

## 目錄 CONTENTS

iv Editor's Note 編者的話

### Articles 論文

01 Part-of-Speech (POS) Tagging Interpreting Corpora: Methods Developed for the Chinese/English Political Interpreting Corpus (CEPIC)

*Jun Pan Fernando Gabarron-Barrios Steven He Billy Tak Ming Wong*

47 Gamification of Interpreter Training: A Future Perspective

*Honghua Wang*

65 “中國故事”英中翻譯之“重新聚焦”模式：以彼得·海斯勒中國題材敘事新聞作品之簡中譯本為例

*吳音然*

77 嚴復《天演論·譯例言》講疏

*Yu Kit Cheung*

95 漢語可比語料庫中名詞化標記的分佈特徵差異及其原因探析

*戴光榮*

113 翻譯即協調：黎翠珍英譯粵劇《霸王別姬》研究

*張鼎程 張旭*

### Book Review 書評

131 Compositional Translatology: Unveiling New Perspectives on Translating Chinese Classics into English

*Xiangyu Yu Yifan Zhu*

### Essay 隨筆

137 關於宗教翻譯的幾點思考：翻譯的宗教性和宗教的翻譯性

*羅選民*

141 稿約凡例 Guidelines for Contributors

# Translation Quarterly

No. 105 2022

香港翻譯學會出版

二〇二二年  
第一百零五期

Published by  
The Hong Kong Translation Society

# 《翻譯季刊》

## *Translation Quarterly*

二零二二年九月 第一百零五期

No. 105, September 2022

版權所有，未經許可，不得轉載。

All Rights Reserved

Copyright © 2022 THE HONG KONG TRANSLATION SOCIETY ISSN 1027-8559-105

\*\*\*\*\*

The Hong Kong Translation Society has entered into an electronic licensing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The full text of *Translation Quarterly* can be found on EBSCO Publishing's databases.

\*\*\*\*\*

*Translation Quarterly* is available for open access on the website of its publisher, the Hong Kong Translation Society, at <https://www.hkts.org.hk/translation-quarterly>. All rights of the journal and materials therein are reserved by the publisher.

Readers may exercise fair use of the materials in *Translation Quarterly* for research, teaching and non-profit making purposes, provided that due and express acknowledgment be made citing the respective publication details, including the author, article title, journal title, issue number and year of publication, and where applicable, the publisher. Permission for reproduction, distribution or display of the journal and materials therein for any other purposes should be sought by writing to the publisher at [tq@hkts.org.hk](mailto:tq@hkts.org.hk).

# 翻譯季刊

*Translation Quarterly*

香港翻譯學會

The Hong Kong Translation Society

## Founding Chief Editor 創刊主編

Liu Ching-chih 劉靖之

## Honorary Chief Editor 榮譽主編

Leo Tak-hung Chan 陳德鴻

## Chief Editor 主編

Li Dechao 李德超

## Associate Editors 副主編

Shelby Chan 陳嘉恩

Li Bo 李波

Liu Kanglong 劉康龍

Song Geng 宋耕

## Editorial Board 編輯委員會

Chan Kit-ying Elsie (Chair) 陳潔瑩 (主席) Poon Hon-kwong Joseph 潘漢光

Cheung Kay Fan Andrew 張其帆 Poupard Duncan 鄧彧

Laviosa Sara 邵璐

Lee Tong King 李忠慶 Sorby Stella 洪蘭星

Li Defeng 李德鳳 Wang Binhua 王斌華

Lung Wai-chu Rachel 龍惠珠 Yan Xiu Jackie 鄒秀

Moratto Riccardo 莫冉 Zheng Binghan 鄭冰寒

## Advisory Board 顧問委員會

Baker Mona Lörscher Wolfgang

Dollerup Cay St. André James 沈安德

Goldblatt Howard 葛浩文 Yang Cheng-shu 楊承淑

Lin Wen-yueh 林文月

## Editorial Manager 編務經理

Liu Zhongzhu 劉中柱

## 編者的話：

本期為《翻譯季刊》第一百零五期，共收錄六篇研究文章、一篇書評及一篇隨筆。潘珺、Fernando Gabarron-Barrios、Steven He 及黃德銘闡述了中英政治口譯語料庫 (CEPIC) 的構建問題。該庫庫容約 650 萬詞，涵蓋普通話、粵語和英語的演講原文及口譯的逐字轉錄。作者對該庫作了嚴格的詞性標注，並附帶語用特徵等注釋。在該文中，作者詳細闡述了 CEPIC 的詞性標注過程，對提高粵語，尤其口語的詞性標注做了有益的探討。王紅華的研究將遊戲方式引入口譯培訓，詳盡介紹如何通過 InterpFighters 將遊戲與口譯培訓融合起來。InterpFighters 是為口譯學生專門設計的遊戲，包括交傳和同傳訓練的影音材料。作者指出，引入 InterpFighters 可培養和發展學生的口譯能力，提高口譯技能，包括世界知識、專業知識、文化知識、記憶技能和人際交往能力。另外，亦可提高學生的能動性，使他們在遊戲過程中更投入並增強自主學習。吳音然一文圍繞“聚焦”的概念，對海斯勒的兩部中國題材作品《江城》、《尋路中國》及簡體中文譯本進行文本分析，指出“中國故事”元素在翻譯過程中伴隨視角變化得以重構，涉及從“局外人”聚焦者到“局內人”聚焦者變化的“重新聚焦”過程。張宇傑一文對嚴復《天演論·譯例言》做了注疏，旁徵博引，對該翻譯學的經典著作做了詳盡的注釋及解釋，讓讀者更好地理解當中的詞句義理。戴光榮採用語料庫方法探討了名詞化在翻譯及原創漢語的差異。作者發現，漢語翻譯語料庫中名詞化標記使用頻率高於漢語母語語料庫，名詞化標記在漢語譯文語料庫中的搭配也更為豐富。張鼎程及張旭考察了黎翠珍的英譯粵劇《霸王別姬》，指出譯者在譯文中有效協調中英兩套詩學規範，充分傳遞了原作的信息和音樂效果。黎譯本具有很強的可讀性、可表演性和可接受性，對粵劇藝術的推廣和傳播是一大助力。本期也收入兩篇書評。余翔宇和朱一凡對潘文國《中籍英譯通論》作了評述，指出該書有效結合翻譯理論及實踐，但某些概念仍存在模糊及明確定義。最後本期也收錄羅選民的隨筆。作者對宗教翻譯做了深入思考，闡述了翻譯的宗教性和宗教的翻譯性的辯證關係。

劉康龍

二零二二年九月



## 目錄 CONTENTS

iv **Editor's Note** 編者的話

### Articles 論文

- 01** Part-of-Speech (POS) Tagging Interpreting Corpora: Methods Developed for the Chinese/English Political Interpreting Corpus (CEPIC)  
*Jun Pan Fernando Gabarron-Barrios Steven He Billy Tak Ming Wong*
- 47** Gamification of Interpreter Training: A Future Perspective  
*Honghua Wang*
- 65** “中國故事”英中翻譯之“重新聚焦”模式：以彼得·海斯勒中國題材敘事新聞作品之簡中譯本為例  
*吳音然*
- 77** 嚴復《天演論·譯例言》講疏  
*Yu Kit Cheung*
- 95** 漢語可比語料庫中名詞化標記的分佈特徵差異及其原因探析  
*戴光榮*
- 113** 翻譯即協調：黎翠珍英譯粵劇《霸王別姬》研究  
*張鼎程 張旭*

### Book Reivew 書評

- 131** Compositional Translatology: Unveiling New Perspectives on Translating Chinese Classics into English  
*Xiangyu Yu Yifan Zhu*

### Essay 隨筆

- 137** 關於宗教翻譯的幾點思考：翻譯的宗教性和宗教的翻譯性  
*羅選民*
- 141** 稿約凡例 Guidelines for Contributors

# Part-of-Speech (POS) Tagging Interpreting Corpora: Methods Developed for the Chinese/English Political Interpreting Corpus (CEPIC)<sup>[1]</sup>

*Jun Pan<sup>1</sup> Fernando Gabarron-Barrios<sup>2</sup> Steven He<sup>3</sup> Billy Tak Ming Wong<sup>4</sup>*

**Address:** <sup>1,2,3</sup>Department of Translation, Interpreting and Intercultural Studies, Hong Kong Baptist University, Hong Kong SAR, China;

<sup>4</sup> Institute for Research in Open and Innovative Education, Hong Kong Metropolitan University, Hong Kong SAR, China

**E-mail:** <sup>1</sup>janicepan@hkbu.edu.hk; <sup>2</sup>fgabarron@life.hkbu.edu.hk; <sup>3</sup>stevenhehaoshen@gmail.com; <sup>4</sup>billywtm@gmail.com

**Correspondence:** Jun Pan

**Citation:** Pan, Jun, Fernando Gabarron Barrios, Steven He, and Billy Tak Ming Wong. 2022. "Part-of-Speech (POS) tagging interpreting corpora: Methods developed for the Chinese/English Political Interpreting Corpus (CEPIC)." *Translation Quarterly* 105: 1-45.

## ***Abstract***

*The Chinese/English Political Interpreting Corpus (CEPIC) is an open access corpus with about 6.5 million word tokens. The corpus features verbatim transcriptions of original speeches and interpretations in Putonghua, Cantonese, and English. It is part-of-speech (POS) tagged and annotated with prosodic and paralinguistic features. This paper reports the procedure of POS tagging the CEPIC. The most problematic step was POS tagging verbatim transcriptions of Cantonese, a popular "minority" language with limited language-related computer resources. Initial trials of existing tools and resources resulted in a low accuracy rate. The situation was complicated by the prosodic and paralinguistic annotations added to the verbatim transcriptions. In the CEPIC project, Stanford CoreNLP 3.9.2 (Manning et al. 2014) was employed. Manually tagged and checked POS data were used as basis to train and enhance performance of the Stanford tagger. The procedure involved conversion between Traditional and Simplified Chinese, and development of regular expressions to fix common issues. Methods to further enhance POS tagging Cantonese data were discussed at the end of the paper. Apart from contributing to interpreting corpora development, the procedure and discussions are believed to be able to shed light on the enhancement of POS tagging for spoken language,*

*in particular “minority” languages.*

## 1. Introduction

Corpus-based interpreting studies (CIS) has grown into a very important subfield of research in the past two decades (Shlesinger 1998; Bendazzoli and Sandrelli 2009; Setton 2011; Straniero Sergio and Falbo 2012; Russo et al. 2018). In spite of its importance, this subfield features very few open access interpreting corpora, except for the European Parliament Interpreting Corpus (EPIC) and its spinoffs (Russo et al. 2012).<sup>[2]</sup> The paucity of such data is mainly attributed to the difficulties of data collection, transcription and annotation of spoken data (Bendazzoli 2018; Bernardini et al. 2018). These difficulties also affect the size of this kind of corpora, which is usually of no more than one million word tokens (see Russo et al. 2018). Furthermore, some combinations of languages are rarer to find than others, e.g. Chinese-English, among existing (open access) corpora (ibid).

Another hinderance to CIS is the low performance of some Natural Language Processing (NLP) methods, such as taggers and lemmatisers, with spoken data, compared to written corpus projects. Westpfahl and Schmidt (2016, 1495), for instance, report a POS tagging accuracy of 81.61%, with TreeTagger (Schmid 1995) on a corpus of spoken German conversational data. The standard tagset/parameter file configuration failed to work due to the lack of typically spoken features, including but not limited to cases such as disruptions, unintelligible words, interjections, hesitation markers, onomatopoeia, etc., in most tagsets and training datasets. Sandrelli and Bendazzoli (2006) discussed in particular the challenges of applying POS tags (Treetagger, Freeling and GRAMPAL) to interpreting data that includes paralinguistic features such as word truncations, mispronounced words, disfluencies, etc., as well as ungrammatical sentences typical to spoken data. They suggested modifying and adding special tags to the existing tagging systems.

This paper, therefore, addresses the issue of Part-of-Speech (POS) tagging of interpreting corpora through introducing methods developed for the Chinese/English Political Interpreting Corpus (CEPIC). Special focus is put on POS tagging verbatim transcriptions of Cantonese, a popular “minority” language with limited language-related computer resources.

## 2. About the CEPIC

In response to the paucity of open access interpreting corpora relating to the Chinese language, the CEPIC was developed.<sup>[3]</sup> The corpus, with over 6 million tokens in size, is an on-line corpus designed for the study of Chinese/English political interpreting and translation (Pan 2019a, 2019b). It consists of verbatim transcripts of speeches delivered by top political figures

from Hong Kong, Beijing, London and Washington DC, as well as their translated/interpreted texts (ibid). Table 1 shows the composition of the CEPIC data by language (Pan 2019a).

Table 1: The composition of the CEPIC by language (Pan, 2019a).

	Word (Word Token)	Unique Word (Type)
Chinese	2,578,911	83,312
Cantonese	1,072,368	61,837
Putonghua	1,506,541	30,320
English	3,815,083	32,748
Total (Chinese + English)	6,393,994	116,060

The main speech types (Table 2) of CEPIC include monologues such as the reading of government reports (subset 1, 3, 5, 7, 9, 11 and 13), as well as dialogues such as question and answer sessions at press conferences (subset 2, 4, 6, 8, and 10), parliamentary debates (subset 12 and 14), as well as remarks delivered at bilateral meetings (subset 15 and 16, ibid).

Table 2: The composition of the CEPIC by speech types (Pan 2019a)

Speech Type		Word (Word Token)
<b>1</b>	HK SAR Policy Addresses (HKPA)	1,290,774
<b>2</b>	Press Conferences of HK SAR Policy Addresses (HK-PAPC)	326,194
<b>3</b>	HK SAR Budget Speeches (HKBS)	1,167,530
<b>4</b>	Press Conferences of HK SAR Budget Speeches (HKBSPC)	419,236
<b>5</b>	PRC Reports on the Work of the Government (PRC-WoG)	782,794
<b>6</b>	Press Conferences of PRC Reports on the Work of the Government (PRCWoGPC)	448,111
<b>7</b>	US State of the Union Addresses (USSoUA)	275,018
<b>8</b>	Press Conferences of US State of the Union Addresses (USSoUAPC)	266,639
<b>9</b>	US Budget Speeches (USBS)	73,115
<b>10</b>	Press Conferences of US Budget Speeches (USBSPC)	328,850
<b>11</b>	UK State Opening Addresses of Parliament (UK-SOoP)	31,006
<b>12</b>	Debates on the UK State Opening Addresses of Parliament (UKSOoPD)	53,941
<b>13</b>	UK Budget Speeches (UKBS)	469,452

14	Debates on the UK Budget Speeches (UKBSD)	376,721
15	Bilateral Meetings between PRC Key Politicians and their Counterparts in US (BMPRCUS)	70,138
16	Bilateral Meetings between PRC Key Politicians and their Counterparts in UK (BMPRCUK)	14,473
<b>Total</b>		6,393,994

Data of CEPIC were collected in two methods: 1) speech transcripts and their translations collected from government websites (raw data); and 2) revised or newly transcribed (when there are no readily available transcripts) verbatim versions of these speeches and their interpreted texts annotated with different prosodic and paralinguistic features that are of concern to the study of spoken language and interpreting, based on audios/videos collected from government websites and TV program archives (annotated data; the details of categories of prosodic and paralinguistic features included in CEPIC are discussed in Pan [2023, forthcoming]). The corpus is POS tagged.

Due to the small number of open access interpreting corpora and the limited size of existing corpora in the field (usually of no more than one million word tokens; see Russo et al. 2018), the transcription and annotation of CEPIC were mainly exploratory. Nevertheless, some of the contributions to CIS by the CEPIC encompass: 1) accessibility; 2) size; 3) language combination; 4) development of Natural Language Processing (NLP) methodologies to work with spoken data; 5) possible solutions to future collection of interpreting corpora by providing templates of metadata collection and solutions to spoken data transcription and annotation, especially for interpreting with the language combination of Chinese (Putonghua and Cantonese) and English; and 6) addressing understudied challenges involved in political interpreting (Pan 2019b).

Table 3 shows features of the CEPIC with a list of translation and/or interpreting corpora featuring political/parliamentary speeches. It displays information about these corpora: whether they include written or spoken data, paralinguistic annotations, total number of tokens, if they are open access, POS tagged, and finally, the segmenters (only for corpora including Chinese) and the POS taggers utilized. A website link is provided in the cases where these corpora are open access. As can be seen from the Table, the annotation of CEPIC is particularly challenging because of its comparative large size of data annotated with paralinguistic features.

Table 3: Translation and/or interpreting corpora featuring political/parliamentary speeches

Name of the corpus	Written /Spoken	Paralinguistic features	Tokens	Open access	POS tagged	POS tagger [4]
The Chinese/English Political and Interpreting Corpus (CEPIC) [5]	W+S	√	6,393,994	√	√	Stanford segmenter and POS tagger
The Chinese-English Conference Interpreting Corpus (CECIC) [6]	W+S	√	544,211	X	√	Treetagger and ICT-CLAS
The Chinese-English Consecutive Interpreting Corpus of Chinese Premier Press Conferences (CEIPPC) [7]	S	X	Over 100,000	X	X	X
The Corpus of Political Speeches [8]	W	X	6,269,359	√	X	X
The Digital Corpus of the European Parliament [9]	W+S	X	1,370,000,000	√	X	X
The European Comparable and Parallel Corpora (ECPC) [10]	W	X	207,513,043 (*Updating)	√ (*Updating)	√	Treetagger and Freeling
The European Parliament Interpreting Corpus (EPIC) [11]	S	√	177,295	X	√	Treetagger and Freeling

European Parliament Translation and Interpreting Corpus (EPTIC) <sup>[12]</sup>	W+S	√	367,341	√	√	Treetagger
European Parliament interpreting corpus Ghent (EPICG)	S	√	220,000	X	√	Treetagger
The WAW corpus	W+S	√	859,790	X	X	X
The UN Parallel Text <sup>[13]</sup>	S	X	Unknown	Subscription	X	X
中国外交话语语料库 <sup>[14]</sup>	W	X	Unknown	√	X	X

### 3. Linguistic resources and POS tagging for minority languages

Corpus data is absolutely necessary for many modern language engineering applications; scarcity in corpus data may gravely reduce the ability of language engineers to develop tools and systems for a certain language (McEnery and Ostler 2000; Sarasola 2000). Corpus data shortage might even have negative consequences for the future. Languages that do not partake in the electronic media are doomed to stagnate, or even atrophy (Ostler 1999, 3). Languages with few available resources are referred to as low density languages, which increasingly became a focus of attention within language engineering research by the end of the millennium. The terms “low density language” and “minority language” are used interchangeably, since both terms are used and accepted by the corpus linguistics and language engineering communities (McEnery et al. 2000). Somers (1998/2004) considers the term “minority language” ironic, because it is used to refer to under-resourced languages, but these languages include several of the world’s top 20 most spoken languages, Cantonese being one of them.

McEnery and Ostler (2000, 12) make a general distinction among languages with a written form, suggesting four broad types, depending on the machine-readable textual material available over time that makes them ready to be studied:

- 1) Official majority languages (e.g. Putonghua in China, English in the UK).
- 2) Official minority languages (e.g. Cantonese in China, Welsh in the UK).
- 3) Unofficial languages (both large, e.g. Kurdish in Turkey, and relatively small, e.g. Sylheti in the UK).

4) Endangered languages (e.g. Guugu Yimidhirr in Australia).

Written Cantonese is included in most NLP tools designed to process Chinese (Table 4 [cf. Table 6]); however, spoken Cantonese is under-represented in these tools. Fortunately, there exist some linguistic resources for spoken Cantonese, such as corpora, dictionaries, grammars, and jyutping romanization.

Table 4: An overview of open access available Cantonese dictionaries

	Written + Spoken Cantonese	Putong -hua	pin1 yin1	jyut6 ping3	Translation(s) to	Number of entries
<a href="#">CC-Cedict</a> [15]	W	√	√	X	English	120,096
<a href="#">CC-Canto</a> [16]	W+S	√	√	√	English	26,419
<a href="#">CC-Cedict-canto-readings</a> [17]	W	√	√	√	X	106,256
<a href="#">HanDeDict</a> [18]	W	√	√	X	German	149,576
<a href="#">CF-Dict</a> [19]	W	√	√	X	French	54,718
<a href="#">CH-Dict</a> [20]	W	√	√	X	Hungarian	16,491

In the methodology section of this paper we will provide details on the process of leveraging dictionaries from the CEDICT project to convert from Traditional (Hong Kong variant) to Simplified Chinese (and back to Traditional), to enhance the results for spoken Cantonese POS tagging. [21] As it can be observed in Table 4, CC-Canto dictionary is especially useful for our purpose, since it includes both written and spoken entries. Dictionaries featured in Table 4 are open access and available to download, their plain text formatting is consistent, separated by a space, and can be transformed into e.g. tsv or csv formats.

During the past two decades work with low density languages has been increasing. For instance, Streiter and William De Luca (2003) compare three approaches to cope with the limitation of linguistic resources (such as corpora and dictionaries) in minority language processing, i.e. rule-based, statistical-based, and example-based approaches. They conclude that an example-based approach has the advantage of requiring less linguistic resources than statistical approaches, and that example-based approaches can be used for most NLP tasks with low density languages.



A minimal corpus linguistic annotation should at least include POS information (McEnery et al. 2000), since it is an essential NLP tool, often considered the initial step in language analysis (Fang and Cohn 2017). Walther et al. (2010) developed a POS tagger for Kurmanji Kurdish. Arkhangelskiy and Medvedeva (2016) developed morphologically annotated corpora for minority languages of Russia, however, they defend that their corpora should not only be POS annotated, but also include full morphological tagging. Lameris and Stymne (2021) focus on the POS annotation of Scots, a language spoken in Scotland and Northern Ireland. Even more challenging specific examples can be found, such as POS tagging informal language used in social media, e.g. tweets in Irish (Lynn et al. 2015), tweets in Malay (Ariffin and Tiun 2018), etc. It can be expected that in the near future there will be even more research on corpus construction and annotation of low density languages, annotated with at least POS tags.

POS tagging has different purposes. In a classical NLP pipeline, tokenizing and sentence splitting are the first steps, then comes POS tagging, which can help the machine to take decisions in next steps, e.g. lemmatising, name entity recognition, and parsing sentence structure. In corpus linguistics, it can help study the usage of e.g. nominalization in spoken data. In translation, term-extraction tools that use a linguistic approach can, based on POS tagged data, identify word combinations that match particular POS patterns (Somers 1998/2004). Given the aforementioned importance, POS tagging was performed on the entire CEPIC data.

## **4. POS Tagging the CEPIC: challenges and solutions**

The languages included in the CEPIC (Cantonese, Putonghua, and English) were POS tagged with Stanford CoreNLP 3.9.2 (Manning et al. 2014), with an initial accuracy of only 69% for Cantonese in the annotated subset (with prosodic and paralinguistic features). The tagsets employed in this project were the Penn Chinese Treebank and the Penn English Treebank (that is the reason why other taggers such as Jieba were discarded, since they utilize different tagsets). Guidelines on POS enhancement for Putonghua and written Cantonese (Xia 2000), and for English (Santorini 1995), were employed before releasing the corpus on the CEPIC website.

Like the EPIC, the annotated subset of the CEPIC includes prosodic and paralinguistic features, and includes typical spoken language structures such as ungrammatical sentences (see Sandrelli and Bendazzoli 2006). Therefore, evaluation and revision of the POS tagging was a necessary step to take.

POS revision of the corpus followed a semi-automatic approach based on manually checked data (consisting of 30 percent of the entire data set), divided in different subsets that represented the whole corpus. The process of revision focused on finding textual patterns that could be found and replaced by using regular expressions. The accuracy rate of the POS tagging improved in general for all data. However, Cantonese (especially in the annotated

subset) presented a low accuracy rate, mainly due to inaccurate segmentation. If segmentation is inaccurate, it is expected that POS tagging will perform poorly.

The challenge of POS tagging CEPIC mainly lied in its way of transcribing Cantonese, which aimed to be a verbatim representation of this official minority language. Transcriptions of spoken forms of Cantonese (referred to as “spoken Cantonese” in this article, i.e. the method employed in the transcription of Cantonese in the annotated subset of CEPIC) and its formalized written forms in Traditional Chinese (referred to as “written Cantonese” in this article, i.e. a writing system similar in vocabulary and sentence structure to written Putonghua and employed in the raw subset of CEPIC) share a complex relation: many features of Cantonese are not found in writing. Cantonese has a considerable amount of literature, oral and written, however, relatively little of this has been translated. Oral (or colloquial) Cantonese includes many words and expressions which are not normally considered acceptable in written Chinese (Matthews and Yip 2011). This vernacular style is also increasingly found in Hong Kong newspapers (Snow 2004).

Figure 1 shows the differences between the raw and annotated data in CEPIC, with an example taken from the reading of the policy address by former Chief Executive Mr. Tung Chee-hwa on 12 January 2005 (Pan 2019a). Even a formal reading of a written text suggests the lexical differences between the “Cantonese Raw” and “Cantonese Annotated” subsets such as “仍然”/“仍舊”(still), “二十”/“廿”(twenty), “我們”/“我哋”(we), “/”/“嘅” (a particle indicating possessiveness; see Appendix 2 for a full list of such particles), etc., whilst the “Putonghua Raw”/“English Raw” and “Putonghua Annotated”/“English Annotated” are almost identical to each other.

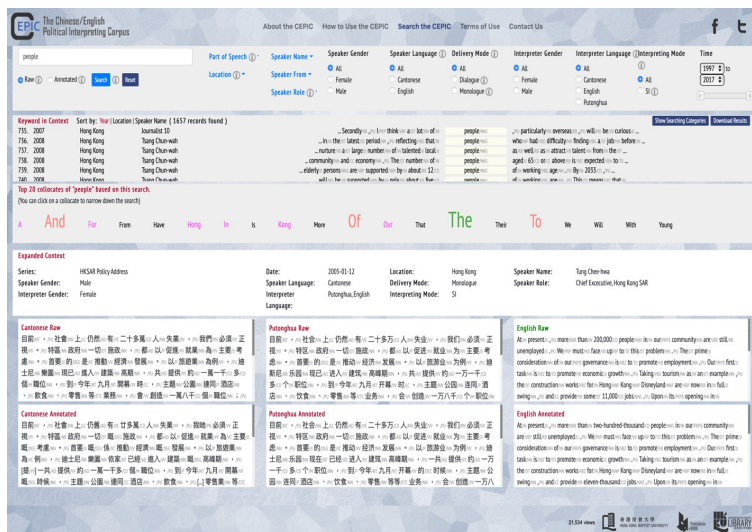


Figure 1: A screen capture of the CEPIC which shows the differences between the raw and annotated data

In this paper we focus primarily on the enhancement of POS tagging for spoken Cantonese (annotated data). We obtained positive results by converting written Cantonese into Simplified Chinese, POS tagging, and lastly converting the text back to Traditional Chinese. Thus, we started with the hypothesis that the approach taken for written Cantonese would also benefit the accuracy of POS annotation for spoken Cantonese.

In the next sections we present: 1) the methodology and tools employed during the development of the CEPIC; and 2) the approaches (rule-based and statistical-based; Streiter and Willian De Luca 2003) to enhance low accuracy in spoken Cantonese POS tags.

## **4.1 Steps taken for POS tagging**

The methodology used for POS tagging CEPIC started with a document conversion, from .docx format to plain text files in Unicode. The next steps consisted of segmentation of the Chinese texts and POS tagging. After that, procedures of POS accuracy enhancement were employed.

### **4.1.1 Document conversion**

Unicode is considered as the future for multilingual text encoding by many language engineers (McEnery and Ostler 2000). During the initial steps of the development of the CEPIC, a team of research assistants and student helpers were recruited and trained to help generate verbatim transcriptions, and to annotate prosodic and paralinguistic features using a manual developed by the first author (Pan [2023, forthcoming]). Transcribers and annotators worked with .docx files which was a document format they were most familiar with. In the end, the corpus consisted of 16 batches with a total of 950 .docx files. These .docx files may include different character encodings, <sup>[22]</sup> which represented a risk that may have led to corrupted characters. In addition, Unicode is needed to process texts with most modern NLP tools. Therefore, AntFileConverter 1.2.1 (Anthony 2017) was employed to convert all files to utf-8 (Unicode standard) plain text files, in batch processing.

### **4.1.2 Segmentation of Chinese**

Tokenization of raw text is “a standard pre-processing step for many NLP tasks” (The Stanford Natural Language Processing Group n.d., para 1). For English, tokenization “usually involves punctuation splitting and separation of some affixes like possessives” (ibid). Other languages, such as Chinese and Japanese, are standardly written without spaces between words, thus, they require “more extensive token pre-processing”, which is usually called segmentation (ibid).

Because of the Chinese language combinations in the CEPIC corpus, two methods were tested: 1) to segment with Jieba segmentation engine, included in SegmentAnt (Anthony, 2017), and add part-of-speech tags with Stanford POS tagger (Toutanova et al. 2003); and 2) to segment with Stanford Word Segmenter (Tseng et al. 2005), <sup>[23]</sup> included in Stanford

CoreNLP toolkit, and at the same time to add part-of-speech tags with Stanford POS tagger, also included in Stanford CoreNLP. In general, the latter approach, segmentation and POS tagging simultaneously, provided better results (see Appendix 1).

Within SegmentAnt (Figure 2) different options for segmentation are available. We used the Jieba segmentation engine and tried different dictionaries, i.e. the “small” and “big” dictionaries provided by Jieba, and noticed that spoken Cantonese vocabulary, e.g. “我哋”, [24] was not included in these dictionaries. At that point we tried adding CC-Cedict and CC-Canto (only the column with Traditional Chinese) as “custom” dictionaries. Results improved considerably, nevertheless, we kept obtaining low levels of segmentation accuracy, and consequently, low levels of POS tagging accuracy for annotated Cantonese data.

SegmentAnt also gives the option to POS tag simultaneously during segmentation (Figure 3). However, this segmenting tool, making use of Jieba and PyNLPIR engine for POS tagging, was initially developed for written Chinese and Japanese and resulted in a low accuracy rate on the spoken Chinese (esp. the Cantonese) data in CEPIC. After a comparison of the accuracy rate and with the intention to use comparable tag sets across all CEPIC language sets, we decided to use the Penn Treebank, which was designed and employed to annotate both spoken and written texts (Marcus et al. 1993; Taylor et al. 2003). Stanford Word Segmenter was used in the second approach, which consisted of segmenting and POS tagging simultaneously with Stanford NLPCore. In addition, Stanford Word Segmenter uses a statistical modelling approach, and it gives the option to train a segmentation model by the user.

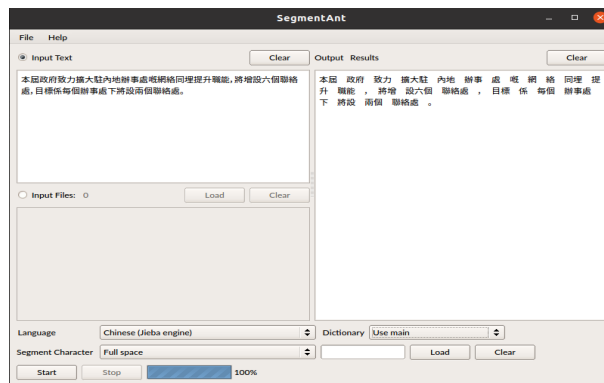


Figure 2: Chinese segmentation with SegmentAnt (Jieba engine)

### 4.1.3 POS tagging

A part-of-speech (POS) tagger is a piece of software that reads text in some language and assigns parts-of-speech tags to each word, such as noun, verb, adjective, etc., although generally computational applications use more fine-grained POS tags like “noun-plural” (Toutanova et al. 2003). Processing segmentation and part-of-speech tagging at the same time is the most efficient option, because the decisions taken by the segmenter help the tagger to decide better

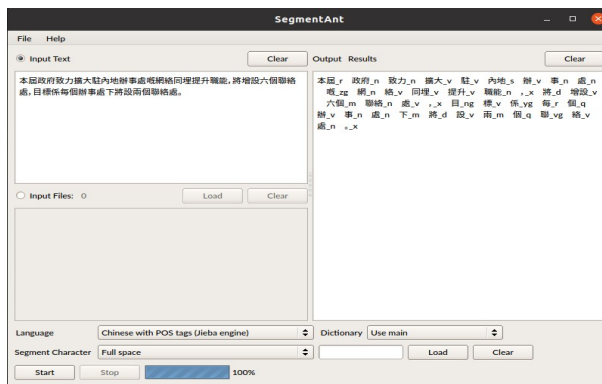


Figure 3: Chinese word segmentation with SegmentAnt and simultaneous POS tagging (Jieba engine)

on the tagging process (Ng and Low 2004).

POS taggers such as TagAnt (Anthony 2015), Figure 4, only annotate POS for European languages. However, SegmentAnt offers the option not only to segment Chinese and Japanese, but also to add POS tags simultaneously, as shown in Figure 3.

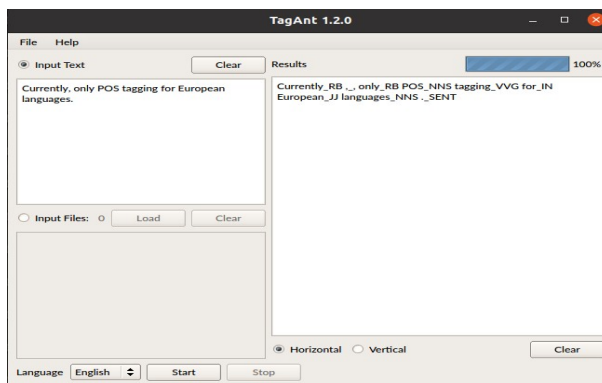


Figure 4: TagAnt Graphical User Interface

However, since SegmentAnt employs Jieba engine that targets written language, and CEPIC includes a substantial amount of spoken language and aims to employ comparable tag sets across the different language sets, an initial method of employing SegmentAnt and consequently Stanford Tagger (built on Penn Treebank and originally developed for both written and spoken language; Santorini 1995; Xia 2000), was employed before Stanford NLP Core was identified.

Stanford Tagger works with both European and Asian languages and learns a log-linear conditional probability model from tagged text, using a maximum entropy method (Toutanova 2000). Tagger results are only comparable when tested not only on the same data and tag set, but with the same amount of training data. Tagging performance increases as training set size

grows, largely because the percentage of unknown words decreases while system performance on them increases.

Table 5 shows the details and accuracy rate of the Stanford taggers for English and Chinese provided by the Stanford Natural Language Processing Group (n.d).<sup>[25]</sup> Whilst the English models performed with a high accuracy rate (about 90% for even unknown words), the two different available models of Chinese reach an accuracy of around 80% to 84% on unknown words. The accuracy rate the taggers performed on the Cantonese sets of the CEPIC data was slightly above 70% for Cantonese Raw and below 60% for Cantonese Annotated (calculated based on manual check of a random sample; see Appendix 1; also Table 6). Oral particles, pronouns, some verbs such as “to be” and “to have”, and certain annotations (truncated words, fillers, hesitations) and other specific features of oral Cantonese, can be considered in this process as unknown words. Eventually, we used the distsim model for the rule-based approach, since it provides a better accuracy for unknown words in both Chinese and English.

Table 5: Accuracy rate of Stanford tagger for the English and Chinese models indicated by the Stanford Natural Language Processing Group (n.d)

Model Name	Details	Performance
wsj-0-18-bidirectional-distsim.tagger	Trained on WSJ sections 0-18 using a bidirectional architecture and including word shape and distributional similarity features. Penn Treebank tagset.	97.28% correct on WSJ 19-21 (90.46% correct on unknown words)
wsj-0-18-left3words.tagger	Trained on WSJ sections 0-18 using the left3words architecture and includes word shape features. Penn tagset.	96.97% correct on WSJ 19-21 (88.85% correct on unknown words)
wsj-0-18-left3words-distsim.tagger	Trained on WSJ sections 0-18 using the left3words architecture and includes word shape and distributional similarity features. Penn tagset.	97.01% correct on WSJ 19-21 (89.81% correct on unknown words)
english-left3words-distsim.tagger	Trained on WSJ sections 0-18 and extra parser training data using the left3words architecture and includes word shape and distributional similarity features. Penn tagset.	N/A
english-bidirectional-distsim.tagger	Trained on WSJ sections 0-18 using a bidirectional architecture and including word shape and distributional similarity features. Penn Treebank tagset.	N/A

wsj-0-18-caseless-left3words-distsim.tagger	Trained on WSJ sections 0-18 left3words architecture and includes word shape and distributional similarity features. Penn tagset. Ignores case.	N/A
english-caseless-left3words-distsim.tagger	Trained on WSJ sections 0-18 and extra parser training data using the left3words architecture and includes word shape and distributional similarity features. Penn tagset. Ignores case.	N/A
chinese-nodistsim.tagger	Trained on a combination of CTB7 texts from Chinese and Hong Kong sources. LDC Chinese Treebank POS tag set.	93.46% on a combination of Chinese and Hong Kong texts (79.40% on unknown words)
chinese-distsim.tagger	Trained on a combination of CTB7 texts from Chinese and Hong Kong sources with distributional similarity clusters. LDC Chinese Treebank POS tag set.	93.99% on a combination of Chinese and Hong Kong texts (84.60% on unknown words)

From Table 6, the only NLP tools that processed correctly oral Cantonese features, e.g. “我哋”, were PyCantonese and N-LTP, although after revising the segmentation and POS tagging in general, for our sample, PyCantonese would work better. However, PyCantonese, which includes a statistical model trained with spoken data from the HKCanCor, lacks support for punctuation symbols. Apparently, during the process of transcription of this corpus, only basic punctuation was used (i.e. dot, comma, exclamation mark, and question mark). Since PyCantonese does not perform well with unknown words and punctuation, Stanford CoreNLP 3.9.2 was employed, following the rule-based approach presented in this paper. The paper will thus end with a discussion on the statistical-based approach to create a model to segment and POS tag spoken Cantonese, which represents an opportunity to use PyCantonese as a reference to enhance the existing limited performance of NLP tools for this “minority” language.

Table 6: Accuracy comparison among available tools for segmentation and POS tagging of oral Cantonese

	GUI	Process Pu-tonghua + Cantonese	Simultaneous segmenting + tagging	Custom trained models	Custom dictionaries	POS Tagset	Acc. on spoken Cantonese CEPIC sample	Acc. on spoken Cantonese CEPIC sample T-S-T*

<b>Stanford CoreNLP 3.9.2</b> <sup>[26]</sup>	No	Yes	Yes	Yes	No	Penn Tree-bank	0.69	<b>0.81</b>
<b>Jieba</b> <sup>[27]</sup>	No	Yes	Yes	No	Yes	NEUSCP	NC**	NC**
<b>NLTK</b> <sup>[28]</sup>	No	No	NA	NA	NA	NA	NA	NA
<b>N-LTP</b> <sup>[29]</sup>	No	Yes	Yes	Yes	No	NEUSCP	NC**	NC**
<b>PyCantonese</b> <sup>[30]</sup>	No	No	No	No	Yes	UPOS	NC**	NC**
<b>PyNLPIR</b> <sup>[31]</sup>	No	Yes	Yes	No	No	Symbol	NC**	NC**
<b>RNNTagger (Treetagger)</b> <sup>[32]</sup>	No	No	NA	NA	NA	NA	NA	NA
<b>SegmentAnt (Jieba)</b> <sup>[33]</sup>	Yes	Yes	Yes	No	Yes	NEUSCP	NC**	NC**
<b>Spacy (zh_core_web_lg)</b> <sup>[34]</sup>	No	Yes	Yes	Yes	No	Penn Tree-bank	0.31	0.36
<b>Spacy (zh_core_web_trf)</b> <sup>[35]</sup>	No	Yes	Yes	Yes	No	Penn Tree-bank	0.33	0.37
<b>Stanza</b> <sup>[36]</sup>	No	Yes	Yes	Yes	No	Penn Tree-bank	0.57	0.58
<b>THULAC</b> <sup>[37]</sup>	No	Yes	Yes	No	Yes	NEUSCP	NC**	NC**
<b>Treetagger</b> <sup>[38]</sup>	No	No	Yes	No	No	NEUSCP	NC**	NC**

\*Traditional converted to Simplified, segmented and POS tagged, and converted back to Traditional.

\*\*Not Comparable

To enhance the accuracy rate, Stanford CoreNLP was employed. It does the segmentation (with a higher accuracy as shown in our data; Appendix 1) and POS tagging simultaneously. Nevertheless, the Cantonese subsets (1, 2, 3 and 4 in Table 2) still show problems (with an accuracy rate increased by 4% for Cantonese Raw and 8% for Cantonese Annotated on the randomly selected sample; Appendix 1).

## 4.2 POS enhancement for the CEPIC

To further enhance the POS performance on the Cantonese subset, different procedures were employed with the Cantonese Raw and Annotated data. For the Cantonese Raw sets, since the wording of traditional Chinese (written form of Cantonese) and simplified Chinese (written form of Putonghua, a “non-minority language” with more well-developed NLP resources) were almost identical, we tried to 1) convert the traditional Chinese texts to simplified Chinese; <sup>[39]</sup> 2) employ Stanford CoreNLP 3.9.2; 3) convert the tagged simplified Chinese back to Traditional Chinese. The POS tagging was immediately improved to above 90% (Appendix 1). After the procedures were taken, both the Cantonese Raw and Annotated texts were compared with the original version and every difference caused by Chinese character variants or other reasons was manually corrected, in order to ensure consistency.

Nevertheless, this approach was not applicable to the Cantonese Annotated data due to the major lexical and syntactic differences between Cantonese and Putonghua. Therefore, segmentation plus POS tagging using Stanford CoreNLP 3.9.2 was employed (also see Appendix



1).

In this section we discuss POS enhancement considered for the CEPIC, a rule-based and statistical-based approach. The former refers to the “approach using resources which are more complex than the output”, which is “deduced from the linguistic resource (rules)”, and the latter uses “internally less structured resources than the expected output”, i.e., those “induced based on observed frequencies” (Streiter and Willian De Luca 2003, 236).

#### 4.2.1 Rule-based approach

During the process of POS tags enhancement we made use of the guidelines for Putonghua and written Cantonese (Xia 2000), and for English (Santorini 1995). A manual revision of about 30% of the total dataset was performed, using subsets that represented the whole corpus. This manual revision was assisted with regular expressions (Appendix 4), in order to avoid cumbersome repetition of revision of textual patterns.

