31-44．

[382] 滕越洋．地方国企高管晋升机制研究[D]．济南：山东大学，2019．

[383] 田志鹏，邝继浩和刘爱玉．社会转型时期大学生职业声望评价——以北京大学本科生调查为例[J]．青年研究，2013（5）：1-10+94．

[384] 王巍巍．口译教学体系中的质量评估——广外口译专业教学体系理论与实践(之五)[J]．中国翻译，2017，38（4）：45-52．

[385] 温芳芳和佐斌．社会分类的概念、线索及影响机制[J]．心理科学，2019，42（02）：395-401．

[386] 温芳芳和佐斌．刻板印象威胁研究[M]．武汉：华中师范大学出版社，2021．

[387] 吴泽萍和青立花.珠三角地区英语口译人才需求调查报告[J].考试周刊,2015(03)：67-69．

[388] 新华社．中共中央国务院中央军委关于授予翟志刚同志“航天英雄”刘伯明、景海鹏同志“英雄航天员”荣誉称号并颁发“航天功勋奖章”的决定[J]．载人航天，2008（4）：12．

[389] 新华社．延迟退休改革文件解读[J]．工友，2024（10）：24-26．

[390] 徐琳宏，林鸿飞，潘宇，等．情感词汇本体的构造[J]．情报学报，2008，27（2）．

[391] 杨承淑，笹岡敦子和詹成．逐步口译中的非语言讯息结构[J]．编译论丛，2011，4（2）．

[392] 杨亚平，王沛，尹志慧，等．刻板印象激活的无意图性及其大脑神经活动特征[J]．心理学报，2015，47（4）：488-502．

[393] 姚艳梅．高校外语专业男生学习弱势的原因及对策[J]．河北大学学报(哲学社会科学版)，2011，36（1）：138-140．

[394] 应小丽和蒋国勇．对高师院校男女生性别比失衡的思考——以浙江省高师院校为例[J]．青年研究，2002（3）：27-31．

[395] 藻灵．科学家颜宁:科研是最美的童话[J]．新青年(珍情)，2014（11）：14-15．
[396] 张东亮．女翻译官张京：再一次凭实力“出圈”[J]．廉政瞭望，2021（6）：46-47．
[397] 张凤兰．专业口译需要字正腔圆吗？[J]．编译论丛，2009，2（1）．
[398] 张海洋和蔡航．头衔的价值——来自网络借贷的证据[J]．经济学(季刊)，2018，17（4）：1525-1556．
[399] 张威和柯飞．从口译用户看口译质量评估[J]．外语学刊，2008（3）：114-118．
[400] 张莹瑞和佐斌．社会认同理论及其发展[J]．心理科学进展，2006（3）：475-480．
[401] 赵延东和 Pedersen J．受访者推动抽样:研究隐藏人口的方法与实践[J]．社会，2007（2）：192-205+208．
[402] 郑世林，周黎安和何维达．电信基础设施与中国经济增长[J]．经济研究，2014，49（5）：77-90．
[403] 中国翻译协会．翻译服务：口译服务要求[S]．2018．
[404] 中国翻译协会．2018 中国语言服务行业发展报告[M]．北京：外文出版社，2019．
[405] 中国翻译协会．2022 中国翻译人才发展报告[R]．北京：中国翻译协会，2022：1-34．
[406] 中国科学院．中国科学院院士章程[J]．中国科学院院刊，1994（3）：227-229．
[407] 周晓杰．网络问卷调查平台问卷星运营研究[D]．湖南大学，2020．
[408] 宗刚，李盼道和孙晨晨．改革开放以来我国职业声望排序及变迁研究[J]．北京工业大学学报(社会科学版)，2016，16（02）：11-17．
[409] 佐斌．刻板印象内容与形态[M]．武汉：华中师范大学出版社，2015．
[410] 佐斌，温芳芳，宋静静，等．社会分类的特性、维度及心理效应[J]．心理科学进展，2019，27（1）：141-148．
[411] 佐斌，温芳芳和朱晓芳．大学生对年轻人和老年人的年龄刻板印象[J]．应用心理学，2007（3）：231-236．
[412] 佐斌，张阳阳，赵菊，等．刻板印象内容模型:理论假设及研究[J]．心理科学进展，2006（1）：138-145．

Appendices

Appendix 1. The questionnaire on the public stereotype of a professional interpreter in China

社会公众对我国口译员的职业印象调查研究 *(titled as "A study on the public perception of the occupational image of interpreters in China")*

(This survey was written in Simplified Chinese, and the accompanying English content is its translation)

您好，我是山东大学翻译学院翻译研究所的李天韵博士，现就中国公众对口译员的职业声望和社会印象进行调研。问卷共 30 题左右，占用您约 8 分钟时间。在您认真填写问卷后，将有机会收到 10 元红包（共 50 份）作为答谢。您的参与将为我国口译员的职业形象研究作出发展贡献。

Dear Sir, madam or to whom it may concern,

I am Tianyun Li, a PhD candidate from the Institute of Translation Studies at the School of Translation Studies, Shandong University. I am conducting a survey on the occupational prestige and social perceptions of interpreters among the Chinese public. The questionnaire consists of approximately 30 questions and will take about 8 minutes of your time. Upon your careful completion of the questionnaire, you will have an opportunity to receive a CNY￥10 red envelope (a total of 50 available) as a token of appreciation. Your participation will contribute to the advancement of research on the professional image of interpreters in China.

1、您的性别 [单选题] *:

What is your gender? [single choice question]

○男 Male	○女 Female	○自定 _____ * Other (please specify):

2、您的年龄 [填空题] *: _____

What is your age? [fill-in-the-blank question]:

3、您的政治面貌 [单选题] *:

What is your political affiliation? [single choice question]:

○群众 Grass-root public	○共产党员 Communist Party member	○民主党派 Democratic party member
○共青团员（28 岁以下） Communist Youth League member (under 28)	○经组织部认定的无党派人士 Non-party personnel recognized by the Organization Department of the CPC	

4、您的长住地区 [填空题] *：_____

Where do you live? [fill-in-the-blank question]

5、您是在校学生吗？ [单选题] *

Are you currently a student? [single choice question]

○我不是学生 I am not a student	○我是学生 I am a student

6、您的职业 [填空题] *：_____

What is your occupation? [fill-in-the-blank question]

（依赖于“您是在校学生吗？”的第 1 个选项）

(answer only if the participant selected "I am not a student" in question 5)

7、您的父亲或母亲从事的职业（取成就较高者） [填空题] *：_____

What is the occupation of your parents or guardians? [fill-in-the-blank question]

（依赖于“您是在校学生吗？”的第 2 个选项）

(answer only if the participant selected "I am a student" in question 5)

8、您的家庭年收入约为（万元） [填空题] *：_____

What is your family’s annual income (in ten thousand yuan)? [fill-in-the-blank question]

（依赖于“您是在校学生吗？”的第 2 个选项）

(answer only if the participant selected "I am a student" in question 5)

9、您的职称（如有） [单选题]：

What is your professional title, if any? [single choice question]

（依赖于“您是在校学生吗？”的第 1 个选项）
(answer only if the participant selected "I am not a student" in question 5)

○正高级 Senior level	○副高级 Deputy senior level	○中级 Intermediate level	○初级 Junior level

10、 您的稳定年收入约为（万元） [填空题] *：_____

What is your annual income (in ten thousand Chinese Yuan)? [fill-in-the-blank question]

（依赖于“您是在校学生吗？”的第 1 个选项）
(answer only if the participant selected "I am not a student" in question 5)

11、 您的最高学历 [单选题] *：

What is your highest level of education? [single choice question]

○小学及以下 Primary school or below	○初中 Middle school	○高中/中专 High school/vocational school	
○专科 Associate degree	○本科 Bachelor's degree	○硕士 Master's degree	○博士 Doctorate

12、 您最高学历的专业 [单选题] *：

What is your field of study for your highest degree? [single choice question]

（依赖于“您的最高学历”的第 4;5;6;7 个选项）
(Dependent on the 4th, 5th, 6th, or 7th option of "What is your highest level of education?")

○文学 Literature	○历史学 History	○哲学 Philosophy	○经济学 Economics	○管理学 Management	○法学 Law
○教育学 Education	○艺术学 Arts	○理学 Science	○工学 Engineering	○医学 Medicine	○农学 Agriculture

13、 您会的外语 [多选题] *：

What foreign language(s) do you speak? [multiple choice question]

□我不会外语 Not any	□英语 English	□法语 French	□俄语 Russian	□德语 German	□日语 Japanese
□韩国语 Korean	□阿拉伯语 Arabic	□西班牙语 Spanish	□其他语言 _____* Other (please specify) _____*		

14、 您的英语水平 [单选题] *:

What is your proficiency in English? [single choice question]

（依赖于“您会的外语”的第 2 个选项）

(Dependent on the 2nd option of "What foreign languages do you speak?")