Since Cantonese is under-represented (minority language) in the current NLP tools, segmentation and POS tagging provided low accuracy results on the Annotated Cantonese set. In addition, the manual checking process identified typical mistakes occurred during the conversion between Traditional and Simplified Chinese. A set of language conversion rules were therefore generated based on the manually checked data and Cantonese dictionaries (Appendix 3).

The most characteristic features of Cantonese that could normally be replaced taking a single action in CEPIC included (Part 2, Appendix 3):

- Personal pronouns
- Verbs
- Oral particles
- Annotations
- Other specific features

##### 1) *Personal pronouns*

Personal pronouns in written and spoken Putonghua, or in written and spoken English, are exactly the same. However, there are very characteristic differences between written and spoken (or colloquial) Cantonese. For example, one of the frequent POS tagging mistakes identified was “我哋” (over 5,000 occurrences, Appendix 3):

- 我们 - (pinyin: wǒmen) meaning “we” in Putonghua.
- 我們 - (jyutping: ngo5 mun4) meaning “we” in written Cantonese.
- 我哋 - (jyutping: ngo5 dei6) meaning “we” in spoken Cantonese.

Mistaken taggers on similar personal pronouns such as “你哋” were searched and corrected in one go (Appendix 3).

## 2) Verbs

Like pronouns, auxiliary verbs caused a similar problem:

-是 - (jyutping: si6) - meaning “to be” in written Cantonese.

-係 - (jyutping: hai6) - meaning “to be” in spoken Cantonese.

-沒有 - (jyutping: mut6 jau5) - meaning “to not have” in written Cantonese.

-冇 - (jyutping: mou5) - meaning “to not have” in spoken Cantonese.

Since Chinese follows an SVO grammatical structure, we observed that failing to segment pronouns and auxiliary verbs at the beginning of sentences accurately led to a sequence of errors in the tagged sentence, e.g.:

我哋冇足夠嘅錢/VV 畀佢哋/PN 做/VV , /PU

Just by finding and replacing text, we could solve issues related to pronouns and the verbs “to be” and “to have”, e.g.:

我哋/PN 冇/VV 足夠嘅錢/VV 畀佢哋/PN 做/VV , /PU

However, it can be noted that some issues still remained, such as “畀佢哋/PN”, instead of “畀/VV 佢哋/PN”.

## 3) Oral particles

Once again, oral particles constitute a specific characteristic featured in Cantonese. Appendix 2 features a complete list of oral particles and their ways of transcription in CEPIC. For example:

-啦 - (jyutping: lai1) - is placed at the end of an imperative sentence, making it sound more like a request than an order.

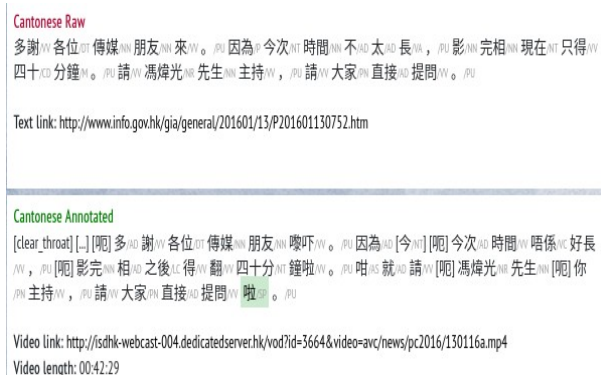


Figure 5: Snapshot taken from the CEPIC corpus, showing the oral particle “啦” included in spoken Cantonese (annotated data), but not in written Cantonese (raw data)

Oral particles were tagged as /SP (sentence-final particle) when they were located at the end of a sentence, and /IJ (interjection) when they were located in a different place.

#### 4) Annotations

Segmenters and POS taggers are not designed to work with annotations included in spoken data. Working with annotations such as fillers, e.g. [呢], [呃], [啊], [er], or self-repairs was problematic. An easy solution was to search for them at text segments where square brackets were tagged as punctuation, e.g. [/PU. However, a more complex issue to tackle was the bidirectionality approach of taggers such as Stanford's, which considers the previous word and the next word in order to take the tagging decision.

Furthermore, during the manual revision of tags, annotators focused eagerly on following the provided instructions, and sometimes, lapses that could be mainly attributed to fatigue would occur, such as tampering with POS tags. Regular expressions assisted in this revision, by checking the whole corpus as a single batch, looking for inconsistencies, and then fixing them, individually, depending on the issue. For example:

我哋/PN 並/vv 非要 AD [干擾]/NR 干預/VV 市場/NN , /PU

-並/vv, the manually revised tag should be uppercase This issue can be found by using the regex:  $\vee \backslash [ a-z ] + \backslash$

-非要 AD is missing its slash “/” symbol after the word and before the POS tag This issue can be found by using the regex:  $\backslash p \{ \text{script=Han} \} [A-Z]^+$

-干預/VV 市場/NN is missing the space between tokens (it should be 干預/VV 市場/NN) This issue can be fixed by using the regex: Find:  $(/[A-Z]^+)(\backslash p\text{script=Han})$   
Replace: \$1 \$2

All these issues could be detected instantly and revised case by case, using the negative lookahead regex:

-  $(!?\backslash p\text{script=Han} + \vee \backslash [A-Z]^+ \backslash)$

#### 5) Other cases

Ideally, all of the “係” should be verbs (VC) and all “喺” should be prepositions (P) in Cantonese. There is a possibility that those who did the transcription did not follow this rule consistently. When performing manual checking, it will be good to correct any ” 喺” to ” 係” when it is actually meant to be a verb, and correct any “係” to “喺” if it is meant to be a preposition. In POS tagging, it can be a rule that we always have “喺/P” and “係/VC”.

After enhancement through manual revision, assisted with regular expressions, written Cantonese still showed less than ideal results, and spoken Cantonese still showed a low level of accuracy, for segmentation and POS tagging.

Given the nature of the annotated data of the CEPIC corpus, we needed to apply some processing before and after POS tagging. Regular expressions were developed to automate this



word/expression and the column with its Putonghua equivalent; 4) concatenating the two aforementioned columns; 5) deleting duplicates; and 6) ordering entries alphabetically. After deleting duplicates, the new custom dictionary had a total of 233,652 entries, which included words and multi-word expressions.

Another forthcoming approach to tackle the problem of low segmentation and POS tagging accuracy for written Cantonese, and especially for spoken Cantonese, will be to train a statistical model, once we can generate and/or have access to enough manually fully revised data. This approach will be time consuming and require collaboration from different projects that include segmented and/or POS tagged spoken Cantonese data.

Joining efforts with existing corpora such as the Hong Kong Cantonese Corpus (HKCanCor) <sup>[41]</sup> (Kang-Kwong and Wong 2015) would be a good starting point. This corpus comprises transcribed conversations recorded between March 1997 and August 1998. It contains recordings of spontaneous speech (51 texts) and radio programs (42 texts), which involve 2 to 4 speakers, with 1 text of monologue. In total, it contains around 230,000 Chinese words. The text is word-segmented (i.e., tokenization is at word-level, and each token can span multiple Chinese characters). Tokens are annotated with part-of-speech (POS) tags and romanized Cantonese pronunciation.

Furthermore, the compilation of HKCanCor, and the manual segmentation and POS tagging, have provided sufficient data to create a statistical model, which is already included in PyCantonese (Jackson 2020), a Python library for Cantonese linguistics and NLP. <sup>[42]</sup> Actually, this is the only NLP tool that, by default, could correctly segment and POS tag e.g. “我哋” as a pronoun, when working with a sample from the CEPIC. Unfortunately, it does not use the Penn Chinese treebank tagset, and therefore it was originally excluded from the list of tools to segment and POS tag the CEPIC.

## 5. Discussion and conclusions

All the experiments with the different available tools for annotating were performed in Google Colab, with Jupyter notebooks, in order to ensure reproducible results. Part of the sample extracted from the CEPIC corpus that was used for these experiments can be found in Appendix 1. This sample was extracted from the HKPAPC (“Press Conferences of HKSAR Policy Addresses”) subcorpus of the CEPIC, specifically from the year 2017; it contains 20 paragraphs with over 500 words. Since only some of the available tools offer the option to split sentences, the sentences of the sample were split as a pre-processing step.

This experiment consisted mainly in comparing the results from two tables: 1) table that features results from adding POS tags with the available tools directly to the spoken Cantonese sample (Acc. on spoken Cantonese CEPIC sample); and 2) table that features results from converting the sample to Simplified Chinese, adding POS tags, and converting the sample

back to Traditional Chinese (Acc. on spoken Cantonese CEPIC sample T-S-T).

The rule-based approach taken in this project to enhance accuracy of segmentation and POS tagging of oral Cantonese provided satisfactory results. As it can be observed in Table 6, accuracy results for Stanford CoreNLP 3.9.2 increased from 69% to 81%.

The tool for calculating the accuracy ratio was *diffliB*,<sup>[43]</sup> a Python module that includes several functions and classes to compare datasets. Besides the function of calculating a percentage comparison ratio, it also provides the option to present results of sequence comparisons in a human-readable format, utilizing symbols to display the differences cleanly. As a reference, we used the sample with 20 sentences and approximately 500 words, after having performed exhaustive manual revision.

To conclude, the rule-based approach provided satisfactory results, however, it required a manual revision after the conversion process, to ensure all characters in the final text were consistent with the characters in the original text.

Certainly, there is a wide gamut of available tools to automatically process and annotate Chinese texts. However, these tools use different POS tagsets, i.e. Chinese Penn Treebank part-of-speech tagset, Chinese NEUSCP part-of-speech tagset, UPOS (Universal POS tagset) and the Chinese symbol part-of-speech tagset. Furthermore, the number of tags included in each tagset might range from 50 to 200. These characteristics hinder comparability of results among the currently available tools for annotation of Chinese texts. In this paper, we have focused on the Penn Chinese treebank tagset, so we have only compared tools that utilize this tagset.

One option, in order to compare the accuracy of all the available tools, would be to convert different tagsets, e.g. conversion from Chinese NEUSCP part-of-speech tagset, or Chinese symbol part-of-speech tagset to Chinese Penn Treebank part-of-speech tagset. However, this would be an inaccurate approach, since different tagsets follow different criteria; for example, in Chinese, the UPOS tagset has only one tag for verbs (VERB), while the Penn Treebank has more than one (VC for copula, VE for 有 as the main verb, and VV for other verbs).

In the case of Cantonese, as a minority language, there exist a limited number of linguistic resources. Fortunately, there is a specific project, HKCanCor, which has already compiled a corpus of oral Cantonese; this corpus has been used to train statistical models for segmentation and POS tagging. It might suffice to manually revise the segmentation and POS tagging of approximately 100,000 tokens, in order to merge this newly revised data with the previously revised data included in HKCanCor, and consequently obtain a larger dataset to update the statistical models of PyCantonese. This approach would likely result in a more robust tool to process spoken Cantonese.

We believe that apart from contributing to interpreting corpora development, the procedure and discussions will shed light on the enhancement of POS tagging for spoken language, in particular “minority” languages at large.

### Acknowledgement

This research was supported by the Early Career Scheme (ECS) of the Research Grants Council (Project No.: 22608716), the Digital Scholarship Grant of the Hong Kong Baptist University, the Faculty Research Grant of the Hong Kong Baptist University (Project No.: FRG2/17-18/046) and the General Research Fund (GRF) of the Research Grants Council (Project No.: 12623122).

### Notes

- [1] Part of the article is based on a conference presentation. Pan, Jun, Gabarron-Barrios, Fernando, & He, Haoshen Steven. 2021. "Part-of-Speech (POS) Tagging Enhancement for the Chinese/English Political Interpreting Corpus (CEPIC)." Paper presented at the Conference on Translation Studies in East Asia: Tradition, Transition and Transcendence (2021EAST). Abstract in Conference E-booklet (2021, 79). Hong Kong: Hong Kong Polytechnic University & Hong Kong Translation Society.
- [2] EPIC is considered the first online open access interpreting corpus (including speeches and interpretations delivered in Italian, English, and Spanish of 177,295 word tokens; <https://catalog.elra.info/en-us/repository/browse/ELRA-S0323/>). It has led to the development of the EPIC suite corpora, including the European Parliament Interpreting Corpus Ghent (EPICG), and the European Parliament Interpreting and Translation Corpus (EP-TIC), information of which is listed in Table 3. Nevertheless, the corpus ceased to be freely available since quite a few years ago.
- [3] <https://digital.lib.hkbu.edu.hk/cepic>
- [4] According to Tian and Lo (2015), Stanford POS tagger and the TreeTagger, among other POS taggers, were recommended for their effectiveness, which seems to reflect the scenario of taggers used by corpora listed in this table.
- [5] <https://digital.lib.hkbu.edu.hk/cepic/>
- [6] Hu and Tao (2013).
- [7] See Wang (2012).
- [8] <https://digital.lib.hkbu.edu.hk/corpus/terms.php> (Ahrens 2015).
- [9] <https://ec.europa.eu/jrc/en/language-technologies/dcep>
- [10] <http://ecpc.xtrad.uji.es/#ECPC-Web> (see Calzada Perez et al. 2006)
- [11] <https://catalog.elra.info/en-us/repository/browse/ELRA-S0323/> (See Monti et al. 2005; Russo et al. 2012)
- [12] <https://corpora.dipintra.it/eptic/> (see Bernardini et al. 2016)
- [13] <https://catalog.ldc.upenn.edu/LDC94T4A>
- [14] [http://www5.zzu.edu.cn/cdd/th\\_sjgy.jsp?urltype=tree.TreeTempUrl&wbtreeid=1106](http://www5.zzu.edu.cn/cdd/th_sjgy.jsp?urltype=tree.TreeTempUrl&wbtreeid=1106)
- [15] <https://cc-cedict.org/editor/editor.php?handler=Download>

- [16] <https://raw.githubusercontent.com/gkovacs/fcix-rime-config/master/cccedict-canto-readings.txt>
- [17] <https://raw.githubusercontent.com/gkovacs/fcix-rime-config/master/cccedict-canto-readings.txt>
- [18] [http://www.handedict.de/chinesisch\\_aeutsch.php?mode=dl\&sid=d80e36eefdb05750bd130ae1f322ca09](http://www.handedict.de/chinesisch_aeutsch.php?mode=dl\&sid=d80e36eefdb05750bd130ae1f322ca09)
- [19] <https://chine.in/mandarin/dictionnaire/CFDICT/>
- [20] <https://chdict.zydeo.net/en/download/>
- [21] <https://en.wikipedia.org/wiki/CEDICT>
- [22] English texts were frequently encoded in ASCII (American Standard Code for Information Interchange), Putonghua texts were generally encoded in GB (国标, Guóbiāo, character encoding for Simplified Chinese), and Cantonese texts in Big5 (character encoding for Traditional Chinese).
- [23] <https://nlp.stanford.edu/software/segmenter.shtml>
- [24] 我哋 - (jyutping: ngo5 dei6) means “we” in spoken Traditional Chinese.
- [25] Stanford POS Tagger Accuracy information can be found in the README.txt file, in “models” folder of the software package: <https://nlp.stanford.edu/static/software/tagger.shtml#Download>
- [26] <http://nlp.stanford.edu/software/stanford-corenlp-full-2018-10-05.zip>
- [27] <https://pypi.org/project/jieba/>
- [28] <https://www.nltk.org/>
- [29] <https://ltp.ai/>
- [30] <https://pycantonese.org/>
- [31] <https://pypi.org/project/PyNLPIR/>
- [32] <https://www.cis.lmu.de/~schmid/tools/RNNTagger/>
- [33] <https://www.laurenceanthony.net/software/segmentant/>
- [34] [https://spacy.io/models/zh\#zh\\_core\\_web1g](https://spacy.io/models/zh\#zh_core_web1g)
- [35] [https://spacy.io/models/zh\#zh\\_core\\_web1rf](https://spacy.io/models/zh\#zh_core_web1rf)
- [36] <https://stanfordnlp.github.io/stanza/index.html>
- [37] <http://thulac.thunlp.org/>
- [38] <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>
- [39] The imbedded conversion tool of Microsoft Word 2010 was employed. A list of typical mistaken words in the conversion process was identified and corrected by a simple “find” and “replace” command (Part 1, Appendix 3).
- [40] <https://www.cityu.edu.hk/csc/install-guide/TCSCTranslate.htm>
- [41] <http://compling.hss.ntu.edu.sg/hkcancor/>
- [42] <https://github.com/jacksonlee/pycantonese>
- [43] <https://docs.python.org/3/library/difflib.html>



## References

- Ahrens, Kathleen. 2015. "Corpus of Political Speeches." Hong Kong Baptist University Library. <http://digital.lib.hkbu.edu.hk/corpus/>.
- Anthony, Laurence. 2017. "AntFileConverter (Version 1.2.1) [Computer Software]." Tokyo, Japan: Waseda University. <http://www.laurenceanthony.net/software>.
- Anthony, Laurence. 2017. "SegmentAnt (Version 1.1.3) [Computer Software]." Tokyo, Japan: Waseda University. <http://www.laurenceanthony.net/software>.
- Anthony, Laurence. 2015. "TagAnt (Version 1.2.0) [Computer Software]." Tokyo, Japan: Waseda University. <https://www.laurenceanthony.net/software>.
- Ariffin, Siti Noor Allia Noor, and Sabrina Tiun. 2018. "Part-of-speech tagger for Malay social media texts." *GEMA Online Journal of Language Studies*, 18 no. 4: 124–142. <https://doi.org/10.17576/gema-2018-1804-09>.
- Arkhangelskiy, Timofey, and Maria Medvedeva. 2016. "Developing morphologically annotated corpora for minority languages of Russia." *CEUR Workshop Proceedings*, 1607: 1–6.
- Bendazzoli, Claudio, and Annalisa Sandrelli. 2009. "Corpus-based Interpreting Studies: Early work and future prospects." *Tradumatica 7. L'aplicació dels corpus linguistics a la traducció*. <http://webs2002.uab.es/tradumatica/revista/num7/articles/08/08art.htm>.
- Bendazzoli, Claudio. 2018. "Corpus-based interpreting studies: Past, present and future developments of a (wired) cottage industry." In *Making Way in Corpus-based Interpreting Studies (New Frontiers in Translation Studies Series)*, 1–19. Singapore: Springer.
- Bernardini, Silvia, Adriano Ferraresi, and Maja Miličević. 2016. "From EPIC to EPTIC — Exploring simplification in interpreting and translation from an intermodal perspective." *Target: International Journal of Translation Studies*, 28 no. 1: 61–86. <https://doi.org/10.1075/target.28.1.03ber>.
- Bernardini, Silvia, Adriano Ferraresi, Mariachiara Russo, Camille Collard, and Bart Defrancq. 2018. "Building interpreting and intermodal corpora: A how-to for a formidable task." In *Making Way in Corpus-based Interpreting Studies (New Frontiers in Translation Studies Series)* 21–42. Singapore: Springer.
- Calzada Pérez, María, Noemí Cucala, and José Martínez Martínez. 2006. "ECPC: European Parliamentary Comparable and Parallel Corpora." *Procesamiento del lenguaje natural*, 37: 349–350.
- Fang, Meng, and Trevor Cohn. 2017. "Model transfer for tagging low-resource languages using a bilingual dictionary." *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*: 587–593.
- Hu, Kaibao, and Tao Qing. 2013. "The Chinese-English conference interpreting corpus: Uses and limitations." *Meta* 58, no. 3: 626–642.
- Jackson, L. Lee. 2020. "PyCantonese: Cantonese Linguistics and NLP in Python." *GitHub*

- repository. <https://github.com/jacksonlee/pycantonese>.
- Kang-Kwong, Luke, and May L.Y. Wong. 2015. "The Hong Kong Cantonese Corpus: Design and Uses." *Journal of Chinese Linguistics* 12: 309–330.
- Lameris, Harm, and Sara Stymne. 2021. "Whit's the Richt Pairt o Speech: PoS tagging for Scots." *Proceedings of the Eighth Workshop on NLP for Similar Languages, Varieties and Dialects. Association for Computational Linguistics*: 39–48.
- Linguistic Data Consortium. n.d. "Transcription guidelines for EARS Chinese telephony conversational speech database", Linguistic Data Consortium, Penn University. Available from <https://tinyurl.com/2p8z42c6>. Accessed 19 July 2021.
- Lynn, Teresa, Kevin Scannell, and Eimear Maguire. 31 July 2015. "Minority language Twitter: part-of-speech tagging and analysis of Irish Tweets." *ACL 2015 Workshop on Noisy User-generated Text 2015 (W-NUT)*. Beijing, China.
- Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. "The Stanford CoreNLP Natural Language Processing Toolkit." *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*: 55–60.
- Marcus, Mitchell, Beatrice Santorini, and Mary Ann Marcinkiewicz. June 1993. "Building a large annotated corpus of English: The Penn Treebank." *Computational Linguistics, Special Issue on Using Large Corpora* 19, no. 2: 313–330. <https://aclanthology.org/J93-2004>.
- Matthews, Stephen, and Virginia Yip. 2011. *Cantonese: A Comprehensive Grammar*. Routledge. <https://doi.org/10.4324/9780203835012>.
- McEnery, Tony, and Nick Ostler. 2000. "A new agenda for corpus linguistics: working with all of the world's languages: Corpus building for minority and lesser studied languages." *Literary and Linguistic Computing* 15, no. 4: 403–430.
- McEnery, Tony, Paul Baker, and Lou Burnard. 2000. "Corpus Resources and Minority Language Engineering." *Proceedings of LREC 2000*. Athens, Greece.
- Monti, Cristina, Claudio Bendazzoli, Annalisa Sandrelli, and Mariachiara Russo. 2005. "Studying directionality in simultaneous interpreting through an electronic corpus: EPIC (European Parliament Interpreting Corpus)." *Meta: journal des traducteurs* [Meta: Translators' Journal] 50, no. 4. <https://www.erudit.org/en/journals/meta/1900-v1-n1-meta1024/019850ar/>.
- Ng, Hwee, and Jin Low. 25-26 July 2004. "Chinese part-of-speech tagging: One-at-a-time or all-at-once? Word-based or character-based?" *Conference: Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP, A meeting of SIGDAT, a Special Interest Group of the ACL, held in conjunction with ACL: 277–284*. Barcelona, Spain.
- Ostler, Nick. 1999. "Language technology and the Smaller Language." *ELRA Newsletter* 4,

no. 2.

- Pan, Jun. 2019a. "The Chinese/English Political Interpreting Corpus (CEPIC)." *Hong Kong: Hong Kong Baptist University Library*. <https://digital.lib.hkbu.edu.hk/ceplic/index.php>.
- Pan, Jun. 2019b. "The Chinese/English Political Interpreting Corpus (CEPIC): A New Electronic Resource for Translators and Interpreters." *Second Workshop on Human-Informed Translation and Interpreting Technology*: 82–88. Doi:10.26615/issn.2683-0078.2019\_010.
- Pan, Jun. 2023, forthcoming. "New Resources for the Study of Political Translation and Interpreting: The Case of the Chinese/English Political Interpreting Corpus (CEPIC)." In *Translating and Interpreting Political Discourse: New Trends and Perspectives*, edited by Jun Pan, Sandra Halverson, and Jeremy Munday. Leiden: Brill.
- Russo, Mariachiara, Claudio Bendazzoli, Annalisa Sandrelli, and Nicoletta Spinolo. 2012. "The European parliament interpreting corpus (EPIC): Implementation and developments." In *Breaking Ground in Corpus-Based Interpreting Studies*, edited by Francesco Straniero Sergio, and Caterina Falbo, 53–90. Bern: Peter Lang.
- Russo, Mariachiara, Claudio Bendazzoli, and Bart Defrancq, eds. 2018. *Making Way in Corpus-based Interpreting Studies*. Series ed. Defeng Li, Series: New Frontiers in Translation Studies. Singapore: Springer Nature.
- Sandrelli, Annalisa, and Claudio Bendazzoli. 2006. "Tagging a Corpus of Interpreted Speeches: the European Parliament Interpreting Corpus (EPIC)." *LREC*: 647–652.
- Santorini, Beatrice. 1995. "Part-Of-Speech Tagging Guidelines for the Penn Treebank Project (3rd revision, 2nd printing)." *Department of Linguistics, University of Pennsylvania*.
- Sarasola, Kepa. 2000. "Strategic Priorities for the Development of Language Technology in Minority Languages." *Proceedings of Workshop on developing language resources for minority languages: re-useability and strategic priorities. Second International Conference on Language Resources and Evaluation*. <http://ixa.si.ehu.es/node/3919?language=en>
- Setton, Robin. 2011. "Corpus-based Interpreting Studies (CIS): Overview and prospects." In *Corpus-based Translation Studies. Research and applications*, edited by Alet Kruger, Kim Wallmach, and Jeremy Munday, 33–75. London: Continuum.
- Shlesinger, Miriam. 1998. "Corpus-based Interpreting Studies as an offshoot of corpus-based Translation Studies". *Meta* 43, no. 4: 486–493.
- Schmid, Helmut. 1995. "Improvements in Part-of-Speech Tagging with an Application to German." *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.
- Snow, Donald. 2004. *Cantonese as Written Language: The Growth of a Written Chinese Vernacular*. Hong Kong: Hong Kong University Press.
- So, Simon Siu-hing. 2002. *A Glossary of Common Cantonese Colloquial Expressions*. Hong Kong: Chinese University Press.
- Somers, Harold. 1998. "Language resources and minority languages." *Language Today* 5: 20–

24.

- Somers, Harold. 2004. "Translation technologies and minority languages." In *Computers and translation: a translator's guide*. John Benjamins Pub. Co.
- Straniero Sergio, Francesco, and Caterina Falbo. 2012. *Breaking Ground in Corpus-based Interpreting Studies*. Bern: Peter Lang.
- Streiter, Oliver, and Ernesto William De Luca. 2003. "Example-based NLP for minority languages: tasks, resources and tools." *Proceedings of the Workshop Traitement automatique des langues minoritaires et des petites langues, 10th conference TALN*. Batz-sur-Mer, France, 233-242.
- Taylor, Ann, Mitchell Marcus, and Beatrice Santorini. 2003. "The Penn Treebank: An Overview." In *Treebanks. Text, Speech and Language Technology*, edited by Anne Abeillé, volume 20. Springer, Dordrecht. [https://doi.org/10.1007/978-94-010-0201-1\\_1](https://doi.org/10.1007/978-94-010-0201-1_1).
- Tian, Yuan, and David Lo. 2015. "A comparative study on the effectiveness of part-of-speech tagging techniques on bug reports." *2015 IEEE 22nd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*, 570-574.
- The Stanford Natural Language Processing Group. n.d. "Stanford Word Segmenter [Computer software]". The Stanford Natural Language Processing Group, Stanford University. <https://nlp.stanford.edu/software/segmenter.shtml>.
- Toutanova, Kristina, and Christopher D. Manning. 2000. "Enriching the knowledge sources used in a maximum entropy part-of-speech tagger." *Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000)*: 63–70.
- Toutanova, Kristina, Dan Klein, Christopher D. Manning, and Yoram Singer. 2003. "Feature-rich part-of-speech tagging with a cyclic dependency network." *Proceedings of HLT-NAACL*: 252–259.
- Tseng, Huihsin, Pi-Chuan Chang, Galen Andrew, Daniel Jurafsky, and Christopher D. Manning. 2005. "A conditional random field word segmenter for sighthan bakeoff 2005." *Proceedings of the Fourth SIGHAN Workshop on Chinese Language Processing. Association for Computational Linguistics*.
- Walther, Géraldine, Benoît Sagot, and Karën Fort. 2010. "Fast Development of Basic NLP Tools: Towards a Lexicon and a POS Tagger for Kurmanji Kurdish." *Proceedings of the 29th International Conference on Lexis and Grammar*. Belgrade, Serbia.
- Wang, Binhua. 2012. "A descriptive study of norms in interpreting: Based on the Chinese-English consecutive interpreting corpus of Chinese premier press conferences." *Meta* 57, no. 1: 198–212.
- Westpfahl, Swantje, and Thomas Schmidt. 2016. "FOLK-Gold—A GOLD standard for part-of-speech-tagging of spoken German." *Proceedings of the 10th LREC Conference (LREC 2016)*: 1493–1499.

Xia, Fei. 2000. "The Part-Of-Speech Tagging Guidelines for the Penn Chinese Treebank (3.0)." *IRCS Technical Reports Series*, 38.

高華年. 1980. 《廣州方言研究》. 香港：商務印書局.

饒秉才, 周無忌, 歐陽覺亞. 1981. 《廣州話方言詞典》. 香港：商務印書局.

# Appendix 1. POS tagging performance comparison between different methods

Table 1: Randomly selected Cantonese Raw sample text (Excerpted from HKSAR Policy Address, 2017)

Segment Ant + Stanford POS Tagger	Stanford CoreNLP	Stanford CoreNLP + Traditional and Simplified Chinese Conversion
<p>主席/NN、/PU 各位/DT 議員/NN、/PU 各位/DT 市民/NN，/PU 今年/NT 是/VC 香港/NR <u>回歸祖國</u>的/DEC 20/OD 周年/NN。/PU 20/CD 年/DT 來/M，/PU 國家/NN 的/DEG 發展/NN 取得/VV <u>舉世矚目</u>的/DEC 成就/NN，/PU 國際/NN 影響力/NN 亦/AD 不斷/AD 提升/VV。/PU 香港/NR 在/P「/PU 一國/DT 兩制/VV」/PU 的/DEG 安排/NN 下/LC，/PU 有/VE「/PU 一國/DT 兩制/NN」/PU 之/DEG 利/NN，/PU「/PU 兩制/NN」/PU 之/DEG 便/NN。/PU 面向/VV 內地/VV 的/DEC 高速/JJ 發展/NN 和/CC 國家/NN 的/DEG 優惠/NN 政策/NN，/PU 本屆/NN 政府/NN 和/CC 香港/NR 社會/NN 好/AD「/PU 一國/DT 兩制/VV」/PU 的/DEC 獨特/NR 雙重/JJ 優勢/NN，/PU 以經濟/DT 發展/VV 之力/NN，/PU 行/VV 改善/VV 民生/NN 之策/NN，/PU 同時/VV 亦/AD 為/VV 國家/NN 建設/NN 貢獻/NN 力量/NN。/PU 在/P 把握/VV 國家/NN 發展/NN 帶來/VV 機遇/NN、/PU 利用/VV 國家/NN 優惠/NN 政策/NN 的/DEG 同時/NN，/PU 我們/VV 要/VV 清楚/AD 認識/VV 到/VV 香港/NR 是/VC 國家/NN 不可/AD 分離/VV 的/DEC 部分/NN，/PU 這/DT 是/VC 法律/NN，/PU 也/AD 是/VC 國際/VV 公認/NN 的/DEC 政治/NN</p>	<p>主席/NN、/PU 各位/DT 議員/NN、/PU 各位/DT 市民/NN，/PU 今年/NT 是/VC 香港/NR 回歸/VV 祖國/NN 的/DEC 20/OD 周年/NN。/PU 20/CD 年/DT 來/M，/PU 國家/NN 的/DEG 發展/NN 取得/VV <u>舉世矚目</u>的/DEC 成就/NN，/PU 國際/NN 影/NN 響/NN 力/NN 亦/AD 不/斷/NN 提/升/VV。/PU 香/港/NR 在/P「/PU 一/CD 國/DT 兩/制/NN」/PU 的/DEG 安/排/NN 下/LC，/PU 有/VE「/PU 一/CD 國/DT 兩/制/NN」/PU 之/DEG 利/NN，/PU「/PU 兩制/NN」/PU 之/DEG 便/NN。/PU 面/向/VV 內/VV 地/VV 的/DEC 高/速/JJ 發/展/NN 和/CC 國/家/NN 的/DEG 優/惠/NN 政/策/NN，/PU 本/DT 屆/NN 政/府/NN 和/CC 香/港/NR 社/會/NN 好/VV「/PU 一/CD 國/DT 兩/制/NN」/PU 的/DEG 獨/特/NR 雙/重/JJ 優/勢/NN，/PU 以/P 經/濟/DT 發/展/VV 之/DEG 力/NN，/PU 行/VV 改/善/VV 民/生/NN 之/DEG 策/NN，/PU 同/時/VV 亦/AD 為/VV 國/家/NN 建/設/NN 貢/獻/VV 力/量/NN。/PU 在/P 把/握/VV 國/家/NN 發/展/NN 帶/來/VV 機/遇/NN、/PU 利/用/VV 國/家/NN 優/惠/NN 政/策/NN 的/DEG 同/時/NN，/PU 我/們/VV 要/VV 清/楚/AD 認/識/VV 到/VV 香/港/NR 是/VC 國/家/NN 不/可/分/離/VV 的/DEC 部/分/NN，/PU 這/DT 既/CC 是/VC 法/律/NN，/PU 也/AD 是/AD 國/際/VV 公/認/NN 的/DEC 政/治/NN 現/實/NN，/PU 香/港/NR 完/全/AD 沒/AD 有/VE</p>	<p>主席/NN、/PU 各位/DT 議員/NN、/PU 各位/DT 市民/NN，/PU 今年/NT 是/VC 香港/NR 回歸/VV 祖國/NN 的/DEG 20/OD 周/年/NN。/PU 20 年/NT 來/LC，/PU 國/家/NN 的/DEG 發/展/NN 取/得/VV <u>舉世矚目</u>的/DEC 成/就/NN，/PU 國/際/NN 影/響/力/NN 亦/AD 不/斷/AD 提/升/VV。/PU 香/港/NR 在/P「/PU 一國兩制/NR」/PU 的/DEG 安/排/NN 下/LC，/PU 有/VE「/PU 一/CD 國/DT 兩/制/NN」/PU 之/DEG 利/NN，/PU「/PU 兩制/NN」/PU 之/DEG 便/NN。/PU 面/向/VV 內/地/NN 的/DEC 高/速/JJ 發/展/NN 和/CC 國/家/NN 的/DEG 優/惠/NN 政/策/NN，/PU 本/DT 屆/DT 屆/M 政/府/NN 和/CC 香/港/NR 社/會/NN 用/好/VV「/PU 一國兩制/NR」/PU 的/DEG 獨/特/JJ 雙/重/JJ 優/勢/NN，/PU 以/P 經/濟/DT 發/展/VV 之/DEG 力/NN，/PU 行/VV 改/善/VV 民/生/NN 之/DEG 策/NN，/PU 同/時/AD 亦/AD 為/VC 國/家/NN 建/設/NN 貢/獻/NN 力/量/NN。/PU 在/P 把/握/VV 國/家/NN 發/展/NN 帶/來/VV 機/遇/NN、/PU 利/用/VV 國/家/NN 優/惠/NN 政/策/NN 的/DEC 同/時/NN，/PU 我/們/PN 要/VV 清/楚/AD 認/識/VV 到/VV 香/港/NR 是/VC 國/家/NN 不/可/分/離/VV 的/DEC 部/分/NN，/PU 這/PN 既/AD 是/VC 法/律/NN，/PU 也/AD 是/AD 國/際/NN 公</p>

現實/NN，/PU 香港/NR 完全/AD 沒有/VV 獨立/VV 或/CC 以/P 任何/DT 形式/NN 分離/VV 的/DEC 空間/NN。/PU 香港/NR 的/DEG 高度/JJ 自治/NN，/PU 是/VC 按照/P 《/PU 基本法/NN》/PU 規定/VV 的/DEC 自治/NN，/PU 不是/CC 完全/AD 自治/VV，/PU 不是/CC 隨意/VV 的/DEC 自治/NN，/PU 也/AD 不是/CC 任何/DT 其他/DT 形式/NN 或/CC 程度/NN 的/DEG 自治/NN。/PU 在/P 複雜/NR 關係/NN 中/LC，/PU 每/DT 一位/CD 香港/NT 市民/NN 都/AD 有/VE 義務/NN 維護/NN 國家/NN 主權/NN、/PU 安全/NN 和/CC 領土/NN 完整/NN。/PU 本屆/AD 特區/VV 政府/NN 會繼續/NN 嚴格/VV 按照/P 《/PU 基本法/NN》/PU，/PU 落實/VV 「/PU 一國兩制/NN」/PU，/PU 亦/AD 會/VV 繼續/NN 維護/NN 人權/NN、/PU 自由/NN、/PU 民主/NN、/PU 法治/NN 和/CC 廉潔/NN 等/ETC 核心/NN 價值/NN，/PU 促進/VV 社會/NN 穩定/VV 和諧/VV。/PU 我/NN 任內/VV 的/DEC 這份/NN 最後/AD 施政/NN 報告/NN，/PU 回顧/VV 了/AS 本屆/NN 政府/NN 的/DEG 施政/NN 成果/NN，/PU 也/AD 報告/VV 了/AS 我/PN 為/VV 香港/NR 中長期/NN 發展/NN 所/MSP 做/VV 的/DEC 準備/NN。/PU 5/CD 年前/NT，/PU 我/PN 提出/VV 「/PU 迎難/VV 而上/VV」/PU，/PU 這/VV 4/CD 個/M 字/NN，/PU 不/AD 僅/VV 代表/VV 我/PN 的/DEG 初心/NN，/PU 也是/AD 香港人/NN 建設/VV 我/PN 們 這個/NN 家/NN 的/DEG 體會/NN，/PU 「/PU 難/NN」/PU 說明/VV 我/PN 們 對/VV 的/DEC 認識/NN，/PU 「/PU 上/VV」/PU 代表/VV 了/AS 我/PN 們/VV 的/DEC 決心/NN 和/CC 行動/NN。/PU 5/CD 年前/NT，/PU 我/PN

獨立/NN 或/CC 以/P 任何/DT 形式/NN 分離/VV 的/DEC 空間/NN。/PU 香港/NR 的/DEG 高度/JJ 自治/NN，/PU 是/VC 按照/P 《/PU 基本法/NN》/PU 規定/VV 的/DEC 自治/NN，/PU 不/AD 是/VC 完全/AD 自治/VV，/PU 不/AD 是/VC 隨意/VV 的/DEC 自治/NN，/PU 也/AD 不/AD 是/VC 任何/DT 其他/DT 形式/NN 或/CC 程度/NN 的/DEG 自治/NN。/PU 在/P 複雜/NR 關係/NN 中/LC，/PU 每/DT 一位/CD 香港/NT 市民/NN 都/AD 有/VE 義務/NN 維護/NN 國家/NN 主權/NN、/PU 安全/NN 和/CC 領土/NN 完整/NN。/PU 本屆/AD 特區/VV 政府/NN 會繼續/NN 嚴格/VV 按照/P 《/PU 基本法/NN》/PU，/PU 落實/VV 「/PU 一國兩制/NN」/PU，/PU 亦/AD 會繼續/NN 維護/NN 人權/NN、/PU 自由/NN、/PU 民主/NN、/PU 法治/NN 和/CC 廉潔/NN 等/ETC 核心/NN 價值/NN，/PU 促進/VV 社會/NN 穩定/VV 和諧/VV。/PU 我/PN 任內/VV 的/DEC 這份/NN 最後/AD 施政/VV 報告/VV，/PU 回顧/VV 了/AS 本/DT 屆/NN 政府/NN 的/DEG 施政/NN 成果/NN，/PU 也/AD 報告/VV 了/AS 我/PN 為/VV 香港/NR 中長期/NN 發展/NN 所/MSP 做/VV 的/DEC 準備/NN。/PU 5/CD 年前/NT，/PU 我/PN 提出/VV 「/PU 迎難/VV 而上/VV」/PU，/PU 這/VV 4/CD 個/M 字/NN，/PU 不/AD 僅/VV 代表/VV 我/PN 的/DEG 初心/NN，/PU 也是/AD 香港人/NN 建設/VV 我/PN 們 這個/NN 家/NN 的/DEG 體會/NN，/PU 「/PU 難/NN」/PU 說明/VV 我/PN 們 對/VV 的/DEC 認識/NN，/PU 「/PU 上/VV」/PU 代表/VV 了/AS 我/PN 們/VV 的/DEC 決心/NN 和/CC 行動/NN。/PU 5/CD 年前/NT，/PU 我/PN

認/VV 的/DEC 政治/NN 現實/NN，/PU 香港/NR 完全/AD 沒有/VV 獨立/VV 或/CC 以/P 任何/DT 形式/NN 分離/VV 的/DEC 空間/NN。/PU 香港/NR 的/DEG 高度/JJ 自治/NN，/PU 是/VC 按照/P 《/PU 基本法/NN》/PU 規定/VV 的/DEC 自治/NN，/PU 不/AD 是/VC 完全/AD 自治/VV，/PU 不/AD 是/VC 隨意/VV 的/DEC 自治/NN，/PU 也/AD 不/AD 是/VC 任何/DT 其他/DT 形式/NN 或/CC 程度/NN 的/DEG 自治/NN。/PU 在/P 複雜/VV 的/DEC 國際/NN 關係/NN 中/LC，/PU 每/DT 一位/CD 香港/NT 市民/NN 都/AD 有/VE 義務/NN 維護/VV 國家/NN 主權/NN、/PU 安全/NN 和/CC 領土/NN 完整/NN。/PU 本/DT 屆/M 特區/NN 政府/NN 會繼續/VV 嚴格/AD 按照/P 《/PU 基本法/NN》/PU，/PU 落實/VV 「/PU 一國兩制/NN」/PU，/PU 亦/AD 會/VV 繼續/VV 維護/VV 人權/NN、/PU 自由/NN、/PU 民主/NN、/PU 法治/NN 和/CC 廉潔/NN 等/ETC 核心/NN 價值/NN，/PU 促進/VV 社會/NN 穩定/NN 和諧/VA。/PU 我/PN 任內/NN 的/DEG 這/DT 份/M 最後/JJ 施政/NN 報告/NN，/PU 回顧/VV 了/AS 本/DT 屆/M 政府/NN 的/DEG 施政/NN 成果/NN，/PU 也/AD 報告/VV 了/AS 我/PN 為/P 香港/NR 中長期/NN 發展/NN 所/MSP 做/VV 的/DEC 準備/NN。/PU 5年/NT 前/LC，/PU 我/PN 提出/VV 「/PU 迎難/VV 而上/VV」/PU，/PU 這/DT 4/CD 個/M 字/NN，/PU 不/AD 僅/VV 代表/VV 我/PN 的/DEG 初心/NN，/PU 也是/AD 香港人/NN 建設/NN 我們/PN 這個/DT 家/NN 的/DEG 體會/NN，/PU 「/PU 難/NN」