○我没有考级 Not any	○大学四级 College English Test Band 4	○大学六级 College English Test Band 6
○专业四级 Test for English Majors Grade 4	○专业八级 Test for English Majors Grade 8	○其他 _____ * Other: _____

15、 您的法语水平 [单选题] *:

What is your proficiency in French? [single choice question]

（依赖于“您会的外语”的第 3 个选项）

(Dependent on the 3rd option of "What foreign languages do you speak?")

○我没有考级 Not any	○大学四级 College French Test Band 4	○大学六级 College French Test Band 6
○专业四级 Test for French Majors Band 4	○专业八级 Test for French Majors Band 8	○其他 _____ * Other: _____
○A1/A2	○B1/B2	○C1/C2

16、 您的俄语水平 [单选题] *:

What is your proficiency in Russian? [single choice question]

（依赖于“您会的外语”的第 4 个选项）

(Dependent on the 4th option of "What foreign languages do you speak?")

○我没有考级 Not any	○大学四级 College Russian Test Band 4	○大学六级 College Russian Test Band 6
○专业四级 Test for Russian Majors Band 4	○专业八级 Test for Russian Majors Band 8	○其他 _____ * Other: _____
○A1/A2	○B1/B2	○C1/C2

17、 您的德语水平 [单选题] *:

What is your proficiency in Russian? [single choice question]

（依赖于“您会的外语”的第 5 个选项）

(Dependent on the 5th option of "What foreign languages do you speak?")

○我没有考级 Not any	○大学四级 College German Test Band 4	○大学六级 College German Test Band 6
○专业四级 Test for German Majors Band 4	○专业八级 Test for German Majors Band 8	○其他 _____ * Other: _____
○A1/A2	○B1/B2	○C1/C2

18、 您的日本语水平 [单选题] *:

What is your proficiency in Japanese? [single choice question]

（依赖于“您会的外语”的第 6 个选项）

(Dependent on the 6th option of "What foreign languages do you speak?")

○我没有考级 Not any	○N1	○N2	○N3	○N4	○N5	○其他 _____ * Other: _____

19、 您的韩国语水平 [单选题] *:

What is your proficiency in Korean? [single choice question]

（依赖于“您会的外语”的第 7 个选项）

(Dependent on the 7th option of "What foreign languages do you speak?")

○我没有考级 ○Not any	○TOPIK 初级 ○TOPIK beginner	○TOPIK 中级 ○TOPIK intermediate	○TOPIK 高级 TOPIK advanced
○专业四级 Test for Korean Majors Band 4	○专业八级 Test for Korean Majors Band 8	○其他 _____ * Other: _____	

20、 您的阿拉伯语水平 [单选题] *:

What is your proficiency in Arabic? [single choice question]

（依赖于“您会的外语”的第 8 个选项）

(Dependent on the 8th option of "What foreign languages do you speak?")

○我没有考级 Not any	○专业四级 Test for Arabic Majors Band 4	○专业八级 Test for Arabic Majors Band 8	○其他 _____ * Other: _____

21、 您的西班牙语水平 [单选题] *:

What is your proficiency in Spanish? [single choice question]

（依赖于“您会的外语”的第 9 个选项）

(Dependent on the 9th option of "What foreign languages do you speak?")

○我没有考级 Not any	○大学四级 College Spanish Test Band 4	○大学六级 College Spanish Test Band 6
○专业四级 Test for Russian Majors Band 4	○专业八级 Test for Spanish Majors Band 8	○其他 _____ * Other: _____
○A1/A2	○B1/B2	○C1/C2

22、 您的外语水平 [填空题] *: _____

What is your proficiency in other foreign languages? [fill-in-the-blank question]

（依赖于“您会的外语”的第 10 个选项）

(Dependent on the 10th option of "What foreign languages do you speak?")

23、 您通常从哪里了解口译员群体 [多选题] *:

From where do you usually learn about the interpreting profession? [multiple choice question]

□新闻报刊（如《参考消息》） Newspapers (e.g., *Reference News*)
□综艺报刊（如《知音》） Entertainment magazines (e.g., *Zhiyin*)
□新闻节目（如《央视新闻》） News programs (e.g., CCTV News)
□综艺节目（如《非你莫属》） Variety shows (e.g., Only You)
□微信文章（如《涵涵妈妈育儿经》） WeChat articles (e.g., Parenting tips by Hanhan's Mom)
□短视频（如抖音、快手、微信视频号等） Short videos (e.g., TikTok, Kwai, WeChat video channels, etc.)
□与口译员本人有接触 Direct contact with interpreters
□与朋友闲聊 Casual conversations with friends

24、 您接触口译员的频率 [单选题] *:

How frequently do you encounter interpreters? [single choice question]

○没有接触过 I have never encountered interpreters
○只接触过一到两次 I have encountered them only once or twice
○每年一到两次 Once or twice per year
○每年三到十次 Three to ten times per year

○每年十次到五十次 Ten to fifty times per year
○每年五十次到一百次 Fifty to one hundred times per year
○每年一百次以上 More than one hundred times per year

25、 您接触口译员的场合是 [多选题] *:

In what settings do you encounter interpreters? [multiple choice question]

（依赖于“您接触口译员的频率”的第 2;3;4;5;6;7 个选项）

(Dependent on the 2nd, 3rd, 4th, 5th, 6th, or 7th option of "How frequently do you encounter interpreters?")

□线下会议 Offline conferences	□网络会议 Online conferences	□线下会展 Offline exhibitions
□个人私交 Personal interactions	□其他 _____ Other: _____	

26、 您接触到的口译员 [单选题] *:

What is the gender distribution of the interpreters **you have encountered**? [single choice question]

（依赖于“您接触口译员的频率”的第 2;3;4;5;6 个选项）

(Dependent on the 2nd, 3rd, 4th, 5th, or 6th option of "How frequently do you encounter interpreters?")

○绝大部分是女性 The vast majority are female
○主要是女性，有少数男性 Mostly females with a few males
○男性和女性均等 An equal number of males and females
○主要是男性，有少数女性 Mostly male with a few females
○绝大部分是男性 The vast majority are male

27、 您印象中口译员 [单选题] *:

What is **your perception** of the gender distribution among interpreters? [single choice question]

○绝大部分是女性 The vast majority are female
○主要是女性，有少数男性 Mostly females with a few males
○男性和女性均等 An equal number of males and females
○主要是男性，有少数女性 Mostly male with a few females
○绝大部分是男性 The vast majority are male

28、 您认为口译员是 [单选题] *:

Do you think interpreting is a profession more suited for: [single choice question]

○更适合女性从事的职业 Females	○男女均适合从事的职业 Both males and females equally	○更适合男性从事的职业 Males

29、 您认为口译员更适合女性是因为 [多选题] *:

Why do you think interpreting is more suited for females? [multiple choice question]

依赖于“您认为口译员是”的第 1 个选项

(Dependent on the 1st option of "Do you think interpreting is a profession more suited for:")

□我接触到的口译员以女性为主 Most interpreters I have encountered are female
□我印象中的口译员以女性为主 My perception is that most interpreters are female
□我认为女口译员能翻译得更准确 I believe female interpreters are more accurate
□我认为女口译员的表达更流畅 I believe female interpreters are more fluent
□我认为女口译员的声音更好听 I believe female interpreters have more pleasant voices

□我认为女性的语言天赋更好 I believe women have a better natural aptitude for languages
□我认为女口译员的形象更好 I believe female interpreters have a better professional image
□其他：_____* Other: _____

30、 您认为口译员男女均适合是因为 [多选题] *：

Why do you think interpreting is suited for both genders equally? [multiple choice question]

依赖于“您认为口译员是”的第 2 个选项

(Dependent on the 2nd option of "Do you think interpreting is a profession more suited for:")

□我接触到的口译员男女都有 I have encountered both male and female interpreters
□我印象中的口译员男女都有 My perception is that both males and females work as interpreters
□我认为男女口译员都翻译得很准确 I believe both male and female interpreters are equally accurate
□我认为男女口译员的表达都能很流畅 I believe both male and female interpreters are equally fluent
□我认为男女口译员的声音都能好听 I believe both male and female interpreters have equally pleasant voices
□我认为男女之间的语言天赋没有差异 I believe there is no difference in language aptitude between genders
□我认为男女口译员都能形象好 I believe both male and female interpreters can present a professional image
□其他：_____* Other: _____

31、 您认为口译员更适合男性是因为 [多选题] *：

Why do you think interpreting is more suited for males? [multiple choice question]

依赖于“您认为口译员是”的第 3 个选项

(Dependent on the 3rd option of "Do you think interpreting is a profession more suited for:")