<p><u>/VV</u>家/<u>NN</u>的/<u>DEG</u>體會/<u>NN</u>，<u>/PU</u>「<u>/PU</u>難/<u>NN</u>」<u>/PU</u>說明<u>/NR</u>我們/<u>NN</u>對/<u>VV</u>問題/<u>VV</u>的/<u>DEC</u>認識/<u>NN</u>，<u>/PU</u>「<u>/PU</u>上/<u>VV</u>」<u>/PU</u>代表/<u>VV</u>了/<u>AS</u>我們/<u>NN</u>的/<u>DEG</u>決心/<u>NN</u>和/<u>CC</u>行動/<u>NN</u>。<u>/PU</u>5/<u>CD</u>年前/<u>NT</u>，<u>/PU</u>我/<u>PN</u>提出/<u>VV</u>了/<u>AS</u>全面/<u>NN</u>而/<u>MSP</u>具體/<u>VV</u>的/<u>DEC</u>競選/<u>NN</u>；<u>/PU</u>5/<u>CD</u>年來/<u>NN</u>，<u>/PU</u>每年/<u>AD</u>我/<u>PN</u>都/<u>AD</u>發表/<u>VV</u>施政/<u>NN</u>匯報/<u>NN</u>，<u>/PU</u>向/<u>P</u>社會/<u>NN</u>報告/<u>VV</u>政綱/<u>NN</u>的/<u>DEC</u>落實/<u>NN</u>進度/<u>NN</u>。<u>/PU</u>今天/<u>NT</u>，<u>/PU</u>我/<u>PN</u>競選/<u>VV</u>政綱/<u>NN</u>的/<u>DEC</u>承諾/<u>NN</u>基本上/<u>AD</u>已經/<u>AD</u>全部/<u>AD</u>落實/<u>VV</u>，<u>/PU</u>我要/<u>AD</u>感謝/<u>VV</u>所有/<u>DT</u>政治/<u>NN</u>任命/<u>NN</u>官員/<u>NN</u>和/<u>CC</u>公務員/<u>NN</u>同事/<u>NN</u>的/<u>DEG</u>努力/<u>NN</u>，<u>/PU</u>以及/<u>CC</u>行政/<u>NN</u>會議/<u>NN</u>、<u>/PU</u>立法會/<u>NN</u>、<u>/PU</u>法定/<u>JJ</u>機構/<u>NN</u>和/<u>CC</u>廣大/<u>NR</u>市民/<u>NN</u>的/<u>DEG</u>支持/<u>NN</u>和/<u>CC</u>配合/<u>NN</u>。<u>/PU</u>今年/<u>NT</u>施政/<u>NN</u>報告/<u>VV</u>的/<u>DEC</u>封面/<u>NN</u>和/<u>CC</u>封底/<u>NN</u>是/<u>VC</u>維港/<u>NR</u>的/<u>DEG</u>照片/<u>NN</u>。<u>/PU</u>維港/<u>NR</u>日景/<u>NN</u>壯麗/<u>VV</u>，<u>/PU</u>夜景/<u>NN</u>璀璨/<u>VA</u>，<u>/PU</u>標誌/<u>VV</u>着/<u>AS</u>一代/<u>LC</u>又/<u>AD</u>一代/<u>LC</u>香港/<u>NR</u>人務實/<u>NN</u>、<u>/PU</u>進取/<u>VV</u>的/<u>DEC</u>奮鬥/<u>NN</u>成果/<u>NN</u>。<u>/PU</u>香港/<u>NR</u>是/<u>VC</u>福地/<u>NN</u>，<u>/PU</u>我們/<u>NN</u>要/<u>VV</u>借福/<u>VV</u>，<u>/PU</u>要/<u>VV</u>珍惜/<u>VV</u>來/<u>VV</u>之/<u>DEC</u>不易/<u>VA</u>的/<u>DEC</u>成果/<u>NN</u>，<u>/PU</u>珍惜/<u>VV</u>香港/<u>NR</u>的/<u>DEG</u>優勢/<u>NN</u>，<u>/PU</u>珍惜/<u>VV</u>階層/<u>NN</u>和諧/<u>VV</u>，<u>/PU</u>更要/<u>AD</u>珍惜/<u>VV</u>機遇/<u>NN</u>、<u>/PU</u>珍惜/<u>VV</u>時間/<u>NN</u>。<u>/PU</u>我們/<u>NN</u>要/<u>VV</u>齊心/<u>AD</u>一意/<u>VV</u>，<u>/PU</u>開創/<u>NN</u></p>	<p>提出/<u>VV</u>了/<u>AS</u>全面/<u>VA</u>而/<u>CC</u>具體/<u>VA</u>的/<u>DEC</u>競選/<u>NN</u>政綱/<u>NN</u>；<u>/PU</u>5/<u>CD</u>年/<u>M</u>來/<u>NN</u>，<u>/PU</u>每/<u>DT</u>年/<u>M</u>我/<u>PN</u>都/<u>AD</u>發表/<u>VV</u>施政/<u>NN</u>匯報/<u>NN</u>，<u>/PU</u>向/<u>P</u>社會報/<u>NN</u>告/<u>VV</u>政綱/<u>NN</u>的/<u>DEC</u>落實/<u>NN</u>進度/<u>NN</u>。<u>/PU</u>今天/<u>NT</u>，<u>/PU</u>我/<u>PN</u>競選/<u>VV</u>政綱/<u>NN</u>的/<u>DEC</u>承諾/<u>NN</u>基本上/<u>AD</u>已經/<u>AD</u>全部/<u>DT</u>落實/<u>NN</u>，<u>/PU</u>我/<u>PN</u>要/<u>VV</u>感謝/<u>VV</u>所有/<u>DT</u>政治/<u>NN</u>任命/<u>NN</u>官員/<u>NN</u>和/<u>CC</u>公務員/<u>NN</u>同事/<u>NN</u>的/<u>DEG</u>努力/<u>NN</u>，<u>/PU</u>以及/<u>CC</u>行政/<u>NN</u>會議/<u>NN</u>、<u>/PU</u>立法會/<u>NN</u>、<u>/PU</u>法定/<u>JJ</u>機構/<u>NN</u>和/<u>CC</u>廣大/<u>NR</u>市民/<u>NN</u>的/<u>DEG</u>支持/<u>NN</u>和/<u>CC</u>配合/<u>NN</u>。<u>/PU</u>今年/<u>NT</u>施政/<u>NN</u>報告/<u>VV</u>的/<u>DEC</u>封面/<u>NN</u>和/<u>CC</u>封底/<u>NN</u>是/<u>VC</u>維港/<u>NR</u>的/<u>DEG</u>照片/<u>NN</u>。<u>/PU</u>維港/<u>NR</u>日景/<u>NN</u>壯麗/<u>VV</u>，<u>/PU</u>夜景/<u>NN</u>璀璨/<u>VA</u>，<u>/PU</u>標誌/<u>VV</u>着/<u>AS</u>一/<u>CD</u>代/<u>M</u>又/<u>AD</u>一/<u>CD</u>代/<u>M</u>香港人/<u>NN</u>務實/<u>NN</u>、<u>/PU</u>進取/<u>VV</u>的/<u>DEC</u>奮鬥/<u>NN</u>成果/<u>NN</u>。<u>/PU</u>香港/<u>NR</u>是/<u>VC</u>福地/<u>NN</u>，<u>/PU</u>我們/<u>NN</u>要/<u>VV</u>借福/<u>VV</u>，<u>/PU</u>要/<u>VV</u>珍惜/<u>VV</u>來/<u>VV</u>之/<u>PN</u>不/<u>AD</u>易/<u>VA</u>的/<u>DEC</u>成果/<u>NN</u>，<u>/PU</u>珍惜/<u>VV</u>香港/<u>NR</u>的/<u>DEG</u>優勢/<u>NN</u>，<u>/PU</u>珍惜/<u>VV</u>階層/<u>NN</u>和諧/<u>VV</u>，<u>/PU</u>更/<u>AD</u>要/<u>VV</u>珍惜/<u>VV</u>機遇/<u>NN</u>、<u>/PU</u>珍惜/<u>VV</u>時間/<u>NN</u>。<u>/PU</u>我們/<u>NN</u>要/<u>VV</u>齊心/<u>AD</u>一意/<u>VV</u>，<u>/PU</u>開創/<u>NN</u>更/<u>AD</u>美好/<u>VA</u>的/<u>DEC</u>未來/<u>NN</u>。<u>/PU</u>我們/<u>NN</u>這/<u>VV</u>一/<u>CD</u>代/<u>M</u>香港人/<u>NN</u>，<u>/PU</u>承前/<u>AD</u>啓後/<u>VV</u>，<u>/PU</u>任重/<u>VV</u>而/<u>MSP</u>道遠/<u>VV</u>。<u>/PU</u>未/<u>AD</u>來/<u>VV</u>半/<u>CD</u>年/<u>M</u>，<u>/PU</u>我/<u>PN</u>和/<u>CC</u>特區/<u>NR</u>政府/<u>NN</u>繼續/<u>VV</u>盡心盡力/<u>VV</u>，<u>/PU</u>繼續/<u>VV</u>全心全意/<u>AD</u>為/<u>VV</u>香港/<u>NR</u>的/<u>DEG</u>整體/<u>NN</u>和/<u>CC</u>長遠/<u>NN</u>利益/<u>NN</u>全力/<u>AD</u>打拼/<u>VV</u>。<u>/PU</u>日後/<u>AD</u>，</p>	<p><u>/PU</u>說明/<u>VV</u>我們/<u>PN</u>對/<u>P</u>問題/<u>NN</u>的/<u>DEG</u>認識/<u>NN</u>，<u>/PU</u>「<u>/PU</u>上/<u>VV</u>」<u>/PU</u>代表/<u>VV</u>了/<u>AS</u>我們/<u>PN</u>的/<u>DEG</u>決心/<u>NN</u>和/<u>CC</u>行動/<u>NN</u>。<u>/PU</u>5年/<u>NT</u>前/<u>LC</u>，<u>/PU</u>我/<u>PN</u>提出/<u>VV</u>了/<u>AS</u>全面/<u>VA</u>而/<u>CC</u>具體/<u>VA</u>的/<u>DEC</u>競選/<u>NN</u>政綱/<u>NN</u>；<u>/PU</u>5年/<u>NT</u>來/<u>LC</u>，<u>/PU</u>每/<u>DT</u>年/<u>M</u>我/<u>PN</u>都/<u>AD</u>發表/<u>VV</u>施政/<u>NN</u>匯報/<u>NN</u>，<u>/PU</u>向/<u>P</u>社會/<u>NN</u>報告/<u>NN</u>政綱/<u>NN</u>的/<u>DEG</u>落實/<u>NN</u>進度/<u>NN</u>。<u>/PU</u>今天/<u>NT</u>，<u>/PU</u>我/<u>PN</u>競選/<u>VV</u>政綱/<u>NN</u>的/<u>DEG</u>承諾/<u>NN</u>基本上/<u>AD</u>已經/<u>AD</u>全部/<u>AD</u>落實/<u>VV</u>，<u>/PU</u>我/<u>PN</u>要/<u>VV</u>感謝/<u>VV</u>所有/<u>DT</u>政治/<u>NN</u>任命/<u>NN</u>官員/<u>NN</u>和/<u>CC</u>公務員/<u>NN</u>同事/<u>NN</u>的/<u>DEG</u>努力/<u>NN</u>，<u>/PU</u>以及/<u>CC</u>行政/<u>NN</u>會議/<u>NN</u>、<u>/PU</u>立法會/<u>NN</u>、<u>/PU</u>法定/<u>JJ</u>機構/<u>NN</u>和/<u>CC</u>廣大/<u>JJ</u>市民/<u>NN</u>的/<u>DEG</u>支持/<u>NN</u>和/<u>CC</u>配合/<u>NN</u>。<u>/PU</u>今年/<u>NT</u>施政/<u>NN</u>報告/<u>NN</u>的/<u>DEG</u>封面/<u>NN</u>和/<u>CC</u>封底/<u>NN</u>是/<u>VC</u>維港/<u>NR</u>的/<u>DEG</u>照片/<u>NN</u>。<u>/PU</u>維港/<u>NR</u>日景/<u>NN</u>壯麗/<u>VV</u>，<u>/PU</u>夜景/<u>NN</u>璀璨/<u>NN</u>，<u>/PU</u>標誌/<u>VV</u>着/<u>AS</u>一/<u>CD</u>代/<u>M</u>又/<u>AD</u>一/<u>CD</u>代/<u>M</u>香港人/<u>NN</u>務實/<u>VA</u>、<u>/PU</u>進取/<u>VA</u>的/<u>DEC</u>奮鬥/<u>NN</u>成果/<u>NN</u>。<u>/PU</u>香港/<u>NR</u>是/<u>VC</u>福地/<u>NN</u>，<u>/PU</u>我們/<u>PN</u>要/<u>VV</u>借福/<u>VV</u>，<u>/PU</u>要/<u>VV</u>珍惜/<u>VV</u>來/<u>VV</u>之/<u>PN</u>不/<u>AD</u>易/<u>VV</u>的/<u>DEC</u>成果/<u>NN</u>，<u>/PU</u>珍惜/<u>VV</u>香港/<u>NR</u>的/<u>DEG</u>優勢/<u>NN</u>，<u>/PU</u>珍惜/<u>VV</u>階層/<u>NN</u>和諧/<u>VA</u>，<u>/PU</u>更/<u>AD</u>要/<u>VV</u>珍惜/<u>VV</u>機遇/<u>NN</u>、<u>/PU</u>珍惜/<u>VV</u>時間/<u>NN</u>。<u>/PU</u>我們/<u>PN</u>要/<u>VV</u>齊心/<u>AD</u>一意/<u>VV</u>，<u>/PU</u>開創/<u>VV</u>更/<u>AD</u>美好/<u>VA</u>的/<u>DEC</u>未來/<u>NT</u>。<u>/PU</u>我們/<u>PN</u>這/<u>DT</u>一/<u>CD</u>代/<u>M</u>香港</p>
---	--	--



<p><u>/NN</u>更/AD美好/VA <u>的DEC</u>未來/NN。/PU  <u>我們/NN</u> <u>這/VV</u> <u>一代/LC</u> <u>香港</u>  / NR 人/NN <u>/PU</u> <u>承/AD</u> 啓  /VV 後/LC，/PU <u>任重/VV</u> 而  <u>MSP</u> <u>道/VV</u> <u>遠NN</u>。/PU <u>未來</u>  /AD 半年/NT，/PU 我/PN 和  /CC 特區/NR 政府/NN <u>會繼續</u>  /NN <u>盡職/VV</u> <u>盡責NN</u>，/PU 繼續  /VV 全心全意/AD 為/VV 香港  / NR 的/DEG 整體/NN 和/CC <u>長</u>  /NN <u>遠NN</u> 利益/NN 全力/AD 打  拼/VV。/PU <u>日NR</u> 後LC，  /PU <u>我會NN</u> <u>繼續/NN</u> <u>本着P</u> 對  /VV 香港/ NR <u>這個NR</u> 家/NN 的  /DEG <u>熱和NN</u> 擔當/NN，  /PU 和/CC <u>700/CD</u> <u>萬M</u> 香港  / NR <u>一齊NN</u>，/PU 繼續/VV 奉  獻/VV 香港/ NR，/PU <u>奉獻國家</u>  / VV。/PU  <u>多謝AD</u> 各位/PN。/PU</p>	<p>/PU 我/PN <u>會繼續NN</u> <u>本着P</u> 對/VV  香港/ NR <u>這個NR</u> 家/NN 的/DEG 熱  愛/NN 和/CC 擔當/NN，/PU 和/CC  700/CD 萬/M 香港人/NN <u>一CD</u> <u>齊</u>  / M，/PU <u>繼續/NN</u> <u>奉VV</u> <u>獻VV</u> 香  港/ NR，/PU <u>奉獻NR</u> 家/NN。  /PU  <u>多謝各位VV</u>。/PU</p>	<p>人/NN，/PU <u>承/AD</u> 後/VV，/PU  <u>任重/VV</u> 道/VV。/PU 未來/NT 半年  /NT，/PU 我/PN 和/P 特區/NN  政府/NN 會/VV 繼續/VV <u>盡職盡</u>  / VV，/PU 繼續/VV 全心全意  /AD 為/P 香港/ NR 的/DEG 整體  /NN 和/CC 長遠/JJ 利益/NN 全力  /AD 打拼/VV。/PU 日後/AD，  /PU 我/PN 會/VV 繼續/VV <u>本着</u>  / VV 對/P 香港/ NR <u>這個DT</u> 家/NN  的/DEG 熱愛/NN 和/CC <u>擔當</u>  / VV，/PU 和/CC 700 萬/CD 香港  人/NN 一齊/AD，/PU 繼續/VV  奉獻/VV 香港/ NR，/PU 奉獻/VV  國家/NN。/PU  多謝/VV 各位/PN。/PU</p>
<p><b>Accuracy rate: 72.11% (388 correct tags/538 tags)</b></p>	<p><b>Accuracy Rate: 74.02% (416 correct tags/562 tags)</b></p>	<p><b>Accuracy rate: 92.15% (505 correct tags/548 tags)</b></p>

\* Note: The underlined-italics highlights indicate wrong word segmentation with wrong tags and the **underlined-bold** ones correct word segmentation with wrong tags.

Table 2: Randomly selected Cantonese Annotated Sample text (Excerpted from HKSAR Policy Address, 2017)

Segment Ant + Stanford POS Tagger	Stanford CoreNLP
<p>主席/NN、/PU 各位/DT 議員/NN、/PU 各位  /DT 市民/NN，/PU 今年/NT <u>係JJ</u> 香港/ NR <u>回</u>  <u>歸NN</u> <u>國VV</u> <u>嘅NN</u> <u>二十/OD</u> <u>周年/NN</u>。/PU 二  十/CD 年/M <u>嚟JJ</u>，/PU <u>國NR</u> <u>家NN</u> <u>嘅NN</u> <u>發VV</u>  <u>展NN</u> 取得/VV <u>舉世矚目NN</u> <u>日NN</u> <u>嘅VV</u> 成就  /NN，/PU 國際/NN 影響力/NN 亦/AD 不 斷  /AD 提升/VV。/PU 香港/ NR <u>嚟JJ</u> 「/PU <u>一國</u>  / VV <u>兩制NN</u>」/PU <u>嘅/NN</u> 安排/NN 下/LC，  /PU 有/VE 「/PU <u>一國/NN</u>」/PU 之/DEG 利  /NN，/PU 「/PU <u>兩制/NN</u>」/PU 之/DEG 便  /NN。/PU 面向/VV <u>內VA</u> 地/DEV <u>嘅VV</u> 高速  / JJ 發展/NN <u>同埋/VV</u> <u>國CD</u> <u>家M</u> <u>嘅NN</u> <u>優VV</u>  <u>惠VV</u> 政策/NN，/PU <u>本國NN</u> 政府/NN <u>同埋</u>  / VV 香港/ NR <u>社會用NN</u> <u>好JJ</u> 「/PU <u>一國JJ</u> <u>兩</u></p>	<p>主席/NN、/PU 各位/DT 議員/NN、/PU 各位  /DT 市民/NN，/PU 今年/NT 係/VV 香港/ NR 回  歸/VV <u>歸NN</u> <u>國嘅NN</u> <u>二十/OD</u> <u>周年/NN</u>。/PU  二十年/NT <u>嚟/NN</u>，/PU <u>國家嘅NN</u> 發展/NN 取  得/VV <u>舉世矚目NN</u> <u>日嘅NN</u> 成就/NN，/PU 國際  /NN <u>影NN</u> <u>響力NN</u> 亦/AD 不 斷 提升/VV。/PU 香  港/ NR 嚟/VV 「/PU <u>一CD</u> <u>國兩制/NN</u>」/PU <u>嘅</u>  / NN 安排/NN 下/LC，/PU 有/VE 「/PU <u>一</u>  / CD <u>國M</u>」/PU 之/DEG 利/NN，/PU 「/PU <u>兩</u>  / 制/NN」/PU 之/DEG 便/NN。/PU 面向/VV <u>內</u>  / VA 地/DEV <u>嘅VV</u> 高速/ JJ 發展/NN <u>同P</u> <u>埋國</u>  / NR 家NN <u>嘅優惠NN</u> 政策/NN，/PU 本/DT 屆  / NN 政府/NN <u>同CC</u> <u>埋VV</u> 香港/ NR 社會/NN 用  好/VV 「/PU <u>一CD</u> <u>國兩制/NN</u>」/PU <u>嘅獨VV</u></p>

制NN」/PU 嘅NN 獨NN 特AD 雙VA 重AD 優  
 /VV 勢NN，/PU 以經濟AD 發展VV 之  
 /NN，/PU 行/VV 改善/VV 民生/NN 之策NN，  
 /PU 同時/VV 亦AD 為/VV 國家/NN 建設/NN  
 貢獻/NN 力量/NN。/PU [...]  
 喺/IJ 把握/VV 國家/NN 發展/NN 帶/VV 嚟/IJ  
 機遇/NN、/PU 利用/VV 國家/NN 優惠/NN 政  
 策/NN 嘅NN 同CC 時NN，/PU 我哋/PN 要  
 /VV 清楚/AD 認識/VV 到/VV 香港/NR 係/IJ 國  
 家/NN 不可/AD 分VV 離NN 嘅NN 部分/NN，  
 /PU 呢/IJ 個VV 係/IJ 法律/NN，/PU 亦/AD 都  
 /AD 係/IJ 國際/VV 認NN 嘅VV 政治/NN 現實  
 /NN，/PU 香港/NR 完全/AD 冇/IJ 獨/VV 或  
 /CC 以/P 任何/DT 形式/NN 分VV 離NN 嘅  
 /VV 空/JJ 間NN。/PU 香港/NR 嘅/VV 高度/JJ  
 自治/NN，/PU 係/IJ 按照/P 《/PU 基本法  
 /NN》/PU 規NN 定VV 嘅NN 自治/NN，/PU  
 唔/IJ 係/IJ 完全/AD 自治/VV，/PU 唔/IJ 係/IJ  
 隨AD 意VV 嘅/VV 自治/NN，/PU 亦AD 都  
 /AD 唔/IJ 係/IJ 任何/DT 其他/DT 形式/NN 或者  
 /CC 程度/NN 嘅/VV 自治/NN。/PU 喺/IJ 複NN  
 雜NN 嘅NN 國NN 際NN 關NN 係/IJ 中/LC，  
 /PU 每/DT 一位/CD 香港市民/NN 都/AD 有/VE  
 義務/VV 維護NN 家/NR 主NN 權NN、/PU 安全/NN 同埋  
 /NN 領土/NN 完整/VA。/PU [...]  
 本國AD 特區/VV 政府/NN 會繼續NN 嚴格/VV  
 按照/P 《/PU 基本法/NN》/PU，/PU 落實/VV  
 「/PU 一國VV 兩制/NN」/PU，/PU 亦/AD 會  
 /VV 繼續/NN 維護/NN 人權/NN、/PU 自由  
 /NN、/PU 民主/NN、/PU 法治/NN 同埋/NN 廉  
 潔/NN 等/ETC 核心/NN 價值/NN，/PU 促進  
 /VV 社會/NN 穩定/VV 和諧/VV。/PU [...] 我  
 任NN 內VV 嘅VV 依P 份NN 最後AD 後LC 施政  
 /NN 報告/VV，/PU 回顧/VV 咗/IJ 本國  
 /NN 政府/NN 嘅/VV 施政/NN 成果/NN，/PU  
 [亦/AD 都/AD]亦/AD 都/AD 報告/VV 咗/IJ 我  
 /PN 為/VV 香港/NR 中長NN 期NN 發展/NN 所  
 /MSP 做/VV 嘅NN 準NN 備NN。/PU 五年  
 /NT 前/LC，/PU 我/PN 提出/VV 「/PU 迎難VV  
 而MSP 上VV」/PU，/PU 呢/IJ 四個CD 字  
 /NN，/PU 不僅/AD 代表/VV 我/PN 嘅/VV 初  
 /AD 心/NN，/PU 亦/AD 都/AD 係/IJ 香港/NR  
 人建設NN 我哋/PN 呢/IJ 個/CD 家/M 嘅NN 體  
 /VV 會NN，/PU 「/PU 難/NN」/PU 說明/VV  
 咗/IJ 我哋/PN 對問題NN 嘅NN 認NN 識

特優/VV，/PU 以/P 經濟發展VV 之  
 /DEC 力/NN，/PU 行/VV 改善/VV 民生/NN 之  
 /DEG 策/NN，/PU 同辦法國NR 家NN 建設 貢獻  
 /VV 力量/NN。/PU [...]  
 喺/NR 把握/VV 國CD 家M 發展/NN 帶發幾  
 /NN 邊VV、/PU 利用/VV 國CD 家M 優惠/NN  
 政策/NN 嘅同時NN，/PU 我PN 咁VV 要/VV 清楚  
 /AD 認識/VV 到/VV 香港/NR 係國NR 家  
 /NN 不可分NN 離離NN 部分/NN，/PU 呢/SP  
 個係VV 法律/NN，/PU 亦/AD 都/AD 係國際  
 /VV 公認NN 政治/NN 現實/NN，/PU 香港  
 /NR 完全/AD 冇獨VV 立VV 或/CC 以/P 任何  
 /DT 形式/NN 分VV 離離NN 空間/VV。/PU 香 港  
 /NR 嘅VV 高度/JJ 自治/NN，/PU 係/VV 按照/P  
 《/PU 基本法/NN》/PU 規定/VV 嘅/VV 自治  
 /NN，/PU 唔係NN 完全/AD 自治/VV，/PU 唔係  
 隨NN 意NN 嘅VV 自治/NN，/PU 亦/AD 都/AD  
 唔係/VV 任何/DT 其他/DT 形式/NN 或者/CC  
 程度/NN 嘅/VV 自治/NN。/PU 喺復NN 雜復雜  
 係NN 中/LC，/PU 每/DT 一/CD 位  
 /M 香港/NR 市民/NN 都/AD 有/VE 義務/VV 維護  
 /NR 家NN 主權/NN、/PU 安全/NN 同CC 埋領  
 /NN 土NN 完整/VA。/PU [...]  
 本國NR 政府/NN 會繼續VV 格VV 按照  
 /P 《/PU 基本法/NN》/PU，/PU 落實/VV  
 「/PU 一國兩制/NN」/PU，/PU 亦/AD 會  
 繼續VV 維護NN 人權/NN、/PU 自由/NN、  
 /PU 民主/NN、/PU 法治/NN 同CC 埋VV 廉潔  
 /NN 等/ETC 核心/NN 價值/NN，/PU 促進  
 /VV 社會/NN 穩定/VV 和諧/VV。/PU [...] 我  
 /PN 任內嘅VV 依P 份NN 最後AD 施政VV 報告  
 /VV，/PU 回顧/VV 本/DT 屆NN 政府  
 /NN 嘅/VV 施政/NN 成果/NN，/PU [亦/AD 都  
 /AD]亦/AD 都/AD 報告/VV 咗/IJ 我/PN 為/VV 香港  
 /NR 中長期NR 發展/NN 所/MSP 做/VV 嘅VV  
 /NN。/PU 五CD 年前/LC，/PU 我/PN 提出  
 /VV 「/PU 迎VV 難而上NN」/PU，/PU 呢/SP  
 四/CD 個字NN，/PU 不AD 僅VV 代表/VV 我  
 /PN 嘅NN 心/NN，/PU 亦/AD 都/AD 係/VV 香港  
 人/NN 建設/VV 我/PN 咁個國家團體會VV，  
 /PU 「/PU 難/NN」/PU 說明/VV 咗/IJ 我/PN 對問題  
 嘅NN 認NN 識/VV，/PU 「/PU 上/VV」/PU 代表  
 /VV 咗/VV 我/PN 咁個NN 心同VV 埋VV 行  
 動/NN。/PU 五CD 年前/LC，/PU 我/PN 提  
 出/VV 咗/NN 全面AD 而/MSP 具體發展VV

<p><u>/NN</u> ,/PU 「/PU上/VV」/PU代表/VV <u>咗/IJ</u> 我哋/PN <u>嚟/VV</u> <u>決VV心NN</u> <u>同埋/VV</u> <u>行動/VV</u> 。/PU 五年/NT前/LC ,/PU 我/PN 提出/VV <u>咗/IJ</u> <u>全面/AD</u> <u>而其VV</u> <u>體VV</u> <u>嘅NN</u> <u>競NN</u> <u>選/VV</u> <u>政NN</u> <u>綱NN</u> ;/PU 五年/NT來/NN ,/PU <u>每年/AD</u> 我/PN 都/AD 發表/VV 施政/NN 匯報/NN ,/PU 向/P 社會/NN <u>報VV政NN</u> <u>綱NN</u> <u>嘅NN</u> <u>落VV</u> <u>實NN</u> <u>進NN</u> <u>展NN</u> 。/PU 今日/NT ,/PU 我/PN <u>競VV政NN</u> <u>綱VV</u> <u>嘅NN</u> <u>承VV</u> <u>諾NN</u> <u>基本/AD</u> 已經/AD 全部/AD 落實/VV ,/PU <u>我要AD</u> 感謝/VV 所有/DT 政治/NN 任命/NN 官員/NN <u>同埋/VV</u> 公務員/NN 同事/NN <u>嘅VV</u> 努力/NN ,/PU 以及/CC 行政/NN 會議/NN 、/PU <u>立法NN</u> <u>會VV</u> 、/PU 法定/JJ 機構/NN <u>同埋/VV</u> 廣大/NN 市民/NN <u>嘅/AD</u> <u>支持/VV</u> <u>同埋/AD</u> <u>配合/VV</u> 。/PU [...]</p> <p>今年/NT <u>施政VV</u> <u>報NN</u> <u>告VV</u> <u>嘅VV</u> 封面/NN <u>同埋/NN</u> <u>封底/VV</u> <u>係/IJ</u> <u>維VV</u> <u>港NR</u> <u>嘅NN</u> 照片/NN 。/PU 維港/NR 日景/NN <u>壯麗/VV</u> ,/PU 夜景/NN 璀璨/VA ,/PU 標誌/VV 着/AS <u>一代/LC</u> 又/AD <u>一代/LC</u> 香港/NR <u>人務實NN</u> 、/PU <u>進NN</u> <u>取VV</u> <u>嘅NN</u> <u>奮NN</u> <u>鬥VV</u> 成果/NN 。/PU 香港/NR <u>係/IJ</u> 福地/NN ,/PU 我哋/PN 要/VV <u>惜福/VV</u> ,/PU 要/VV 珍惜/VV <u>來VV</u> <u>之/DEC</u> <u>不易VA</u> <u>嘅VV</u> 成果/NN ,/PU 珍惜/VV 香港/NR <u>嘅VV</u> <u>優NN</u> 勢/NN ,/PU 珍惜/VV 階層/NN 和諧/NN ,/PU 更加/AD 要/VV 珍惜/VV 機遇/NN 、/PU 珍惜/VV 時間/NN 。/PU 我哋/PN 要/VV <u>齊/AD</u> <u>一意/VV</u> ,/PU 開創/VV 更/AD 美好/JJ <u>嘅NN</u> <u>未AD</u> <u>來VV</u> 。/PU [...]</p> <p>我哋/PN <u>依P</u> <u>一代/LC</u> 香港/NR 人/NN ,/PU <u>承前AD</u> <u>啓後VC</u> ,/PU <u>任重VV</u> 而/MSP <u>道/VV</u> <u>遠NN</u> 。/PU <u>未來/AD</u> 半年/NT ,/PU <u>我哋/NR</u> 特區/NR 政府/NN <u>會繼續NN</u> <u>盡職VV</u> <u>盡責/VV</u> ,/PU 繼續/VV 全心全意/AD 為/VV 香港/NR <u>嚟/VV</u> <u>整DT</u> <u>體M</u> <u>同埋/NN</u> <u>長VV</u> <u>遠NN</u> 利益/NN 全力/AD 打拼/VV 。/PU <u>日NR</u> <u>後LC</u> ,/PU <u>我曾NN</u> <u>繼續/NN</u> <u>本着P</u> 對/VV 香港/NR <u>依個NR</u> 家/NN <u>嘅NN</u> <u>熱NN</u> <u>愛VV</u> <u>同埋/NN</u> 擔當/NN ,/PU 同埋/VV 七百萬/NT 香港/NR <u>人</u> <u>一齊/VV</u> ,/PU <u>繼續/VV</u> 奉獻/VV 香港/NR ,/PU <u>奉獻/VV</u> <u>國家VV</u> 。/PU <u>多謝AD</u> 各位/PN 。/PU</p>	<p>政綱/NN ;/PU <u>五CD</u> <u>年M</u> <u>來NN</u> ,/PU 每/DT 年/M 我/PN 都/AD 發表/VV 施政/NN 匯報/NN ,/PU 向/P <u>社會報NN</u> <u>告政VV</u> <u>綱NN</u> <u>落</u> <u>實VV</u> <u>NN</u> <u>展NN</u> 。/PU 今日/NT ,/PU 我/PN <u>競</u> <u>選VV</u> <u>政NN</u> <u>綱NN</u> 承諾/NN 基本上/AD <u>已/AD</u> <u>經VV</u> 全部/DT 落實/NN ,/PU 我/PN 要/VV 感謝/VV 所有/DT 政治/NN 任命/NN 官員/NN <u>同P</u> <u>埋VV</u> 公務員/NN 同事/NN <u>嘅VV</u> 努力/NN ,/PU 以及/CC 行政/NN 會議/NN 、/PU 立法會/NN 、/PU 法定/JJ 機構/NN <u>同P</u> <u>埋VV</u> <u>大JJ</u> 市民/NN <u>嘅VV</u> <u>支持/VV</u> <u>同P</u> <u>埋VV</u> 配合/NN 。/PU [...]</p> <p>今年/NT 施政/NN 報告/NN <u>嘅VV</u> 封面/NN <u>同P</u> <u>埋VV</u> <u>封底VV</u> <u>係維NR</u> <u>港NR</u> <u>嘅NN</u> 照片/NN 。/PU 維港/NR <u>日NR</u> <u>景麗NR</u> ,/PU 夜景/NN 璀璨/VA ,/PU 標誌/VV 着/AS <u>一/CD</u> 代/M 又/AD <u>一/CD</u> 代/M 香港人/NN 務實/NN 、/PU <u>進</u> <u>取VV</u> <u>鬥VV</u> 成果/NN 。/PU 香港/NR 係/VV 福地/NN ,/PU <u>我PN</u> <u>哋AD</u> 要/VV <u>惜福/VV</u> ,/PU 要/VV 珍惜/VV <u>來VV</u> <u>之</u> <u>不易VV</u> 成果/NN ,/PU 珍惜/VV 香港/NR <u>嘅勢NN</u> ,/PU 珍惜/VV 階層/NN 和諧/NN ,/PU 更加/AD 要/VV 珍惜/VV 機遇/NN 、/PU 珍惜/VV 時間/NN 。/PU 我/PN 哋/AD 要/VV <u>齊/AD</u> <u>一意/VV</u> ,/PU 開創/VV 更/AD 美好/JJ <u>嘅未來/VV</u> 。/PU [...]</p> <p>我/PN <u>哋依VV</u> <u>一/CD</u> 代/M 香港人/NN ,/PU <u>承前AD</u> <u>啓後VV</u> ,/PU <u>任重VV</u> 而/MSP <u>道遠/VV</u> 。/PU <u>未AD</u> <u>來VV</u> 半/CD 年/M ,/PU 我/PN 同/P 特區/NR 政府/NN <u>會繼續NN</u> <u>盡職盡責/VV</u> ,/PU 繼續/VV 全心全意/AD 為/VV 香港/NR <u>嚟VV</u> <u>整體NR</u> <u>同CC</u> <u>埋長遠NR</u> 利益/NN 全力/AD 打拼/VV 。/PU <u>日後AD</u> ,/PU 我/PN <u>會繼續NN</u> <u>本着P</u> 對/VV 香港/NR <u>依個NR</u> 家/NN <u>嚟愛/VV</u> <u>同CC</u> <u>埋管NN</u> ,/PU <u>同P</u> <u>埋VV</u> <u>七百CD</u> <u>萬M</u> 香港人/NN <u>一/CD</u> <u>齊M</u> ,/PU 繼續/VV 奉/VV 獻/VV 香港/NR ,/PU <u>奉獻/VV</u> <u>國家NR</u> 家/NN 。/PU 多/AD 謝/VV 各位/PN 。/PU</p>
<p>Accuracy rate: 55.91% (331 correct tags/592 tags)</p>	<p>Accuracy rate: 63.70% (344 correct tags/540 tags)</p>

\* Note: The *underlined-italics* highlights indicate wrong word segmentation with wrong tags and the **underlined-bold** ones correct word segmentation with wrong tags.

## Appendix 2. List of oral particles in spoken Cantonese used in the transcription of the CEPIC

Pronunciation	Meaning	Ref. 1*	Ref. 2*	Ref. 3*	Ref.4*	Suggestion
mou5	No 沒有	冇／無	冇			冇
ng4	Not 不	唔	唔			唔
saai3	Utter 完全	晒	晒	晒		晒
gam2	then 那麼	噉	噉	咁		咁
gam3	that 如此	咁	咁			咁
haih	yes 是	係	係	係		係
ngaam1	right	啱	啱			啱
sai2	Use/ need 使用／需要／用（不）用	使	使			使
ge3 ge2	expression/ of 語氣詞：的／（誰）的（東西）	嘅	嘅	嘅		嘅
mai5	don't 不要	咪	咪			咪
wan2	find 找	搵	搵			搵
go2	that 那	嗰	嗰	嗰		嗰
fāan	back/ return 回	番	番	翻		返／翻
je5	thing 東西	嘢	嘢			嘢
lai1 lai3 lai4	語氣詞：表示已經完成〔同：了〕	啦	啦	啦		啦
tai2	see/watch 看	睇	睇	睇		睇
saai1	waste 浪費	噉	噉			噉
zo2	to indicate past action 了	咗	咗	咗		咗
lai4 lei4	come/ to 來	嚟	嚟	嚟		嚟
dei6	plural marker for pronouns 們		哋	哋		哋
di1	a bit 一點	啲	啲	啲		啲
lo1	語氣詞		囉	囉		囉
bo3	語氣詞		噃	噃		噃
tim1	語氣詞		添	添		添

aa3	used before the name/person being addressed	阿	亞／阿	阿		阿
aa1 aa2 aa6 aa3 aa4	語氣詞：疑問、肯定／囑咐／驚嘆		呀	啊／呀	啊	啊
gaa3 gaa4 gaa2	語氣詞：疑問、肯定		㗎	㗎		㗎
aa1	語氣詞：同意／退讓，申辦／追問		吖			吖
naa4	語氣詞		喺	喺		喺
me1	語氣詞：疑問、反問	咩	咩	咩		咩
gwaa3	語氣詞：疑問，不確定		啱	啱		啱
gang2	something for sure 當然		梗			梗
lei4 ne1 ni1	this 這		呢	呢		呢
hai2	at (somewhere/ an event) 在		喺	喺		喺
ze1	only 而已／用作疑問		啫	啫		啫
zaa3 zaa2	only 而已／用作疑問		咋	咋		咋
bei2	to give 給	俾		畀		畀
dok6	there 那兒			嘍		嘍
ma1 ma4	語氣詞			嘛		嘛
ng6 ng2 ng5	query/ out of expectation/ answer 語氣詞：疑問／出乎意料／回答表示認同				嗯／唔	嗯
ak1 aak2 o1 o2	surprise/ understanding 語氣詞：表示驚奇／了解				喔	喔

\* Ref. 1: So, Simon Siu-hing. *A Glossary of Common Cantonese Colloquial Expressions*. Chinese University Press, Hong Kong, 2002.

\* Ref. 2: 饒秉才, 周無忌, 歐陽覺亞. *廣州話方言詞典*. 商務, 香港, 1981.

\* Ref. 3: 高華年. *廣州方言研究*. 商務, 香港, 1980.

\* Ref. 4: Transcription guidelines for EARS Chinese telephony conversational speech database.

## Appendix 3. Rule-based approach to POS enhancement

### Part 1. A list of word conversion between traditional and simplified Chinese

計畫 to 計劃	雇員 to 僱員	受雇 to 受僱	管道 to 渠道
專案 to 項目	水準 to 水平	受雇 to 受僱	社群 to 社羣
網路 to 網絡	儘快 to 盡快	鑒於 to 鑑於	跡象 to 迹象
資訊 to 訊息	憑著 to 憑着	編制 to 編製	佈 to 布
程式 to 程序	平臺 to 平台	接著 to 接着	週期 to 周期

瞭解 to 了解	協定 to 協議	隨著 to 隨着	厘定 to 釐定
軟體 to 軟件	品質 to 質量	彙聚 to 匯聚	著手 to 着手
裡 to 裏	夥同 to 伙同	著眼 to 着眼	鉤 to 鈎
雇主 to 僱主	週邊 to 外圍	占 to 佔	衛生 to 衛生

## Part 2. A list of common mistakes in POS tagging Cantonese Annotated data

Incorrect tag	Correct tag	Number of corrections
嘅/NN	嘅/DEC	11,558
呢/SP (?!啲)	呢/DT \$1 <sup>a</sup>	6,975
我/PN 啲(?!)	我哋/PN \$1	5,435
我/PN 啲/VV	我哋/PN	1,327
CD 萬/M	萬/CD	1,144
咗/NN	咗/AS	1,005
係/NN	係/VC	677
喺/NN and 喺/VV	喺/P and 係/VC <sup>b</sup>	375
嗰 and 嗰啲	嗰/DT and 嗰啲/DT	258
舊/OD 年/M	舊年/NT	93
你/PN 啲(?!)	你哋/PN	62
隨着/AD	隨着/P	27
同埋/NN	同埋/CC	5
上個/VV	上個/NT	2
下個/VV	下個/NT	2
依家/VV	依家/NT	2
今/OD 年/M	今年/NT	0
由/P 於/P	由於/P	0
呢/SP 一次/AD	呢/DT 一/CD 次/M	0
廿四/NN	廿四/CD	0

Notes:

a. Note that 呢 should be tagged as 呢/DT.

b. Note that ideally, all of the 係 should be verbs (VC) and all 喺 should be prepositions (P) in Cantonese. There is a possibility that those who did the transcription did not follow this rule consistently. So if you are performing manual checking, it will be good to correct any 喺 to 係 when it is actually meant to be a verb, and correct any 係 to 喺 if it is meant to be a preposition. In POS tagging, it can be a rule that we always have 喺/P and 係/VC.

The following steps were employed for automatic replacement to enhance the accuracy of automatic replacement:

- 1) Click "Replace..." in any word or excel document, type in "嘅/\*" in the box "Find what" and then type in "嘅/DEG" in the box "Replace with"
- 2) Click "Replace..." in any word or excel document, type in "嘅/DEG \*/PU" in the box "Find what" and then manually tag those "嘅/DEG" as "嘅/SP"
- 3) Click "Find..." in any word or excel document, type in "V\* 嘅/DEG" in the box "Find what", if there are not many cases, manually check if it is DEC or AS. If there are more than 2,000 cases, do the following:
  - a. Click "Find..." in any word or excel document, type in "V\* 嘅/DEG" in the box "Find what" and then manually tag those "嘅/DEG" as "嘅/DEC"
  - b. Click "Find..." in any word or excel document, type in "V\* 嘅/DEC \*NN" in the box "Find what" and then manually tag those "嘅/DEC" as "嘅/AS"
  - c. Click "Find..." in any word or excel document, type in "V\* 嘅/DEC \*PN" in the box "Find what" and then manually tag those "嘅/DEC" as "嘅/AS"

### Examples of Common Mistakes

NN	VV	Incorrect segregation	Digits	Proper Nouns	Others
1. 月/NN 月/NT	1. 仍舊/VV 仍舊/AD	1. 已/AD 經 已經/AD	萬/M, 萬/NN 萬/CD	1. 中华/NR 人民/NN 共和国/NN 中华人民共和国/NR	1. 同/P 同/CC
2. 個/NN 個/M	2. 重點/VV 重點/NN	2. 我/PN 們 我們/PN	億/M, 億/NN 億/CD	2. 落马/VV 洲/NN 落马洲/NR	2. 同/NN 埋/VV 同埋/CC
3. 嘢/NN 嘢/SP	3. 機會/VV 機會/NN	3. 仍/AD 舊/AD 仍舊/AD		3. 荷/NR 里/LC 活/VV 荷里活/NR	3. 幾年/AD 幾/CD 年/M
4. 與/NN 與/CC	4. 經濟/VV 經濟/NN	4. 社/NN 會 社會/NN		4. 九/CD 龙/NN 九龙/NR	4. 呢/SP 個 呢個/DT 呢/DT 個
5. 遠/NN 遠/AD	5. 原則/VV 原則/NN	5. 開/VV 始/VV 開始/VV		5. 东/NN 九/CD 东九/NR	5. 即係/CS 即係/AD
6. 舊年/NN 舊年/NT	6. 機遇/VV 機遇/NN	6. 促/VV 進 促進/VV		6. 强/JJ 积金/NN 强积金	6. 繼續/NR 繼續/AD
7. 即係/NN 即係/AD	7. 宜家/VV 宜家/NT	7. 維/VV 持/VV 維持/VV			
8. 教/NN 師 教師/NN 教/VV 師 教師/NN	8. 時候/VV 時候/NT	8. 一/CD 啲 一			

9. 係咪/NN 係  
咪/VC
10. 加強/NN  
加強/VV
11. 這些/NN  
這些/DT
12. 進行/NN  
進行/VV
13. 自從/NN  
自從/P
14. 大/JJ 約  
/NN 大約/JJ
15. 入/VV 面  
/NN 入面/LC
16. 委/NR 員  
會 委員會/NN
17. 財/NN 政  
司 財政司/NN
18. 政/NN 務  
司 政務司/NN
19. 即/NN 係  
/VC 即係/AD
20. 長/VV 者  
/NN 長者/NN
21. 委/VV 員  
會/NN 委員會  
/NN
22. 預/VV 算  
/VV 案/NN 預  
算案/NN
- 候/NN
9. 因為/VV 因  
為/AD
10. 佢哋/VV 佢  
哋/PN
11. 跟/P 住/VV  
跟住/AD
12. 時間/VV 時  
間/NN
13. 令/VV 到  
/VV 令到/VV
14. 隨/VV 著  
/AS 隨著/P
15. 嚟/VV 講  
嚟/MSP 講/VV
- 啲/CD
9. 可/VV 持/VV  
續發展 可持續  
發展/VV
10. 一/CD 國兩  
/M 制/NN 一  
/CD 國/N 兩  
/CD 制/NN
- /NR
7. 荔枝/NN 窩  
/NN 荔枝窩  
/NR
8. 亞/NR 投行  
/NN 亞投行  
/NR
9. 荔景/VV 荔  
景/NR
10. 油/NN 尖  
/AD 旺/VA 油  
尖旺/NR
11. 深水埗/NN  
深水埗/NR
12. 創/VV 科局  
/NN 創科局  
/NR
13. 珠/NR 三角  
/NN 珠三角  
/NR
14. 十二五/CD  
十二五/NR
15. 九/CD 龍  
/NN 東/NN 九  
龍東/NR
16. 新居/NN 屋  
/NN 新居屋  
/NR
17. 非/AD 綠/JJ  
表/NN 非綠表  
/NR
18. 馬頭/NR 圍  
道/NN 馬頭圍  
道/NR
19. 西/NN 九  
/CD 西九/NR
20. 勾/VV 地表  
/NN 勾地表  
/NR
21. 賣/VV 地  
/DEV 賣地/NR
22. 沙士/NN 沙  
士/NR



## Appendix 4. Regular Expressions used for POS enhancement

A regular expression, regex or regexp (sometimes called a rational expression), is a sequence of characters that define a search pattern. Usually this pattern is used by string searching algorithms for “find” or “find and replace” operations on strings, or for input validation. It is a technique that developed in theoretical computer science and formal language theory.

Perl one-liners

```
perl -pi -e 's/<regex_pattern_to_find>/<replacement>/g' *.txt
```

A one-liner program is textual input to the command-line of an operating system shell that performs some function in just one line of input. The one-liner used in this project is very simple, this way, for example, translators without a strong programming background, can use it efficiently, and build scripts with many searches and replacements, to process large amounts of text in batch process.

This is the one-liner used in the European Comparable and Parallel Corpus (ECPC; Calzada Perez et al. 2006) project, and now used in the Chinese/English Political Interpreting Corpus (CEPIC) project:

```
perl -pi -e 's///g' *.txt
```

e.g.: perl -pi -e 's/<regex to find text pattern>/<regex to replace the found pattern>/g' <filename>

“perl” - command executes the perl language in our Unix terminal, to process the rest of the line.

“-pi” - the “-p” argument makes sure the code gets executed on every line, and that the line gets printed out after that. The “-i” argument makes sure that the files to be processed are edited in-place, meaning Perl opens the files, executes the substitution for each line, prints the output to a temporary file, and then replaces the original file. In this project we use “-p” and “-i” together, thus becoming “-pi”.

“-e” - the “-e” argument allows us to specify the Perl code to be executed right on the command line.

“s” - the search argument.

“g” - the global argument. With this option the search will be in the whole document.

between the apostrophes (') we place the data to be processed. After the first slash we place the “find” and after the second slash we place the “replace”.

“\*.txt” - the asterisk, like in a wildcard search, specifies that we want to make changes in all the files that end with .txt extension; this expression will consider all the .txt files within the present working directory. Alternatively, we could just simply write the name of one file (example.txt). But with this method (\*.txt) we will process all the .txt files in a directory at once (batch mode).

Now let us see an example of the one-liner, with one regex within the “search” field, and another regex within the “replace” field:

```
perl -pi -e 's/^\s(\w)/$1/g' *.txt
```

In this example we could find a space at the beginning of a line, before a letter, and we would replace by just the letter, without the space.

Script before POS tagging

To convert all the text in just one line, delete paragraphs (Perl with Shell Terminal will work better if we convert all the text in just one line, by replacing \n with ZZZXPYYWW).

```
perl -pi -e 's/(\n|r)/ZZXPYYWW/g' *.txt
```



Workaround: start by deleting all /n and then use the ZZXXDOTYYWW\_NNP \_ . as a reference for paragraph jump

```
perl -pi -e 's/(n|r)+//g' *.txt
perl -pi -e 's/\s*ZZXXDOTYYWW\NR\.\VPU\s*/ZZXXPPYYWW/g' *.txt
perl -pi -e 's/\s*ZZXXDOTYYWW\[A-Z]+\.\VPU\s*/ZZXXPPYYWW/g' *.txt
perl -pi -e 's/\s*ZZXXDOTYYWW\.\V[A-Z]+\s*/ZZXXPPYYWW/g' *.txt
# change brackets into Library format
perl -pi -e 's/</[/g' *.txt
perl -pi -e 's/>/]/g' *.txt
```

Fix the dots annotations (pauses)

```
perl -pi -e 's/[VPU TWELVEDOTS\[A-Z]+ \VPU\s*/[.....]/g' *.txt
perl -pi -e 's/[VPU NINEDOTS\[A-Z]+ \VPU\s*/[.....]/g' *.txt
perl -pi -e 's/[VPU SIXDOTS\[A-Z]+ \VPU\s*/[.....]/g' *.txt
perl -pi -e 's/[VPU THREEDOTS\[A-Z]+ \VPU\s*/[...]/g' *.txt
perl -pi -e 's/([])\VPU\s*/$1/g' *.txt
perl -pi -e 's/s*([])\VPU/$1/g' *.txt
```

Delete tags for annotations, after angle brackets

```
perl -pi -e 's/([])#PU\s*/$1/g' *.txt
perl -pi -e 's/s([])#PU/$1/g' *.txt
```

To fix annotations spaces (before and after annotations)

```
perl -pi -e 's/([])\s+/$1/g' *.txt
perl -pi -e 's/([]) /$1/g' *.txt
```

Brackets to appropriate format (two possible formats, depending on having automatic spacing, Chinese, or not, English)

```
perl -pi -e 's^(VPU^(VLRB/g' *.txt
perl -pi -e 's^(VPU^)\VRRB/g' *.txt
perl -pi -e 's^ (VPU^(VLRB/g' *.txt
perl -pi -e 's^ ) \VPU^)\VRRB/g' *.txt
```

Now just left to check that all annotations are POS tagged and properly formatted, including the most problematic ones <several\_words> e.g. [/JJR property\_of\_rights/NNS ]

```
perl -pi -e 's^[interpreterVN.]/[interpreter]/g' *.txt
perl -pi -e 's^[interpreter2VN.]/[interpreter2]/g' *.txt
perl -pi -e 's^[laughVVV]/[laugh]/g' *.txt
perl -pi -e 's^[interruptionVN.]/[interruption]/g' *.txt
perl -pi -e 's^[applauseVN.]/[applause]/g' *.txt
perl -pi -e 's^[raiseVN.\s_VPU\svoiceVN.]/[raise_voice]/g' *.txt
perl -pi -e 's^[coughVVV]/[cough]/g' *.txt
perl -pi -e 's^[clearVN.\s_VPU\sthroatVN.]/[clear_throat]/g' *.txt
perl -pi -e 's^[inaudibleVN.]/[inaudible]/g' *.txt
```

```

perl -pi -e 's^[interruption\N.\s_\VPU\sooh\N.\.]/[interruption_oooh]/g' *.txt
perl -pi -e 's^[interruption\N.\s_\VPU\scheers\N.\.]/[interruption_cheers]/g' *.txt
perl -pi -e 's^[interruption\N.\s_\VPU\slaugh\N.\.]/[interruption_laugh]/g' *.txt
perl -pi -e 's^[interruption\N.\s_\VPU\shear\N.\s_\VPU\shear\N.\.]/[interruption_hear_hear]/g' *.txt
perl -pi -e 's^[interruption\N.\s_\VPU\sin audible\N.\.]/[interruption_inaudible]/g' *.txt
perl -pi -e 's^[㗎\IJ]/[㗎]/g' *.txt
perl -pi -e 's^[㗎\IJ]/[㗎]/g' *.txt
perl -pi -e 's^[㗎\IJ]/[㗎]/g' *.txt
perl -pi -e 's^[㗎\SP]/[㗎]/g' *.txt
perl -pi -e 's^[er\NN]/[er]/g' *.txt
perl -pi -e 's^[um\NN]/[um]/g' *.txt
perl -pi -e 's^[em\NN]/[em]/g' *.txt
perl -pi -e 's^[eh\NN]/[eh]/g' *.txt
perl -pi -e 's^[video\N.\s_\VPU\sstop\N.\.]/[video_stop]/g' *.txt
perl -pi -e 's^[twist\N.\.]/[twist]/g' *.txt
perl -pi -e 's^[code\N.\s_\VPU\sswitching\N.\.]/[code_switching]/g' *.txt

```

To fix annotations spaces (before and after annotations) for library

```

perl -pi -e
's/([interpreter]\|interpreter2)\|laugh\|interruption\|applause\|raise_voice\|cough\|clear_thro
at\|\.\+\|[inaudible]\|interruption_oooh\|interruption_cheers\|interruption_laugh\|interruption_h
ear_hear\|interruption_inaudible\|er\|um\|em\|eh\|video_stop\|twist\|code_switching\)/
$1/g' *.txt

```

To eliminate spaces between annotations

```

perl -pi -e 's/( ) (\)/$1$2/g' *.txt # for English
perl -pi -e 's/( ) (\)/$1$2/g' *.txt # for Chinese

```

To eliminate last space after each sentence

```

perl -pi -e 's/(\. )s+(ZZXXPYWW)/$1$2/g' *.txt
perl -pi -e 's/(\. )s+(ZZXXPYWW)/$1$2/g' *.txt

```

To eliminate possible spaces at the very end of the document

```

perl -pi -e 's/(\. )s+$/ $1/g' *.txt
perl -pi -e 's/(\. )s+$/ $1/g' *.txt

```

Oral Particles in Cantonese

```

perl -pi -e 's/(S)(冇|無|唔|晒|噉|咁|係|啱|使|咪|搵|個|番|翻|返|翻|嘢|啦|睇|啱|左|嚟|咁|啲|囉|囉|添|阿|
亞|啊|呀|啲|吖|噏|咩|咁|梗|呢|啲|啲|咋|俾|畀|嘍|嘛|唔|噏|噏)\[A-Z]+\|ZZXXPYWW/g' *.txt

```

Issues with numbers

```

perl -pi -e 's/我\PN 咁\IJ/我咁\PN/g' *.txt
perl -pi -e 's/減輕\NN/減輕\VV/g' *.txt

```

Stanford Command Line examples

Stanford Tagger (only POS tagging)

```
./stanford-postagger.sh models/chinese-distsim.tagger <input_file>.txt >> <output_file>.txt Stanford  
CoreNLP (segmentation/tokenization, sentence split, and POS tagging, in a pipeline)  
java -cp "*" -Xmx2g edu.stanford.nlp.pipeline.StanfordCoreNLP -props StanfordCoreNLP-  
chinese.properties  
-annotators tokenize,ssplit,pos -file<input_file>.txt -outputFormat conll -output.columns word,pos -  
output.prettyPrint false
```

Command Line options: <https://stanfordnlp.github.io/CoreNLP/cmdline.htm>

To process one file, use the -file option followed by a filename. To process a list of files, use the -fileList parameter, e.g.:

```
java -cp "*" -Xmx2g edu.stanford.nlp.pipeline.StanfordCoreNLP [ -props myprops.props ] -fileList  
filelist.txt
```

By default, output files are written to the current directory. You may specify an alternate output directory with the flag -outputDirectory. If your input is already tokenized and one sentence per line, then you should use the flags: -tokenize.whitespace  
-ssplit.eolonly.

Regular expressions for CEDICT dictionaries conversion to tsv file

To delete noise

Find:

```
(#|#).*?\n
```

Replace: (nothing)

For words with more than one option, e.g.

一個做好，一個做醜 一个做好，一个做丑 [yi2 ge5 zuo4 hao3, yi2 ge5 zuo4 chou3] {jat1 go3 zou6  
hou2, jat1 go3 zou6 cau2} /one plays the good guy, the other plays the bad/

Find:

```
^\p{script=Han}+, (\p{script=Han}+) (\p{script=Han}+), (\p{script=Han}+)( \[.+$)
```

Replace:

```
$1 $3$5\n$2 $4$5
```

cccanto-websdist.txt

Find:

```
^(.*?)(.*?)\[(.*?)\]\{(.*?)\}
```

\(.\*?\)\/\$ Replace:

```
$1\t$2\t$3\t$4\t$5
```

cedict\_ts.u8, dfdict.u8, chdict.u8, handedict.u8

Find:

```
^(.*?)(.*?)\[(.*?)\]\(.*?\)\/$
```

Replace:

\$1\t\$2\t\$3\t\$4

cccedict-canto-readings.txt

Find:

^(.\*?)(.\*?)\[(.\*?)\]

\{(.\*)\}\$ Replace:

\$1\t\$2\t\$3\t\$4



# Gamification of Interpreter Training: A Future Perspective

*Honghua Wang*

**Address:** School of Translation and Foreign Languages, The Hang Seng University of Hong Kong, Hang Shin Link, Siu Lek Yuen, Shatin, New Territories, Hong Kong

**E-mail:** ansonwang@hsu.edu.hk

**Correspondence:** Honghua Wang

**Citation:** Wang, Honghua. 2022. "Gamification of Interpreter Training: A future perspective." *Translation Quarterly* 105: 47-63.

## ***Abstract***

*This study depicts a picture of gamification in interpreter training and proposes for a marriage between games and interpreter training through the illustration of InterpFighters. InterpFighters is a game designed for interpreting students, which incorporates video and audio materials for consecutive interpreting and simultaneous interpreting training. The introduction of InterpFighters serves two purposes. The first is to cultivate and develop interpreting competence of students and to enhance their interpreting skills including world knowledge, subject-matter knowledge, cultural knowledge, memory skills and interpersonal skills. These training elements correspond to different parts in the game, namely, Quests, Leader Board and Special Mission. The second purpose is to increase learners' motivation and to make them willing to spend more time in this game, which equals more voluntary and spontaneous training. This part is realized through the fun factors InterpFighters provides, including Pets, Guild and House. Focus group interviews have been conducted to collect students' feedback about this platform. Three themes are found from the thematic analysis of the interview findings, including usefulness of the platform, practical reward system, and the addition of narrative and competitiveness. This study holds important pedagogical implications and can help learners to embark on the road of autonomous learning.*

## **1. Introduction**

Although the global pandemic has changed the demand for languages services from region to region as well as the working modes of interpreters, in general the interpreting job



market is steady. For instance, Hong Kong saw an increasing recruitment of official language officers over the past five years whose duties include both translation and interpreting. According to the US Bureau of Labour Statistics, the growth rate of employment of interpreters and translators will be 20 percent from 2019-2029, much higher than the average of all occupations (Bureau of Labour Statistics 2021). The Migration Advisory Committee has recommended the inclusion of interpreters in the Shortage Occupation List 2020 and indicated that the need of court interpreters was increasing by seven percent every year in UK. (Migration Advisory Committee 2020).

In the meantime, the translation and interpreting programmes worldwide have been growing by leaps and bounds in both numbers and postgraduate levels (B. Wang and Mu 2009; Orlando 2016). This growth has invited increasing research interests in this regard (e.g. Ehrlich and Napier 2015; Orlando 2016; Yan et al. 2018). Recently, more researchers started to scrutinize traditional interpreter training methods (e.g. Šveda 2021; Šveda and Djovčoš 2021; Tolosa-Igualada and Echeverri 2019). For instance, Šveda and Djovčoš (2021) stated that the COVID-19 pandemic had influenced interpreting practice and training. Indeed, this pandemic has presented challenges to interpreter training which is now either conducted in an on-line mode or in a hybrid mode. On another note, how to motivate interpreting students and the relationship between motivation and interpreting performance have always been a perplexing issue to researchers (e.g., Cai and Dong 2015; Ho 2020; Hodáková 2021; Liu and Yu 2019; Melicherčíková and Dove 2021; Mraček and Vavroušová 2021; Wu 2016). Wu (2016) believed that motivation was “an important factor influencing students’ beliefs, goals and involvement in the process of interpreter training” (15). Through a survey with 160 interpreting trainees, Wu (2016) highlighted the dynamic, situated and multifaceted nature of motivation and pointed out the existence of a moderate level of demotivation in learners. Wu (2016) also reported that interpreting trainees lacked self-directed learning strategies and that the Ideal Self, namely “the representation of the attributes that one desires to possess as an interpreter” played an important role in sustaining learners’ motivation (16). This study put forward three demotivating factors, teaching methods, perceived supports and perceived competence. Liu and Yu (2019) identified four factors related to the (de)motivation in a case study, significant others (teachers, supervisor, and peers), self-perceived ability (in the aspect of language proficiency and disciplinary knowledge), coursework, and curriculum (related to the curriculum design and teaching method). On the one hand, the subject believed that coursework had a positive effect on his study; on the other hand, he expressed concerns about the heavy workload leading to his demotivation. Melicherčíková and Dove (2021) found a statistically significant correlation between motivation and interpreting performance (both consecutive interpreting and simultaneous interpreting). As can be seen from the above literature, it has become an even more demanding task at this moment to keep learners motivated and avoid demotivation.

## 2. Technology and interpreter training

The information and communication technology (ICT) has been playing an increasingly important part in many aspects of our lives with the activity of interpreting being no exception. The presence of technology in interpreting is not something new (e.g. Annalisa 2015; Class and Moser-Mercer 2013; Fantinuoli 2018a; Hansen and Shlesinger 2007; Sandrelli and Jerez 2007). Fantinuoli (2018a) believed that the past century witnessed “two major technological breakthroughs” in interpreting with the first being the use of wired system and the second being the Internet (2). According to Fantinuoli (2018a), the third breakthrough was knocking at our door, which was put forward as “the technological turn in interpreting” (3). One evidence of this technological turn is that the importance of technology is highlighted in many competence models or requirements for interpreters and translators including the Competence Framework 2017 of European Masters in Translation and The National Accreditation Authority for Translators and Interpreters’ (NATTI) certification system. In line with this trend, the application of technology in interpreter training received increasing attention (e.g., Costa et al. 2014; Ehrlich and Napier 2015; Fantinuoli 2018a, 2018b; Şahin 2013; Sandrelli 2005; Sandrelli and Jerez 2007; Xu and Deng 2015).

Research in this field can be roughly categorized into the following two main strands. The first category provided an overview of the history and development in the past decades (e.g., Sandrelli and Jerez 2007; Xu and Deng 2015). For instance, Sandrelli and Jerez (2007) summarized the history of Computer Assisted Interpreter Training (CAIT) under the background and influence of Computer Assisted Language Learning and the three major approaches. The first approach was the integration of speech repositories into interpreter training. Sandrelli and Jerez (2007) further introduced the *IRIS database* at the University of Trieste, which was “no longer being developed or enhanced” (278), *Marius* database from the University of Granada, the European Union Speech Repository and the incorporation of authoring tools in interpreter training from *Interpr-IT*, the first of its kind in CAIT, and *Interpretations* for training beginning learners in simultaneous interpreting to *Black Box 3.0*, “the first commercially available authoring program developed specifically for interpreter training” (288). Sandrelli and Jerez (2007) also pointed out the direction for future development, the application of Virtual Learning Environments (VLEs). Xu and Deng (2005), on the other hand, provided an overview of studies on ICT and interpreter training in China and analysed the pedagogical aspects such as training models and principles, training materials, training strategies and assessment as well as ICT tools. Xu and Deng (2005) believed that the development of technology brought about “dramatic changes and impressive progress” to interpreter training scenario in China over the past decade (230). Despite this, this study identified several issues for future research including less macro-perspective, more empirical studies and increasing focus on the development of professional technical tools.

The second line of research pertains to the application of technology in interpreter train-

ing (e.g., Braun et al. 2013; Carsten et al. 2020; Sandrelli 2005; Şahin 2013). This category focuses on the introduction and incorporation of technological tools in interpreting classrooms. For instance, Sandrelli (2005) elaborated on the development of Black Box, an authoring system, its functions and features as well as directions for future improvement. Sandrelli (2005) also discussed briefly the real challenge of developing a web-based VLE for interpreter training. Şahin (2013) introduced the integration of *Second life* (SL), a popular on-line virtual world, into interpreter training. Şahin (2013) believed that *Second life* could fulfil many pedagogical purposes through avatars, the virtual representations of players (i.e. teachers and students). Specifically, SL enabled learners to learn different types of interpreting such as consecutive interpreting, liaison interpreting and simultaneous interpreting in this quasi-authentic environment; teachers and students could also communicate with each other on a group or an individual basis. Şahin (2013) further illustrated the application of SL with a sample lecture. Braun et al. (2013) elaborated on the application of virtual environments into interpreter training with the IVY (Interpreting in Virtual Reality) project in Europe. The IVY project made use of SL and audio/video corpora to create environment and materials respectively for interpreter training with a focus on business and community interpreting. Authentic materials were used from the ELISA and BACKBONE corpora with new corpora created for additional languages. This project incorporated monologue and dialogue materials as well as multilingual components. There were three modes in IVY environment, namely “interpreting practice” mode for students, “exploration” mode for clients of interpreters and “learning activity” mode for both students and clients. Braun et al. (2013) also discussed findings from the evaluation of IVT project, which identified positive feedback from the participants. Carsten, Ciobanu and Mankauskiene (2020) investigated open interpreter training resources with ORCIT (Online Resources for Conference Interpreter Training as a case study. This study evaluated the effectiveness and popularity of ORCIT with a questionnaire and Google Analytics. The results showed that ORCIT was confirmed to be useful and helpful by trainers, trainees or practitioners and was regarded as instrumental in strengthening learners’ interpreting skills. This study highlighted the importance of evaluating ICT tools and introduced a five-level Holistic Evaluation Framework for Technology-enhanced learning.

### **3. Game and interpreter training**

In interpreter training, it is not always easy to keep learners motivated in learning due to stress (Hansen and Shlesinger 2007) or frustration. With mixed teaching mode (both on-line teaching and face-to-face teaching) or solely on-line teaching due to the pandemic, it becomes even harder to increase students’ learning motivation and monitor their training hours and progress. Some researchers touched upon the impact of technology on motivation in interpreting learning. For instance, Hansen and Shlesinger (2007) pointed out that the e-learning

platforms with video-based materials for self-study contributed to improvement in students' performance as well as increased motivation and more efforts in learning. Şahin (2013) argued that interpreting trainee's motivation can be increased through the application of virtual worlds in interpreter training.

On the other hand, some researchers explored the possibility of integrating game elements or game theory into interpreting (e.g. Alyunina 2019; Wang 2017) in view of the deficiency of traditional teaching methods in motivating students. Alyunina (2019) introduced the usage of a game "Deaf Interpreter" in the consecutive interpretation classroom. This game was developed from a Russian game "Deaf Phone" for children. Students were asked to stand in a chain and took turns to interpret what he or she heard into another language (the language pair concerned was Russian and English) after written translation training and videos interpretation training related to the topic of medicine. The game "Deaf Interpreter" intended to train such skills as listening comprehension, memory, shadowing and note-taking. Wang (2017) discussed interpreting strategies from the perspective of game theory and defined the game between the interpreter and initiator/ client as "a dynamic, cooperative and variable-sum one" (104). Recent years witnessed more exploration of virtual reality in interpreter training, another manifestation of "the influence of the gaming world" (Sandrelli 2015).

Despite that the above studies opened another door in this field, further explorations were needed to make the best out of this marriage between game and interpreter training or the trend of gamification in interpreter training if put in another way. For one thing, the existing studies failed to cover a relatively comprehensive aspect of interpreter training. For example, Alyunina (2019) lent support to the usefulness of games in educational settings to increase learners' interests and participation. However, the game "Deaf Interpreter" can only train learners in limited aspects and be used as an in-class activity mostly rather than a self-study activity. For another, it is still a challenge to sustain learners' motivation, for which we can have many things to borrow from games. Games attract lots of people regardless of their age, gender and occupation and the games market keeps expanding. Wijman (2021) pointed out that the Global Games Market Report 2021 of Newzoo estimated that the year 2020 saw the highest growth in the revenue of the games market since Newzoo started this statistics in 2012. The expected number of players would reach 2.9 billion at the end of 2021 (ibid). Many studies explored functions and attractions of games. For example, Tocci (2012) described five types of game appeals, namely accomplishment, imagination, socialization, recreation and subversion. Specifically, accomplishment was defined as "appeals involving extrinsic and intrinsic rewards"; imagination referred to "appeals involving pretending and storytelling"; socialization was related to "appeals involving friendly social interaction"; recreation meant "appeals for adjusting physical, mental, or emotional state"; subversion dealt with "appeals involving breaking social or technical rules" (ibid). The integration of virtual worlds can undoubtedly be regarded a breakthrough in interpreter training. However, virtual worlds such as SL lack "typ-

ical gaming objective” (Britannica 2017), which may not be able to keep learners’ interests very long.

Under this backdrop, the current study called for gamilization of interpreter training as gamification or applying a simple game or integrating virtual worlds in training cannot satisfy the needs of holding learners’ interests, keeping them motivated and embarking on the road of autonomous learners. Therefore, the following game, *InterpFighters* is designed to fill in this gap.

#### 4. *InterpFighters*: a proposed platform

*InterpFighters* is developed as a game with incorporation of training interpreting competences. This platform aims to provide a learner-centred environment which allows learner “to choose from whom and with whom to learn, and to mix and match learning opportunities (Moser-Mercer 2008, 10), the construction of which echoes Moser-Mercer (2008)’s call for making the best use of virtual learning in interpreter training in the future.

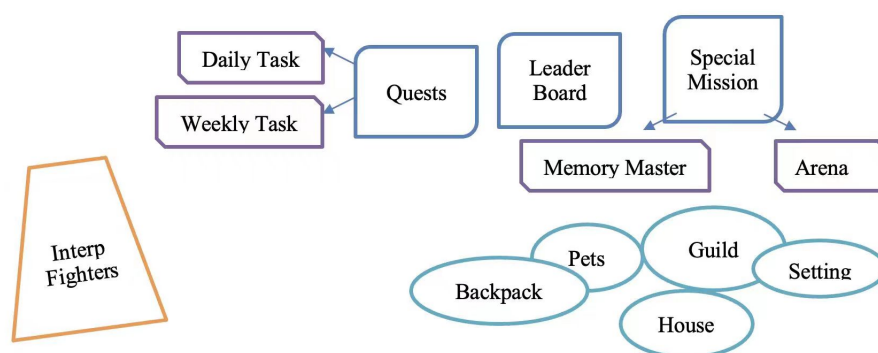


Figure 1: The interface of *InterpFighters*

The goal of interpreter training is to equip learners with sufficient knowledge, skills and competences for future interpreting tasks. There is an abundance of literature regarding this (e.g., Kalina 2000; Lee 2019; Tiselius and Hild 2017). Lee (2019) defined interpreting competence as “a set of knowledge and skills that are required for an interpreter to perform an interpreting task” and pointed out that interpreting competence and interpreter competence were often used interchangeably (84). Tiselius and Hild (2017) argued that we should differentiate competence from expertise and called for the development of both hard competences such as language knowledge, general knowledge and soft competences in translation and interpreting. According to Kalina (2000), interpreting competence was “a highly complex successful interaction of the various skills” (26), which involved text-processing strategies including comprehension strategies and text production strategies. Interpreting competence is a rela-

tively under-researched area due to its overlapping key concepts with translation competence and accordingly being included under the latter as well as a mixture of terms such as abilities, skills and expertise used by researchers arisen from the emphasis on cognitive approaches (Lee 2019).

However, recently it seems that interpreting competence has received more attention (e.g., Oraki and Tajvidi 2020; Schaeffer et al. 2020; Tiselius and Hild 2017; Wang et al. 2020). For example, Wang et al. (2020) referred to interpreting competence as “the interlingual and intercultural mediation ability of instantaneously transferring utterances from a source language into a target language, using language proficiency, related world knowledge, and interpreting strategies”. This study developed the CSE (China Standards of English)-Interpreting Competence Scales which comprised of 369 descriptors. Schaeffer et al. (2020) designed the Translation and Interpreting Competence Questionnaire (TICQ) in three languages (English, Spanish and German). The instruments were proven to be valid in measuring translation and interpreting competence and can be flexibly adapted to meet different research purposes.

In addition, some professional associations of interpreters and accreditation bodies put forward their competence requirements or skills needed for interpreters. For example, the Australian National Accreditation Authority for Translators and Interpreters (NAATI) elaborated on the competencies requirements for interpreters in its report of the Interpreter KSA (Knowledge, skills and attributes) 2016. NAATI (2016) proposed the following competency: language competency, intercultural competency, research competency, technological competency, thematic competency, transfer competency, service provision and ethical competency. The National Standard Guide for Community Interpreting by the Healthcare Interpretation Network stipulated skills and competencies separately for community interpreters, namely, interpreting competence, linguistic competence, research and technical competence, interpreting skills, language skills and interpersonal skills (HIN n.d.). International Association of Conference Interpreters (AIIC) also listed required skills for conference interpreters in its website, including linguistic knowledge, cultural knowledge, listening ability, concentration, memory, general knowledge, pressure management, professional ethics, etc.

Based on the above literature and in view of the significance of interpreting competence, *InterpFighters* is possessed of two features, competence development and motivation sustaining. The competence development comprises of training of the following aspects, knowledge, memory, interpreting skills, interpersonal skills as well as team spirit. The second feature relates to investment facilitators, namely, the fun elements or entertainment factors, the purpose of which is to keep learners motivated and interested in using this platform.

### **A guide to *InterpFighters***

When users/students or teachers first log in this platform, they can create an account and an avatar. Students are free to select the gender and name of the avatar. They need to report to their teacher about their avatar once the account has been set up. Make-up and cos-

tume functions will be available. Students can select hair styles, skin colors and costumes for their avatars through the points they gain after they complete certain tasks. A newb (an inexperienced gamer) guide will be activated for each new user, in which there are detailed descriptions about each task or activity. This guide can also be opened later if need be. For each quest (objective-based activity in a game), there will be a task brief. The main interface of *InterpFighters* is shown in figure 1.

As can be seen above in the main interface of *InterpFighters*, there are three big icons on the top right corner, namely, *Quests*, *Leader Board* and *Special Mission* as well as five small icons on the bottom right corner, *Guild*, *Pets*, *Backpack*, *House* and *Setting*. The *Quests* consist of *Daily Task* and *Weekly Task*. *Daily Task* is in a form of multiple choice questions, the purpose of which is to train students in the aspect of knowledge. The questions in the *Daily Task* are related to firstly, world knowledge, referred to as encyclopedic knowledge in W. Wang et al. (2020), or cultural, historical and political knowledge (intercultural competency) in NAATI (2016), and then subject knowledge, phrased as thematic competency in NAATI (2016), as well as general knowledge about interpreting including basic concepts, theories and requirements. The importance of knowledge cannot be emphasized too much in interpreting learning. Only equipped with knowledge about the current affairs, the related subject and the interpreting profession can interpreters deliver a satisfactory job. Kalina (2000, 6) highlighted the importance of “contextual, situational, world knowledge” as well as cultural knowledge. The United Nations stipulates that interpreters should have world affairs knowledge, be familiar with different subjects and the technical terms involved as well as cultural awareness. However, it is usually the case that interpreting trainers cannot afford to spend too much time in helping trainees to expand the knowledge mentioned above. Therefore, *Daily Task* can serve as a useful mechanism for learners to accumulate knowledge and self-check their progress.

The *Weekly Task* is as the name indicates to be completed on a weekly basis. There are two Instance Zones or virtual scenarios: Consecutive Interpreting Zone (CIZ) and Simultaneous Interpreting Zone (SIZ). These two zones use training materials from an audio and video corpora which will be constructed and updated regularly by trainers for students to practise with. The training materials will cover a wide range of fields, including business, law, public health, history, culture, politics, tourism, food and environment. Both zones offer a virtual environment that simulates the real interpreting scenarios. In SIZ, the avatar will enter a booth and can interact with the recorder, microphone and headset and record his/her performance. He or she can also see the attendees and speaker of the event. In CIZ, the avatar can choose to sit next to or behind his/her client. And across the table are another speaker and his/her interpreter, which is a typical CI setting.

The *Leader Board* displays winners’ list from top 1 to 10 in the *Daily Task*, *Weekly Task* and *Special Mission*. The players on the *Leader Board* will be rewarded with points or in-game currency. The *Leader Board* serves as a vehicle for the cultivation of a sense of com-

petitiveness, and eagerness in learning and confidence. This echoes the attributes “desire to excel”, “being confident” and “willingness to learn” in NAATI (2016). According to NAATI (2016, 19), “desire to excel” refers to being “motivated to improve and perform work tasks consistently to a high level”; “willingness to learn” means “actively and continuously improving knowledge and skills”; “confidence” is “self-assurance on one’s personal judgement and ability”.

The *Special Mission* consists of two parts, *Memory Master* and *Arena*. *Memory Master* refers to a set of memory training games which will be updated on a weekly basis. These games may include memory matching games or sequence memory games, the forms of which will vary from time to time. Many scholars have reserved a place for memory in their interpreting competence model. For example, W. Wang et al. (2020) categorized retrieving memory into cognitive abilities for their construct of interpreting competence. Memory skills were listed under transfer competency of the interpreter KSA in NAATI (2016). Since memories play an important part in the successful completion of interpreting tasks, memory training is always included in interpreting classrooms. For example, Mouallem (2015) held that memory training should start from the beginning and allocated a medium amount of time to memory exercises in the business interpreting module. Kornakov (2000) highlighted the importance of enhancing learners’ self-confidence in their memory as most students reported to have short-term and long-term memory issues. Therefore, *Memory Master* was designed in a similar vein to help students to have some practices in this regard. The winner of the *Memory Master* will be announced through the system message daily and be rewarded with points; individual messages will be sent to winners as well. *Arena* is made up of two types, namely individual challenge for single players and group challenge for a team of three to five players. Different from *Weekly Task*, challenges in *Arena* will either centre on specific topics or be related to current affairs, which will take the form of sight translation, consecutive interpretation or simultaneous interpretation. Teachers will give grades to individual or group performance after the *Arena* time is over through the platform. Winners will get in-game currency which will double the points they can gain from *Daily Task* and *Weekly Task*.

The following are illustrations of the other functions. Any player can establish a *Guild* with a certain amount of in-game currency and others can join it with the leader’s permission. Members of the same *Guild* can contribute to the fund of the Guild after successful completion of tasks and missions, which will help the Guild to upgrade its level. Every Monday members will receive a kind of “bonus” from the *Guild*. The higher the level of the *Guild* is, the better the bonus will be. The bonus may be in-game currency, costumes for the avatar, food for the pets, decoration tools for the house or other rewards. Players can also communicate with each other in the Guild or outside the *Guild* through the chat function (text message or audio message). The purpose of the *Guild* is to enhance interpersonal skills and team spirit. Interpersonal skills play an important part for the successful completion of an interpreting task as interpreters may



need to communicate with their clients and team members as well as leaders of the organization or company or institution that they are serving in. Interpersonal competence is regarded as one of the sub-competences of interpreting competence (Lee 2019). Interpersonal skill, communication skill and interpreting business skill correspond to service provision competency in NAATI (2016). Therefore, the cultivation of interpersonal skills is also important for interpreting trainees. Many studies have established a positive link between video games and communication skills (e.g. Khanmurzina et al. 2020; Barr 2017). Khanmurzina et al. (2020) found that through playing games, student gamers improved their communication skills and learned to work as a team besides learning English. Barr (2017) identified the good impact of playing commercial games on “communication ability, adaptability and resourcefulness in adult learners” (96). On the above ground, the *Guild* function may be helpful to interpreting trainees in this regard.

The *Backpack* function is for the players to store his belongings in this virtual world, such as points, in-game currency, decoration tools, pets, other rewards etc. *Pets* and *House* serve as motivation facilitator. With more points gained, players can upgrade the level of their avatar. When the Avatar reaches certain levels, players can adopt a pet. The pet will grow up together with its owner. The more points its owner gains from participation in the tasks and missions, the faster the pet will grow. The higher the level of the pets is, the more special rewards they can bring to their owners. *House* is a virtual home of the avatar. The avatar can choose the style of his or her house and purchase materials to decorate it with in-game currency. The design of these two functions is to attract learners to spend more time in this platform and motivate them to finish the CI and SI tasks.

## 5. Qualitative findings

To explore the feasibility and enhance the effectiveness of this platform, two focus group interviews were conducted. The participants were students from a translation program of a tertiary institution in Hong Kong. They were in their third year of study and had completed two to three interpreting modules. Each interview consisted of three participants and lasted for around 50 minutes. The interviews were conducted in Chinese, the participants’ mother tongue. The interview data were transcribed and translated into English by the researcher. Thematic analyses were performed based on the approach described in Yan and Horwitz (2008). Three themes were identified, usefulness of the platform, practical reward system, and the addition of narrative and competitiveness.

### 5.1 Usefulness of the platform

The majority of students believed that this platform was interesting and would be helpful to interpreting training. Some of them also expressed concerns about the influence of people’s

different tastes for games on the usage of this platform. In general, participants were supportive of this platform.

*I think this platform includes lots of functions and activities. As far as I am concerned, it is definitely good to use this platform to learn interpreting even though I am not a fan of games. This enables us to learn and entertain at the same time.*

*I think the platform is very attractive. To be specific, I find the House function very appealing. I like decorating this virtual home, such as buying sofa and cloth. I am not sure about those who like the type of “fighting” games. Maybe they will regard this too feminine. But I believe this is a very good platform. I think we will have more interest in this platform than in the current on-line teaching method. This is very useful. On-line teaching makes it difficult for teachers to monitor students’ progress especially when students do not need to turn on the camera. However, this platform enables us to practice on our own and receive feedback from teachers as well. I believe this platform to be a very good supplementary to classroom teaching. The reason is that we can choose to practice based on our learning needs. Therefore, we can have more training after class if we want to do so.*

## 5.2 Practical reward system

The participants proposed for a practical reward system as they believed it to be instrumental to their learning motivation in addition to the points they could get from finishing activities. The rewards they put forward had a wide range in terms of value. Based on this finding, a mixture of virtual and real rewards can be considered in the future implementation of this platform. If some practical rewards could be offered to those with top ranking in the *Leader Board*, it can motivate two types of students to continue to use this platform, namely those who have initiatives in using it from the beginning and those who are not interested in the platform at first. It will be great to have such rewards as computers, tissues or other practical stuff.