□我接触到的口译员以男性为主 Most interpreters I have encountered are male
□我印象中的口译员以男性为主 My perception is that most interpreters are female
□我认为男口译员能翻译得更准确 I believe male interpreters are more accurate
□我认为男口译员的表达更流畅 I believe male interpreters are more fluent
□我认为男口译员的声音更好听 I believe male interpreters have more pleasant voices
□我认为男性的语言天赋更好 I believe men have a better natural aptitude for languages
□我认为男口译员的形象更好 I believe male interpreters have a better professional image
□其他：_____* Other: _____

32、 您印象中女口译员的相貌是 [单选题] *:

What is your perception of the appearance of female interpreters? [single choice question]

依赖于“您认为口译员是”的第 1;2 个选项

(Dependent on the 1st or 2nd option of "Do you think interpreting is a profession more suited for:")

○非常漂亮 Very beautiful	○有点漂亮 Somewhat beautiful	○普通 Average	○有点难看 Somewhat unattractive	○非常难看 Very unattractive

33、 您印象中男口译员的相貌是 [单选题] *:

What is your perception of the appearance of male interpreters? [single choice question]

依赖于“您认为口译员是”的第 2;3 个选项

(Dependent on the 2nd or 3rd option of "Do you think interpreting is a profession more suited for:")

○非常英俊 Very handsome	○有点英俊 Somewhat	○普通 Average	○有点难看 Somewhat	○非常难看 Very

	handsome		unattractive	unattractive

34、 您印象中女口译员的身高是 [单选题] *:

What is your perception of the height of female interpreters? [single choice question]

依赖于“您印象中口译员”的第 1;2;3;4 个选项

(Dependent on any of the 1st, 2nd, 3rd, or 4th options of "What is your perception of the gender distribution among interpreters?")

○160cm 及以下 160cm or below	○160-165cm	○165-170cm	○170-175cm	○175cm 及以上 175cm or above

35、 您印象中男口译员的身高是 [单选题] *:

What is your perception of the height of male interpreters? [single choice question]

依赖于“您印象中口译员的”第 2;3;4;5 个选项

(Dependent on any of the 2nd, 3rd, 4th, or 5th options of "What is your perception of the gender distribution among interpreters?")

○165cm 及以下 165cm or below	○165-170cm	○170-175cm	○175-180cm	○180cm 及以上 180cm or above

36、 您印象中口译员的体型是 [单选题] *:

What is your perception of the body type of interpreters? [single choice question]

○非常瘦弱 Very slim	○有些瘦弱 Somewhat slim	○不胖不瘦 Average	○有些肥胖 Somewhat overweight	○非常肥胖 Very overweight

37、 您印象中口译员的年龄是 [单选题] *:

What is your perception of the age of interpreters? [single choice question]

○25 岁以下 Under 25	○25-35 岁 25-35	○35-45 岁 35-45	○45-55 岁 45-55	○55 岁以上 Over 55

38、 您印象中口译员的学历是 [单选题] *:

What is your perception of the educational background of interpreters? [single choice question]

○博士 Doctorate	○硕士 Master's degree	○本科 Bachelor's degree	○专科 Associate degree

○高中/中专 High school/vocational school	○初中 Middle school	○小学及以下 Primary school or below	

39、 您印象中口译员的年收入是（万元） [填空题] *:
What do you think is the annual income of interpreters (in ten thousand Chinese Yuan)? [fill-in-the-blank question]

_____

40、 您了解到的口译员群体主要于 [多选题] *
Where do you think most interpreters work? [multiple choice question]

□企事业单位全职（如政府部门、国企、外企） Full-time positions in organizations (e.g., government departments, state-owned enterprises, foreign enterprises)
□翻译公司全职（就职于固定的翻译公司） Full-time positions in translation companies
□自由职业（每次口译单独结费） Freelance interpreters (paid per interpretation task)
□兼职口译员（有固定工作，做口译补贴收入） Ad hoc interpreters (holding a full-time job, interpreting to supplement income)

41、 您认为口译员的政治地位比[矩阵量表题] *:
How do you perceive the **political status** of interpreters compared to the following professions? [matrix scale question]

	低很多 Much lower	低一点 Slightly lower	差不多 About the same	高一点 Slightly higher	高很多 Much higher
比保姆 A domestic worker	○	○	○	○	○
比法官 A judge	○	○	○	○	○
比国企经理 A state-owned enterprise manager	○	○	○	○	○

比会计 An accountant	○	○	○	○	○
比记者 A journalist	○	○	○	○	○
比经济学家 An economist	○	○	○	○	○
比科学家 A scientist	○	○	○	○	○
比网络工程师 A network engineer	○	○	○	○	○
比音乐家 A musician	○	○	○	○	○
比银行职员 A bank clerk	○	○	○	○	○

42、 您认为口译员的经济收入比[矩阵量表题] *:

How do you perceive the **income** of interpreters compared to the following professions? [matrix scale question]

	低很多 Much lower	低一点 Slightly lower	差不多 About the same	高一点 Slightly higher	高很多 Much higher
比工会主席 A trade union chairman	○	○	○	○	○
比个体户 A self-employed person	○	○	○	○	○
比网络工程师 A network engineer	○	○	○	○	○
比军人 A military officer	○	○	○	○	○
比私企老板 A private business owner	○	○	○	○	○

比保姆 A domestic worker	○	○	○	○	○
比律师 A lawyer	○	○	○	○	○
比程序员 A programmer	○	○	○	○	○
比歌星 A singer	○	○	○	○	○
比演员 An actor	○	○	○	○	○

43、 您认为口译员的社会声望比[矩阵量表题] *:

How do you perceive the **social prestige** of interpreters compared to the following professions? [matrix scale question]

	低很多 Much lower	低一点 Slightly lower	差不多 About the same	高一点 Slightly higher	高很多 Much higher
比银行职员 A bank clerk	○	○	○	○	○
比音乐家 A musician	○	○	○	○	○
比保姆 A domestic worker	○	○	○	○	○
比经济学家 An economist	○	○	○	○	○
比会计 An accountant	○	○	○	○	○
比网络工程师 A network engineer	○	○	○	○	○
比电影导演 A film director	○	○	○	○	○
比飞行员 A pilot	○	○	○	○	○
比大学教授	○	○	○	○	○

A university professor					
比银行行长 A bank manager	○	○	○	○	○

44、 您认为口译员的**综合地位**比[矩阵量表题] *:

How do you perceive the **overall status** of interpreters compared to the following professions? [matrix scale question]

	低很多 Much lower	低一点 Slightly lower	差不多 About the same	高一点 Slightly higher	高很多 Much higher
比银行行长 A bank manager	○	○	○	○	○
比律师 A lawyer	○	○	○	○	○
比保姆 A domestic worker	○	○	○	○	○
比科学家 A scientist	○	○	○	○	○
比经济学家 An economist	○	○	○	○	○
比私企老板 A private business owner	○	○	○	○	○
比网络工程师 A network engineer	○	○	○	○	○
比记者 A journalist	○	○	○	○	○
比歌星 A singer	○	○	○	○	○
比个体户 A self-employed person	○	○	○	○	○

45、 请用五个形容词概括您对口译员**群体特征的印象**[矩阵文本题] *:

Please use five adjectives to summarize your **impression of the interpreter community** [matrix text question]

1.	_____
2.	_____
3.	_____
4.	_____
5.	_____

46、 如您有意向参与红包发放，请填写您的支付宝号码： [填空题]

If you would like to participate to win a 20 Chinese Yuan red envelope, please provide your Alipay account number: [fill-in-the-blank question]

_____

The end of the questionnaire

Appendix 2. Informed consent form for the collection of interpreting video materials beforehand recording

口译视频材料收集知情同意书

(This informed consent form was written in Simplified Chinese, and the accompanying English content is its translation)

XX 口译员老师：
Dear Interpreter [Name],

您好，我是山东大学翻译学院博士生李天韵，就收集您的口译视频材料进行说明和保证，请您知悉：

Greetings! I am Tianyun Li, a Ph.D. candidate at the School of Translation Studies, Shandong University. I would like to provide the following explanation and assurances regarding the collection of your interpreting video materials:

一、 您被邀请对医学类视频《未来医学：预防是疾病治疗最好的方法》进行交替传译，并允许我收集您在交替传译过程中的摄像头数据，包括口译过程、口译笔记、口译译文等。

You are invited to provide consecutive interpretation for the medical video titled “Future medicine: Prevention is the best method for treating diseases”, and I am requesting your permission to collect data of your consecutive interpreting via your camera, including your video of interpreting, your notes of interpreting, and your rendition.

二、 收集该口译视频材料仅用于教学研究目的，本材料将面向翻译类/大学英语类本科生和硕士生开放，便于其了解具有医学学科背景的口译员进行交替传译的过程、评析口译风格。同时，请允许我使用您已公开的职业信息，如职业背景、资质认证等。

The collection of this interpreting video material will be solely for research and teaching purposes. The material will be made accessible to undergraduate and master’s students majoring in translation & interpreting studies and those attend college English courses for academic purposes, to help them understand the consecutive interpreting process of interpreters with a medical background and to analyse interpreting styles. Additionally, I ask for your permission to use publicly available professional information about you, including your career background and your professional certifications.