*If we can use the points gained from the platform to exchange for something real, that will be very nice. It will be a great motivation for us to use the platform for longer time. It can attract those who are not very keen on this platform and sustain their interests. I think seasonal products from Apple will be a good choice including Apple watch, iPhone and AirPods. It will depend on your budget of course.*

*I think stationery will do including beautiful notebooks. Or a coupon of 20 HKD is also fine. A coupon of 50 HKD will be better.*

*I think the attractive reward includes food and coupons from supermarkets such as PARKn-SHOP.*

*I would consider gift cards or premium membership for Netflix, YouTube and Spotify a nice choice. Those who really like playing games tend to get more enjoyment from audio and video products.*

### 5.3 The addition of narrative and competitiveness

Some students mentioned two additional features to the platform including adding a storyline and enhancing competitiveness. They believed that these two features would be beneficial to their motivation and interests. This finding provided food for thought in the enhanced version of this platform.

*I think a narrative can be added to this platform. Sometimes what games attract me most is not how to play but the storyline. To put it another way, players must finish certain tasks before the next chapter is unveiled. It seems that translation students like stories very much. A narrative can attract them to finish more tasks in the platform.*

*If this platform can also be used in other universities, that will be great. The reason is that our classmates may know each other very well. I will not get the feeling of “I am so amazing” even I rank 1st. However, if we compete with students from other institutions such as the Chinese University of Hong Kong and The Education University of Hong Kong, we will be more motivated to use this platform. We may think that we can “fight against” good students from other universities, which is so very nice.*

## 6. Pedagogical Implications

This study puts forward a platform, *InterpFighters*, which is, to the best of our knowledge, the first endeavor to put interpreting training in a gaming environment. The platform is a hybrid of practice and pleasure as well as exercises and entertainment. Pedagogically speaking, this study has the following implications. To start with, the roles of trainees and trainers are transformed. Learners are pushed to the center of the stage. They are motivated to take control of the training activities which are traditionally in charge by teachers or trainers in the classrooms. This platform afforded learners the flexibility of selecting training schedule and the type of exercises they want to have more practice with. Teachers, on the other hand, retreat to the backstage, play a half invisible role, and serve as administrator of the platform. If teachers would like to mingle with students, they can join the game as players as well. Nevertheless, teachers remain to be graders providing feedback to students through the platform. As administrators, teachers will monitor the platform and update materials database on a regular basis with an invisible hand.

Secondly, the introduction of this platform will contribute to the development of interpreting competence of students. This platform centers on training of the following skills, memory retention skills, subject-related knowledge, world knowledge, and interpersonal skills as well as interpreting skills. Different from the traditional training methods, this platform instills game dynamics into training helping to boost competitiveness and confidence among students. The reward system of this game and regular feedback from trainers will be very helpful in sustaining learners' motivation or their willingness to learn. Only intrinsically motivated can

learners maximize their learning outcomes. Interactions among students and teachers will also be enhanced.

Thirdly, the use of this platform will be beneficial to enhance both interpreting trainees' and trainers' technology skills, which constitutes an important part of interpreter training. With an aim to have a good ranking in the *Leader Board* and enjoy the rewards afterwards, learners are encouraged to make full use of their research skills to search for terminology and relevant information about a certain subject and build their knowledge pool.

## 7. Conclusions

This study puts forward a concept of gamification and advocates for its application in interpreter training. The gamification of interpreter training can take many forms. One way is to combine training tasks in one game similar to *InterpFighters*. Another possible way is to design a series of games for different training purposes, which warrant further research.

As discussed above, this game *InterpFighters* can be regarded as a novel way to develop and enhance interpreting competence and skills of students, which also paved way for autonomous learners. Self-training or self-study is an important part to interpreting trainees. Hansen and Shlesinger (2007) pointed out video-based self-study helped to improve students' motivation and increase their efforts in learning which resulted in better performance. The qualitative findings lend support to the feasibility and usefulness of this platform. In addition, these findings also point out motivational factors from the perspectives of students, which will be helpful for the refinement of this platform.

However, despite the advantages *InterpFighters* may bring to interpreting teaching, it cannot cover each aspect of interpreting competence. We still need traditional teaching methods for certain parts of interpreter training. For instance, ethical competence or ethical knowledge, a very important part for future interpreters, has its rightful place in interpreter training (Mouallem 2015). Professional ethics may be better taught and explained in class by teachers. With *InterpFighters* as an example, we may also explore possibilities of designing an e-platform or a game to teach professional ethics specifically in the future. More empirical data may be collected later to further refine and revise *InterpFighters*.

## References

- Alyunina, Yulia. 2019. "Deaf Interpreter-Scenario for Teaching Russian-English Consecutive Interpretation in a Game Way." Paper presented at EDULEARN19 Conference, Palma, July 1-3.
- Annalisa, Sandrelli. 2015. "Becoming an Interpreter: The Role of Computer Technology."

- MonTI. Monografías de Traducción e Interpretación 2*: 111-138.
- Barr, Matthew. 2017. "Video Games can Develop Graduate Skills in Higher Education Students: A Randomised Trial." *Computers & Education* 113: 86-97.
- Braun, Sabine, Catherine Slater, Robert Gittins, Panagiotis Ritsos and Jonathan C. Roberts. 2013. "Interpreting in Virtual Reality: Designing and Developing a 3D Virtual World to Prepare Interpreters and their Clients for Professional Practice." In *New Prospects and Perspectives for Educating Language Mediators*, edited by Don Kiraly, Silvia Hansen-Schirra, and Karin Maksymyski, 93-120. Tuebingen: Gunter Narr.
- Britannica, The Editors of Encyclopaedia. 2017. "Second Life". Encyclopedia Britannica, September 6. Accessed July 16 2021. <https://www.britannica.com/topic/Second-Life>
- Bureau of Labor Statistics. 2021. "Job Outlook". Accessed June 18 2021. <https://www.bls.gov/ooh/media-and-communication/interpreters-and-translators.htm>.
- Cai, Rendong, and Yanping Dong. 2015. "Interpreter Training and Students of Interpreting in China." *Journal of Translation Studies* 16(4): 167-191.
- Carsten, Svetlana, Dragoş Ciobanu, and Dalia Mankauskiene. 2020. "The Challenge of Evaluating Open Interpreter Training Resources: Case Study of ORCIT." *The Interpreter and Translator Trainer* 1-16.
- Class, Barbara, and Barbara, Moser-Mercer. 2013. "Training Conference Interpreter Trainers with Technology—A Virtual Reality." *Quality in Interpreting: Widening the Scope* 1:293-313.
- Costa, Hernani, Gloria Corpas Pastor, and Isabel Durán Muñoz. 2014. "Technology-Assisted Interpreting." *MultiLingual* 143(25): 27-32.
- Ehrlich, Suzanne, and Jemina Napier. 2015. *Interpreter Education in the Digital Age: Innovation, Access, and Change*. Washington: Gallaudet University Press.
- Fantinuoli, Claudio. 2018a. "Interpreting and Technology: The Upcoming Technological Turn." In *Interpreting and Technology*, edited by Claudio Fantinuoli, 1-12. Berlin: Language Science Press.
- Fantinuoli, Claudio, ed. 2018b. *Interpreting and Technology*. Berlin: Language Science Press.
- Hansen, Inge Gorm, and Miriam Shlesinger. 2007. "The Silver Lining: Technology and Self-study in the Interpreting Classroom." *Interpreting* 9(1): 95-118.
- HIN. n.d. 2020. "National Standard Guide for Community Interpreting Services." Accessed July 21, 2020. <https://www.imiaweb.org/uploads/pages/672..pdf>
- Ho, Chen-En. 2020. "Tapping into Interpreting Students' Motivation to Engage in Targeted Translation Practice." *Current Trends in Translation Teaching and Learning E* 7: 39-76.
- Hodáková, Soňa. 2021. "Motivational Structure and the Interpreter's Personality." In *Changing Paradigms and Approaches in Interpreter Training*, edited by Pavol Šveda, 185-208. New York: Routledge.
- Kalina, Sylvia. 2000. "Interpretation Competences as a Basis and a Goal for Teaching." *The*

*Interpreters' Newsletter* 10: 3-32.

- Khanmurzina, Rimma R., Elena I., Cherdymova, Tatyana Yu, Guryanova, Rita A., Toriia, Evgenia M, Sukhodolova, and Larisa I., Tararina. 2020. "Computer Games Influence on Everyday Social Practices of Students-Gamers." *Contemporary Educational Technology* 11(1): 11-19.
- Kornakov, Peter. 2000. "Five Principles and Five Skills for Training Interpreters." *Meta: journal des traducteurs/Meta: Translators' Journal*, 45(2): 241-248.
- Lee, Jieun. 2019. "Competence, Interpreting". In *Routledge Encyclopedia of Translation Studies*, 3rd edition, edited by Mona Baker, and Gabriela Saldanha, 84-89. London: Routledge.
- Liu, Chunhong, and Chengyuan Yu. 2019. "Understanding Students' Motivation in Translation Learning: A Case Study from the Self-Concept Perspective." *Asian-Pacific Journal of Second and Foreign Language Education* 4(1): 1-19.
- Melicherčíková, Miroslava, and Michael E., Dove. 2021. "Interpreter Trainees' Performance—Motivation, Quality and Assessment (An Empirical Study)." In *Changing Paradigms and Approaches in Interpreter Training*, edited by Pavol Šveda, 209-228. New York: Routledge.
- Migration Advisory Committee. 2020. "Review of the Shortage Occupation List: 2020." Accessed July 21 2021. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/927352/SOL\\_2020\\_report\\_final.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/927352/SOL_2020_report_final.pdf)
- Mraček, David, and Petra Mračková, Vavroušová. 2021. "Self-Reflection Tools in Interpreter Training: A Case Study Involving Learners' Diaries." In *Changing Paradigms and Approaches in Interpreter Training*, edited by Pavol Šveda, 229-247. New York: Routledge.
- Moser-Mercer, Barbara. 2008. "Skill Acquisition in Interpreting: A Human Performance Perspective." *The Interpreter and Translator Trainer* 2(1): 1-28.
- Mouallem, Summer. 2015. "Interpreter Training: Language, Culture or More? Course Design: The Experience of The University of Central Lancashire, Preston, UK." *European Scientific Journal, ESJ* 11: 185-199.
- NAATI. 2016. "NAATI Interpreter Certification: Knowledge, Skills and Attributes." Accessed July 17 2021. [https://www.naati.com.au/media/1262/interpreter-ksa-paper-final\\_290216\\_2.pdf](https://www.naati.com.au/media/1262/interpreter-ksa-paper-final_290216_2.pdf).
- Oraki, Abouzar, and Gholam-Reza Tajvidi. 2020. "Training Translators and Interpreters: The Need for a Competence-Based Approach in Designing University Curricula." *Iranian Journal of English for Academic Purposes* 9(2): 42-56.
- Orlando, Marc. 2016. *Training 21st Century Translators and Interpreters: At the Crossroads of Practice, Research and Pedagogy*. Vol. 21 of *Transkulturalität – Translation – Transfer*. Berlin: Frank & Timme GmbH.
- Şahin, Mehmet. 2013. "Virtual Worlds in Interpreter Training." *The Interpreter and Translator Trainer* 7(1): 91-106.

- Sandrelli, Annalisa. 2005. "Designing CAIT (Computer-Assisted Interpreter Training) Tools: Black Box." Paper presented at MuTra 2005—Challenges of Multidimensional Translation, Saarbrücken, May 2-6.
- Sandrelli, Annalisa, and Jesús de Manuel, Jerez. 2007. "The Impact of Information and Communication Technology on Interpreter Training: State-of-the-art and Future Prospects." *The Interpreter and Translator Trainer* 1(2): 269-303.
- Schaeffer, Moritz, David Huepe, Silvia Hansen-Schirra, Sascha Hofmann, Edison Muñoz, Boris Kogan, Eduar Herrear, Agustín Ibáñez and Adolfo M. García. 2020. "The Translation and Interpreting Competence Questionnaire: An Online Tool for Research on Translators and Interpreters." *Perspectives* 28(1): 90-108.
- Šveda, Pavol, ed. 2021. *Changing Paradigms and Approaches in Interpreter Training: Perspectives from Central Europe*. New York: Routledge.
- Šveda, Pavol, and Martin Djovčoš. 2021. "Introducing Interpreter Training in Central Europe." In *Changing Paradigms and Approaches in Interpreter Training*, edited by Pavol Šveda, 7-21. New York: Routledge.
- Tiseliu, Elisabet, and Adelina Hild. 2017. "Expertise and Competence in Translation and Interpreting." In *The Handbook of Translation and Cognition*, edited by John W. Schwieter, and Aline Ferreira, 425-444. Hoboken: John Wiley & Sons.
- Tocci, Jason. 2012. "Five ways games appeal to players" Gamasutra. April 19. Accessed August 12 2021. [https://www.gamasutra.com/view/feature/168807/five\\_ways\\_games\\_appear\\_to\\_players.php](https://www.gamasutra.com/view/feature/168807/five_ways_games_appear_to_players.php)
- Tolosa-Igualada, Miguel, and Álvaro Echeverri. 2019. "Because Something should Change. Translator and Interpreter Training: Past, Present and Future." *Monti* 11: 29-46.
- Wang, Binghua, and Lei Mu. 2009. "Interpreter Training and Research in Mainland China: Recent Developments." *Interpreting* 11(2): 267-283.
- Wang, Shuai. 2017. "On Interpreting Strategies with a Perspective of Game Theory." Paper presented at 3rd International Conference on Social, Education and Management Engineering (SEME 2017), Shanghai, November 26-27.
- Wang, Weiwei, Yi Xu, Binhua Wang, and Lei Mu. 2020. "Developing Interpreting Competence Scales in China." *Frontiers in Psychology* 11: 481.
- Wijman, Tom. 2021. "Global games market to generate \$175.8 billion in 2021; despite a slight Decline, the market is on track to surpass \$200 billion in 2023." Newzoo, May 6. Accessed July 2 2021. <https://newzoo.com/insights/articles/global-games-market-to-generate-175-8-billion-in-2021-despite-a-slight-decline-the-market-is-on-track-to-surpass-200-billion-in-2023/>
- Wu, Zhiwei. 2016. "Towards Understanding Interpreter Trainees' (De) Motivation: An Exploratory Study." *The International Journal for Translation & Interpreting* 8(2): 13-25.
- Xu, Mingwu, and Juntao Deng. 2015. "Integrating Information and Communication Technol-

- ogy into Interpreter Training in China: State-of-the-Art and Prospect.” *Chinese Journal of Applied Linguistics* 38(2): 219-233.
- Yan, Jackie Xiu, and Horwitz, E. K. 2008. “Learners’ Perceptions of How Anxiety Interacts with Personal and Instructional Factors to Influence their Achievement in English: A Qualitative Analysis of EFL Learners in China.” *Language Learning* 58(1): 151-183.
- Yan, Xiu, Jun Pan, and Honghua, Wang. 2018. *Research on Translator and Interpreter Training - A Collective Volume of Bibliometric Reviews and Empirical Studies on Learners*. Singapore: Springer.





# “中國故事” 英中翻譯之 “重新聚焦” 模式： 以彼得·海斯勒中國題材敘事新聞作品之簡體中譯本為例

吳音然

**Address:** Department of Translation, The Chinese University of Hong Kong, Hong Kong, China

**E-mail:** yinranwu@cuhk.edu.hk

**Correspondence:** Yinran Wu

**Citation:** Wu, Yinran. 2022. “‘Re-focalization’ in the E-C Translation of ‘China Stories’: A Case Study of the Simplified Chinese Translations of Peter Hessler’s China-Themed Narrative Journalistic Works.” *Translation Quarterly* 105: 65-76.

## ***Abstract***

‘Re-focalization’ in the E-C Translation of ‘China Stories’: A Case Study of the Simplified Chinese Translations of Peter Hessler’s China-Themed Narrative Journalistic Works (by Yinran Wu)

*Contemporary China-themed narrative journalistic writings depict the lives of ordinary Chinese people against the backdrop of China’s social transformation through the perspectives of western writers. The Chinese translation of these China stories highlights an underexplored translation phenomenon. Drawing upon the simplified Chinese translations of Peter Hessler’s two book-length narrative journalistic works (i.e., River Town and Country Driving), this paper explores how the Chinese characters and events are re-developed from a different viewing angle in the target narratives than the source narratives, with the goal of revealing the underlying translation mechanism of this phenomenon. Using the narratological term ‘focalization’ as a heuristic notion, the paper traces changes in linguistic markers between the source and target texts, unveiling a ‘re-focalization’ process where the focalizer’s status is altered from ‘outsider’ to ‘insider’. Under the impact of the Chinese translator, the translated narratives, re-focalized through an ‘insider’, are found to echo the Chinese people’s shared understandings of their community and reflect their potential attachment to the Chinese characters.*

## 一、引言

本文所言“中國故事”是以中國為題材書寫的當代英文敘事新聞作品。敘事新聞(narrative journalism)是新聞體裁的一個新興分支，屬非虛構創作範疇，通常借用文學敘述手法，透過敘述者或人物的主觀視角，來刻畫真實事件，又稱為文學新聞(literary journalism)(Hartsock 2000, 6; Van Krieken and Sanders 2019, 12)。本文所關注的敘事新聞作品，主要於千禧年前後問世，多以 20 世紀下半葉中國改革變遷為背景，刻畫尋常百姓的命運起伏。其中，美國作家彼得·海斯勒(中文名：何偉)所著“中國三部曲”是其中的代表之作。2010 年後，此類體裁作品陸續譯成簡體、繁體中文，譯本一經推出，反響巨大，引起中文讀者對中國題材西方作品空前濃厚的閱讀興趣。

這類作品的中譯隱含一個尚待探索的翻譯現象。原作者在中國旅居多年，深諳中國百姓的生活細節，其用英文書寫的中國故事，再由熟悉西方文思的中文譯者譯成中文。作者與譯者對於彼此文化的理解與可能產生的共鳴，似乎讓這類作品的翻譯更多涉及如何將英文寫就的中國人事還原為中文表達的問題。然而，作者對中國社會生活帶有人文關懷的細緻記敘，仍透露出作者自身特有的觀察視角，也難免帶有西方價值理念的痕跡。而中文譯者，則與源文本故事所述人物、事件身處同一社會時空，譯者對同胞及其所處生活環境的既有認識，可能潛在影響目標文本敘述“中國故事”的視角。

在為數不多關注海斯勒作品中譯的研究中，個別研究側重評價作品中的文化元素回復本來面貌的還原問題(王瓊 2016, 15-16)，另或與繁體中文譯本比較，著重分析政治敏感話題在簡體譯本中的過濾或改寫(Zhang 2018)，而尚未揭示此類作品中譯過程中由視角轉變產生的複雜變化。

海斯勒作品的敘述對象多為中國內地的人物與事件，為探究與“中國故事”存在親近關聯的譯者及其影響下的視角變化，本文故以簡體譯本為主要分析對象。在圍繞相關文本分析結果的基礎上，本研究提出“重新聚焦”模式，以此闡釋這類體裁作品中譯過程中伴隨視角轉變的重構問題。

## 二、西方作者視角下的中國故事

英文敘事新聞作品源自美國 19 世紀末(Connery 1990, 4; Hartsock 2000, 14)，而中國題材英文敘事新聞的早期創作，則可追溯至 20 世紀 30、40 年代，以埃德加·斯諾(Edgar Snow)等美國記者的作品為代表。2010 年後，譯成簡體中文的中國題材英文敘事新聞書寫(見表 1)，多關注中國內地特定地域背景下的普通人，或打工者等社會群體，並以此為著眼點，來解讀改革變遷時代的中國。有些作品涉及講述草根民眾或社會邊緣群體與社會現實抗爭的故事，僅在台灣出版繁體中文譯本，如海斯勒“中國三部曲”第二部《甲骨文》(*Oracle Bones*)、伊恩·詹森(Ian Johnson)所著《野草》(*Wild Grass*)。

中國題材英文敘事新聞採用“融入式報導”(immersion reporting)模式(Connery

表 1: 中國題材當代英文敘事新聞簡體中文譯著

作者	源文本	目標文本	譯者
Peter Hessler	<i>River Town: Two Years on the Yangtze (2001)</i>	《江城》(2012)	李雪順
	<i>Country Driving: A Journey from Farm to Factory (2010)</i>	《尋路中國：從鄉村到工廠的自駕之旅》(2011)	
	<i>Strange Stones: Dispatches from East and West (2013)</i>	《奇石：來自東西方的報導》(2014)	
Leslie T. Chang	<i>Factory Girls: From Village to City in a Changing China (2008)</i>	《打工女孩：從鄉村到城市的變動中國》(2013)	張坤/吳怡瑤
Michael Meyer	<i>The Last Days of Old Beijing: Life in the Vanishing Backstreets of a City Transformed (2009)</i>	《再會，老北京：一座轉型的城，一段消逝的老街》(2013)	何雨珈
	<i>In Manchuria: Journeys Across China's Northeast Frontier (2015)</i>	《東北遊記》(2017)	

1990,16; Kramer 1995, 22-23)。與依托新聞機構的中國報導有所不同，作者不再單一憑藉西方對中國的刻板印象，站在遠處審度中國，而是以“參與式觀察者”的姿態深入中國社會生活之中，近距離接觸中國百姓；在此基礎上，再透過作者於文本中的敘述代理，即敘述者“我”的視角，來講述所見所聞。在中國內地生活的十年間，海斯勒遇到了形形色色的中國百姓，這些人構成了“中國故事”的主體，提供了“深入人文中國的途徑”(Cohen 2013, 271)。

儘管以百姓生活為敘述起點，這類作品的立意如史景遷(Johnathan Spence)所說，仍在於揭示更大範疇內的“地理與歷史空間”(Spence 2006)。這般以小見大的創作思路，令作者對中國社會的解讀，仍不免反映出美國新聞界精英達成共識的“恆久價值”(Gans 1979, 42-45)，如“個人主義”、“利他民主”(altruistic democracy)等西方讀者可辨識的價值符號(Song 2012, 798; Song and Lee 2015, 4)。不少作品還透過“小鎮田園主義”(small-town pastoralism)的恆久價值(Gans 1979, 49)，來審視中國改革開放以來的社會變遷格局，關注經濟發展過程中的隱憂與矛盾，譬如城市與村莊的衝突或自然環境退化問題。海斯勒的《尋路中國》即勾勒出城市化進程中大量年輕人從西部農村出走的現象，而《江城》則透過三峽工程的建設，來審視長江的命運以及江

城居民面對巨變的心態。

總體而言，英文敘事新聞作品對中國的書寫，介入了西方作者的“主體性”（Harrers and Broersma 2014, 643），既雜糅了敘述者在個體層面上對中國百姓的觀察，又反映出敘述者透過西方價值對中國社會的解讀。兩種視角的融合，貫穿在中國題材英文敘事新聞的創作中。上述視角在簡體中譯本中究竟發生了何種變化，後文將引入敘述學“聚焦”概念以進一步闡釋。

### 三、聚焦與聚焦者

“聚焦”（focalisation）的概念由傑哈德·熱奈特（Gérard Genette）（1980, 89）提出，在敘述學中，通常指敘述時觀察人物、事件的特定視角（申丹、王麗亞 2010, 88）。這個觀察視角，與敘述者採用的“據點”（vantage point），或者說觀察位置有關。例如，呈現同樣一場比賽，賽場上的球員、場邊的攝像機以及若干年後回憶比賽的球迷，由於觀察位置不同，所採用的視角亦會不同。敘述學對上述觀察位置的討論，主要圍繞文本中的觀察實體展開。米克·巴爾（Mieke Bal）、施洛米絲·里蒙-凱南（Shlomith Rimmon-Kenan）把觀察實體一定程度擬人化，稱之為“聚焦者”（focalizer; focalizer）（Bal 2009, 149; Rimmon-Kenan 2002, 75）。當然，聚焦者並非故事中切實存在的主體，但讀者往往可以依循相關敘述，推測出聚焦者所具有的某種屬性與特徵，例如根據對球員心理的揣測及外在行為的描述，可推測出一個具有場外評論員特徵的聚焦者。鑒於此，聚焦者可謂隱含在文本之中而由文本內容及語言特徵所建構的“特殊存在”。

在敘事文本中，聚焦者可與敘述者重合，即敘述者從自身視角出發，來聚焦人物、事件。此外，聚焦者還可與故事人物重合，即敘述者透過故事人物的感知來聚焦（申丹 2006, 521）。例如，敘述“射失點球”的事件，既可通過場外記者在文本中的敘述代理，即敘述者，來觀察雙方球員和現場觀眾的外在反應，又可透過球員的感知，來聚焦失球瞬間的內在心理活動。在上述例子中，與敘述者重合的聚焦者，由於處於“射失點球”的事件之外，其觀察焦點便不同於身處事件之內，而與故事人物重合的聚焦者。這種差異，不單歸咎於時空位置的不同，更與聚焦者的身份屬性存在密切關聯，可以說，“聚焦”是帶有主觀色彩的活動（Bal 2009, 8/145）。

里蒙-凱南認為，人們往往把狹義上的“觀察”理解為視覺、聽覺等相關聯的感知活動，但“觀察”實際上包含不同層面，通常還蘊含人的思維、情感、意識形態立場等內容（Rimmon-Kenan 2002, 72-84; Toolan 2001, 60）。譬如，在聖修伯里所著《小王子》中，成年人眼中的一頂漁夫帽，在小孩眼中則成了蛇吞吃大象的場景，這個差異便源自成年人與小孩思維的差異（Saint-Exupéry 2000, 1-3）。又如，同樣一雙眼睛，人們可能調動感官系統，看到一雙藍眼睛，也可能透過內心情感，看到一雙憂鬱的眼睛，還可能採用價值評判，看到一雙渴望自由的眼睛。可以說，聚焦者對人物、事件的觀察，是一個受聚焦者“身心因素影響的過程”（psychosomatic process）（Bal 2009, 145）。在聚焦者主觀因素的影響下，一定視域範圍內的人物、事件等聚焦對象，會呈現出不同特徵，並投射出具體形象。

翻譯活動關涉時空變遷，以及社會文化環境的轉變。在原作者與譯者的影響下，源文本和目標文本所分別構建出的聚焦者，亦可能在思維、情感、意識形態立場上存在潛在差異。在圍繞海斯勒作品簡體中譯本為案例的分析中，本研究發現在“中國故事”譯成中文時，譯本中聚焦者的屬性確實出現了變化，揭示出一個“重新聚焦”的過程。

## 四、海斯勒作品中譯中的“重新聚焦”

海斯勒書寫中國的敘事新聞作品當中，有三部譯成簡體中文，即《江城》、《尋路中國》以及《奇石》。其中，《奇石》是海斯勒短片敘事新聞作品集。鑒於《奇石》中的部分文章並不關於中國或中國人，而圍繞其他文化背景下的人物創作，本文重點圍繞《江城》、《尋路中國》這兩部敘事新聞長篇作品，就中國百姓的個體及群體敘事展開探究，包括《江城》中涪陵居民的故事，以及《尋路中國》中農村移民的故事。

為考察源文本、目標文本聚焦者觀察中國人物、事件分別採用的視角，本研究以標示聚焦的“語言標記”（linguistic markers）（Rimmon-Kenan 2002, 84-86; Van Krieken and Sanders 2019, 13）作為文本分析工具。里蒙-凱南（1989, 149）認為，“聚焦本身是非語言的；然而和作品文本中其他任何東西一樣，聚焦也是靠語言表現出來的”。本文通過比較分析源文本、目標文本中的語言標記（Wu 2017, 28-29; 吳音然 2019, 67-76），以進一步探討兩個文本中的敘述者聚焦中國人物、事件時，在思維、情感以及意識形態立場上的細微變化。分析發現，中國題材當代英文敘事新聞中譯隱含一個“重新聚焦”的過程，涉及聚焦者背景知識的轉換以及聚焦者與聚焦對象情感連結的建立，此過程歸根結底指向由“局外人”聚焦者到“局內人”聚焦者的轉變。

### 4.1 聚焦者背景知識的轉換

源文本敘述者“我”是一個在中國旅居數年的外國人，與這個敘述者重合的聚焦者，對中國社會的認知在某些方面與中國百姓有所不同。當譯成中文時，目標文本聚焦者視野里呈現出的中國人物、事件，轉而折射出中國百姓對其所歸屬社群的集體認知。

聚焦者背景知識的轉換，可由《尋路中國》簡體中譯本里的下述規律予以闡釋。該作品的主角是離開農村、去往城市的“移民”（migrants）。貫穿在源文本中的“migrants”一詞，凸顯出這個群體的流動屬性，即由熟悉環境去往陌生環境的空間位置變化。在描述這個群體的行為時，源文本還頻繁採用了一組標示“離開”的語言標記，例如“migrate”、“leave the village/countryside/home”、“be gone”等。這些語言標記共同反映出源文本聚焦者主要透過空間位置變換聚焦，將這個群體視為遠離故土的“出走者”。

目標文本中，聚焦者的上述觀察重心出現了偏移。在多數情況下，“migrants”譯為“外出/外來/進城務工人員”，間或譯為“外出務工者”、“流動務工人員”或“在外做工的人”。儘管這些標記仍包含空間位置變化的成分，但更多標示出該群體在城市

就業的勞務屬性，而“人員”二字還透露出務工者體制內的群體身份。此外，譯本處理標示“離開”的語言標記時，傾向添加“打工”、“務工”等字眼，使觀察焦點進一步產生變化，可通過下例闡釋：

### 例 1

源文本：

…where the farming had always been tough and the lure of **migration** was all but irresistible. Sometimes it felt as though people had left in a rush. (p. 127)

目標文本：

這些地方一直很難種莊稼，人們對**外出務工**的誘惑似乎難以抗拒。在這些地方，人們似乎離去得匆忙。(頁 119)

上例中，源文本探討了農民離開西部農村的客觀原因。其中，“migration”標示出聚焦者透過空間視角，來聚焦農村人“遠走他鄉”的行為，暗示農村人的離開更多受外力驅使：原地地耕作條件艱難，所以被迫出走。目標文本中，“migration”譯為“外出務工”，其中，“外出”二字通常指因事去往外地，與“務工”一詞配合，轉而突出了離開的去向與目的。儘管該表述符合中國百姓對打工群體的慣有認識，但隨著語言標記變化，目標文本聚焦者關注的不再是離開本身，而是離開農村後，前往城市尋求務工機會的動機。由此一來，聚焦者鏡頭下的群體不再是“出走者”，轉而呈現出“務工者”的形象。類似的聚焦變化，還體現在下例之中：

### 例 2

源文本：

“All of the young people **leave** our village,” she said. “Nobody stays here anymore. I’m not planning to go back.” …Among uneducated **female migrants**, jobs tend to be sharply divided according to looks…(p.90).

目標文本：

“我們村所有的年輕人都**出去了**，”她說，“沒人願意呆在這個地方。我也沒有打算回去。”……對那些沒有讀過多少書的**進城務工女性**而言，幹什麼工作完全憑長相而定……（頁 84）

源文本：

Often I stopped in a village and saw only the very old, the disabled, and the very young, because **migrants** left their children behind to be raised by grandparents (p.91).

目標文本：

我在各個村莊停車的時候，看到的只有老年人、殘疾人、還有就是年齡很小的**小傢伙**。**那些人進城務工**後，把孩子留在家里，由爺爺奶奶或者外公外婆撫養。(頁 85)

在以上譯例中，目標文本對兩處“migrant”（遷離農村的人/遷居異地的人）的翻譯，均顯化了進城務工的目的。此外，“leave our village”譯成“出去了”；“出去”二字不僅隱去了離開的對象“村莊”，還反映出文中人物“她”的視角，即年輕人走出農村尋找機會不過是稀鬆尋常的舉動。加之增譯“務工”等字眼，聚焦者的觀察重心隨之出現偏移：源文本中“農村人口流失”的現象轉變成為目標文本中“農村人進城務工”的現象。

上述聚焦變化，在更深的層面上反映出源文本、目標文本聚焦者在知識結構上的不對稱。從源文本來看，聚焦者的觀察多少帶有“小鎮田園主義”價值觀的痕跡。這一價值取向關照工業城市發展進程中，以小鎮為代表的“渺小事物”（smallness）于維繫人際紐帶、人文傳統、自然生態方面具有的意義，而其中關於“田園”的意象還代表對過往生活的追憶（Gans 1979, 48-50）。圍繞此價值觀寫就的一類美國新聞報道，即側重反思隨城市擴張而田園鄉村逐漸消亡的現象（Gans 1979, 49）。反觀源文本所述故事，西北村莊人丁漸稀，大量年輕人為追趕東部沿海發展浪潮而切斷了與故土的紐帶，聚焦者鏡頭下的這一群體形象恰也投射出“小鎮田園主義”的影子。

此外，源文本聚焦者眼下的移民浪潮還似與美國19世紀下半葉的“鍍金時代”（The Gilded Age）遙相呼應。在《尋路中國》源文本中，中國經濟急速增長的景象貫穿全篇，海斯勒曾在文中藉以“鍍金時代”類比中國內地的“改革時代”。“鍍金時代”背景下，美國工業高速發展，大量歐洲移民為擺脫日益嚴峻的政治、經濟以及宗教矛盾，遠渡重洋來到美國這片“應允之地”（Cashman 1993, 74）。源文本聚焦者將遠走他鄉的中國農民看作“移民”，亦似把這個群體類比為經濟增長、城鄉矛盾加劇之下的時代產物。

相較而言，目標文本聚焦者鏡頭下的“務工者”，則反映出中國百姓對打工群體的既有認知：這群人走出農村、去往城市，是為謀求更好的出路，更多是順應時勢而為的主動選擇。而來到城市的農村人，在為自身爭取工作機會的同時，也參與了城鄉發展進程。官方敘事即把這個群體視為“推動中國經濟和社會結構變革的巨大力量”（中國農民工問題研究總報告起草組 2006）。因此，儘管面臨離土離鄉的特殊境遇，打工群體具有的勞動屬性，通常是界定其身份的重要面向。

可見，源文本、目標文本聚焦者分別透過不同的背景知識對同一社會群體進行觀察。而這一背景知識的差異，不僅僅在於高低深淺，而還涉及立場不同：當源文本聚焦者透過美國文化腳本，來觀察中國社會時，目標文本聚焦者則透過中國社會的集體認知進行聚焦。這個集體認知，不單關涉官方話語，還蘊含中國百姓的人生經歷，及其對生活環境形成的內在認知，包括對過往的集體回憶，以及對社會變遷的親身感受（吳音然 2019, 78-165）。

## 4.2 聚焦者與聚焦對象情感連結的建立

源文本、目標文本聚焦者對中國百姓及社會的觀察，還蘊含聚焦者與聚焦對象（即中國人物）在情感上的親疏遠近之別。其中，目標文本聚焦者視域內呈現出的故事人物形象，反映出包括譯者、讀者在內的中國百姓，與故事人物之間因共同的文化、政治身份產生的情感連結。



上述區別可由《江城》譯本中四川方言的使用加以詮釋。在源文本中，人物對話大多採用標準英語，僅在少數情況下包含對四川方言的直譯、音譯，或在行文中提示相關人物說的是方言。然而，在翻譯標準英語表述的人物對白時，譯者採用了內地讀者可辨識的四川方言，主要包括代詞“啥”、語氣助詞“哦”、“啲”、動詞“曉得”、副詞“好”等。這些方言多見於農民、船主、餐館老闆等人物的話語中。

通常，方言的使用直接標示人物的“聲音”，即人物說話的風格，但方言也可間接反映聚焦者觀察人物的視角。具有川味風格的對話通常帶有“逗趣調笑”性質，透露出川渝地區人物樂天、幽默、狡黠的性格特徵（王中 2015, 218-219）。從這個意義上看，四川方言可標示聚焦者透過心理視角對人物內在特質的觀察，令聚焦者視野里投射出鮮活生動的人物形象。例如：

### 例 3

源文本：“His feet are so big—look at those enormous shoes!” (p.353)

目標文本：“他那雙腳也大得很嘛！你看他穿得是好大一雙鞋啲！”（頁 382）

在方言的加工下，上例中鄉間農婦對敘述者這個外國人帶有打趣意味的評論，令源文本聚焦者的鏡頭下“自得其樂”的農婦形象更為活靈活現。《江城》的譯者李雪順是涪陵當地人，與作者筆下川渝地區的故事人物存在特有的親近關聯，隨著四川方言在目標文本中穿插使用，這份親近關聯也投射到了目標文本聚焦者與聚焦對象的關係之中。

此外，四川方言還參與塑造出具有能動性的故事人物，使中文譯者/讀者與故事人物之間的情感連結以更微妙的方式，融入目標文本聚焦者的觀察之中。源文本聚焦者不時透過社會環境審視“人”，某些故事人物可謂是中國社會群體或現象的縮影，用於反映社會議題。下例中對長江命運不聞不問的船夫，便代表了對身外時事漠不關心的一般中國百姓。這樣的人物可視為一種抽象的“類型”（type）。源文本多用標準英語表述人物對話，較少顯示說話人的情緒、性格等個性化特徵，這在一定程度上也參與塑造出類型化角色。然而，四川方言則打破了人物身上的共性，聚焦者視域中轉而呈現出個性鮮活、帶有地域特徵的人物。可由下例進一步闡釋：

### 例 2

源文本：

“The three Gorges aren’t too risky if you understand the river, though. Of course, if you don’t know the river, it’s difficult, but we’ve been through there many times. And after all those trips it’s not so interesting anymore……” …But this Fuling boatsman is still a boatsman; his interest is shipping, not the lore and history and poetry of the river. (p.129)

目標文本：

“不過，如果熟悉水情，三峽也沒啥子危險。當然，如果不懂，那就難開多了。不過，我們都走過好多次了。走來走去，也沒啥樂趣了……” ……不

過，這個涪陵船夫總歸只是個船夫，他感興趣的是貨運，而不是這條江河的知識、歷史和詩意。(頁 141-142)

片段最末敘述者對船夫的評論，表現出只顧營生，而對三峽漠不關心的船夫形象，而船夫的言論也從側面印證了他對江河深層意義的不理解。目標文本中的四川方言，則牽制了源文本聚焦者對船夫的上述觀察。其中，“沒啥子危險”除引人聯想江川地區百姓的爽朗的情性外，還將源文本對航行狀況的客觀陳述(“aren't too risky”)，轉化為船夫面對急流險灘毫不膽怯的心態。而表示“往、到”的動詞“走”，還把航船活動轉變成無甚稀奇的日常行為，映襯出船夫行船多日、船技嫺熟。儘管片段最後，目標文本聚焦者仍以批判的眼光審視船夫，但其視野里也呈現出一個豐富立體的川江船夫形象。

可以說，中國人物在源文本中處於被觀察的客體位置，人物的一舉一動是經源文本敘述者的目光及其所認同的價值理念過濾之下形成的產物。然而，通過包括四川方言在內的語言標記變化，目標文本聚焦者仿佛將中國故事人物，視為與中國主體類似的行為主體，將被觀察的人物“他”等同於中國主體“我”。上一節中，被迫離開農村的“出走者”重塑為主動尋求機會的“務工者”，這也反映出聚焦者似將中國主體的自我意識投射到故事人物中，從而在鏡頭下塑造出更具能動性的人物形象。聚焦者觀察中發生的上述微妙轉變，揭示出中文譯者/讀者與中國故事人物達成的自我認同心理，以及在更深層面上存在的情感紐帶。

### 4.3 “局外人”聚焦者到“局內人”聚焦者的轉變

由上文分析可見，以海斯勒作品為代表的中國題材當代敘事新聞中譯，在認知與情感層面上，隱含一個重新聚焦的過程，而這個過程在根本上關係到源文本、目標文本聚焦者屬性的差異。

具體來說，源文本聚焦者對中國人物、事件的觀察，在認知層面上，折射出西方價值觀(如“小鎮田園主義”)。而聚焦者鏡頭下的類型化角色，還反映出聚焦者似與人物保持一定距離，間或透過社會環境而非人物的內在屬性，來解讀人物及其行為。在這個意義上，源文本聚焦者佔據“局外人”的觀察立場。相較而言，目標文本聚焦者則呈現出“局內人”屬性。聚焦者對中國人物、事件的觀察，一方面反映出中國百姓對中國社會及相應社會群體(如打工群體)的集體認知，而另一方面，還折射出中文譯者/讀者與中國故事人物間潛在的情感連結。

聚焦者從“局外人”到“局內人”的轉變，與譯者的“主體性”密切相關。正是基於譯者對中國百姓、社會的認知和情感，中國百姓的所思所感得以融入“局內人”聚焦者的觀察之中。儘管目標文本敘述者，在中文讀者腦海中仍浮現外國人的“面貌”(persona)，但與敘述者關聯的聚焦者發生了變化，轉而呈現出“局內人”的特徵。這個由敘述者中分離出的“局內人”聚焦者，可視為擁有類似人格特徵的中文譯者在目標文本中的“化身”(Wu 2017, 40)。

當然，“局內人”聚焦者對中國社會所作觀察，通常與官方立場趨向一致，或多或少表現出社會規訓的痕跡，這似乎並非完全是譯者個體選擇的產物，而還可能關涉出

版編輯、審查者作出的安排。然而，帶有雜合性質的“局內人”聚焦者，仍與譯者存在最親密的聯繫。聚焦者鏡頭下的“務工者”，以及具有內在生命力的江川百姓，均反映出譯者對語言標記的調整並非全然基於對中國社會政治環境的理性考量，而還包含中國百姓的固有認識、情感訴求，及在這些因素影響下的下意識選擇。因此，“局內人”聚焦者與所謂的“守門人”（gatekeeper）不盡相同，後者通常從屬於特定機構，就引入目標文化的譯語內容把關，對源文本中與目標文化意識形態衝突的內容有意識地加以過濾。不同於這類擁有機構身份而面目模糊的“守門人”，“局內人”聚焦者則更多凝結了譯者的身心因素，尤其是譯者在認知與情感層面上，與所歸屬群體及社會時空存在的親近關聯。

上述以“局內人”聚焦者為主體的“重新聚焦”過程，歸根結底是一種與目標文化密切相關的翻譯現象，但翻譯過程中產生的重構，並不全然等同於安德列·勒弗維爾（André Lefevere）提出的“翻譯改寫”（Lefevere 1992a/1992b）。在審視“文本形象”在翻譯中產生的變化時，“翻譯改寫”理論更為強調譯者受到目標文化系統相關因素的外部制約，以及來自目標文化當權者及新興勢力的影響。相較而言，“重新聚焦”則關注譯者的能動性，即譯者身心因素在翻譯重構過程中產生的影響。不過，上述意義上譯者的能動性，也未必等同於蒙娜·貝克（Mona Baker）“重新敘述”思想（re-narration）（Baker 2006）中的運動家（activist）屬性，後者更為關注譯者為達成特定政治意圖（political agenda）而有意識採取的重敘決策。與此不同，“重新聚焦”側重譯者在思維、情感等因素影響下的下意識選擇，以及由此在翻譯中產生的細微變化。

## 五、結語

本文圍繞“聚焦”的概念，對海斯勒的兩部中國題材英文敘事新聞作品及簡體譯本進行分析，將這類“中國故事”翻譯過程中伴隨視角變化產生的重構，解讀為涉及“局外人”聚焦者到“局內人”聚焦者變化的“重新聚焦”過程。在譯者的影響下，目標文本中與“外國人”敘述者關聯的聚焦者，在一定程度上仿佛透過中國百姓的眼光來觀察中國。

上述“重新聚焦”模式的提出，還可為詮釋其他類型的“中國題材異語書寫”的中譯機制，或王宏印（2015/2016）、梁志芳（2013）等所提“回譯”現象帶來一定啟示。這類作品的中譯，是一種與目標文化密切相關的翻譯現象，但翻譯過程中產生的重構，並不同於勒弗維爾所言“翻譯改寫”。伴隨“局內人”聚焦者生成的“重新聚焦”更強調譯者在身心因素影響下敘述自我的本能，即藉由他者講述的中國故事，來表達中國百姓對所生活土地的理解與情感。

## 參考文獻

Baker, Mona. 2006. *Translation and Conflict: A Narrative Account*. London: Routledge.

- Bal, Mieke. 2009. *Narratology: Introduction to The Theory of Narrative*, translated by Christine van Boheemen. Toronto: University of Toronto Press.
- Cashman, Sean Dennis. 1993. *America in the Gilded Age*. New York and London: New York University Press.
- Cohen, Paul. 2013. "Peter Hessler: Teacher, Archaeologist, Anthropologist, Travel writer, Master storyteller." *The Journal of Asian Studies* 72 (2): 251-272.
- Connery, Thomas Bernard. 1990. "A Third Way to Tell the Story: American Literary Journalism at the Turn of the Century." In *Literary Journalism in the Twentieth Century*, edited by Norman Sims, 3-20. Oxford: Oxford University Press.
- Gans, Herbert. 1979. *Deciding What's News: A Study of CBS Evening News, NBC Nightly News, Newsweek, and Time*. New York: Pantheon Books.
- Genette, Gérard. 1980. *Narrative Discourse: An Essay in Method*, translated by Jane E. Lewin. Oxford: Blackwell.
- Harbers, Frank, and Marcel Broersma. 2014. "Between Engagement and Ironic Ambiguity: Mediating Subjectivity in Narrative Journalism." *Journalism* 15 (5): 639-654.
- Hartsock, John. 2000. *A History of American Literary Journalism: The Emergence of a Modern Narrative Form*. Amherst: University of Massachusetts.
- Hessler, Peter. 2010. *Country Driving: A Journey from Farm to Factory*. New York: Harper Collins.
- Hessler, Peter. 2001. *River Town: Two Years on the Yangtze*. New York: Harper Collins.
- Kramer, Mark. 1995. "Breakable Rules for Literary Journalists." In *Literary Journalism: A New Collection of the Best American Nonfiction*, edited by Norman Sims and Mark Kramer, 21-36. New York: Ballantine Books.
- Lefevere, André. 1992a. *Translation, Rewriting, and the Manipulation of Literary Fame*. London: Routledge.
- Lefevere, André. 1992b. *Translation/History/Culture: A Sourcebook*. London: Routledge.
- Rimmon-Kenan, Shlomith. 2002. *Narrative Fiction: Contemporary Poetics*. London: Routledge.
- Saint-Exupéry, Antoine. 2000. *The Little Prince*, translated by Howard Richard. Orlando: Harcourt.
- Song, Yunya. 2012. "Shifting Journalistic Paradigms of American Correspondents on Contemporary China: The Case of Orville Schell." *Public Relation Review* 38 (5): 796-798.
- Song, Yunya, and Chin-Chuan Lee. 2015. "The Strategic Ritual of Irony: Post-Tiananmen China as Seen Through the 'Personalized Journalism' of Elite US Correspondents." *Media, Culture & Society*, 37 (8): 1176-1192.
- Spence, Johnathan. 2006. "Letters from China," review of *Oracle Bones*, by Peter Hessler, *New York Times*, April 30, 2006, <https://www.nytimes.com/2006/04/30/books/review/30spence.html>.

- Toolan, Michael. 2001. *Narrative: A Critical Linguistic Introduction*. London: Routledge.
- Van Krieken, Kobie, and Jose Sanders. 2019. "What is Narrative Journalism? A Systematic Review and an Empirical Agenda." *Journalism* 22(6): 1393-1412.
- Wu, Yinran. 2017. "Exploring Translators' Impact on Translated Narrative: A Model of Refocalization." *MANUSYA Journal of Humanities* 23: 23-44.
- Zhang, Pinling. 2018. "Same Languages but Different Ideologies in Mandarin Chinese Translations of Peter Hessler's Two Books." Paper presented at The 6th IATIS International Conference on Translation and Cultural Mobility, Hong Kong, August 2018.
- 里蒙-凱南 Rimmon-Kenan, Shlomith. 1989. 《敘事虛構作品：當代詩學》 = *Narrative Fiction: Contemporary Poetics*. 姚錦清譯。北京：三聯。
- 李雪順譯。2011. 《尋路中國：從鄉村到工廠的自駕之旅》。上海：上海譯文出版社。
- 李雪順譯。2012. 《江城》。上海：上海譯文出版社。
- 梁志芳。2013. "文化回譯" 研究—以賽珍珠中國題材小說《大地》的中譯為例。《當代外語研究》8：51-55。
- 申丹。2006. 〈視角〉。趙一凡、張中載、李德恩編，《西方文論關鍵字》，511-527。北京：外語教學與研究出版社。
- 申丹、王麗亞。2010. 《西方經典敘事學：經典與後經典》。北京：北京大學出版社。
- 王宏印。2015. 從“異語寫作”到“無本回譯”——關於創作與翻譯的理論思考。《上海翻譯》3：1-9。
- 王宏印。2016. 朝向一種普遍翻譯理論的“無本回譯”再論——以《大唐狄公案》等為例。《上海翻譯》1：1-9。
- 王瓊。2016. 探議“譯敘空間”的概念。《譯苑新壇》8：11-18。
- 王中。2015. 《方言與 20 世紀中國文學》。合肥：安徽教育出版社。
- 中國農民工問題研究總報告起草組。2006. 《中國農民工問題研究總報告》 <http://paper.usc.cuhk.edu.hk/Details.aspx?id=5780>
- 吳音然。2019. 《回譯與重敘研究：以彼得·海斯勒“中國三部曲”英中翻譯研究為例》。博士論文，香港浸會大學。

# 嚴復《天演論·譯例言》講疏

Yu Kit Cheung

**Address:** Centre for Translation and Intercultural Studies, The University of Manchester, The United Kingdom

**E-mail:** yukit.cheung@manchester.ac.uk

**Correspondence:** Yu Kit Cheung

**Citation:** Cheung, Yu Kit. 2022. “Exegesis of Yan Fu’s ‘Note on Translation Strategy’ in his Chinese Translation of T.H. Huxley’s *Evolution and Ethics*.” *Translation Quarterly* 105: 77-94.

## *Abstract*

Exegesis of Yan Fu’s ‘Note on Translation Strategy’ in his Chinese Translation of T.H. Huxley’s *Evolution and Ethics* (by Yu Kit Cheung)

*Yan Fu is an historical figure of interdisciplinary importance, whose ‘Note on Translation Strategy’ (‘The Note’ hereafter) in Tianyan lun, his fin de siècle Chinese translation of Thomas H. Huxley’s Evolution and Ethics marked the beginning of modern Chinese translation theories. Whilst his theory has received overwhelming academic attention and has been a staple work in collections of essays and courses on Chinese translation theories, the annotated versions of ‘The Note’ – predominantly in the selected essays of Yan Fu – demand amendments in one way or another.*

*This paper contributes to the research of ‘The Note’ by offering a fresh and full annotated edition of the text. The annotations will contextualise the readers to approach this demanding text from the lens of conventional classical Chinese scholarship as well as from the perspective of Translation Studies where appropriate. It breaks with precedent that rather than mere glosses of difficult words or phrases, it has engaged a wide range of contextual material, including various editions of Tianyan lun and Yan Fu’s letters, poems, and essays in the annotations.*

## 一、弁言

嚴復《天演論·譯例言》堪稱現代中國翻譯理論的嚆矢，影響深遠。當中提出了翻譯三難「信」、「達」、「雅」，尤為人熟悉。然而，嚴氏只提出了信達雅三難，卻沒有申論其涵義；百多年來就信達雅之討論不輟，而「雅」的意思尤晦。其實，不惟信達雅三難之涵意疏紊，〈例言〉全文亦不易，一則全文以文言文寫成，二則要讀通全文，所需要的學問範疇亦頗廣，包括文字訓詁、經學、英文文法、古書徵引文句慣例等。

二〇一五年春，筆者與 Roberta Raine 教授於香港嶺南大學翻譯系主講「翻譯史」，凡十三講，筆者其中一講專論嚴復。〈譯例言〉文字艱澀，故搜討現成注釋、英譯，冀便於學生學習。然而，〈例言〉注解不多，主要的有周振甫（1959）、沈蘇儒（1998）、牛仰山（2006），而英譯至今則只有徐兆墉刊於香港中文大學的翻譯雜誌 *Renditions* 的版本（Hsu 1973）。四者各有所長，然而或注釋未夠深入，或釋義譯文猶未盡當，或只着眼探討信、達、雅之內涵；至於〈例言〉所涉及的英文文法概念，全文詳注更付闕如。筆者因此不揆譎陋，翻檢字書載籍，參考《天演論》不同版本以及嚴復其他著作及書信<sup>[1]</sup>，妄為作注，並拈錄與〈例言〉相關的資料，包括嚴復生平、其主要翻譯、《天演論》主要版本、《天演論》目次，附於注後，草成是篇，冀有裨學生，亦冀有益於中國近代翻譯史研治之萬一。

本疏有兩大特色。第一，筆者根柢訓詁經史與翻譯學（Translation Studies），詳析難字難句，並指出嚴氏個別遣詞引文之謬。第二，除字書載籍外，筆者亦參攷《天演論》不同版本以及嚴復的其他著作和書信。

〈譯例言〉之版本有二<sup>[2]</sup>。「慎始基齋木刻本」末段誌嚴復付梓木刻之緣起。至於其他版本蓋非木刻，出版時被刪汰。本〈疏〉據「慎始基齋本」。

〈譯例言〉全文連日期署名計 1,065 字。本〈疏〉編排如次：

### 一、弁言

### 二、〈譯例言〉全文

### 三、〈譯例言〉注釋

### 四、附錄

附錄一、嚴復生平畧

附錄二、嚴復主要翻譯

附錄三、嚴譯《天演論》主要版本

附錄四、《天演論》目次

二〇一五年三月於香港嶺南大學翻譯系

二〇一六年，筆者發表了〈嚴復《天演論·譯例言》譯事三難與「三曰」之關聯探微〉（張宇傑 2016），二〇一九年又寫了〈「雅」的真諦〉（張宇傑 2019），資料均據本〈疏〉，而二文可與本〈疏〉互為表裏。

二。二二年七月又識於英國曼徹斯特大學  
翻譯與跨文化研究中心

## 二、〈譯例言〉全文

譯事三難信達雅。求其信已大難矣。顧信矣不達。雖譯猶不譯也。則達尚焉。海通已來。象寄之才。隨地多有。而任取一書。責其能與於斯二者。則已寡矣。其故在淺嘗一也。偏至二也。辨之者少三也。今是書所言。本五十年來西人新得之學。又為作者晚出之書。譯文取明深義。故詞句之間。時有所真到坳益。不斤斤於字比句次。而意義則不倍本文。題曰達旨。不云筆譯。取便發揮。實非正法。什法師則有云。學我者病。來者方多。幸勿以是書為口實也。

西文句中名物字。多隨舉隨釋。如中文之旁支。後乃遙接前文。足意成句。故西文句法。少者二三字。多者數十百言。假令仿此為譯。則恐必不可通。而刪削取徑。又恐意義有漏。此在譯者將全文神理融會於心。則下筆抒詞。自然互備。至原文詞理本深。難於共喻。則當前後引襯。以顯其意。凡此經營。皆以為達。為達即所以為信也。

易曰脩辭立誠。子曰辭達而已。又曰言之無文。行之不遠。三曰乃文章正軌。亦即為譯事楷模。故信達而外。求其爾雅。此不僅期以行遠已耳。實則精理微言。用漢以前字法句法。則為達易。用近世利俗文字。則求達難。往往抑義就詞。毫釐千里。審擇於斯二者之間。夫固有所不得已也。豈鈞奇哉。不佞此譯。頗貽艱深文陋之譏。實則刻意求顯。不過如是。又原書論說。多本名數格致。及一切疇人之學。倘於之數者向未問津。雖作者同國之人。言語相通。仍多未喻。矧夫出以重譯也耶。

新理踵出。名目紛繁。索之中文。渺不可得。即有牽合。終嫌參差。譯者遇此。獨有自具衡量。即義定名。顧其事有甚難者。即如此書上卷導言十餘篇。乃因正論理深。先敷淺說。僕始繙卮言。而錢唐夏穗卿曾佑病其濫惡。謂內典原有此種。可名懸談。及桐城吳丈摯父汝綸見之。又謂卮言既成濫詞。懸談亦沿釋氏。均非能自樹立者所為。不如用諸子舊例。隨篇標目為佳。穗卿又謂如此。則篇自為文。於原書建立一本之義稍晦。而懸談懸疏諸名。懸者玄也。乃會撮精旨之言。與此不合。必不可用。於是乃依其原目。質譯導言。而分注吳之篇目於下。取便閱者。此以見定名之難。雖欲避生吞活剝之誚。有不可得者矣。他如物競天擇儲能效實諸名。皆由我始。一名之立。旬月踟躕。我罪我知。是存明哲。

原書多論希臘以來學派。凡所標舉。皆當時名碩。流風緒論。泰西二千年之人心民智係焉。講西學者所不可不知也。茲於篇末。畧載諸公生世事業。粗備學者知人論世之資。

窮理與從政相同。皆貴集思廣益。今遇原文所論。與他書有異同者。輒就譴陋所知。列入後案。以資參考。間亦坳以己見。取詩稱嚶求。易言麗澤之義。是非然否。以俟公論。不敢固也。如曰標高揭己。則失不佞懷鉛握槧。辛苦逐譯之本心矣。



是編之譯。本以理學西書。繙轉不易。固取此書。日與同學諸子相課。迨書成。吳丈擊甫見而好之。斧落徽引。匡益實多。顧惟探蹟叩寂之學。非當務之所亟。不願問世也。而稿經新會梁任父、沔陽盧木齋諸君借鈔。皆勸早日付梓。木齋郵示介弟慎之於鄂。亦謂宜公海內。遂災棗梨。猶非不佞意也。刻訖寄津覆斟。乃為發例言。并識緣起如是云。

光緒二十四年歲在戊戌四月二十二日嚴復識於天津尊疑學塾

據清光緒盧氏慎始基齋刻本

### 三、〈譯例言〉注釋

譯事三難<sup>[3]</sup>：信達<sup>[4]</sup>雅<sup>[5]</sup>。求其信已大難矣，顧信矣不達，雖譯猶不譯也，則達尚<sup>[6]</sup>焉。海通已來<sup>[7]</sup>，象寄<sup>[8]</sup>之才，隨地多有，而任取一書，責<sup>[9]</sup>其能與<sup>[10]</sup>於斯二者，則已寡矣。其故在淺嘗一也<sup>[11]</sup>；偏至二也<sup>[12]</sup>；辨之者少三也<sup>[13]</sup>。今是書所言，本五十年來西人新得之學，又為作者晚出之書。譯文取明深義<sup>[14]</sup>，故詞句之間，時有所俱<sup>[15]</sup>到<sup>[16]</sup>坳<sup>[17]</sup>益，不斤斤<sup>[18]</sup>於字比<sup>[19]</sup>句次，而意義則不倍本文<sup>[20]</sup>。題曰達一信<sup>[21]</sup>，不云筆譯，取便發揮<sup>[22]</sup>，實非正法。什法師則有云：<sup>[23]</sup>「學我者病。」<sup>[24]</sup>來者方多<sup>[25]</sup>，幸<sup>[26]</sup>勿以是書為口實也。<sup>[27]</sup>

西文句中名物字<sup>[28]</sup>，多隨舉隨釋，如中文之旁支，後乃遙接前文，足意成句。<sup>[29]</sup>故西文句法，少者二三字，多者數十百言<sup>[30]</sup>。假令<sup>[31]</sup>仿此為譯，則恐必不可通，而刪削取徑<sup>[32]</sup>，又恐意義有漏。此在譯者將全文神理融會於心，則下筆抒詞，自然互備。至原文詞理本深，難於共喻<sup>[33]</sup>，則當前後引襯，以顯其意。凡此經營，皆以為達，為達即所以為信也。

《易》曰「脩辭立誠。」<sup>[34]</sup>子曰「辭達而已。」<sup>[35]</sup>又曰「言之無文<sup>[36]</sup>，行之不遠。<sup>[37]</sup>」三曰<sup>[38]</sup>乃文章正軌<sup>[39]</sup>，亦即為譯事楷模。故信達而外，求其爾雅<sup>[40]</sup>。此不僅期以行遠已耳<sup>[41]</sup>。實則精理微言，用漢以前字法句法，則為達易；用近世利俗文字<sup>[42]</sup>，則求達難。往往抑義就詞，毫釐千里。審擇於斯二者之間<sup>[43]</sup>，夫固有所不得已也，豈鈞奇哉<sup>[44]</sup>！不佞<sup>[45]</sup>此譯，頗貽艱深文陋之譏<sup>[46]</sup>，實則刻意求顯，不過如是。又原書論說，多本名數格致<sup>[47]</sup>，及一切疇人之學<sup>[48]</sup>，倘於之<sup>[49]</sup>數者向未問津<sup>[50]</sup>，雖作者同國之人，言語相通，仍多未喻，矧<sup>[51]</sup>夫出以重譯<sup>[52]</sup>也耶？

新理踵出<sup>[53]</sup>，名目紛繁，索之中文，渺不可得，即有牽合，終嫌參差。譯者遇此，獨有自具衡量<sup>[54]</sup>，即義定名<sup>[55]</sup>。顧其事有甚難者，即如此書上卷〈導言〉十餘篇，乃因正論理深<sup>[56]</sup>，先敷淺說。僕<sup>[57]</sup>始繙<sup>[58]</sup>「卮言」<sup>[59]</sup>，而錢唐夏穗卿<sup>曾佑</sup>病其濫惡<sup>[60]</sup>，謂內典原有此種<sup>[61]</sup>，可名「懸談」。及桐城吳丈擊父<sup>汝綸</sup>見之<sup>[62]</sup>，又謂「卮言」既成濫詞，「懸談」亦沿釋氏<sup>[63]</sup>，均非能自樹立者所為，不如用諸子舊例<sup>[64]</sup>，隨篇標目為佳。穗卿又謂如此，則篇自為文，於原書建立一本之義稍晦。而「懸談」「懸疏」諸名<sup>[65]</sup>，懸者玄也，乃會撮精旨之言<sup>[66]</sup>，與此不合，必不可用。於是乃依其原目，質譯「導言」<sup>[67]</sup>，而分注吳之篇目於下，取便閱者。此以見定名之難，雖欲避生

吞活剝之誚<sup>[68]</sup>，有不可得者矣。他如「物競」<sup>[69]</sup>「天擇」<sup>[70]</sup>「儲能」「效實」諸名，皆由我始<sup>[71]</sup>。一名之立，旬月<sup>[72]</sup>踟躕<sup>[73]</sup>。我罪我知<sup>[74]</sup>，是存明哲<sup>[75]</sup>。

原書多論希臘以來學派，凡所標舉，皆當時名碩<sup>[76]</sup>。流風<sup>[77]</sup>緒論<sup>[78]</sup>，泰西二千年之人心民智係焉<sup>[79]</sup>，講西學者所不可不知也。茲<sup>[80]</sup>於篇末，畧載諸公生世事業，粗備學者知人論世之資<sup>[81]</sup>。

窮理與從政相同<sup>[82]</sup>，皆貴集思廣益。今遇原文所論，與他書有異同者，輒<sup>[83]</sup>就讎陋所知<sup>[84]</sup>，列入後案，以資參考。間亦附以己見，取《詩》稱嚶求<sup>[85]</sup>，《易》言麗澤之義<sup>[86]</sup>。是非然否，以俟公論<sup>[87]</sup>，不敢固也<sup>[88]</sup>。如曰標高揭己<sup>[89]</sup>，則失不佞懷鉛握槧<sup>[90]</sup>，辛苦逐譯<sup>[91]</sup>之本心矣。

是編之譯，本以理學<sup>[92]</sup>西書，繙轉不易<sup>[93]</sup>，固取此書<sup>[94]</sup>，日與同學諸子相課。迨書成<sup>[95]</sup>，吳丈摯甫見而好<sup>[96]</sup>之，斧落徽引，匡益實多。顧惟<sup>[97]</sup>探蹟<sup>[98]</sup>叩寂<sup>[99]</sup>之學，非當務之所亟<sup>[100]</sup>，不願問世也。而稿經新會梁任父<sup>[101]</sup>、沔陽盧木齋諸君借鈔，皆勸早日付梓<sup>[102]</sup>，木齋郵示介弟慎之於鄂<sup>[103]</sup>，亦謂宜公海內，遂災棗梨<sup>[104]</sup>，猶非不佞意也。刻訖寄津覆靺<sup>[105]</sup>，乃為發例言，并識<sup>[106]</sup>緣起如是云。

光緒二十四年歲在戊戌四月二十二日嚴復識於天津尊疑學塾<sup>[107]</sup>

據清光緒盧氏慎始基齋刻本

## 四、附錄

### 附錄一、嚴復生平畧

嚴復，清文宗咸豐三年癸丑（一八五三年）生，中華民國十年辛酉（一九二一年）卒，享壽六十九歲。

王蘧常《嚴幾道年譜》：

先生諱復，初名體乾，入馬江船政學堂，易名宗光，字又陵。登仕始改今名，字幾道。晚號滄懋老人。別署天演宗哲學家。又別號尊疑尺蠖。姓嚴氏，福建侯官人也。（1936，1）

《清史稿·文苑三》：

嚴復，初名宗光，字又陵，一字幾道，侯官人。早慧，嗜為文。閩督沈葆楨初創船政，招試英俊，儲海軍將才，得復文，奇之，用冠其曹，則年十四也。既卒業，從軍艦練習，周歷南洋、黃海。日本窺臺灣，葆楨奉命籌防，挈之東渡訶敵，勘測各海口。光緒二年，派赴英國海軍學校肄戰術及砲臺建築諸學，每試輒最。侍郎郭嵩燾使英，賞其才，時引與論析中西學術同異。學成歸，北洋大臣李鴻章方大治海軍，以復總學堂。二十四年，詔求人才，復被薦，召對稱旨。論繕所擬萬言書以進，未及用，而政局猝變。越二年，避拳亂南歸。

是時人士漸傾向西人學說，復以為自由、平等、權利諸說，由之未嘗無利，脫靡所折衷，則流蕩放佚，害且不可勝言，常於廣眾中陳之。復久以海軍積勞敘副將，盡棄去，入貲為同知，累保道員。宣統元年，海軍部立，特授協都統，尋賜文科進士，

充學部名詞館總纂。以碩學通儒徵為資政院議員。三年，授海軍一等參謀官。復殫心著述，於學無所不窺，舉中外治術學理，靡不究極原委，抉其失得，證明而會通之。精歐西文字，所譯書以瓌辭達奧旨。

其《天演論》自序有曰：「仲尼【筆者案「尼」讀「泥」】之於六藝也，易、春秋最嚴……二千年來，士徇利祿，守闕殘，無獨闢之慮，是以生今日者，乃轉於西學得識古之用焉。」凡復所譯著，獨得精微皆類此。

世謂紓【案林紓】以中文溝通西文，復以西文溝通中文，並稱「林嚴」。辛酉秋，卒，年六十有九。著有文集及譯《天演論》、《原富》、《羣學肄言》、《穆勒名學》、《法意》、《羣己權界論》、《社會通詮》等。(趙爾巽等撰 1942, 1544 上、中)

引《清史稿·文苑傳》說嚴復生平者不多，故詳引之如上，以備參攷。

### 附錄二、嚴復主要翻譯

年份	年紀	譯作	原作	案語
1895 [108]	四十三歲	天演論	T.H. Huxley, <i>Evolution and Ethics</i>	嚴 據〈侯官嚴先生年譜〉：「未數月而脫稿。」（見王棊主編 1986, 1548）
1897	四十五歲	原富	Adam Smith, <i>An Inquiry into the Nature and Causes of the Wealth of Nations</i>	
1897	四十五歲	群學肄言	Herbert Spencer, <i>Study of Sociology</i>	
1899	四十七歲	羣己權界論	John Stuart Mill, <i>On Liberty</i>	
1903	五十一歲	穆勒名學 (未完)	John Stuart Mill, <i>A System of Logic</i>	
1903	五十一歲	社會通詮	Edward Jenks, <i>A History of Politics</i>	
1906	五十四歲	法意	C.L.S. Montesquieu, <i>De l'esprit des Lois</i>	
1909	五十六歲	名學淺說	William Stanley Jevons, <i>Primer of Logic</i>	

除上列譯品八種外，嚴復還有譯文約二十篇，俱收汪征魯、方寶川、馬勇主編（2014, 469-629）。茲不一。

### 附錄三、嚴譯《天演論》主要版本

1. 味經本（1895），扉頁題「光緒乙未春三月陝西味經售票處重刊」
2. 手稿本（1896），有〈自序〉作於光緒丙申重九。書中有「丁酉四月十七日刪節」、「丁酉六月初三日刪改訖」等語。

3. 慎始基齋本（1898）「慎始基齋叢書」
4. 富文本（1901），扉頁印有「光緒辛丑仲春富文書局石印」字樣
5. 吳京卿節本天演論（1903）
6. 商務印書館（1905），1931年收入「嚴譯名著叢刊」

#### 附錄四、《天演論》目次

《天演論》正文前，有吳汝綸（1840-1903）序、嚴復自序、〈譯例言〉。茲錄《天演論》篇目如下：

吳序

自序

#### 譯例言

#### 上卷導言十八篇

察變第一 廣義第二 趨異第三 人為第四 互爭第五  
人擇第六 善敗第七 烏託邦第八 汰蕃第九 擇難第十  
蜂羣第十一 人羣第十二 制私第十三 恕敗第十四 最旨第十五  
進微第十六 善羣第十七 新反第十八

#### 下卷論十七篇

能實第一 憂患第二 教源第三 嚴意第四 天刑第五  
佛釋第六 種業第七 冥往第八 真幻第九 佛法第十  
學派第十一 天難第十二 論性第十三 矯性第十四 演惡第十五  
羣治第十六 進化第十七

## 注釋

- [1] 有關《天演論》之主要版本，請參「附錄三」。
- [2] 〈譯例言〉第四段，即「新理踵出……是存明哲」一段本無，乃後增。王蘧常《嚴幾道年譜》云：「譯例蚤成於譯竣時。故摯父作序已見之。蓋後加者移隸於此。」（1936，40）
- [3] 嚴復雖言「譯事」，卻並非泛指所有翻譯。〈譯例言〉裏提出的翻譯觀和翻譯法，只針對《天演論》而言。嚴譯《原富·譯事例言》：「……是譯與《天演論》不同。下筆之頃。雖於全節文理。不能不融會貫通為之。然於辭義之間。無所顛倒附益……」（1930b，6）。可見嚴氏繙譯《原富》的策略與譯《天演論》的不同。Mary Snell-Hornby 說：「This phenomenon, whereby a theorist makes global observations on translation in general, but actually means only one, often narrow area of it, still characterizes translation studies today - to the detriment of a general theory of translation.」（1988/1995，14）。嚴氏把二者混為一談，或因為《天演論》是第一部自己

選擇、有系統的譯品。有關嚴復的主要譯品，請參「坵錄二」。Snell-Hornby 所言之「a theorist makes global observations on translation in general, but actually means only one」，大抵即此。

[4] 「手稿本」〈譯例〉：「是譯以理解明白為主。詞語顛倒增減。無非求達作者深意。然未嘗離宗也。」（汪征魯等 2014，8）有關《天演論》的主要版本，詳見「坵錄三」。信：不「離宗」，不離「作者深意」。達：通也。第二段「必不可通」的「通」即達意，傳達原文的意思。因此第二段說「題曰達信」。Venuti (1995, 182)「信」作 *fidelity*，是；「達」作 *clarity or comprehensibility*，非也。

[5] 雅：爾雅。見〈譯例言〉第三段：「求其爾雅」。爾雅，與第三段「近世利俗文字」相對。「利俗文字」注解詳下。劉熙《釋名·釋典藝》：「爾雅。爾、昵也。昵、近也。雅、義也。義、正也。」（1985，100）「爾雅」即近古、近正。陳湛銓疏〈詩大序〉時指出，「雅」是「夏」之借字，聲音相近故也。夏即中國，故云正。（鄧又同編 1989，131-132）「爾雅」的文字即「用漢以前字法句法」（見原文第三段），即先秦儒家經典的文字。

或以為「雅」單指雅麗、「文字水平高」（沈蘇儒 1998，34）等。徐兆墉英譯作「*elegance*」（Hsu 1973，4），而英文翻譯論著提及「雅」時多譯之為「*elegance*」（Fong 1995，582；Wang 1995，999；Ma 1995，382；Wu 1995，529）。Venuti (1995, 182) 作「*elegance or fluency*」，Elizabeth Sinn 作「*elegance of style*」(1995, 441)，Liu Miqing 作「*gracefulness (elegance)*」(1995, 2)，Huang Yushi 作「*elegancy*」(1995, 278)。這些詮釋、翻譯都不算錯，因為先秦儒家經典的文字確然上乘（詳見本注第五段），然而卻僅指出了「雅」之一隅，抓不住「雅」的真義。魏文帝曹丕《典論·論文》：「夫文本同而末異。蓋奏議宜雅……詩賦欲麗。」（蕭統選，李善注 1936，1128）可見「雅」與「麗」是有分別的。嚴氏之「雅」重點在遣詞是否正統，亦請參注 42。嚴復在第三段以《左傳》「言之無文，行之不遠」證譯文「雅」之重要，是斷章取義，不倫不類。讀者宜明辨之。詳參注 36。

嚴復認為要傳達原文深意，文字便需爾雅，因為《天演論》所言是第三段所謂的「精理微言」。《天演論·自序》云：「赫胥黎氏此書之旨，本以救斯賓塞任天為治之末流，其中所論，與吾古人有甚合者。」（1898/1995，3 上）嚴氏認為《天演論》的主旨，就是「吾中土所前有」（同上）的「大義微言」（1898/1995，2 下）。嚴氏舉例以證歐洲學術所得之「名理公例」（同上）即《易》理。《易》開物成務，絜淨精微。因此他說「考道之士，以其所得於彼者，反以證諸吾古人之所傳，乃澄湛精瑩（後略）」（同上）。

既是大義微言，便要用先秦儒家經典的文字來翻譯，纔可以流傳久遠，因為古人多認為這些經典的文字最上乘。今天還知先哲的道理，全賴上乘的文字。劉勰《文心雕龍·原道》云：「道沿聖以垂文，聖因文以明道。」（劉勰 1937，3 下）《文心雕龍·徵聖》：「夫作者曰聖，述者曰明。陶鑄性情，功在上哲。夫子文章，可得而聞，則聖人之情，見乎文辭矣。先王聖化，布在方冊；夫子風采，溢於格言。是以遠稱唐世，則煥乎為盛；近褒周代，則郁哉可從：此政化貴文之徵也。」（劉

颺 1937, 4 上) 先秦文字之正猶拉丁文之在歐洲。嚴復〈致梁卓如書〉：「此文【案拉丁文】及希臘文，乃西洋文學根本，猶之中國雅學……」（南洋學會研究組編 1959, 120）。

先秦經典文字之上乘在於其「文」，就是經過修飾。【雅的文字是經過修飾，但有修飾的文字並不必然就是雅。讀者當明辨之。】《文心雕龍·原道》：「文之為德也大矣，與天地並生者何哉？夫玄黃色雜，方圓體分，日月疊璧，以垂麗天之象；山川煥綺，以鋪理地之形；此蓋道之文也。」（劉颺 1937, 1 上）況且，由上文可見，文就是天之道。《天演論》可比《易》理，而《易》闡天之道，譯筆自然要雅要文才是正道，用詞俚俗便有乖正道。

因此，一九〇二年，嚴復致函梁啟超云：「竊以謂文辭者，載理想之羽翼，而以達情感之音聲也。是故理之精者不能載以粗獷之詞，而情之正者不可達以鄙倍之氣。中國文之美者，莫若司馬遷、韓愈。而遷之言曰：『其志潔者，其稱物芳。』愈之言曰：『文無難易，惟其是。』僕之於文，非務淵雅也，務其是耳」（汪征魯等 2014, 122）。日正為「是」，「是」就是「正」的意思。嚴復選詞折衷於正。《天演論》之所載既可比《易》道，當以先秦經典所傳道之文字來翻譯才算正。嚴復之言正，猶今語言學謂「語域合適」（appropriate register）。正則安，不正便招險。《周易》言之極詳。正而安乃儒家思想之門蘊也。

其實，嚴復「雅」的概念由他極端的「歸化」（domesticating）翻譯策略而來。粗翻「手稿本」卷初稿，便屢見其翻譯每每不顧原文，逕引中國古書（汪征魯等 2014, 9-33）。吳汝淪不同意嚴復的翻譯手法。丁酉（1897 年）二月初七日，吳汝淪致嚴復書曰：「若自翻譯為一書，則可縱意馳騁，若以譯赫氏之書為名，則篇中所引古書古事，皆宜以原書所稱西方者為當，似不必改用中國人語。以中事中人，固非赫氏所及知。法宜如晉、宋名流所譯佛書，與中儒著述，顯分體制，似為入式。」（王栻主編 1986, 1560）

黃遵憲也不同意嚴復銳意以先秦古文來翻譯。光緒二十八（1902）年秋黃氏致嚴復函云：「以四千餘歲以前創造之古文，所謂六書，又無衍聲之變，孳生之法，即以書寫中國中古以來之物之事之學，已不能敷用，況泰西各科學乎……今日以為二十世紀之世界矣，東西文明，兩相接合，而譯書一事，以通彼我之懷，闡新舊之學，實為要務……公以為文界無革命，弟以為無革命而有維新。如四十二章經，舊體也，自鳩摩羅什輩出，而內典別成文體，佛教益行矣。」（陳錚編 2005, 434-436）

[6] 尚：崇尚。全句謂「達」也應重視。

[7] 已來：即「以來」。

[8] 象寄之才：譯者。《禮記·王制》：「中國戎夷。五方之民。皆有性也……五方之民。言語不通。嗜欲不同。達其志。通其欲。東方曰寄。南方曰象。西方曰狄鞮。北方曰譯。」（阮元校勘 1980, 總 1338）案：鞮，音「低」。《周禮》裏譯者稱「象胥」：「象胥掌蠻夷閩貉戎狄之國。使掌傳王之言而諭說焉。以和親之。若以時入賓。則協其禮與其辭言傳之。」（《秋官·司寇第五》）（阮元校勘 1980, 總

899)

- [9] 《說文》卷六下貝部：「責。求也。」(許慎撰，徐鉉校定 1963，130 下)
- [10] 與：合符。
- [11] 淺嘗：對翻譯的認識不夠深。
- [12] 偏至：偏頗極端，不周全。指泥守「信」或「達」，而不知二端均重要。
- [13] 辨之者少：能分辨「信」與「達」的很少。周振甫(1959，103)注曰：「辨別譯事三難的人少，不了解翻譯的甘苦。」牛仰山注曰(2006，149)：「指能辨別譯事三難者少，亦即不懂翻譯的甘苦。」沈蘇儒(1998，34)云：「不能真正理解原文。」三注似皆失之。
- [14] 取：求取。
- [15] 偵，同「顛」。
- [16] 到：同「倒」
- [17] 坳：「坳」乃「附」的正字。《說文》卷十三下土部：「坳。益也。」(許慎撰，徐鉉校定 1963，288 上)。
- [18] 斤斤：過分着意。
- [19] 比：靠近的意思，去聲，音「臂」。
- [20] 倍：《說文》卷八上入部：「倍。反也。」(許慎撰，徐鉉校定 1963，166 上) 本文：即原 (source text)。
- [21] 指：今天「旨」的正字。《說文》卷一。下心部：「指。意也。从心旨聲。」(許慎撰，徐鉉校定 1963，217 上)
- [22] 取便：方便。
- [23] 什法師：鳩摩羅什(344-413)，翻譯過不少佛經，著名的有《金剛經》。
- [24] 學我者病：學我的方法的便會出問題。筆者案「學我者病」之出處不詳。余尋檢載籍不果。嘗就此請教香港恆生大學佛典翻譯專家蕭世昌博士。蕭博士亦謂嘗尋檢出處未果。蕭博士教務繁忙，撥冗幫忙，敬申謝忱。
- [25] 方：正在。
- [26] 幸：希望。
- [27] 口實：話柄、藉口。全句謂「希望不要以這本書為藉口」。
- [28] 名物字：即名詞 (noun)。這是英文語法概念，當時還沒有統一的叫法。
- [29] 即英文的「關係從句」(relative clause)。
- [30] 言：一言可以為一字，亦可為一句。此處指一字，猶「五言詩」、「七言詩」之「言」。
- [31] 假令：假如。
- [32] 刪去枝節，只述大意，即 paraphrase。
- [33] 喻：曉也。明白意。
- [34] 嚴復由此起一連引了三句儒家經典的引文，對應信、達、雅三難。其實這三句引文與信達雅三難的關聯不大。原因主要有二。第一，古人屬文引經斷章取義，由來已久，例子很多。可參錢鍾書《管錘編》第一冊「四六、襄公二十四年斷章取

義」(1979, 224)。第二,嚴復對三曰的掌握根本不扎實。其舊學根基屢被批評。饒宗頤評嚴復之讀《史記》云:「又陵治古書,不涉訓詁,不免皮傳。」(2003, 105)羅豐祿(1850-1903)更說,「嚴又陵文章何足數,不過熟習史漢,試舉十三經中能精通一經乎?」(黃克武 2005, 11-13)。有關三曰與三難關聯的詳細分析,請見拙文(張宇傑 2016)。

「脩辭立誠」出《易·乾文言》。《易·乾文言》:「九三曰。君子終日乾乾。夕惕若。厲。无咎。何謂也。子曰。君子進德修業。忠信。所以進德也。脩辭立其誠。所以居業也。」(阮元校勘 1980, 總 15)乾卦九三處境險厲。〈乾文言〉此段教人在野不為所用時處極位之道。有關〈乾〉九三的象數分析,詳參李道平撰,潘雨廷點校(1994, 30-32)。

脩,通「修」。「修」在經文中屢作「脩」。修、飾也。辭,文辭,統言言語、文章。《說文》卷三上言部:「誠。信也。」(許慎撰,徐鉉校定 1963, 52 下)「信。誠也。」(許慎撰,徐鉉校定 1963, 52 上)「誠」、「信」二字轉注。嚴復以「修辭立誠」以證「信」之重要即此,與譯文要信實關聯不大。

[35] 《論語·衛靈公》:「子曰。辭達而已矣。」孔曰:「凡事莫過於實。辭達則足矣。不煩文豔之辭。」(阮元校勘 1980, 總 2519)朱熹注曰:「辭取達意而止。不以富麗為工。」(1936, 69)。清儒王夫之《讀四書大全說》:「『達』有兩義,言達其意而意達於理也。」(1975, 447)「辭達而已矣」並非只謂辭不需要文采,而是「文」與「質」要相副。換言之,言辭上要體現儒家的中道,與嚴復以「達」為「傳達」的意思不合。

[36] 《左傳·襄公二十五年》:「(前略)文子曰。其辭順。犯順不祥……仲尼曰。《志》有之。言以足志。文以足言。不言誰知其志。言之無文,行而不遠。晉為伯鄭入陳。非文辭不為功。慎辭也。」(阮元校勘 1980, 總 1985)《正義》曰:「足猶成也。」「慎辭哉」下云「樞機之發。榮辱之主。」(總頁同)孔穎達疏曰:「《傳》言子產善為文辭於鄭有榮也。」(總頁同)引文所說的是行文要修飾才能行遠,但「雅」則指遣詞要正統,忌用當時新興的「利俗文字」。引文與「雅」亦不對應。

[37] 《左傳·襄公二十四年》:「大上有立德。其次有立功。其次有立言。雖久不廢。此之謂不朽。」(阮元校勘 1980, 總 1979)義證:顧炎武〈初刻日知錄自序〉:「蓋天下之理無窮。而君子之志於道也。不成章不達。」(顧炎武著,黃汝成集釋 1936, 1)《日知錄》卷十九〈立言不為一時〉:「天下之事。有言在一時。而其效見於數十百年之後者。」(同上, 448))嚴復〈題侯疑始填詞圖冊〉:「心聲精者傳。韻語亦天演。」(嚴復 1977, 3)

[38] 三曰:這三句引文。這三句話與三難的關聯其實不強。詳參注 34-36。

[39] 正軌:正途。

[40] 爾雅:近正。詳見注 5。

[41] 吳汝綸序:「抑汝綸之深有取是書。則又以嚴子之雄於文。以為赫胥黎氏之指趣。得嚴子乃益明。凡吾聖賢之教,上者,道勝而文至。其次,道稍卑矣,而文猶足以久。獨文之不足,斯其道不能以徒存。」又曰:「今學者方以時文、公牘、說部