三、 同时，我会对本视频材料进行研究，包括探究具备学科背景口译员与语言专业

背景口译员的翻译风格差别、受众是否看重口译员的学科背景或身份等。

I will also conduct research using this video material, focusing on exploring the differences in interpreting styles between interpreters with a subject-specific background and those with a language-specific background, and examining whether the audience values the interpreter's subject-matter expertise or professional identity.

四、 该视频材料将仅由我本人保存，同时保证不会上传至任何公共视频平台，也不会传送给任何其他人士。我将遵守教学和研究守则合理谨慎使用您提供的视频材料。

The video material will be stored exclusively by me and will not be uploaded to any public video platforms, nor will it be shared with any third parties. I guarantee to handle the material with care, adhering to teaching and research ethics, and will use the material responsibly.

五、 参加本研究的补偿为：XX 元人民币。补偿于视频录制结束后立刻发放。如果您提供的视频材料帮助我完成了研究突破，我将在追加补偿 XX 元人民币作为感谢。

Compensation for your participation in this study will be XX Chinese Yuan. This compensation will be provided immediately after the video recording is completed. If your contribution to the video material results in significant breakthroughs in my research, I will provide an additional compensation of XX Chinese Yuan as a token of appreciation.

六、 如果您对您提供视频的教学及研究成果感兴趣，需要了解更进一步的教学或研究进展，可随时联系我为宜。

If you are interested in the teaching and research outcomes from the use of your video, or if you wish to learn more about the progress of the teaching or research, please feel free to contact me at any time.

李天韵

Tianyun Li

2021 年 6 月 13 日

13th June 2021

Appendix 3. Speech transcription content of experimental video

实验使用口译视频的语音转写

This presentation by Dr Ross Walker was delivered in English, while the accompanying Chinese content is its literal translation.

Ni hao ma (Greetings)! My name is Dr Ross Walker, and I'd like to thank you for the honour of being a speaker at the very prestigious *Yike Talks* in your wonderful country. Imagine a world where there is no disease, imagine a world where everyone is healthy and full of energy. I believe that is the future of medicine. What allows me to make these bold predictions? I have been a doctor for just under 40 years. I've been a heart specialist for over 30 years, but we all have a journey and often along the journey, our hardest lessons in life are our best teachers. And I was privileged that my hardest lesson came when I was only 17 years old. This man here, this young man here, was my best friend in high school, and one week before our final exams, he drowned and died. I was devastated. There was no way I could study. I thought I would fail my exams, but his grandfather, who was one of the greatest ever Australians, a man called Sir Laura Lodottes, said to me at the funeral, "Ross, you and Jack were going to do medicine together. You have to do it for both of you." And I've been practicing medicine for the last 40 years for myself and for Jack. I've been speaking in wonderful forums like this to try and promote the message of good health.

你好吗？我的名字是罗斯·沃克医生，非常感谢你们给予我这份荣誉，让我在贵国非常有声望的《一刻 Talks》上作为演讲者。想象一个没有疾病的世界，想象一个每个人都健康并且充满活力的世界。我相信这就是医学的未来。是什么让我能够做出这些大胆的预测呢？我从医将近 40 年了。我已经做了 30 多年心脏专科医生，但我们每个人都有一段旅程，并且往往在旅途中，我们生命中最艰难的教训是我们最好的老师。而我有幸在我 17 岁时就经历了我最艰难的教训。这个人，这个年轻人，是我高中时最好的朋友，在我们期末考试前一周，他溺水身亡。我非常难过，根本没办法学习。我以为我会考试不及格，但是他的祖父——他是有史以来最伟大的澳大利亚人之一，名叫劳拉·洛多特爵士——在葬礼上对我说："罗斯，你和杰克本来要一起学医的。你必须为了你们两个人去做。"于是我过去 40 年一直为自己和杰克行医。我一直在像这样的美好论坛上演讲，努力推广健康的理念。

So what is going to happen in the future? Over the past few decades there, there's been

enormous efforts around the world to try to reverse aging. What about the major advances in medicine? When I started medicine 40 years ago, you had a one in three chance of dying within a month if you had a heart attack. Now it's well less than 5% because of the advances in the treatment of heart problems. We now have beautiful images of the heart, beautiful images of the heart that can be captured with just an intravenous injection. You can get a magnificent picture to know exactly what's going on in the heart. We also now have, instead of major bypass surgery, we can put stents across blocked arteries. We're now putting valves from the arteries in the leg, replacing heart fails to stop major operations. So, things like major operations on the heart will be a thing of the past in the near future.

那么未来会发生什么呢？在过去的几十年里，全球都在为逆转衰老做出巨大的努力。那么医学方面的重大进展呢？当我 40 年前开始行医时，如果你患有心脏病发作，你有三分之一的可能会在一个月内死亡。而现在，这个概率已经远远低于 5%，因为心脏问题治疗方面取得了进展。我们现在拥有漂亮的心脏图像，只需通过静脉注射就能捕捉到这些美丽的心脏图像。你可以得到一张非常清晰的图片，准确了解心脏的状况。我们现在也有了替代大型冠状动脉搭桥手术的方法，我们可以在堵塞的动脉上放入支架。我们现在通过腿部的动脉放入瓣膜，来替换失效的心脏瓣膜，以避免大型手术。所以像这样的心脏大型手术在不久的将来将成为过去。

We're also now 3D printing organs, and we're going to see transplantation stopping. When we perfect the 3D transplantation, sorry 3D printing of the heart, kidneys, and livers, this is going to be happening in the relatively near future. We've heard a bit about cancer treatments today. The old treatment for cancer involved major operations, chemotherapy, and radiotherapy, quite effective, but they also damaged natural and normal parts of the cell. Whereas these days, what we're doing now is taking people's own cancers, personalized cancer management, and creating a vaccine against the cancer. You may not realize this, but a cancer cell has an invisible shield around it to stop the immune system from recognizing it, and that's why the cancer grows and grows and grows. We are now retraining the immune system with immunotherapy to kill that shield and retrain the immune cells to kill the cancer. Recently, a study was done on laboratory animals that had advanced melanoma, and with the combination of a personalized vaccine against the melanoma and immunotherapy, there was a 100% cure. I believe cancer will be cured at some stage over the next 10 to 20 years, if not sooner.

我们现在还在进行 3D 打印器官的研究，并且我们将看到器官移植将停止。当我们完善了 3D 移植——抱歉，是 3D 打印心脏、肾脏和肝脏的技术，这将在相对不远的将来实现。今天我们也听到了关于癌症治疗的一些内容。过去的癌症治疗包括大型手术、化疗和放疗，这些方法相当有效，但它们也会损害细胞的自然和正常部分。而如今，我

们所做的是利用患者自身的癌细胞，进行个性化的癌症管理，并为癌症开发疫苗。你可能没有意识到，癌细胞周围有一个看不见的屏障，阻止免疫系统识别它，这就是为什么癌症会不断生长、扩散。我们现在通过免疫治疗重新训练免疫系统，消除这个屏障，并重新训练免疫细胞来杀死癌细胞。最近，对患有晚期黑色素瘤的实验动物进行了一项研究，通过个性化的黑色素瘤疫苗和免疫治疗的结合，达到了 100%的治愈率。我相信癌症将在未来 10 到 20 年，甚至更早的时间内被治愈。

Mapping the cancer genome is also very important for personalized medicine. What I was saying before, and also mapping the genes of people to see what their risk is for Alzheimer's disease, because the earlier you start treatment, the better off that person is. But the best treatment for all of these things is prevention. My whole career has been based around the concept of prevention. But here's the problem in medicine: we start with the ambulance parked at the bottom of the cliff, waiting for people to fall off. So instead of fixing the rails at the top of the cliff — prevention — what do we do in medicine? We make faster ambulances. It's like the business world waiting for someone to go bankrupt, and then they give them financial advice. It’s always the best treatment is prevention.

绘制癌症基因图谱对于个性化医学来说也非常重要。正如我之前所说的，绘制人类的基因图谱也是为了了解他们患阿尔茨海默病的风险，因为治疗开始得越早，患者的状况就会越好。但所有这些问题的最佳治疗方法就是预防。我整个职业生涯都是围绕预防这一理念展开的。但医学中的问题在于：我们把救护车停在悬崖底下，等着人们从悬崖上掉下来。所以，医学中我们没有去修好悬崖上的护栏——也就是预防，而我们在做什么呢？我们让救护车开得更快。这就像商业世界中等到有人破产了，才给他们提供理财建议。最好的治疗永远是预防。

And there are five keys to prevention. Number one, you cannot be healthy if you have any addictions. 360 million people in China smoke, and 60% of male doctors in China smoke. You cannot be healthy and smoke, you cannot be healthy and drink too much alcohol, you cannot be healthy and use illegal drugs. So, number one is to quit addictions. Number two is good quality sleep. Seven to eight hours of good quality sleep is as good for your body as not smoking, so cultivate every night of your life a good quality sleep habit. Number three is good quality nutrition. And nutrition is simple. It's called "eat less and eat more naturally." That's all you need to know about nutrition. Number four is what I call the second best drug on the planet: three to five hours every week of some form of exercise. If you have three to five hours every week of exercise, you reduce your risk for cardiovascular disease, cancer, Alzheimer's disease, depression, and diabetes by 30%. You reduce your risk for osteoporosis by 50%. You drop your blood pressure and you sleep better. There is no pharmaceutical drug on the planet that is as

powerful as exercise.

预防有五个关键要点。第一，你如果有任何成瘾行为，就不可能健康。在中国有 3.6 亿人吸烟，60%的男性医生吸烟。你无法吸烟还保持健康，你无法饮酒过量还保持健康，你无法使用非法药物还保持健康。所以，第一个关键点是戒除成瘾行为。第二是高质量的睡眠。每晚七到八小时的高质量睡眠对你的身体来说，就像不吸烟一样有益，因此培养每天晚上都拥有良好的睡眠习惯。第三是高质量的营养。营养很简单，就是“少吃一点，吃得更加自然。”这就是你需要了解的全部营养知识。第四，我称之为地球上第二好的药物：每周三到五小时的运动。如果你每周能进行三到五小时的运动，你将降低 30%的心血管疾病、癌症、阿尔茨海默病、抑郁症和糖尿病的风险。你还可以将骨质疏松的风险降低 50%。你的血压会下降，睡眠质量也会提高。没有任何药物能像运动那样强大。

But the best drug on the planet is a thing called happiness. Happiness is more important than anything else. And do you know that if you combine those five things — no addictions, good sleep, good nutrition, exercise, and happiness — you reduce your risk for all diseases by somewhere between 70 to 80%. If I give you a pill to lower your cholesterol, I reduce your risk for a heart attack by 20%. So why on earth wouldn't anyone follow the preventive strategy? So I believe the future of medicine is extremely exciting, but the best treatment for heart disease, the best treatment for cancer, the best treatment for Alzheimer's disease, and all of our other modern killers, is not to get them in the first place. It's called prevention. Don't wait for the future. The time to start is right now.

但是，世界上最好的药物是一种叫做“幸福”的东西。幸福比任何其他事情都重要。你知道吗，如果你将这五件事结合起来——没有成瘾行为、良好的睡眠、良好的营养、运动和幸福——你可以将所有疾病的风险降低大约 70%到 80%。如果我给你一颗降低胆固醇的药片，它可以将你心脏病发作的风险降低 20%。那么，为什么没有人愿意遵循这种预防策略呢？所以，我相信医学的未来极其令人兴奋，但治疗心脏病、治疗癌症、治疗阿尔茨海默病以及所有其他现代致命疾病的最佳方法，不是等到它们发生，而是避免它们发生。这就是所谓的预防。不要等到未来才行动，开始的时间就是现在。

Appendix 4. The speech transcript of the interpreter's rendition

实验使用口译译文的语音转写内容

(This rendition was presented in Simplified Chinese, while the accompanying English content is its literal translation, including the filling pauses)

呃，大家好，我的名字叫做 Ross Walker，我今天非常荣幸能够受邀在贵国的一刻 talks 上向大家做演讲。大家想象一下，想象一个没有疾病的世界，想象人人都能健康，想象人人都充满活力，这，就是未来的医学。是什么让我，呃，做出如此，啊，有信心如此大胆的预测呢？呃，我呢已经行医超过 40 年了，我也有 30 年以上的心脏专科医生的经验，（换气声）那么这一路走来，其实每一个人呢（换气声）都会在自己的人生旅途中，当中学到自己，（呃）学到自己最艰难的一些教训，那么（换气声）我就是人生当中最艰难的教训就是自己的老师（咂嘴声）。那么我在 17 岁的时候（停顿）是学到了（停顿）最（停顿）最艰难的教训，（换气声）那么图中的这个年轻人，大家可以看到这个年轻人，他是我高中时最好的朋友，在我们期末考试前一周，他因溺水而过世了。当时我，呃，心（停顿，呃）当时我非常的心痛，当时我无心学习我觉得我可能会考试会，考试会失败掉我可能无法通过考试，（换气声）但是他的，爷爷，一个非常伟大，非常伟大的澳大利亚人，呃，叫做 Mars Lodotte 爵士，你们两个呢本来应该是一起去学习的，一起去寻医的，但是将来呢你要一个人走下去了，你要为你们两个人而行医，所以说在我 40 年的，（轻咳）行医历程当中，我，不仅是为我自己，我也是为杰克，我，我进行共同的行医。那么在，这，在这些年当中，我在很多非常好的舞台上做过演讲，像，今天的一样。

Umm, hi, my name is Ross Walker, today I am very honoured to be invited to give a speech at your esteemed country's *YiKe Talks* event. Everyone, imagine, imagine a world without disease, imagine everyone can be healthy, imagine everyone full of vitality, this is the future of medicine. What gives me, uh, the confidence to make such, ah, bold predictions? Uh, I've been practicing medicine for over 40 years, and I have over 30 years of experience as a cardiologist. (breathing sound) So, along the way, actually everyone (breathing sound) will, in their life journey, learn their own, (uh) learn their own most difficult lessons. Well (breathing sound), for me, the hardest lesson in life is my teacher (mouth clicking sound). When I was 17 years old (pause), I learned (pause) the (pause) hardest lesson (breathing sound), well, the young man in the picture, everyone can see this young man, he was my best friend in high school, and one

week before our final exams, he passed away due to drowning. At that time, I, uh, heart (pause, uh), at that time, I was very heartbroken, I couldn't focus on studying, I thought I might fail the exam, I might not pass the exam (breathing sound), but his grandfather, a very great, very great Australian, uh, named Sir Mars Lodotte, told me, "You two were supposed to study together, seek medicine together, but in the future, you have to walk this road alone, you have to practice medicine for the two of you." So, in my 40 years of (light cough) practicing medicine, I not only practiced for myself but also for Jack, I practiced medicine for both of us. Over the years, I have spoken on many prestigious stages, just like today.

世界上的医学家做出了巨大的努力。那么（停顿）我们现在可以明确的是，在图中我们可以非常明确的看到一个非常，漂亮的一个心脏的一个图片，那么其实呢，在过去几十年当中，医学有了巨大的进展，那么在过去，如果说，（咂嘴声）如果说你有心脏病发作的话，那么在一个月之内你会有 1/3 的，概率，过世，但是现在呢这个概率被降到了 5%以下，为什么因为，在心脏治疗方面我们有了重大的进展，（换气声）那么我们在图中可以看到一幅非常漂亮的图片，这是一个（呃），这是一个心脏的照片，（换气声）那么现在呢仅仅通过静脉注射，那么我们就可以搞清楚心脏有哪些问题（换气声），那么现在呢我们不再做，更多的搭桥手术，（换气声）我们现在可以通过支架来治疗重大的心脏疾病，（换气声）我们可以通过腿部的入，我们可以通过腿部的（呃嗯）手术做介入，然后（嘶）给心脏做（呃）做瓣膜的置换，（换气声）那么这些都是在心脏领域做出的巨大的成就我们现在，不再做过多的心脏的大手术。

The world's medical professionals have made great efforts. Well (pause), now we can clearly see, in the picture, we can clearly see a very, beautiful, picture of a heart. In fact, in the past few decades, there has been great progress in medicine. In the past, if (mouth clicking sound), if you had a heart attack, you would have a one-third chance of dying within a month, but now that chance has dropped below 5%. Why? Because we've made significant progress in heart treatment. (breathing sound) Now, in the picture, you can see a very beautiful image, this is a (uh), this is a picture of a heart (breathing sound). Now, simply through intravenous injection, we can figure out what's wrong with the heart (breathing sound). Now, we no longer do, many bypass surgeries (breathing sound), we can now use stents to treat major heart diseases (breathing sound), we can go through the leg, we can go through the leg (uh) surgery to do interventions, then (hiss), for the heart, we do (uh) valve replacement (breathing sound). These are the major achievements in cardiology; we no longer perform too many large heart surgeries.