為學，而嚴子乃欲進之以可久之詞。」（嚴復 1898/1995，1 下）

- [42] 吳汝綸序：「吾則以謂今西書之流入吾國，適當吾文學靡敝之時，士大夫相矜尚以為學者，時文耳、公牘耳、說部耳。」（同上）「利俗文字」大抵是吳指的時文、公牘、說部（小說、筆記等）。時文即八股文。嚴復與梁啟超論自己所譯之《原富》曰：「吾譯正以待多讀中國古書之人……慕藏山不朽之名譽，所不必也。苟然為之，言龐意纖，使其文之行于時，若蜉蝣旦暮之已化。此報館之文章，亦大雅所諱也。」（汪征魯等 2014，121）一九一八年十月二日，嚴復致四子嚴璿家書云：「中文教習所出之題，自是時式，無怪吾兒詫為未見。須知時下報館文章，什九皆此類也。兒不知題目中『研習』二字，在教員不過用為『誦讀』二字之替代，以為教覺新鮮文明，合於維新之教育，乃不謂吾兒將此二字認真……見調膚詞，全行勾抹。」（汪征魯等 2014，532）

民國八年（一九二〇年）七月廿日，嚴復致嚴璿家書：「至於國文……與昔賢規矩，及兒在書房所習者，大不相侔。我們既入學校，而國文分數，又有升班關係，自不得不勉強從俗，播弄些新名詞之類，依教員所言，繳卷塞責。」（南洋學會研究組編 1959，156）由這兩封信可見利俗文字還指「報館的文章」，即當時「維新」而冒起的新名詞。

嚴復序《馬氏文通要例啟蒙》曰：「吾黨舒城陶散生，通中西文，得是書，大通其說，乃以近俗語言為發凡例。蓋《文通》者，說文字言語之原則公例者也……使其信於古文辭而異於近俗言語者，非《文通》矣。」（孫應祥、皮后鋒編 2004，164）

- [43] 閒：今天「間」的正字。

[44] 釣奇：謀取巨利。（羅竹風主編 1993，1205）此處引伸為「擡高譯文的價值」。

[45] 不佞：自謙詞，猶不才、不敏，是先秦習語。（詳見楊伯峻 1991，865）

[46] 艱深文陋：用艱深深奧的文字文飾淺陋的譯文。蘇軾〈答謝民師書〉：「揚雄好為艱深之辭，以文淺易之說。」（蘇軾 1936，621）

[47] 格致：物理、化學等自然科學。《禮記·大學》：「古之欲明明德於天下者……先修其身。欲修其身者，先正其心。欲正其心者，先誠其意。欲誠其意者，先致其知。致知在格物。」（阮元校勘 1980，總 1673）統言之，「格物致知」指了解事物的道理。【案析言之指做事處世要把握中庸之道。】物理、化學等探研自然世界的規律，故稱「格致之學」。

[48] 疇人之學：天文曆算。《漢語大詞典》第七卷「疇人」下云：「古代天文歷算之學，有專人執掌，父子世代相傳為業，稱為『疇人』。亦指精通天文歷算的學者。」（羅竹風主編 1991，1406）

[49] 於之數者：在這幾方面。王引之《經傳釋詞》卷九：「之，是也。」（1966，198）

[50] 問津：尋訪探求。陶淵明〈桃花源記〉：「南陽劉子驥。高尚士也。聞之。欣然規往。未果。尋病終。後遂無問津者。」（遼欽立校注 1979，166）

[51] 矧：況且。粵音「診」；普音：shěn。

[52] 《漢書》卷十二「平帝紀第十二」：「元始元年春正月，越裳氏重譯獻白雉一，黑雉

- 二，詔使三公薦宗廟。」（班固撰，顏師古注 1962，348）師古曰：「越裳，南方遠國也。譯謂傳言也。道路絕遠，風俗殊隔，故累譯而後乃通。」（同上，349）「重譯」即翻譯學之 *relay translation*。惟《天演論》乃嚴復直接由英文譯成中文。此處當指翻譯而已。
- [53] 新理踵出：新的理論湧現。《說文》卷二下足部：「踵。追也。」（許慎撰，徐鉉校定 1963，46 下）卷二下辵部：「追。逐也。」（許慎撰，徐鉉校定 1963，41 下）
- [54] 「衡」、「量」皆讀平聲。《書·舜典》：「同律度量衡。」（阮元校勘 1980，總 127 中）「衡」是秤。《說文》卷八上重部：「量。稱輕重也。」（許慎撰，徐鉉校定 1963，169 下）
- [55] 即，就也。靠近的意思。
- [56] 乃：於是。
- [57] 僕：我。自謙詞。
- [58] 繙：即「翻」。過去「翻譯」常作「繙譯」。
- [59] 卮言：謙詞，謂自己的話支離破碎，不成系統。《莊子·雜篇·寓言》：「卮言日出。和以天倪。」卮，音「之」。清郭慶藩《莊子集釋》：「《字略》云：『卮：圓酒器也。』王云：『夫卮器，滿即傾，空則仰，隨物而變，非執一守固者也。施之於言，而隨人從變，已無常主者也。』司馬云：『謂支離無首尾言也。』」（郭慶藩撰，王孝魚點校 1961，948）
- [60] 夏曾佑（1863-1924），今浙江杭縣人，著有《中國古代史》。杜維運教授有〈夏曾佑傳〉，參杜維運（1994，125-133）。張豈之主編《民國學案》第二卷亦有「夏曾佑學案」（2005，55-79）。
- [61] 內典：佛經之謂。
- [62] 吳汝綸（1840-1903），字摯甫，清安徽桐城人。「父」「甫」通。《清史稿·文苑三》（趙爾巽等撰 1942，1543 下-1544 上）有傳。《清儒學案》卷一八九〈摯甫學派〉下云：「自望溪倡古文義法，劉、姚繼之，桐城一派遂為海內正宗，綿延二百年，而摯甫為之殿。」（徐世昌等編，沈芝盈、梁運華點校 2008，7283）或以為嚴復拜吳氏習桐城古文。實則嚴僅私淑於吳汝綸而已。詳見黃克武（2005，25）。
- [63] 釋氏：即佛教。佛教以釋迦牟尼為宗，故亦云釋氏。
- [64] 諸子：《漢書·藝文志》卷三〈諸子略〉：「諸子十家，其可觀者九家而已。皆起於王道既微，諸侯力政，時君世主，好惡殊方，是以九家之術，蠡出並作，各引一端，崇其所善，以此馳說，取合諸侯……今異家者各推所長，窮知究慮，以明其指，雖有蔽短，合其要歸，亦六經之支與流裔。使其人遭明王聖主，得其所折中，皆股肱之材已」（班固編撰，顧實講疏 1984，166-167）。戰國時政局動盪，不同思想家各自提出安定天下之計，冀各國君主接納。較著名的有《莊子》、《老子》、《列子》、《韓非子》等。諸子在傳統文獻四部分類【案經、史、子、集】中自居一部，即第三部「子部」。余嘉錫《古書通例》卷一：「古書多摘首句二字以題篇……顧炎武曰：『《三百篇》之詩人，大率詩成，取其中一字二字三四字以名篇（後畧）。』……古人之著書作文，亦因事物之需要，而發乎不得不然，未有先

命題，而強其情與意曲折以赴之者。故《詩》《書》之篇名，皆後人所題。諸子之文，成於手著者，往往一意相承，自具首尾，文成之後，或取篇中旨意，標為題目。」(1985, 28) 案可見不惟諸子隨篇標目，其他先秦古書亦然。

[65] 同「懸談」。

[66] 乃：是。

[67] 質：質樸，與「文」相對。「質譯導言」，謂簡單地譯為「導言」。

[68] 誚：責讓。

[69] **Struggle for existence**。詳見〈天演論中西名表〉，載嚴復(1930a, 1)。上海商務印書館有「嚴譯名著叢刊」，收嚴譯八種，《天演論》乃其一，一九三〇年初版。〈天演論中西名表〉附書後。

[70] 即 **Selection**。參〈天演論中西名表〉(嚴復 1930a, 1)。

[71] 商務印書館「嚴譯名著叢刊例言」曰：「嚴先生之譯名，為力求典雅故，多為讀者所不能明瞭，且與近日流行之譯名不盡同，本叢刊在每冊之末，均附有譯名對照表，一面將原文列出，一面將近日流行之名詞，附列於後，使讀者易於明瞭。」見商務印書館編譯所：〈嚴譯名著叢刊例言〉，載嚴復(1930a, 頁闕)。

[72] 旬月：一個月。

[73] 踟躕：即猶豫。

[74] 我罪我知：即「罪我知我」。罪我：怪責我(不以正法翻譯)。知我：明白我(不依正法翻譯的苦衷)。《孟子·滕文公下》：「春秋。天子之事也。是故孔子曰。知我者。其惟春秋乎。罪我者。其惟春秋乎。」(阮元校勘 1980, 總 2714)

[75] 以存明哲：猶明哲保身。《書·說命》【「說」，同「悅」】：「知之曰明哲。明哲實作則。」(阮元校勘 1980, 總 174)

[76] 名碩：著名學者。

[77] 流風：流下來的風氣。《孟子·公孫丑上》：「紂之去武丁。未久也。其故家遺俗。流風善政。猶有存者。」(阮元校勘 1980, 總 2684)

[78] 緒論：餘論。《楚辭·九歌·涉江》：「乘鄂渚而反顧兮。欸秋冬之緒風。」王注：緒，餘也。(洪興祖撰，白化文等點校 1983, 129)

[79] 泰西：猶極西，舊指西方歐美各國。

[80] 茲：現在。

[81] 知人論世：要了解某人便要知道其身處的時代環境。《孟子·萬章下》：「頌其詩。讀其書。不知其人可乎。是以論其世也。是尚友也。」(阮元校勘 1980, 總 2746)

[82] 窮理：考究事物的道理。

[83] 輒：總是、就。

[84] 譎陋：淺陋。譎，淺也。

[85] 《詩·小雅·伐木》：「嚶其鳴矣。求其友聲。」(阮元校勘 1980, 總 410)《正義》曰：「君子雖遷於高位，不可以忘其朋友。」(阮元校勘 1980, 總 410)

[86] 麗澤之義：即朋、友【同門為朋，同志曰友】互相切磋之意。「麗澤」一詞出《易·兌·象》：「麗澤兌。君子以朋友講習。」麗，偶也。平聲，音離。陳湛銓《周易講

- 疏》：「朋友講習。以仁義道藝名節相引導。則悅樂無極。無絲毫之害。故兌象悅懌。以朋友講習闡揚之。」(2014, 41-42)
- [87] 俟：等待。
- [88] 固：固守。
- [89] 標高揭己：韓愈〈祭河南張員外文〉：「君德渾剛。標高揭己。有不吾如。唾猶泥滓。」閻琦注曰：「以上四句謂署【即張員外】個性耿直，胸不能容物。標高揭己，猶言高自標置。揭，高舉也。」(閻琦校注 2004, 470)
- [90] 懷鉛握槧：「即謂攜帶筆簡，以備隨時記錄、著述。鉛，鉛粉；槧，木板。皆古代書寫用具。」(羅竹風主編 1991, 785) 槧：音暹。
- [91] 逡譯：即翻譯。《說文》卷二下辵部：「逡。遷徙也。」(許慎撰，徐鉉校定 1963, 40 下)「逡」今多作「移」。「逡」是正字。
- [92] 理學：即「理科」，與「文科」相對，指物理、化學、數學等科，與宋明理學無涉。《說文》卷一上玉部：「理。治玉也。」(許慎撰，徐鉉校定 1963, 12 上)《說文解字注》一篇上：「理。治玉也……是理為剖析也。玉雖至堅。而治之得其觥理以成器不難。謂之理……在物之質曰肌理。曰腠理。曰文理。得其分則有條而不紊、謂之條理。」(段玉裁 1981, 15 下-16 上)物理、化學、數學等科目探索自然界的定律，古云「理科」。
- [93] 繙轉不易：不容易翻譯。
- [94] 固：通「故」，於是。
- [95] 迨，即「逮」，及也。《說文》卷二下辵部：「唐逮。及也……臣鉉等曰：或作迨。」(許慎撰，徐鉉校定 1963, 40 下)全句意為「及至書譯完後」。
- [96] 好：喜歡，動詞，去聲。
- [97] 惟：想到。泛思曰惟。《說文》卷十下心部：「惟。凡思也。从心佳聲。」(許慎撰，徐鉉校定 1963, 218 下)
- [98] 蹟：幽深難見。《周易·繫辭上》：「聖人有以見天下之蹟。而擬諸其形容。」(阮元校勘 1980, 總 79)
- [99] 叩寂：叩拜寂寞，猶冷門之意。陸機〈文賦〉：「課虛無以責有。叩寂寞而求音」(蕭統選，李善注 1936, 351)。
- [100] 亟：急。《說文》卷十三下二部：「亟。敏疾也。从人从口，从又从二。二，天地也。」(許慎撰，徐鉉校定 1963, 286)亟、急，一聲之轉。當務之所亟即當務之急。《孟子·盡心上》：「知者無不知也。當務之為急。仁者無不愛也。急親賢之為務。」(阮元校勘 1980, 總 2771)全句意謂「付印《天演論》不是當前所急於的」。
- [101] 父：當作「公」。梁任公，即梁啟超。
- [102] 付梓：付印。古時雕版刻書以梓木為上，後因稱書籍刊印為「付梓」。(羅竹風主編 1986, 1127)
- [103] 介弟：尊稱別人之弟。鄂：湖北。
- [104] 棗梨：謂雕版印刷。舊時多用棗木或梨木雕刻書版，故稱。(羅竹風主編 1989,

1108) 災棗梨：自謙辭。謂自己付刻的書沒有價值，浪費了棗木梨木。成語有「災梨禍棗」。

[105] 斟：通「校」。校對。

[106] 識：即「誌」。「意」、「志」、「識」三字古通。「志」作動詞時，今多作「誌」。《三國志》、《漢書·藝文志》、蒲松齡《聊齋志異》皆作「志」。

[107] 光緒二十四年，西曆一八九八年。是年（戊戌年）有「戊戌變法」，即「百日維新」。事敗，譚嗣同等「戊戌六君子」被殺。同年，北京大學前身京師大學堂成立。三年前，即一八九五年，中日甲午戰爭。中國敗，簽《馬關條約》。

[108] 味經本的刊印年份頗具爭議。湯志鈞推斷這個版本實則刻於 1897 年或 1898 年，1898 年印出。參湯志鈞（2000，75）。茲姑錄如上，以備一說。

## 參考文獻

- Fong, Gilbert. 1995. "Literary Translation: Translated Literature in Pre-Modern China." In *An Encyclopaedia of Translation: Chinese-English • English-Chinese*, edited by Chan Sin-wai and David Pollard, 580-590. Hong Kong: The Chinese University Press.
- Hsu, C.Y. 1973. "General Remarks on Translation." In *Renditions* (1): 4-6.
- Huang, Yushi. 1995. "Form and Spirit." In *An Encyclopaedia of Translation: Chinese-English • English-Chinese*, edited by Chan Sin-wai and David Pollard, 277-287. Hong Kong: The Chinese University Press.
- Liu, Miqing. 1995. "Aesthetics and Translation." In *An Encyclopaedia of Translation: Chinese-English • English-Chinese*, edited by Chan Sin-wai and David Pollard, 1-13. Hong Kong: The Chinese University Press.
- Ma, Zuyi. 1995. "History of Translation in China." In *An Encyclopaedia of Translation: Chinese-English • English-Chinese*, edited by Chan Sin-wai and David Pollard, 373-387. Hong Kong: The Chinese University Press.
- Sinn, Elizabeth. 1995. "Yan Fu." In *An Encyclopaedia of Translation: Chinese-English • English-Chinese*, edited by Chan Sin-wai and David Pollard, 429-447. Hong Kong: The Chinese University Press.
- Snell-Hornby, Mary. 1988/1995. *Translation Studies: An Integrated Approach*. Amsterdam; Philadelphia: John Benjamins.
- Venuti, Lawrence. 1998. *The Scandals of Translation: Towards an Ethics of Difference*. London; New York: Routledge.
- Wang, Zuoliang. 1995. "Translation standards." In *An Encyclopaedia of Translation: Chinese-English • English-Chinese*, edited by Chan Sin-wai and David Pollard, 999-1003. Hong Kong: The Chinese University Press.
- Wu Jingrong. 1995. "Chinese-English Dictionaries." In *An Encyclopaedia of Translation: Chinese-English • English-Chinese*, edited by Chan Sin-wai and David Pollard, 519-532. Hong Kong: The Chinese University Press.

- 牛仰山選注。2006。《嚴復文選》。天津：百花文藝出版社。
- 王夫之。1975。《讀四書大全說》。北京：中華書局本。
- 王引之。1966。《經傳釋詞》。香港：太平書局。
- 王栻主編。1986。《嚴復集》。北京：中華書局。
- 王蘧常。1936。《嚴幾道年譜》。上海：商務印書館。
- 朱熹。1936。《四書章句集注》。上海：世界書局。
- 朱熹。1961。《詩集傳》。香港：中華書局。
- 余嘉錫。1985。《古書通例》。上海：上海古籍出版社。
- 李道平撰，潘雨廷點校。1994。《周易集解纂疏》。北京：中華書局。
- 杜維運。1994。〈夏曾佑傳〉。《歷史的兩個境界》125-133。臺北：東大圖書公司。
- 汪征魯、方寶川、馬勇主編。2014。《嚴復全集》。福州：福建教育出版社。
- 沈蘇儒。1998。《論信達雅：嚴復翻譯理論研究》。臺北：商務印書館。
- 阮元校勘。1980。《十三經注疏：附校勘記》。北京：中華書局影印。
- 周振甫。1959。《嚴復詩文選》。北京：人民文學出版社。
- 南洋學會研究組編。1959。《嚴幾道先生遺著》。新加坡：南洋學會。
- 段玉裁。1981。《說文解字注》。第二版。上海：上海古籍出版社。
- 洪興祖，白化文等點校。1983。《楚辭補注》。北京：中華書局本。
- 孫應祥、皮后鋒編。2004。《嚴復集補編》。福州：福建人民出版社。
- 徐世昌等編，沈芝盈、梁運華點校。2008。《清儒學案》。北京：中華書局。
- 班固撰，顏師古注。1962。《漢書》。北京：中華書局。
- 班固編撰，顧實講疏。1984。《漢書藝文志講疏》。上海：上海古籍出版社。
- 張豈之主編。2005。《民國學案》。長沙：湖南教育出版社。
- 張宇傑。2016。嚴復《天演論·譯例言》譯事三難與「三曰」之關聯探微。《亞太跨學科翻譯研究（第二輯）》(2): 50-65。北京：清華大學出版社。
- 張宇傑。2019。〈「雅」的真諦〉。1月2日，《經濟通 ET NET》，[www.etnet.com.hk/w/wt/lifestyle/internationalaffairs/lingnan/57824](http://www.etnet.com.hk/w/wt/lifestyle/internationalaffairs/lingnan/57824)
- 許慎撰，徐鉉校定。1963。《說文解字：附檢字》。北京：中華書局。
- 郭慶藩撰，王孝魚點校。1961。《莊子集釋》。北京：中華書局。
- 陳湛銓。2014。《周易講疏》。香港：商務印書館。
- 陳錚編。2005。《黃遵憲全集》。北京：中華書局。
- 湯志鈞。2000。再論康有為與今文經學。《歷史研究》(6): 72-79。
- 遼欽立校注。1979。《陶淵明集》。北京：中華書局。
- 黃克武。2005。走向翻譯之路：北洋水師學堂時期的嚴復。《中央研究院近代史研究所集刊》(49): 1-40。
- 楊伯峻。1991。《春秋左傳注》[修訂本]。北京：中華書局。
- 趙爾巽等撰。1942。《清史稿》。出版城市闕：聯合書店。
- 劉熙。1985。《釋名》。北京：中華書局。
- 劉勰。1937。《文心雕龍》。上海：商務印書館。

- 鄧又同編。1989。《香港學海書樓陳湛銓先生講學集》。香港：學海書樓。
- 蕭統選，李善注。1936。《文選》。香港：商務印書館。
- 錢鍾書。1979。《管錐編》。北京：中華書局。
- 閻琦校注。2004。《韓昌黎文集注釋》。西安：三秦出版社。
- 羅竹風主編。1986。《漢語大詞典》(第一卷)。上海：上海辭書出版社。
- 羅竹風主編。1991。《漢語大詞典》(第七卷)。上海：上海辭書出版社。
- 羅竹風主編。1993。《漢語大詞典》(第十一卷)。上海：漢語大辭典出版社。
- 羅竹風主編。1989。《漢語大詞典》(第四卷)。上海：上海辭書出版社。
- 嚴復。1898/1995。《天演論》[慎始基齋刻本(1898年)]。載《續修四庫全書》，第1297冊。上海：上海古籍出版社。
- 嚴復。1930a。《天演論》[嚴譯名著叢刊]。上海：商務印書館。
- 嚴復。1930b。《原富》。上海：商務印書館。
- 嚴復。1977。《瘡懣堂詩集》。臺北：臺灣商務印書館。
- 蘇軾。1936。《蘇東坡全集》。上海：世界書局。
- 饒宗頤。2003。〈書《清史稿·文苑傳》後〉。編輯委員會著編。《饒宗頤二十世紀學術文集》，第二十冊，105。臺北：新文豐。
- 顧炎武著，黃汝成集釋。1936。《日知錄集釋》。上海：世界書局。

# 漢語可比語料庫中名詞化標記的分佈特徵差異及其原因探析

戴光榮

**Address:** School of Interpreting and Translation Studies, Guangdong University of Foreign Studies, China

**E-mail:** carldy@163.com

**Correspondence:** Guangrong DAI

**Citation:** DAI, Guangrong. 2022. "A Study on Distributional Difference of Nominalization Markers in Chinese Comparable Corpora and Its Causes." *Translation Quarterly* 105: 95-111.

## ***Abstract***

A Study on Distributional Difference of Nominalization Markers in Chinese Comparable Corpora and Its Causes (by Guangrong Dai)

*The phenomenon of nominalization encompasses a wide range of aspects and has garnered academic attention, albeit rarely discussed in research on translated language. This paper delves into four Chinese nominalization markers (NMs) within a Chinese comparable corpus, primarily focusing on the distributional differences of NMs between the two corpora. It endeavors to identify the underlying rationales behind the prevalence of nominalization in translated Chinese. The results reveal that the translated Chinese corpus exhibits a higher frequency of NMs and a broader range of NMs' collocations compared to the native Chinese corpus. Furthermore, the grammatical categorization of verbs and adjectives in translated Chinese becomes intricate after undergoing the process of nominalization. Nominalization serves to fulfill distinct textual and genre functions by means of the internal integration of clauses, thereby rendering the text more refined and objective, while also retaining the professionalism of the texts.*

## 一、引言

漢語語法的最大特點是沒有嚴格意義上的形態變化（呂叔湘 1999，7），英語中的名詞單複數之分可以通過後綴來區分，動詞的不同時態、語態等可以通過不同的形態變化得到體現，形容詞與副詞比較級及最高級也可以通過形態變化得以彰顯。漢語中的詞類辨別，離開上下文語境，就沒有英語詞類那樣簡單，同一個“管理”，既可以



做名詞，如“企業的管理”，又可以做動詞“管理各類事務”。我們經常面臨這樣的語言現象：處於主賓語位置上的動詞、形容詞，已經失去了動詞與形容詞的主要功能，即認為主語賓語位置上的動詞和形容詞具有名詞的性質，就是一種名詞化。人類語言的一個普遍性是，動詞或者形容詞通過某種語法手段可以轉化為名詞（石毓智 2010, 259）。本文在前賢研究的基礎上，依據朱德熙（1983）的研究，關注漢語中四個名詞化標記（的、所、者、之），通過檢索它們在漢語可比語料庫的使用情況，展開全面分析，重點探討三個問題，其一，漢語名詞化標記在翻譯漢語語料庫中的使用特徵，是否不同于原創漢語語料庫中的使用特徵；其二，可比語料庫中漢語名詞化使用差異主要體現在哪些方面，有何規律性特徵；其三，漢語譯文中名詞化高頻使用的原因探析。

## 二、名詞化定義及其實現方式

名詞化作為一種跨範疇操作，指的是從非名詞結構中派生出名詞的語法過程 (Comrie and Thompson 1985, 349)。名詞化有多種類型，如動詞或動詞短語轉化為名詞 (refuse → refusal; arrive → arrival; receive → reception; criticise → criticism)，形容詞或形容詞短語轉化為名詞（如 red → redness; true → truth），小句派生出名詞短語（如 She answered the letter. → Her answering of the letter.）。這些語法過程，都可稱之為名詞化 (Crystal 2008, 328; Li and Thompson 1981, 75; Quirk et al. 1985, 1288)。

每一種語言都存在名詞化現象，不同的語言，其名詞化實現的方式多種多樣。相比漢語來說，英語中的名詞化現象更為普遍，學界也從多方面對英語中的名詞化現象進行了探討，如結構語言學、轉換生成語法、功能語言學、認知語言學 (Baker and Vinokurova 2009; Bierwisch 2009; Billig 2008; Halliday 1994/2000; Langacker 1991; Martin 2008; Noonan 1997; Salanova 2007; 王立非、劉英傑 2011; 張高遠 2008)。

有許多學者對印歐語系以外語言中的名詞化現象進行了探討，比較典型的研究如 Yap and Grunow-Hårst (2010) 關於亞洲語言名詞化用法的探討，結合類型學和歷時研究的觀點，追溯了名詞化結構如何隨著時間的推移從指稱用途發展到非指稱用途，後者包括修飾功能（如關係化和狀語化）；在某些情況下，名詞化結構進一步發展為限定從句，有時被重新分析為獨立的結構，具有鏡像、證據、認知、態度或其他說話者情緒立場的含義。這些發展構成了通用構式從命題性用途擴展到語法和語用用途的普遍趨勢的一部分 (Yap and Grunow-Hårsta 2010)。Yap, Grunow-Hårsta and Wrona (2011) 分析了亞洲約 60 種語言中的名詞化現象，包括漢藏語系內的漢語 (Sinitic languages 古漢語與現代漢語、地方方言如粵語等)、藏緬語 (Tibeto-Burman languages 如達芒語、景頗語)、伊朗語 (Iranian languages)、韓語與日語 (Korean and Japanese languages)、南島語 (Austronesian languages)、巴布亞語 (Papuan languages)，他們的研究將名詞化大致分為形態、句法、語義、語用和歷時等五大邏輯系統 (Gerner 2012)。

漢語名詞化的研究起步也很早（參閱黎錦熙 1924/2007；黎錦熙、劉世儒 1960；朱德熙等 1961），研究成果也較為豐富，主要集中于對原創漢語的研究，對漢語翻譯

文本中的名詞化現象探討不多（戴光榮、左尚君 2017）。

### 三、漢語名詞化及其標記

漢語中是否存在名詞化現象，學界對此進行了很長時間的探討。早期主張名詞化的認為，從意義上看，主賓語位置上的動詞、形容詞由行為、性狀轉為事務；這些動詞形容詞具有名詞的語法特點（如可受定語修飾、與名詞組合為聯合結構等）；這些動詞形容詞失去了原為動詞形容詞的一部分或全部語法特點（黎錦熙、劉世儒 1960），因此漢語是存在名詞化現象的（高名凱、計永佑 1963）。也有學者認為漢語不存在名詞化現象，如朱德熙、盧甲文、馬真（1961）認為“漢語的語言事實完全不支援名物化的說法”，但是這些觀點隨著時間的推移而發生了變化，朱先生在後來的研究中，也認為“漢語的動詞、形容詞本身可以做主賓語，也可以名詞化以後做主賓語”（朱德熙 1983）。

在漢語中，名詞化包括在動詞、動詞短語、句子或包括動詞在內的句子的一部分後面加上“的”（Li and Thompson 1981, 575；石毓智 2010），比如“吃的”可以指食物，“做研究的”可以指從事研究工作的人。除了詞彙層面的名詞化，漢語中也存在句法層面的名詞化（陸燦、潘海華 2013；石定栩 2004, 2011；徐新宇 2011）。本文重點探討詞彙層面的名詞化現象，即朱德熙（1983）一文中所分析“的、所、者、之”四個名詞化標記，分析它們在漢語可比語料庫中的使用情況。

根據尚忠華等人的《現代漢語頻率詞典》統計（Xiao, Rayson and McEnery 2009）發現，這四個名詞化標記在頻率詞典中的排序非常靠前，其歸一化頻率、原始使用頻率及離散係數也非常高，如表 1 所示，“的”排名第一，“之”排名 84，“所”排名 113，“者”排名 271，可以說這四個詞在現代漢語中的顯示度非常高。

表 1: 四個名詞化標記在頻率詞典中的相關統計資料

	排序	歸一化頻率（每百萬詞）	離散係數	使用頻率
的	0001	236106	0.95	223484
之	0084	5125	0.77	3951
所	0113	3646	0.8	2906
者	0271	1795	0.67	1211

我們來看看這四個名詞化標記各自不同的用法。

根據現有的研究，現代漢語施事名詞化標記“的”是通過類比古漢語名詞化標記“者”而發展起來的（Yap and Wang 2011, 66）。由中國社會科學院語言研究所詞典編輯室編纂的《現代漢語詞典》（第六版，2012），將“的”的詞性標注為“助詞”，主要用法有 6 種，其中特別標注其作為名詞化標記的用法：用來構成沒有中心詞的“的”字結構，代替上文所說的人或物（“這是我的，那才是你的”），指某一種人或物（“送報的、我愛吃辣的”），表示某種情況、原因等。在《現代漢語八百詞》中，“的”可以構成“的”字短語用來修飾名詞。現代漢語中，除連詞、助詞、嘆詞外，各種詞語都可

構成“的”字短語修飾名詞(呂叔湘 1999, 157)。我們根據前賢的研究,將“的”的用法結構歸納如下(表 2):

表 2: “的”結構分類及例證

類型	結構	例證
a	名 + 的 + 名	朋友的飛機票、集體的力量、民族的未來
b	動 + 的 + 名	唱的歌、研究的問題、下車的地點
c	形 + 的 + 名	美好的生活、新鮮的食材、堅強的戰士
d	副 + 的 + 名	歷來的習慣、暫時的困難、一貫的表現 (限於少數幾個雙音節副詞)
e	介詞短語 + 的 + 名	對問題的看法、關於文學的知識 (介詞限於‘對、對於、關於’)
f	象聲詞 + 的 + 名	當當的鐘聲、嗖的一個箭步
g	小句/四字語 + 的 + 名	你寄來的信、工業發展的速度、兩全其美的解決辦法

再來看漢語中“者”與“所”的用法。有研究表明,漢語中的輕名詞(light noun)“者”和“所”分別演化為施事和受事名詞化標記(Yap and Wang 2011)。

自西元前 700 年至西元前 100 年,古漢語名詞化標記“者”充當專門的施事名詞化標記,隨後擴大表達從屬功能,如“夫執與者為誰 Who is the one driving the carriage”。“者”作為一個多功能的輕量名詞,在語法上通過演變而成為一個施事名詞化標記,並逐步擴展到其他用法,通過名詞化以及副詞化表達不同的功能(含句末情緒標記)(Yap and Wang 2011, 78)。“者”在《現代漢語八百詞》中標記為“尾碼”,表示有某種信仰、從事某種工作或有某種特性的人,也可用於指代事物或人,構成名詞,用法如下表 3 所示:

表 3: “者”結構分類及例證(參閱呂叔湘 1999, 655-656)

類型	結構	例證
a	名 + 者	筆者、馬列主義者、醫務工作者
b	動/形 + 者	記者、讀者、消費者、強者、長者
c	前/後 + 者	在這兩個條件中,前者是主要的,後者是次要的
d	數 + 者	二者必居其一,三者都很重要

古漢語名詞化標記“所”是由地名名詞衍生而來的一種專門的受事名詞化標記。從古漢語的名詞“所”,通過語法化演變而成名詞化標記,可以看到名詞化、相對化和屬格化之間的聯繫,通過名詞化還衍生出來其他用途(包括說話人態度標記和被動語態標記)(Yap and Wang 2011, 99)。

在現代漢語中,“所”主要用在及物動詞之前,使“所 + 動”成為名詞性短語(多用於書面語),大致分為如下幾類:

- (1) 加“的”修飾名詞(被修飾的名詞在意念上是前面動詞的受事):我所認識的人、他所瞭解的情況、他是廣大觀眾所熟悉的一位老演員、所談的道理、所用的方法、所需的費用

(2) 加“的”代替名詞：我所知道的就是這些 (= 我所知道的事情)、他所說的未必確實 (= 他所說的話)、實驗結果同我們所預期的完全一致、他所談的不過是些生活瑣事、所考慮的正是這一點

(3) 不加“的”代替名詞：據我所知 (= 我所知道的情況)、果然不出我們所料、所見所聞、所論甚詳

現代漢語中的“之”來自古漢語中的代詞與結構助詞，大致包括如下幾種含義：

(1) 代替人或事物：影片情節十分悲慘，觀眾無不為之感動；對於任何膽敢來犯的侵略者，必須堅決、徹底、乾淨、全部殲滅之

(2) 用法大致與現代漢語中“的”字相當：赤子之心、光榮之家、公民有受教育和參加政治活動之權利

(3) 在一個小句的主語的謂語之間加進“之”字，使整個短語變為名詞性：這次技術革新運動範圍之廣泛，影響之深遠，都是前所未有的（呂叔湘 1999，672-674）。

## 四、漢語名詞化在可比語料庫中的分佈情況

為了更好地分析漢語中名詞化用法特徵，我們對上文探討的四個名詞化標記作為研究對象，對它們在現代漢語可比語料庫（LCMC 與 ZCTC）當中進行檢索。

有關蘭卡斯特漢語語料庫（LCMC）及浙大漢語譯文語料庫（ZCTC）的相關介紹，請參閱（尚忠華 2012；尚忠華、戴光榮 2010）。這兩個語料庫作為漢語可比語料庫，無論從語料取樣標準，還是語料標注方法，都是漢語研究的好幫手（LCMC 與 ZCTC 為各包括 100 萬詞的漢語平衡語料庫）。這兩語料庫的選材範圍及其比例構成如表 4 所示：

本節首先對四個名詞化標記在漢語可比語料庫的總體分佈情況進行檢索，然後針對其各自具體用法，分別進行分析，探討它們在可比語料庫中的使用特徵差異。

它們在漢語原創文本語料庫（LCMC）中的使用情況，可從圖 1 看出十五類不同文體中這些名詞化標記的使用頻率：

從圖 1 可以看出，四個名詞化標記在科學與學術文體中使用最多，傳記與散文類次之，其歸一化頻率（每十萬詞的使用頻率）結果有一定出入，學術文體仍然排在首位，新聞評論位居第二，如圖 2 所示：

接著來看漢語譯文文本語料庫（ZCTC）中名詞化標記的使用情況。圖 3 顯示四個名詞化標記在漢語譯文語料庫中的使用頻率：

其使用頻率結果顯示，漢語譯文語料庫中，學術文體使用最多，傳記與散文類次之，其歸一化頻率（每十萬的使用頻率）結果有一定差異，宗教文體排首位，學術文體位居第二，如圖 4 所示：

將漢語原創文本語料庫（LCMC）與漢語譯文語料庫（ZCTC）中名詞化標記的分佈情況做一個對比，其異同可以從如圖 5 所示的折線圖看得出來：

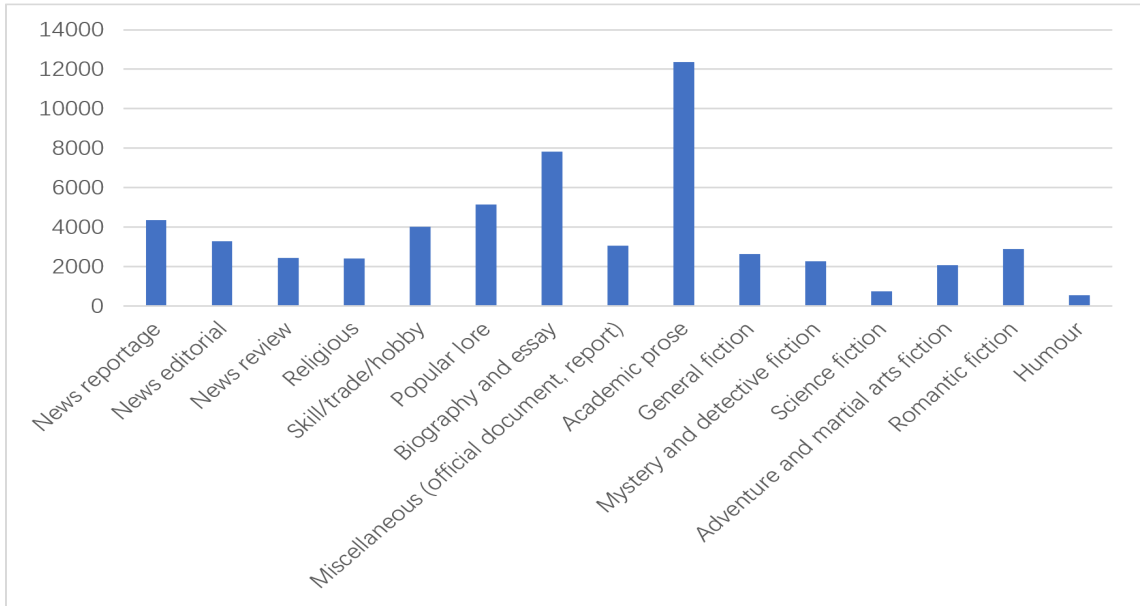


圖 1: LCMC 中十五類不同文體中名詞化標記的使用頻率

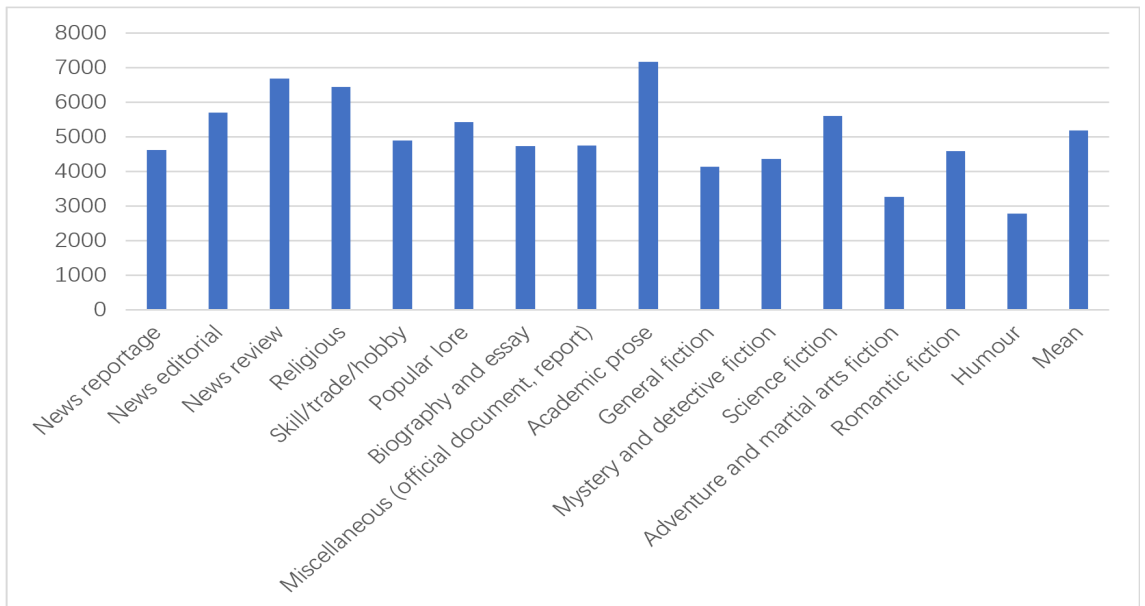


圖 2: LCMC 十五類不同文體中名詞化標記歸一化頻率（每十萬詞的使用頻率）

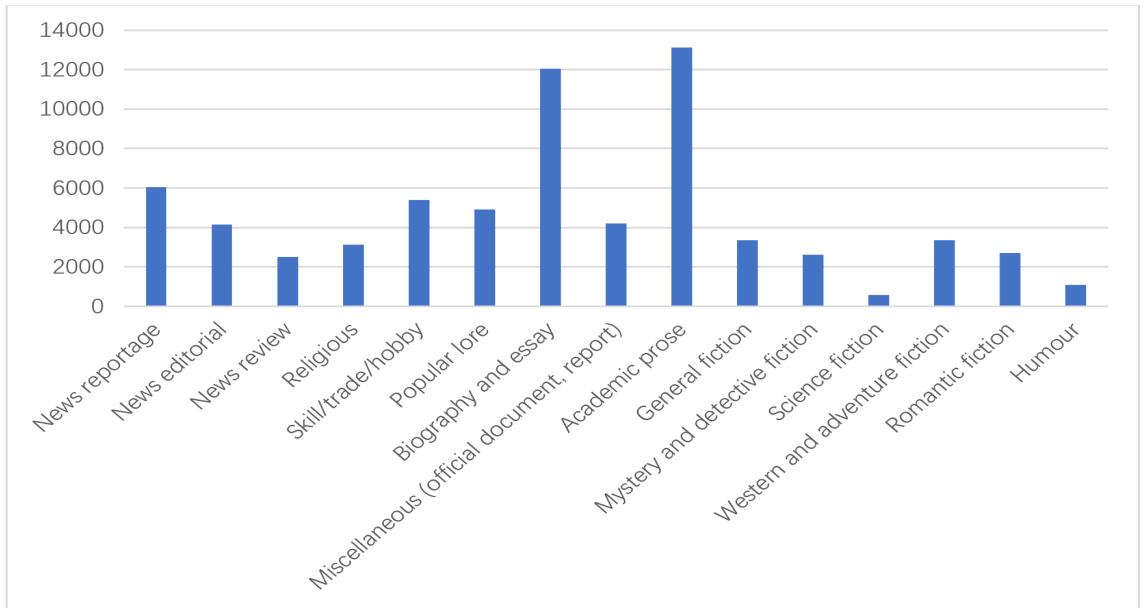


圖 3: ZCTC 十五類不同文體中名詞化標記的使用頻率統計圖

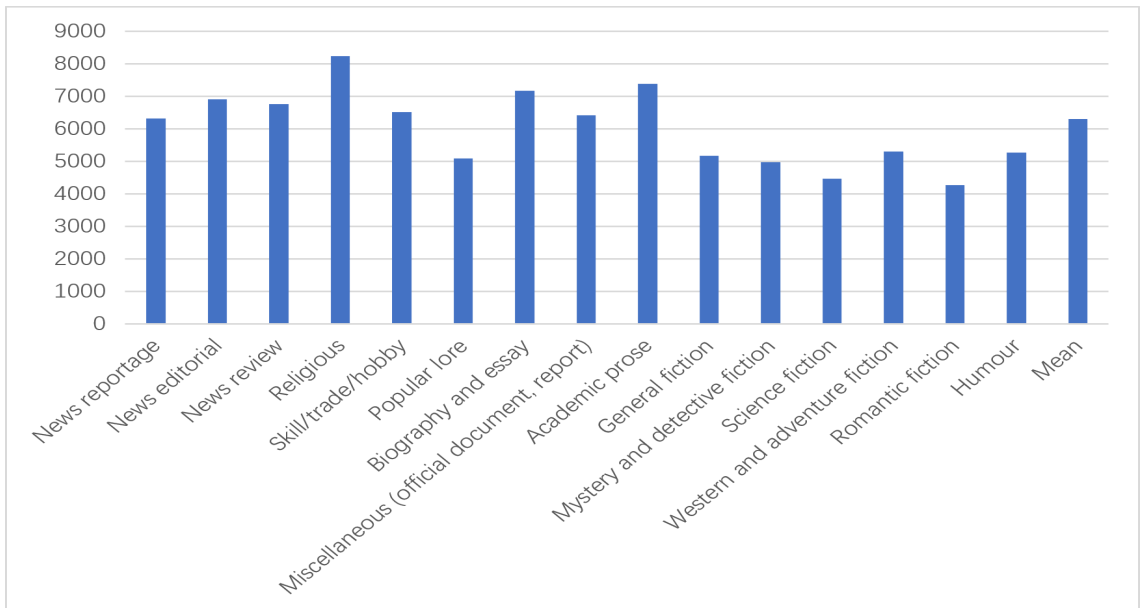


圖 4: ZCTC 十五類不同文體中名詞化標記歸一化頻率 (每十萬詞的使用頻率)