我们现在也看到了 3D 打印的器官，那么在将来，随着 3D 打印器官技术的愈加成熟和完善，我们可以预期到（呃）我们可以（呃嗯）将，将不再会有移植手术，在未来呢会有 3D 打印的心脏、肾脏等等。（长换气声）那么在未来那么对于癌症，也对于癌症

的治疗也会有新的方法，在过去旧的癌症的治疗的方法是通过更大的，是通过大手术是通过化疗是通过放疗（长换气声），当然这些治疗方式都是非常有效的（短换气声），但是在这些治疗有效的同时，他们也会损伤到也会杀害，也会损伤到影响到（吸气声）正常的细胞。那么，在未来呢我们可以，我们可以找到患者的癌症细胞，然后给他们进行个性化的治疗（换气声），然后开发出（哦）取到患者的癌症细胞，开发出（轻咳）这些患者个性化的癌症的疫苗，大家可能没有意识到在癌症细胞周围有一层，看不到的（呃）一个屏蔽，那么这就是为什么，癌症呢会逃过免疫细胞的杀灭，这就是为什么免疫系统对于癌症会视而不见。（换气声）那么在将来呢我们会重新训练免疫系统，让他们能够识别癌症、杀灭癌症。那么在一个（犹豫）在一个实验当中，这个实验动物它们都是有（呃）晚期黑色素瘤的实验动物（换气声），通过免疫治疗（停顿，呃嗯）通过免疫治疗，那么这些动物呢达到了完全的治愈，（换气声）所以说我们预期呢在 10~20 年或者更（嗯）不是非常久远的未来是能够在癌症治疗方面实现这一点的。

Now we are also seeing 3D-printed organs, and in the future, as 3D-printed organ technology becomes more mature and perfected, we can expect (uh) that we will no longer have transplant surgeries. In the future, there will be 3D-printed hearts, kidneys, and so on. (long breath) In the future, there will be new methods for cancer treatment too. In the past, cancer treatment was through bigger, it was through large surgeries, chemotherapy, and radiation (long breath). Of course, these treatments were very effective (short breath), but while they were effective, they also damaged or killed, they also damaged and affected (inhale) normal cells. In the future, we can, we can find a patient’s cancer cells, then give them personalized treatments (breathing sound), then develop, oh, take the patient’s cancer cells, develop (light cough) personalized cancer vaccines for these patients. You may not realize that around cancer cells there is an invisible (uh) shield, and that’s why cancer escapes the immune system's killing, that’s why the immune system doesn’t see cancer. (breathing sound) In the future, we will retrain the immune system to recognize cancer, to kill cancer. In an (hesitation), in an experiment, this experiment on animals, the animals (uh) had advanced melanoma (breathing sound), through immunotherapy (pause, uhmm), through immunotherapy, these animals reached complete cure (breathing sound). So we expect that in 10 to 20 years, or (uh) in the not too distant future, we will achieve this in cancer treatment.

就像我之前讲到的一样（呃）基因图谱的绘制对于一个人来说对于一个人来说也是非常重要的，（换气声）能够了解他在癌症方面是否有一些风险，那么同样重要的是我们可以通过基因，了解到，一个人患有阿茨海默病的风险，那么治疗的，发现的越早，治疗的越早，那么这个效果就越好，但是，最重要的其实是预防、预防呢贯穿我整个的职业生涯。那么现在的医学我的感觉就像是我们把车停在悬崖的、悬崖峭壁的下面，然后

等着人从悬崖上掉下来。（换气声）那么我们呢现在没有把悬崖上的栏杆建起来，我们现在做的呢是把救护车的速度越提越快。所以说（呃嗯）这在另外一个比方就是在、就是在、就是在从商业的角度讲，现在呢大家等人破产了以后，我们再给他提一些理财建议，所以说为什么我们，不进行提前的预防呢？

Just like I said earlier (uh), mapping a person's genome is also very important to understanding if they have any risk for cancer (breathing sound), and it's just as important that we can, through genes, understand a person's risk of developing Alzheimer's disease. The earlier it's detected and treated, the better the result, but the most important thing is prevention. Prevention has been a theme throughout my entire career. The modern medicine feels like we park an ambulance at the bottom of a cliff, a cliff, waiting for people to fall off the cliff (breathing sound), thus we, now we haven't built the fence at the top of the cliff yet. What we're doing now is just speeding up the ambulance. So (uhmm) this is another metaphor, that is, that is, that is from, from a business perspective, now, people wait until someone goes bankrupt, then we give them some financial advice, so why don't we, do preventive measures in advance?

那么接下来要讲的是预防我认为呢，我认为呢要维持要保持健康的话，那么其实有 5 个非常重要的点，那么第一点呢就是戒掉成瘾的行为，那么在中国有超过 3.6 亿人吸烟在中国有 60%以上的男性一生吸烟，就是吸烟的话你是不可能健康的。成瘾性行为包括吸烟、喝酒、非法用药。所以说如果想保持健康的话，第一个重要的要素呢就是戒掉成瘾性行为。那么第二点呢就是（呃嗯）第二点就是高质量的睡眠，（换气声）那么高，那么每天每晚 7~8 个小时的高质量的睡眠，（换气声）它的重要性和戒烟是一样的所以说一定要保持，每天 7~8 小时的高质量的睡眠。（换气声）第三点，就是高质量的营养。那么高——就是高质量的营养，其实说起来非常简单大家记住一句话就可以，吃的少，吃的更加自然。那么第四点、最重要的我认为是我把它叫做这个星球上第二、重要的、第二重要的事情，这个叫做每周要运动 3~5 小时，（换气声）如果说能够每周运动这是这个，我把它叫做这个星球上（呃）排名第二的、治疗，那么这就是（短停顿）每周运动 3~5 小时，如果说每周能运动 3~5 小时的话，可以有效的（嗯）降低（呃）冠脉疾病，这个 CAD 就是心脏冠脉疾病，降低癌症，降低阿兹海默疾病降低抑郁降低，糖尿病发生率能够降低 30%（换气声）。同样的每周运动 3~5 小时可以降低骨质疏松的发生，（换气声）可以降低血压，可以保证、可以使睡眠的质量更好。

So next, I want to talk about prevention. I think, I think that to maintain, to keep healthy, there are actually 5 very important points. The first point is to quit addictive behaviours. So, in China, there are over 360 million people who smoke and over 60% of men in China smoke at some point in their lives. So if you smoke you simply cannot be healthy. Addictive behaviours include smoking, drinking alcohol, and illegal drug use. So if you want to stay healthy, the first

important factor is to quit addictive behaviours. Now the second point is (uhmm), the second point is high-quality sleep (breathing sound), so, high—so 7 to 8 hours of high-quality sleep every night (breathing sound). Its importance is the same as quitting smoking so it's very important to maintain 7 to 8 hours of high-quality sleep per day (breathing sound). The third point is high-quality nutrition. So high—high-quality nutrition. In fact is actually very simple, you can just remember one sentence: eat less, eat more naturally. Now the fourth point the most important one, I think I name it the second most important thing on this planet. This is called exercising 3 to 5 hours a week (breathing sound). If you can exercise weekly, this is, I call it the second most important (uh) treatment on this planet. This is (short pause) exercising 3 to 5 hours a week. If you can exercise 3 to 5 hours a week, it can effectively (uh) reduce (uh) coronary artery disease, this CAD, which is coronary artery disease, reduce cancer, reduce Alzheimer's disease, reduce depression, reduce the incidence of diabetes by 30% (breathing sound). Similarly, exercising 3 to 5 hours a week can reduce the occurrence of osteoporosis (breathing sound), can lower blood pressure, can ensure, can improve sleep quality.

但是是，但是我认为最好的药物世界上最好的药物就是 happiness，就是幸福或者快乐，那么幸福、幸福感要比任何其他的事物都更加的重要。（咂嘴声）那么如果说我们能够把我刚才讲的这 5 点都能够做到的话，戒掉成瘾性行为，（停顿，呃）有高质量的睡眠，有高质量的营养的话，它就能够把你的胆固醇降 20%，那么为什么不去采取这些预防的措施呢？我相信未来的医学呢会是非常令人激动的，但是最重要的是、但是最重要的是（换气声）预防、预防心脏疾病、癌症、阿兹海默病以及现代的其他的、高致死的一些疾病。（换气声）所以说这些最重要的就是预防把他们在（停顿）就是把他们的发生(呃)，把它们让它们不要发生，通过预防的手段来实现，所以说大家不要等待。那么治愈疾病的治疗疾病的最佳时间就是现在。谢谢大家。

But, is I believe the best medicine, the best medicine in the world is happiness, that is, happiness or joy, and happiness, the feeling of happiness, is more important than anything else. (mouth clicking sound) So, if we can achieve the 5 points I just talked about, quitting addictive behaviours, (pause, uh), getting high-quality sleep, having high-quality nutrition, it can lower your cholesterol by 20%. So, why not take these preventive measures? I believe that the future of medicine will be very exciting, but the most important thing, the most important thing (breathing sound) is prevention—prevention of heart disease, cancer, Alzheimer's disease, and other modern, highly fatal diseases. (breathing sound) So the most important thing is prevention, to stop them from (pause), to stop them from happening (umm), to prevent them through preventive measures. So, everyone don't wait. The best time to cure a disease, the best time to treat a disease is now. Thank you.

Appendix 5. Test for users' listening comprehension

口译质量感知测试题

(This test paper was written in Simplified Chinese, and the accompanying English content is its translation)

学号 [填空题] *: _____
Student ID [fill-in-the-blank question]

1. 手术是治疗疾病的最好方法 [判断题] *:
Surgery is the best method for treating diseases [true/false question]

正确 True	错误(正确答案) False (correct answer)

2. 当下心脏病发作后致死率是三分之一 [判断题] *:
The current mortality rate after a heart attack is one-third [true/false question]

正确 True	错误(正确答案) False (correct answer)

3. 通过动脉注射可以看到心脏的健康状态 [判断题] *:
The health of the heart can be observed through an arterial injection [true/false question]

正确 True	错误(正确答案) False (correct answer)

1. 搭桥手术是治疗心脏疾病的现代方法 [判断题] *:
Bypass surgery is a modern method for treating heart diseases [true/false question]

正确 True	错误(正确答案) False (correct answer)

2. 癌症疫苗是可以研发的 [判断题] *:
It is possible to develop a cancer vaccine [true/false question]

正确(正确答案) True (correct answer)	错误 False

6. 癌细胞周围有阻止免疫系统探测的屏障 [判断题] *:

Cancer cells are surrounded by a barrier that prevents the immune system from detecting them [true/false question]

正确(正确答案) True (correct answer)	错误 False

7. 每周锻炼两到三小时可以降低心脏病、癌症风险 [判断题] *:

Exercising two to three hours a week can reduce the risk of heart disease and cancer [true/false question]

正确 True	错误(正确答案) False (correct answer)

8. 二十年内有望完全治愈癌症 [判断题] *:

Cancer is expected to be completely cured within twenty years [true/false question]

正确(正确答案) True (correct answer)	错误 False

9. 喝可乐上瘾不会影响健康 [判断题] *:

Being addicted to drinking Coca Cola does not affect health [true/false question]

正确 True	错误(正确答案) False (correct answer)

10. 最好的药物是充足的睡眠 [判断题] *:

The best medicine is adequate sleep [true/false question]

正确 True	错误(正确答案) False (correct answer)

The end of the test

Appendix 6. Questionnaire on users' perceptions of interpreting quality

用户口译质量感知的问卷

(This questionnaire was written in Simplified Chinese, and the accompanying English content is its translation)

本问卷数据仅供实验分析使用，不作任何其他用途。
The data from this questionnaire is used solely for experimental analysis and will not be used for any other purposes.

学号 [填空题] *:
Student ID [fill-in-the-blank question]

_____

支付宝号码（用于奖励优秀回答） [填空题] *:
Alipay account (for rewarding outstanding responses) [fill-in-the-blank question]

_____

1、您的性别 [单选题] *

○男 Male	○女 Female	○其他（请说明） _____ * Other (please specify): _____

2、您的年龄 [填空题] *:
What is your age? [fill-in-the-blank question]

_____

3、您是哪里人 [填空题] *:
Where are you from? [fill-in-the-blank question]

_____

4、您的专业 [单选题] *:
What is your major? [single choice question]

○文学 Literature	○历史学 History	○哲学 Philosophy
○经济学 Economics	○管理学 Management	○法学 Law
○教育学 Education	○艺术学 Arts	○理学 Science
○工学 Engineering	○医学 Medicine	○农学 Agriculture

5、您是否会英语？ [单选题] *:

Do you speak English? [single choice question]

○会 Yes	○完全不会 No

6、您的英语水平 [单选题] *:

What is your level of English proficiency? [single choice question]

依赖于“您是否会英语？”的第 1 个选项

(Dependent on the 1st option of "Do you speak English?")

○我还没有考级 I have not taken a proficiency test	○大学四级 College English Test Band 4	○大学六级 College English Test Band 6
○专业四级 Test for English Majors Band 4	○专业八级 Test for English Majors Band 8	○ 其 他 （ 请 说 明 ）_____ * Other (please specify): _____

7、您是否会其他外语（不是英语）？ [单选题] *:

Do you speak any other foreign languages (excluding English)? [single choice question]

○会（请说明语种） _____ * Yes (please specify which language): _____	○不会 No

8、您其他外语的水平 [单选题] *:

What is your level of proficiency in other foreign languages? [single choice question]

○我还没有考级 I have not taken a proficiency test	○请列举证书 _____ * Please list certificates: _____

以下是正文部分。问卷不记名，请如实评价。

The following is the main section of the questionnaire. Responses are anonymous, please evaluate truthfully.

9、翻译的信息很全面 [单选题] *:

The interpreter's information was complete [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

10、 翻译的语速很均匀 [单选题] *:

The interpreter's speaking speed was consistent [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

11、 翻译的用词很专业 [单选题] *:

The interpreter's word choice was professional [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

12、 我信任翻译的内容 [单选题] *:

I trust the content provided by the interpreter [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

13、 翻译漏了很多信息 [单选题] *:

The interpreter missed a lot of information [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

14、 翻译讲起来磕磕巴巴的 [单选题] *:

The interpreter's delivery was hesitant [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

15、 翻译的医学术语翻得有问题 [单选题] *:

The interpreter made mistakes in translating medical terms [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

16、 翻译翻得让我不太放心 [单选题] *:

I didn’t feel confident in the interpreter’s translation [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

17、 翻译把外国专家的话都翻出来了 [单选题] *:

The interpreter translated everything the foreign expert said [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

18、 翻译的话语节奏很稳 [单选题] *:

The interpreter’s speech had a steady rhythm [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

19、 翻译的医学名词表达很到位 [单选题] *:

The interpreter conveyed medical terms accurately [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

20、 我可以很放心地依赖翻译 [单选题] *:

I can confidently rely on the interpreter [single choice question]

非常不同意 Strongly disagree	○1	○2	○3	○4	○5	非常同意 Strongly agree

21、 你觉得这位译员的翻译质量如何？ [填空题] *:

How do you rate the overall quality of this interpreter's rendition? [fill-in-the-blank question]

_____

22、 你觉得这位译员的语言表达如何？ [填空题] *:
How do you rate this interpreter's language expression? [fill-in-the-blank question]

_____

23、 你觉得这位翻译的医学专业程度如何？ [填空题] *:
How do you evaluate this interpreter's level of medical expertise? [fill-in-the-blank question]

_____

24、 这位翻译是否让你感到信任，为什么？ [填空题] *:
Did this interpreter make you feel confident? Why? [fill-in-the-blank question]

_____

The end of the questionnaire

致谢

我们生活的时代正在发生深刻且惊人的变化，七年前，我笃定地认为自然语言领域的机器学习技术只能止步于“处理”而非“理解”，但生成式预训练模型的发展如汹涌的钱塘江大潮冲垮了我肤浅的一厢情愿。神经网络翻译技术解放了千家万户的书面外语沟通，大语言模型则不禁让我们口译人对明天心生隐忧。从文登的口译客户请我校对豆包翻译的邮件译文一刻起，我知道有些旧日光景怕是一去不复返了。科技的高速发展让我疲于适应每一个崭新的今天，以至于不敢放松心情去从容地迷茫。在与时间赛跑的过程中，我从英语专业学生、口译方向研究生变化为口译研究人员，明天我就要站上三尺讲台去面对一双双求知的双眼。也许另一个平行时空中的我有同样的性格、同样的活力，但因为身份不同，发展路径便可能大相径庭。

在人工智能让大家彻底摆脱我们之前，继续研究我们口译员对人与人之间联结的影响还是有必要的。因此首先我想感谢我的导师崔英教授。彼时在香港中文大学（深圳）的分会场中，我的听众只有寥寥四五人，而让一个素昧平生的年轻人成为自己的第一位博士生，我要衷心地感谢崔老师无私地把这样宝贵的机会给予给我。在山威和导师相处的时光中，我时刻被崔老师严谨的治学精神、庄重的处世态度和真诚的待人风格所感染。读博不仅是对科研能力的提升，更是对道德品性的规劝，使我成为言出必行、事必躬亲的正直的人。导师的耳提面命使我抖擞，谆谆教诲使我明理，时至今日，我每每回想起导师在读期间对我的教导与告诫，便不觉尝试收起心中的顽劣心猿。许多教诲脑子已经记住了，但真正用心参悟还需时日。导师对我心性的规劝是我毕生的财富，提携之恩、教诲之恩，无以为报，唯有没齿不忘。

我也想感谢亚当·密茨凯维奇大学的 Agnieszka Chmiel 教授，同时感谢国家留基委的中波互换奖学金对我的资助，感谢研究生院李天一老师在申请过程中付出的诸多辛劳。波兰联培一年使我学术能力得到了质变的提升。除此之外，Chmiel 教授乐观阳光的笑容和她的金色头发、浅蓝色眼眸、洋红色小西装、圣诞树风格小领带一样洋溢着欢脱的温暖。我也有过需要鼓励的沮丧时刻，但即便我没表达出来，也能收到她言语的春风，使低落的心情一下子缓解了。在波兰还遇到了许多真诚的朋友，比如直爽的 Agnieszka Lijewska 助理教授、总是另辟蹊径的 Kostiantyn Mazur 博士、自律刻苦的拳击搭档 Christopher、给予过波兰语口译帮助的 Paulina Rozkrut 等，希望大家一切都好。

我要感谢参与自动字幕实验的 23 名波兰口译员同事、在质量感知实验中提交有效回复的 452 名东华理工大学的学生被试及 103 名山东大学的学生被试，以及感谢在刻板印象问卷调研中提交有效回复的 237 名被调研者。如果没有他们，这篇论文及博士期间发表的论文都只能落于纸上谈兵，谢谢他们提供了真实有效的数据支撑。数据收集过程得到了东华理工大学外国语学院廖华英教授、刘悦明院长、陈勇副教授及山东大学翻译

学院贾文峰博士、王苏华博士等老师的支持，谢谢他们有效保证了数据质量，减少了数据收集可能产生的变数。

感谢我的科研启蒙人刘敏华教授，有幸曾修读到刘教授亲自任教的同声传译实务及口译研究方法论两门课程。香港浸会大学攻读硕士虽只有一年，但彼时我在这里打下了良好的科研基础，能从国际视角去看待口译研究，对我成长发展的裨益之处无需赘言。除此之外，刘教授求真务实的科研态度深刻地影响了我的价值观。做学术不应是为自己包揽名利的工具，而是我们一起推动口译学科朝着更完善的方向发展的过程。感谢良师在前方，景行行止地让我保持着笃定的信心。

感谢预答辩、外审及正式答辩阶段中通读了我的论文的审稿专家，包括张峻峰院长、谌莉文教授、Agnieszka Chmiel 教授、庄会彬教授、王湘云教授、朱浩然副教授及两位匿名外审专家，阅读一本两百页的文稿需要投入很多工作量，谢谢诸位专家为论文完善所付出的时间与精力。

感谢 Daniel Gile 教授、Agnieszka Chmiel 教授、刘敏华教授、Michaela Albl-Mikasa 教授、Kilian Seeber 教授、Helle Dam 教授、张其帆副教授等产出杰出论文的先行学者们，他们身体力行地推动了口译学科的发展定型和边界拓展。我也要感谢在世界各处的青年口译学者们，包括邵张旻子博士、刘益光研究员、况霍凌霄博士、冯佳副教授、谢芮博士、陈世荣博士等国内青年学者，也包括 Paulina Rozkrut、Francesca Maria Frittella、César González、Bianca Prandi 等在欧洲认识的伙伴们，阅读大家的论文每每使我充满动力，它像纽带一般将我们精神联结在一起。

感谢翻译学院常伟东副书记在读期间给我的帮助。常书记是我的辅导员，在读期间许多事件都是由书记出面协调，最后得到了妥善的解决，我要感谢书记在处理事情上的管理智慧。同时感谢两任研究生秘书王海燕老师、周雨晴老师和单小芳老师在行政上提供的许多帮助。日常接触中也感谢张雄博士、贾文峰博士、王苏华博士、梁远副教授、孙田丰博士等老师的友好和体贴。同时感谢张玲玲老师等文华苑宿舍管理员，在我住宿期间提供了许多便利和帮助，让文华苑成为我可靠而安心的庇护所。

在威海校区三年和洪家楼校区的一年光阴中，我和我的同窗好友们结下了难忘的情谊。我要特别感谢我的两位室友王军礼老师和马吉德老师，我们从面试认识开始就结下了深厚的友谊，两位老师在日常生活中给予了我相当多的包容和关心。我的同一届好友们，韩树坤、赖海威、额尔德木吐、吴铭、刘青松、苏瑞丽等十一位同学，大家为达到毕业标准都付出了相当多的心血，也让我们更加珍视来之不易的缘分，保持彼此关系的淳朴善良，感谢有大家的一路同行。感谢学弟学妹王一涵、王腾飞、牟舒煜、张艺馨、王文彤、柴静等提供的许多帮助。

巍峨高山，深沉厚重。我要感谢山东大学的文化精神对我的耳濡目染，使我领悟到做事做人的道理。五载时光历历在目，虽已三十而立，十年一剑还须磨，学无止尽。行

将步入社会，我自觉尚未做好准备，但比起刚从波兰返校时幼稚冒进的精神状态已判若两人。千磨万击还坚劲，相信经过锤炼的我，能用更坚韧的心理状态、更克制的行事态度去面对接下来的挑战。

我要感谢泰州图书馆 24 小时自习室，疫情期间无法留校的寒假我在这里写出了文献综述的一部分，感谢罗森的自热便当填补了彼时我的辘辘饥肠。我还要感谢上海嘉定吉利研发中心的研发和测试人员，你们造出的机动车安全可靠，已经是我待人接物的左膀右臂。

感谢我的家人，祖父祖母、父亲母亲、姑父姑妈和我们大家庭的亲属们在我的成长过程中无条件地赋予了很多行动支持。在刻板印象调研收集中，他们都为问卷发放和回收贡献了许多力量，使我在彼时就有信心研究定能顺利完成。最后的感谢给我的爷爷，他像大树般遮风挡雨多年，面对不公从未忍气吞声，刚毅地为晚辈的健康成长创造了安然稳健的环境。而今长辈亲属们年事已高，祝愿他们身体健康，晚年无虞。

攻读学位期间发表的学术论文目录

1. Li, T. & Chmiel, A. (2024). Automatic subtitles increase accuracy and decrease cognitive load in simultaneous interpreting. *Interpreting*. https://doi.org/10.1075/intp.00111.li
2. 李天韵 & 崔英 (2022). 内隐社会认知路径的口译质量感知偏差研究. *外国语*45 (5), 100–109. https://jfl.shisu.edu.cn/article/id/670
3. Cheung, A. K. F. & Li, T. (2022). Machine aided interpreting: an experiment of automatic speech recognition in simultaneous interpreting. *Translation Quarterly* 104 (2), 1–20.
4. Li, T. & Fan, B. (2020). Attention-sharing initiative of multimodal processing in simultaneous interpreting. *International Journal of Translation, Interpretation, and Applied Linguistics* 2 (2), 42–53. https://doi.org/10.4018/IJTIAL.20200701.oa4
5. Cui, & Li, T. (2023). The effectiveness of second-person reference in enhancing reader memory in English-Chinese advertisement translation: An empirical study. *Translation Spaces* 12 (1), 144–166. https://doi.org/10.1075/ts.21033.cui

学位论文评阅及答辩情况表

<table>
<tr><td rowspan="6">论文评阅人</td><td colspan="2">姓　名</td><td>专业技术职　务</td><td>是否博导（硕导）</td><td colspan="2">所 在 单 位</td><td>总体评价</td></tr>
<tr><td colspan="2">匿名评阅人</td><td></td><td></td><td colspan="2"></td><td>合格</td></tr>
<tr><td colspan="2">匿名评阅人</td><td></td><td></td><td colspan="2"></td><td>良好</td></tr>
<tr><td colspan="2">匿名评阅人</td><td></td><td></td><td colspan="2"></td><td>合格</td></tr>
<tr><td colspan="2"></td><td></td><td></td><td colspan="2"></td><td></td></tr>
<tr><td colspan="2"></td><td></td><td></td><td colspan="2"></td><td></td></tr>
<tr><td rowspan="9">答辩委员会成员</td><td colspan="2">姓　名</td><td>专业技术职　务</td><td>是否博导（硕导）</td><td colspan="3">所 在 单 位</td></tr>
<tr><td>主席</td><td>谌莉文</td><td>教授</td><td>博导</td><td colspan="3">浙江工商大学</td></tr>
<tr><td rowspan="7">委
员</td><td>Agnieszka Chmiel</td><td>教授</td><td>博导</td><td colspan="3">亚当·密茨凯维奇大学</td></tr>
<tr><td>王湘云</td><td>教授</td><td>博导</td><td colspan="3">山东大学</td></tr>
<tr><td>庄会彬</td><td>教授</td><td>博导</td><td colspan="3">山东大学</td></tr>
<tr><td>朱浩然</td><td>副教授</td><td>博导</td><td colspan="3">华中科技大学</td></tr>
<tr><td></td><td></td><td></td><td colspan="3"></td></tr>
<tr><td></td><td></td><td></td><td colspan="3"></td></tr>
<tr><td></td><td></td><td></td><td colspan="3"></td></tr>
<tr><td colspan="3">答辩委员会对论文的总体评价</td><td>优秀</td><td>答辩秘书</td><td>王苏华</td><td>答辩日期</td><td>2025 年 5 月 30 日</td></tr>
<tr><td>备注</td><td colspan="7"></td></tr>
</table>