表 4: The genres covered in LCMC and ZCTC

Code	Genre	Number of samples	Proportion
A	Press reportage	44	8.8%
B	Press editorials	27	5.4%
C	Press reviews	17	3.4%
D	Religious writing	17	3.4%
E	Skills, trades and hobbies	38	7.6%
F	Popular lore	44	8.8%
G	Biographies and essays	77	15.4%
H	Miscellaneous (official documents, reports)	30	6%
J	Science (academic prose)	80	16%
K	General fiction	29	5.8%
L	Mystery and detective fiction	24	4.8%
M	Science fiction	6	1.2%
N	Adventure fiction	29	5.8%
P	Romantic fiction	29	5.8%
R	Humour	9	1.8%
	Total	500	100%

這些名詞化標記在兩個語料庫中的分佈折線圖走勢基本吻合，說明漢語翻譯文本與漢語原創文本中的名詞化使用情況比較相似，漢語譯文符合漢語原創語言的語言規範。如果對十五類文體中這四個名詞化標記的使用情況做一個詳細的分析，就可以看出可比語料庫中名詞化標記在不同文體中的差異，如表 5 所示：

從上表可以看出，在十五類文體中，除了 F(Popular lore)、M(Science fiction)、P(Romantic fiction) 三類之外，其餘 12 類文體中的名詞化標記在漢語譯文中的使用頻率均高於在漢語原創文本中的使用頻率。是什麼原因導致名詞化標記在漢語譯文中的使用頻率高於在漢語原創文本中的使用頻率，需要我們對這些標記進行分類探討。

## 五、名詞化標記在漢語可比語料庫中使用情況的分類探討

上文對可比語料庫中四個名詞化標記做了總體使用情況的分析。在檢索過程中我們發現，四個名詞化標記各自使用情況還是存在很大差異。本節對四個標記做分類探討。

先來看“的”結構在可比語料庫中的分佈情況，如表 6 所示：

從統計分析中可以看出，前三種結構“名+的+名、動+的+名、形+的+名”在漢語譯文語料庫中使用的頻率明顯高於漢語母語語料庫。我們來看看語料庫中該結構的相關例證：

1. 因此，應用電子信息技術改造傳統產業和國民經濟各部門，以節能、節材和提高產品附加值來推動國民經濟的發展，促進經濟信息化，將粗放型經濟轉變到集約型的軌道上來，是實現我國經濟迅速增長的關鍵，是應該

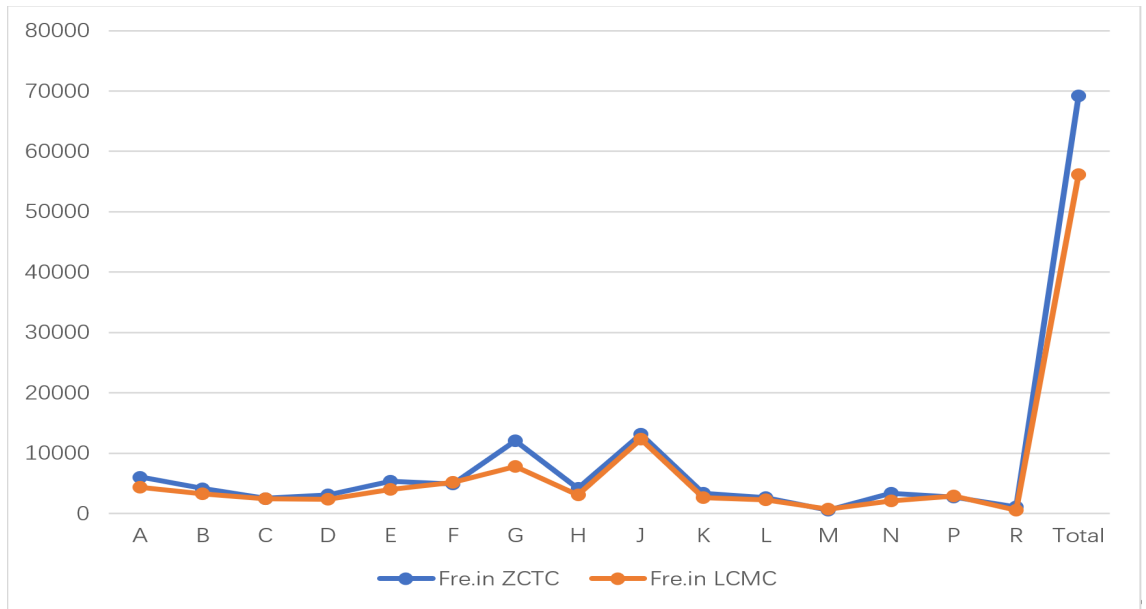


圖 5: 漢語可比語料庫名詞化標記分佈圖折線圖

長期堅持的方針。(LCMC)

2. 對於現代商品經濟，科學技術是起搏的心臟，是周身流動的血液。(LCMC)

3. 不是股市上升為發明創造帶來更多資金，而是股市下跌使新開公司可以得到的風險資本減少。(ZCTC)

4. 於是也就不難想像，戈爾會親自坐到電腦前，對總統的發言稿進行最後潤色，或是及時地敦促總統提前錄好廣播講話內容，以便安安穩穩地去大衛營度假。(ZCTC)

再來看“所”結構在可比語料庫中的分佈情況，如表 7 所示：

從上表資料可以看出，“所”結構在漢語譯文語料庫中使用的頻率明顯高於漢語母語語料庫。通過查閱漢語譯文語料庫中的相關例證，我們發現，“所”結構大多用於科學技術、學術文體、新聞評論等文體中，如：

5. 瑞典皇家學院決定將 2000 年諾貝爾物理學獎授予三位科學家和發明家，他們的工作，特別是他們所發明的快速電晶體、鐳射二級管和集成線路（晶片）奠定了現代資訊技術的基礎。(ZCTC)

6. 如果分子生物學家所作的基因研究正確的話，那麼這一發現已使科學家進入 400 至 600 萬年前的進化階段，猿類和人類就是在這一時期分道揚鑣的。(ZCTC)

7. 因此，從就業的角度來看，東南亞地區所需的是勞動力吸納型的技術，而非勞動力節約型的技術。(ZCTC)



表 5: 可比語料庫中不同文體名詞化標記使用分佈對比差異統計分析

	Freq. in LCMC	Freq. in ZCTC	Log-likelihood	Sig.	
A	4366	6040	243.16	0.000	***
B	3291	4147	84.92	0.000	***
C	2437	2509	0.19	0.664	
D	2405	3119	80.98	0.000	***
E	4009	5390	181.18	0.000	***
F	5154	4918	10.19	0.001	**
G	7834	12057	834.10	0.000	***
H	3071	4196	156.53	0.000	***
J	12361	13133	12.24	0.000	***
K	2637	3362	76.11	0.000	***
L	2276	2614	18.06	0.000	***
M	741	588	20.31	0.000	***
N	2084	3368	284.13	0.000	***
P	2901	2705	10.53	0.001	**
R	548	1080	168.29	0.000	***
Total	56115	69226	1163.17	0.000	***

表 6: “的” 結構在可比語料庫中的分佈情況

“的” 結構類型	Freq. in LCMC	Freq. in ZCTC	Log-likelihood	Sig.	
a(名 + 的 + 名)	8762	9280	7.46	0.006	**
b(動 + 的 + 名)	3212	4125	99.13	0.000	***
c(形 + 的 + 名)	4459	5758	144.57	0.000	***
d(副 + 的 + 名)	25	42	4.08	0.043	*
e(介 + 的 + 名)	112	123	0.35	0.556	
f(象聲 + 的 + 名)	9	15	1.42	0.234	

表 7: “所” 結構在可比語料庫中的分佈情況

“所” 結構類型	Freq. in LCMC	Freq. in ZCTC	Log-likelihood	Sig.	
a(名/代 + 所 + 動 + 的)	275	627	135.25	0.000	***
b(動 + 名/代 + 所 + 動)	3212	4125	99.13	0.000	***

8. 在中國，從任何角度所看到的變化都不是統計數字所能代表的。(ZCTC)

“者” 在 LCMC 中的使用頻率為 496，ZCTC 中為 676，漢語譯文語料庫中的“者”使用頻率顯著性高於漢語母語語料庫 (LL=24.82, sig.=0.000)。在漢語譯文語料庫中，“者”廣泛用於比較正式的文體，如新聞評論：

9. 第一，新貿易理論者將工業組織引入貿易理論，將貿易理論從完全競爭模式的束縛中解放出來。

10. 規模範圍經濟意味著基本公共服務提供者的數量將會減少。

11. 只要種咖啡是賺錢的, 種植者就會繼續生產, 但是因為價格將被壓低至生產的平均成本, 沒有人會一夜暴富。

值得指出的是, 漢語譯文語料庫中“者”的搭配模式相比漢語母語語料庫中“者”的搭配(搭配範圍 L1: R1)來說, 要豐富得多, 如圖 6、圖 7 所示。LCMC 中的搭配有 487 種(如: 違犯者、愛鳥者、購房者、信佛者), ZCTC 有 608 種(如: 泛神論者、擁護者、溺水者、實用主義者)。

Node	Frequency	Z-score	^
违犯	11	104.8	
爱鸟	7	73.0	
购房	4	59.0	
练功	8	52.7	
尊	9	49.4	
就餐	2	46.7	
溜号	2	46.7	
攻击	7	39.5	
信佛	2	38.1	
买入	2	38.1	
昏睡	3	37.4	
观察	18	37.2	
肥胖	9	34.8	
毁谤	1	33.0	
逾越	1	33.0	
寻医	1	33.0	
应聘	1	33.0	▼

487 collocates

Save Query Show

Controls

Forms

圖 6: “者”在 LCMC 中的搭配

本文作者曾對漢語可比語料庫(LCMC 與 ZCTC)中的“之”進行過檢索, 因未能窮盡“之”的用法, 故對 LCMC 與 ZCTC 進行了分詞及詞性標注調整, 發現“之”在 LCMC 與 ZCTC 中的用法分別如圖 8、圖 9 所示, 單獨的“之”統計資料與先前的研究有一點出入, 請參閱(戴光榮 2013, 101)。

我們將“之”的用法分為兩類: 一類是單獨的“之”用法, 一類是“之+其他成分”的用法。這兩類結構在可比語料庫中的分佈情況統計如表 8:

表 8: “之”結構在可比語料庫中的分佈情況

“之”及其結構類型	Freq. in LCMC	Freq. in ZCTC	Log-likelihood	Sig.	
之	1205	874	58.61	0.000	***
之+其他	1285	1827	86.02	0.000	***

對於單獨的“之”字用法, 先前有做了初步的探討, 限於篇幅, 此處不展開探討, 可參閱 Dai (2016, 176-181)。查閱可比語料庫中的“之”字及其結構的用法, 可以看出書面語用法居多, 如下面例證中的“王者之風”、“失敗乃成功之母”、“進化之路”

Node	Frequency	Z-score
泛神论	12	98.7
主义	26	55.8
拥护	6	54.0
原教旨主义	7	51.4
溺水	4	51.0
无神论	4	51.0
叛逃	3	49.4
实用主义	7	48.3
继任	4	43.0
批评	15	40.8
诘问	2	40.3
乐观主义	2	40.3
自然主义	2	40.3
征服	7	39.8
创始	3	38.2
说谎	4	37.9
不详	4	37.9
捍卫	5	36.7

608 collocates

Save Query

Show  Controls  Forms

圖 7: “者” 在 ZCTC 中的搭配

Word	Frequency	Forms
之	1205	2
之一	140	1
之上	23	1
之下	41	1
之中	173	1
之内	21	1
之前	103	1
之友	1	1
之后	264	1
之和	4	1
之外	51	1
之所以	39	1
之类	47	1
之谊	1	1
之间	352	1
之际	25	1

圖 8: “之” 及其結構類型 (LCMC)

Word	Frequency	Forms
之	874	3
之一	202	1
之上	29	1
之下	45	1
之中	139	1
之內	30	1
之前	271	1
之後	404	1
之和	1	1
之外	96	1
之所以	38	1
之類	35	1
之間	514	1
之際	23	1

圖 9: “之” 及其結構類型 (ZCTC)

等：

12. 這樣咄咄逼人的王者之風多少令我們不由自主地緊張。
13. 一般都認為失敗是不好的事情，但我們認為失敗是一個學習過程，失敗乃成功之母，在所有的事上都知道正確的答案在人生中並不是最重要的事情。
14. 如果這種新猿種確實是在森林中進化的，科學家將不得不重新思考這些假設，正如懷特所說的，科學家應當考慮“在進化之路上推動我們的並不是大草原”。

綜合以上對四個名詞化標記的分類探討，我們可以看出，相比在漢語母語語料庫來說，它們在漢語譯文語料庫中的用法更為靈活，包括搭配範圍擴大，使用頻率增加（單獨的“之”字用法除外，但是“之”的其他搭配結構擴大了很多，詳情參閱 Dai 2016），可以說這些名詞化標記在漢語譯文環境裡，其正式化、書面語化等語篇功能得到了更加彰顯。

## 六、漢語名詞化研究的思考

從上文對現代漢語四個名詞化標記的探討，我們可以對漢語名詞化研究做更深層次的思考：

第一，動詞、形容詞經過名詞化之後，從詞類劃分角度來看，該如何歸類？這個問題學界有不同的看法。從名詞化相對的詞類劃分來看，過去的詞類理論都認為詞類的劃分建立在充要條件的基礎之上，符合某些充要條件的詞屬於某一類，否則不屬於這一類詞。在分類標準上，這樣的詞類理論嚴格區分“語法性質”和“語法特徵”。按照這種理論來看，詞類是離散的，詞類與詞類之間有明確的界限，一個詞要麼屬於這一類，要麼不屬於這一類（沈家煊 1999，250）。如果從認知語言學來看，語言各個層

面上的單位形成了一個連續統，要找出那些僅為此類詞所有而為他類詞所無的語法特徵，在現實中還是比較困難的。詞類的原型範疇理論 (Langacker 1987; Taylor 1995) 認為一類詞的內部具有不勻稱性，有些成員是典型成員，有些則是非典型成員。一個詞類的確定是憑一些自然聚合在一起的特徵，但它們並非充要條件。一個詞類的典型成員，具備這些特徵的全部或大部分，非典型成員則只具備這些特徵的一小部分。詞類的邊界不是明確的而是模糊的，詞類與詞類之間不是離散的而是連續的，因此我們不應機械理解“詞有定類”這一觀點（袁毓林 1995；沈家煊 1999）。

第二，從上文看出，名詞化標記常用於書面語體，如學術話語、新聞評論、宗教等。也有研究探討了名詞化在不同語體中的特徵（曲英梅 2009；曲英梅、楊忠 2009；楊信璋 2006）。研究顯示，名詞化結構通常以短語形式來表達一句句子，其表現形式體現詞語密度增大、結構言簡意賅、內部組織嚴密、邏輯關係明確等特性 (Halliday 1994/2000, 350-351)，因此 Halliday 將名詞化的過程比喻為“打包”，表示過程的動詞和表示屬性的形容詞被名詞化後，原來的簡單小句被“打包”成片語，或兩三個小句被“打包”成一個小句，詞彙的密度增加，用詞量減少，從而增加了小句的信息量和詞彙密度，使語篇更為簡潔正式。

相比口語語體來說，書面語將大量詞項塞進各個小句，通過詞彙密度的增大而容納更多的資訊，內容相比來說變得更為複雜。書面語的詞項面臨更少的小句來容納它們，它們仍然是整體語法結構的一部分，它們被併入了名詞片語中，而名詞片語是語法中用以包裝高密度詞項的首要來源。

第三，誠如 Halliday 所說，名詞化是非常普遍的現象，它代表著人類認知朝事物性發展的方向 (Halliday 1998)，是語法隱喻最強有力的資源。通過名詞化方式，過程和特性被重新措辭為名詞，它們在名詞片語中擔當事物角色而不是小句中的過程或屬性。從名詞化詞彙的使用頻率、詞義選擇、語篇定位等方面，名詞化結構具有語篇精煉和客觀功能、語篇專業功能、語篇權威增強功能以及語類交際目的體現功能，能夠揭示出名詞化結構的分佈與語篇交際目的之間的密切關係（徐新宇 2011）。

從上文例證（如來自科技類文體的例 5、例 6 等）可以看出，這些名詞化表達屬於 Halliday 所說的隱喻表達，它們扮演了雙重角色：一方面，它為建構專業術語的不同層次提供了可能性；另一方面把複雜段落壓縮成為名詞形式用來作主位，從而使論證過程逐步展開成為可能。

第四，名詞化通過把小句變為名詞或者名詞短語，從而使表過程的動詞和屬性的形容詞具有名詞特性。名詞化通過小句內部的整合，從而實現其特定的語篇功能及語類功能（張高遠 2008；張高遠、王克非 2004; 2008a; 2008b）。

## 七、結論

本文通過探討漢語可比語料庫中名詞化現象的差異，深入分析四個名詞化標記在不同文體中的分佈特徵，對每一個名詞化標記在漢語母語語料庫與漢語譯文語料庫中的使用特徵進行了分類探討。

通過對比分析，我們發現，在漢語可比語料庫中，名詞化標記於漢語譯文語料庫的使用頻率遠高於漢語母語語料庫中的使用頻率，名詞化標記在漢語譯文語料庫中的搭配範圍也更為豐富。漢語譯語文本中的動詞、形容詞經過名詞化之後，面臨詞性定奪的難題。通過資料分析及例證梳理，可以看到漢語譯文語料庫中的名詞化標記常用於書面語體，如科技文本、學術話語、新聞評論等。從功能與認知的角度分析，名詞化通過小句內部的整合，實現其特定的語篇功能及語類功能，讓語篇內容更為精煉、表述更為客觀，更加彰顯語篇的專業性。

### 銘謝

本文所涉研究由國家社科基金一般項目（項目編號：22BYY042）資助。

## 參考文獻

- Baker, M. C., and Vinokurova, N. 2009. "On Agent Nominalizations and Why They Are Not like Event Nominalizations." *Language* (3): 517-556.
- Bierwisch, M. 2009. "Nominalization—Lexical and Syntactic Aspects." In *Quantification, Definiteness, and Nominalization*, edited by A. Giannakidou & M. Rathert, 281-319. Oxford: Oxford University Press.
- Billig, M. 2008. "The Language of Critical Discourse Analysis: the case of nominalization." *Discourse & Society* (6):783-800.
- Comrie, B., and Thompson, S. A. 1985. "Lexical nominalization." In *Language Typology and Syntactic Description, Volume III: Grammatical categories and the lexicon*, edited by T. Shopen, 349-398. Cambridge: Cambridge University Press.
- Crystal, D, ed. 2008. *A Dictionary of Linguistics and Phonetics (Sixth Edition)*. Oxford: Blackwell Publishing.
- Dai, G. 2016. *Hybridity in Translated Chinese: A Corpus Analytical Framework*. Singapore: Springer.
- Gerner, M. 2012. "The Typology of Nominalization." *Language and Linguistics* (4):803-844.
- Halliday, M. A. K. 1994/2000. *An Introduction to Functional Grammar (Second edition)*. Beijing: Foreign Language Teaching and Research Press.
- Halliday, M. A. K.. 1998. "Things and Relations: Regrammatizing Experience as Technical Knowledge." In *Reading Science: Critical and functional perspectives on discourses of science*, edited by J.R.Martin & R. Veel, 185-236. London: Routledge.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar. Volume I. Theoretical Prerequisites*. Stanford: Stanford University Press.
- Langacker, R. W. 1991. *Foundations of Cognitive Grammar: Volume II: Descriptive Application*. Stanford, California: Stanford University Press.
- Li, C. N., and Thompson, S. 1981. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley & London: University of California Press.

- Martin, J. R. 2008. "Incongruent and proud: de-vilifying 'nominalization' ". *Discourse & Society* (6): 827-836.
- Noonan, M. 1997. "Versatile Nominalizations." In *Essays on Language Function and Language Type*, edited by J. Bybee, J. Haiman, & S. A. Thompson, 373-394. Amsterdam: John Benjamins.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London & New York: Longman.
- Salanova, A. P. 2007. *Nominalizations and aspect*. Ph.D. dissertation. Massachusetts Institute of Technology.
- Taylor, J. R. 1995. *Linguistic Categorization: Prototypes in Linguistic Theory (Second Edition)*. Oxford: Oxford University Press.
- Xiao, R., Rayson, P., and McEnery, T. eds.. 2009. *A Frequency Dictionary of Mandarin Chinese: Core Vocabulary for Learners*. New York: Routledge.
- Yap, F. H., & Grunow-Hårsta, K.. 2010. "NonReferential Uses of Nominalization Constructions." *Language and Linguistics Compass* 3(1):1-22.
- Yap, F. H., Grunow-Hårsta, K., and Wrona, J. eds. 2011. *Nominalization in Asian Languages: Diachronic and typological perspectives*. Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Yap, F. H., and Wang, J. 2011. "From light noun to nominalizer and more: The grammaticalization of zhe and suo in Old and Middle Chinese." In *Nominalization in Asian languages: diachronic and typological perspectives*, edited by F. H. Yap, K. Grunow-Hårsta, & J. Wrona, 61-107. Philadelphia: John Benjamins Publishing Co.
- 戴光榮。2013。《譯文源語透過效應研究》。上海：上海交通大學出版社。
- 戴光榮、左尚君。2017。翻譯漢語中的名詞化特徵：一項基於語料庫的實證研究。亞大跨學科翻譯研究 4：97-110。
- 高名凱、計永佑。1963。從“動詞形容詞的名物化”問題說到漢語的詞類問。北京大學學報(人文科學版)2：49-66。
- 黎錦熙。1924/2007。《新著國語文法》。長沙：湖南教育出版社。
- 黎錦熙、劉世儒。1960。語法再研討——詞類區分和名詞問題。中國語文 1：5-8。
- 陸燦、潘海華。2013。從英漢比較看漢語的名物化結構。外語教學與研究 5：643-656。
- 呂叔湘。1999。《現代漢語八百詞(增訂本)》。北京：商務印書館。
- 曲英梅。2009。《基於語料庫的英漢動名化對比研究》。長春：東北師範大學博士學位論文。
- 曲英梅、楊忠。2009。英漢動詞名化運算式的名化度對比分析。外語學刊 3：66-69。
- 沈家煊。1999。《不對稱和標記論》。南昌：江西教育出版社。
- 石定栩。2004。〈名詞化、名物化與“的”字結構〉。黃正德編。《中國語言學論叢第三輯》，78-92。北京：北京語言大學出版社。

- 石定栩。2011。《名詞和名詞性成分》。北京：北京大學出版社。
- 石毓智。2010。《漢語語法》。北京：商務印書館。
- 王立非、劉英傑。2011。我國英語名化研究三十年：回顧與思考。外國語 6：56-63。
- 肖忠華。2012。《英漢翻譯中的漢語譯文語料庫研究》。上海：上海交通大學出版社。
- 肖忠華、戴光榮。2010。尋求“第三語碼”：基於漢語譯文語料庫的翻譯共性研究。外語教學與研究 1：52-58。
- 徐新宇。2011。英語名物化結構的語篇功能分析——以國際貿易銷售合同為。現代外語 2：143-150。
- 楊信璋。2006。名詞化在語體中的作用：基於小型語料庫的一項分析。外語電化教學 4：3-7。
- 袁毓林。1995。詞類範疇的家族相似性。中國社會科學 1：154-170。
- 張高遠。2008。《英漢名詞化對比研究——認知功能取向的理論解釋》。北京：中國社會科學出版社。
- 張高遠、王克非。2004。名詞化研究的新路徑。外國語 6：45-50。
- 張高遠、王克非。2008a。X-ER 派生詞的構造類型、語義擴展及其指稱功能實現的理據。外國語 6：9-15。
- 張高遠、王克非。2008b。非施事-er 名詞與中動結構關係考察。現代外語 1：26-37。
- 朱德熙。1983。自指和轉指：漢語名詞化標記“的、者、所、之”的語法功能和語義功能。方言 1：16-31。
- 朱德熙、盧甲文、馬真。1961。關於動詞形容詞“名物化”的問題。北京大學學報(人文科學版)4：51-64。





# 翻譯即協調：黎翠珍英譯粵劇《霸王別姬》研究

張鼎程<sup>1</sup> 張旭<sup>2</sup>

**Address:** <sup>1</sup> Faculty of Languages and Linguistics, University of Malaya, Malaysia; <sup>2</sup> School of Foreign Studies, Guangxi Minzu University, China

**E-mail:** <sup>1</sup> zdczyhking@gmail.com; <sup>2</sup> zhangxuhk@163.com

**Correspondence:** Xu Zhang

**Citation:** Zhang, Dingcheng, and Xu Zhang. 2022. "Translation as Mediation: C.C. Jane Lai's English Translation of the Cantonese Opera *Farewell My Concubine*." *Translation Quarterly* 105: 113-130.

## ***Abstract***

Translation as Mediation: C.C. Jane Lai's English Translation of the Cantonese Opera *Farewell My Concubine* (by Dingcheng Zhang and Xu Zhang)

*Is Cantonese Opera translatable? Can verbal music be reproduced in translation? What defines an effective translation of performative texts? This study, drawing upon theories from modern translation studies, examines the work of Hong Kong bilingual translator C.C. Jane Lai and her translation of the Yueju opera "Farewell My Concubine." It investigates how the translator endeavored to reproduce verbal music while adhering to the norms of playability, readability, and acceptability. The primary focus of the study is on the strategies and norms employed and the aesthetic impact achieved in the receptor culture.*

## 一、引言

中國文化走出去是當今主流意識形態重點打造的文化工程，也是社會熱議的話題。作為中國文化重要組成部分的粵劇（Cantonese Opera 或 Yueju Opera）能否被譯介出去並為西方讀者/觀眾所接受和欣賞，除了高超的戲曲表演藝術外，還取決於其劇本的翻譯。

粵劇作為嶺南地區最重要的藝術形式之一，於 2006 年經國務院批准列入第一批國家級非物質文化遺產名錄；2009 年又被聯合國教育、科學及文化組織列入“人類非物質文化遺產代表作”名錄，逐漸為不同地區和文化背景的人們所知。粵劇藝術對外

傳播，除了其自身的表演外，其文本翻譯也至關重要。早在 2007 年，香港粵劇發展基金便資助大專院校或學者，推行粵劇劇碼簡介及劇本翻譯計畫，意在將一批優秀的粵劇作品推向世界。

2013 年 10 月，當代香港著名雙語作家兼譯家黎翠珍應邀英譯了香港已故粵劇編劇家葉紹德（1930—2009）的劇本《霸王別姬》（*Farewell My Concubine*），收入吳鳳平、陳鈞潤合編的《葉紹德粵劇劇本精選》（*Yip Shiu Tak's Cantonese Opera Manuscripts*），由香港大學教育學院中文教育研究中心出版。收入該集的還有《辭郎洲》（*Island of Farewell*）、《朱弁回朝》（*Zhu Bian Returns to Court*）、《十五貫》（*Fifteen Strings of Cash*）、《李後主》（*Li The Later Ruler*）。全書採取漢英對照形式，各劇前有劇本考證和賞析。該書出版得到香港粵劇發展基金資助。本文試以黎翠珍英譯《霸王別姬》為研究對象，考察她是如何詮釋劇本，協調中英兩套詩學規範，並以高超的譯筆充分展現中國文化形象，再現言語的音樂美，進而實現譯本的可表演、可演說和可接受之目的。

## 二、源語劇本概覽與譯者翻譯觀說略

粵劇起源於兩廣地區，又稱“廣東大戲”，發源於佛山，以粵方言演唱，是漢族傳統戲曲之一。粵劇形成於廣東，後傳入廣西、香港、澳門、臺灣，在東南亞和美洲各國有華僑居住的地方均有粵劇演出。粵劇的基本聲腔為梆簧，並保留了弋陽腔與昆腔的部分曲牌以及南音、粵謳、木魚、龍舟、板眼等廣東民間說唱的曲調和民歌、樂曲、時代曲、小調等民間小曲。腳色行當有文武生、小生、正印花旦、二邦花旦、醜生和武生。其唱詞結構基本上是七字句、帶冠七字句和十字句等，現在則普遍用長句和自由句格。唱腔分大喉、平喉、子喉三種：大喉高亢激越，為武生專用；平喉穩重厚實，為小生、文武生專用；子喉清圓委婉，為旦角專用。伴奏樂器有管弦樂器及打擊樂器。其板式變化體以一對“上、下樂句”為基礎，基本板式為中板，另有慢中板、慢板、快板、快中板、散板、連板、鬮板和清板等，在變奏中突出節拍、節奏變化的作用。粵劇的聲腔清悠婉麗、優美動聽，同時不同流派有著不同的唱腔風格（蘭青 2014，68-69）。

20 世紀 40 年代，英語粵劇開始在香港亮相。1947 年，香港華仁書院的華仁戲劇社主席黃展華開始創作英語粵劇，把很多傳統粵劇改用英語來表演。在新加坡，敦煌劇坊主席黃仕邦用英文創作粵劇。在馬來西亞，鐘莉莉用英文編寫出《龍鳳配》，從此開了馬來亞英語粵劇的先河。為了推廣粵劇，海外粵劇團做出大膽嘗試，他們把粵劇帶到社區、學校演出，配上該國文字字幕，以讓當地觀眾看懂。隨著粵劇成功“申遺”後名聲大振，粵劇在世界的影響進一步擴大。此後，英譯粵劇也逐漸進入研究者的視野（潘福麟 2000；曹廣濤、彭可欣、張派豪 2018；李燕霞、曾衍文 2018；陳宇 2020；楊浩然、張映先 2021）。

《霸王別姬》是一出名劇，典出《史記·項羽本記》。故事講的是楚漢相爭時，西楚霸王項羽不敵劉邦，被圍垓下。霸王看著虞姬和烏騅馬，慷慨悲歌，虞姬在一旁為項

羽跳了人生最後一支舞。當項羽準備突圍時，虞姬向霸王訣別，自刎而死。該故事後被改編成戲劇，長期在舞臺上演出，經久不衰。

葉紹德所編粵劇《霸王別姬》是根據梅蘭芳演出的《霸王別姬》改編而成，於1983年由勵群粵劇團於新光戲院首演。梅劇劇本由齊如山、吳震修根據明朝沈采的昆曲《千金記》，並參考楊小樓、尚小雲排演的京劇《楚漢爭》（據清逸居士所編京劇《楚漢爭》排演，而清逸居士《楚漢爭》亦據《千金記》）編成初稿，最早於1922年由梅蘭芳與楊小樓主演。後在長期演出中改編壓縮，形成梅蘭芳晚年的演出定本（吳鳳平、陳潤鈞2013，518）。葉劇與梅劇雖然演的是同一個歷史事件，但是在內容和藝術上有較大差異。總體而言，梅劇為京劇，有九場，情節相對簡單；葉劇為粵劇，有六場：《漂母飯信》《鴻門會》《追賢》《九裡山》《兵敗垓下》《霸王別姬》，其情節有很大不同，主要表現在：一、葉劇在梅劇的基礎上增加不少情節，如第二場《鴻門會》；二、葉劇刪除了梅劇中一些情節，刪除的情節主要有四個：項羽見到韓信的罵榜，項羽出兵時出現的旗斷馬嘶，李左車激項羽進埋伏圈，項羽自刎。三、葉劇在與梅劇大致近似的情節進行了渲染和微調。同時，葉劇在藝術表現形式上也有很大改動。主要表現在：一、人物形象增多，並進行了角色調整；二、人物刻畫更加細緻；三、在場景佈置、人物科介、語言等方面描寫更詳盡。

現代翻譯理論尤其是功能主義和社會學途徑的研究者普遍認為，譯者在翻譯轉換中起著至為重要的作用。在語言交流過程中，所有的翻譯實際上是一種協調行為（mediation），也就是，“譯者一定程度地干預了轉換過程，並在處理文本的過程中將自己的知識和信念（即意識形態）注入其中”（Hatim and Hason 1997，147）。自此，在詮釋和翻譯研究中人們關注的重點就從文本轉到生產這些文本的人的一端。另外，蘇珊·巴斯奈特在《翻譯研究》一書中也反復強調，文本功能是翻譯的中心問題（Bassnett 2002,131）。她尤其重視戲劇文本的可表演（playability or performability）、可言說（speakability）和讀者/觀眾的接受（acceptance）問題。如果依此來審視黎翠珍英譯粵劇《霸王別姬》，將會有一番新的發現。根據《精選》中楊偉誠《序》和吳鳳平《前言》交代，該書所選譯本主要是為了供粵劇團體做字幕之用，為在英語世界推廣粵劇文化和粵劇做出貢獻（吳鳳平、陳潤鈞2013，1-2）。既然是做字幕之用，首先文本的資訊功能最為重要，其次才是戲曲文詞的音樂特質，這點在面對外國讀者或聽眾時更是如此。因此，要將這些資訊準確充分地傳遞過去，同時著眼於觀眾的接受，譯者必須協調兩套詩學規範，進而重寫出一部戲劇來。這點在黎譯本中就有明顯的體現。

由於擁有豐富的戲劇翻譯和表演經驗，黎翠珍歷來非常重視言語的音樂性。她曾說過：“聲音是很重要的，是一種言語音樂（verbal music），文字的聲音是一種很有趣的東西，不用經過人的腦袋而能夠走進人的心裡”（黎翠珍2005，21）。她在翻譯詩歌、小說、戲曲等不同文類作品時都盡其所能來再現原作的音樂效果，在翻譯粵劇《霸王別姬》時更是如此。

黎翠珍總是謙稱自己不懂音樂，但這並不意味著她不重視語言的音樂性。當年筆者在研究她英譯過士行《鳥人》時曾請教過她是如何傳譯劇本的音樂性問題。2019年11月13日，她在給我的回信中說：“《鎖五龍》那一段‘西皮快板’我是跑去買了錄

音帶，跟著學唱了半天才下筆譯的。後來譚葉紹德編的粵劇《霸王別姬》也有作這類參考，只是做得沒有這麼細了。”話雖如此，她在翻譯《霸王別姬》之前曾特意找來粵劇版光碟，反復地觀賞每一個場景和唱腔，然後再動筆翻譯。這和她以往翻譯《原野》《鳥人》《嬌情》等劇本不同，因為翻譯那些劇本時重點是注意劇本的可表演、可演說和可接受性。而《霸王別姬》則不同，因為它原本是能演唱的，這就要求她在注重劇本內容的同時更要注重原作的音樂效果之傳譯，而這是相當困難的。不過，黎翠珍卻通過自己的詮釋和演繹，推出一部頗具特色的譯作。

### 三、戲曲術語音譯與釋義

據黎翠珍的《譯者注》介紹，本次翻譯採用同書中陳鈞潤（Rupert Chan）在翻譯《李後主》時的許多標準，同時區別對待舞臺提示、道白與唱腔；歷史人物名稱、地名均採取普通話音譯；詞牌名、調名則用廣東話音譯以示區別（吳鳳平、陳潤鈞 2013，581）。

首先是曲牌名和曲調名、粵劇術語翻譯。鑒於術語翻譯必須遵循簡潔的原則，黎譯多是採取音譯加補償的辦法。

粵劇術語	黎譯	術語釋義
白欖	Speaks: Baatlaam	百欖：一種有節奏、有腳韻和句逗比較整齊的說白。
七字清下句	Sings: Chatjiching lower line	七字清：每句七字，分兩頓，每頓六拍。第一頓四字，第二頓三字。每句均由漏底板後起唱（叮位起唱），板位收，無過序。
唱《雨淋鈴》	Sings: To the tune of Yulumling “Rain on the bells”	《雨淋鈴》：唐教坊曲名，也作《雨霖鈴》。至宋演化為詞調名，保留了原曲哀怨的風格，也名《雨淋鈴慢》。雙調，103字，仄韻。
乙反南音	Sings: Yifaan Namyum	南音：七字句格，包括乙反南音或流水南音，分為起式、正文、收式三部分，每四句一首。乙反南音基本唱法與普通南音相同，區別於反對調是將普通（正線）南音調中的63唱成74，行腔哀怨、淒涼、如泣如訴。
轉正線	Sings: Changes to jingsin	正線：粵劇音樂的定調，有正線和反線兩種：正線相等於C調（1=C2）有時因為演員的音域需要，正線也可以（1=d2）來定調。
轉二王下句	Changes to sing: Yiwong lower line	二王：又稱二黃，粵劇屬皮黃系統，基本曲調成為椰子（西皮）二王（二黃）。

反線中板 下句	Sings: Yiwong Saubaan	反線：粵劇唱腔音樂根據絃樂器定弦命名的一種。反線包括的主要調式有五聲 C 宮調等梆子板腔及七聲 C 徵調。中板：是梆子體系中最常用的一種曲牌，它的唱腔音樂比較豐富，可以用不同的調式去表達不同的情感。分正線、反線、反乙、有序等。反線中板的特點是淒怨纏綿。
木魚	Sings: Muk' yu	木魚：粵語地區一種獨立的曲藝品種，生旦同腔。句法和龍舟、南音基本相同。但它沒有龍舟那麼口語化，每句用字不太多，偏近于南音的句子。起式只用二句，行腔較婉轉抒情，在粵劇（曲）中較多用於哀怨的訴述情節，也有時用於交代、敘述情節。
滾花下句	Sings: Gwanfa lower line	滾花：粵劇梆黃腔調的一種唱法。可能從早期粵劇的快板演變而成。接近於滾調，長短快慢可以自由變化。根據用法的不同，又有“長句滾花”“滾花煞板”等類別。
反線《蕭蕭班馬鳴》	Sings: Faansin Siusubaanma Ming “Horse-neighing”	《蕭蕭班馬鳴》：粵劇小曲名，仿牌子曲創作。
快排子頭	Sings: Fast Paijtau	排子頭：原本鑼鼓名稱，即鑼鼓打完排子頭。牌子原出昆曲噴呐或簫的調子。
慢板下句	Sings: Maanbaan lower line	慢板：粵劇梆子音樂板式結構一部分（另三個為散板、中板和芙蓉）。慢板部分有慢板、河調慢板、快慢板。
快二流	Sings: Fast Yilau	快二流：板式的一種，1/4 拍，快速。

鑒於戲曲是兼具時間和空間二重性的綜合藝術，它的表演稍縱即逝，不容許觀眾有更多的時間來揣摩劇本在細枝末節上的意義，尤其是面對文化底蘊深厚的曲詞牌、唱腔等術語時，最簡便的方法便是音譯。從嚴格意義上來說，音譯就是不譯。音譯又是那種極端的充分性翻譯，這種翻譯只有在某一文化處於極度強勢或為別國普遍認同的情況下才能廣為接受。實際上，它是譯者不得已間而採取的策略，頻繁地使用就會讓讀者產生陌生感甚至拒斥心理。而在翻譯《霸王別姬》的過程中，好在譯者都有補償措施。譬如在粵劇術語前加上“Sings”，表示是唱腔；或在音譯前加上“to the tune of …”，表示曲牌名。至於一些已有的譯名，譯者便直接借用，如《春風得意》譯作“to the tune of ‘Rejoicing in the Spring Breeze’”，將《婢女淚》譯作“The Maid’s Tears”，將《百花亭鬧酒》譯作“Carousing in the Hundred Flower Pavilion”，將《水仙子》譯作“to the tune of ‘Narcissus’”，等等。從讀者接受角度看，這種釋義的做法也便於讀者/觀眾理解和接受。

由於整部戲曲中詞牌、調名和其他術語眾多，自然增加了讀者接受的困難。故而像同書中的陳鈞潤在翻譯《李後主》時就在每場譯文之後加了若干注釋。可惜黎翠珍未能借鑒這種做法。另外，對於那些音譯，編者如能在書末另加一個音譯術語對照表，並附簡單的釋義，則更方便讀者閱讀和理解。

#### 四、音樂性念白之傳譯

如果說歌舞性、虛擬性和程式化是中國戲曲的三大基本特徵，那麼唱、念、做、打則是戲曲表演的四種藝術手段。其中與語言緊密相連的有兩點：一、唱，指表演中的歌唱；二、念，指表演中具有音樂性的念白，二者互為補充，構成戲曲表演藝術兩大要素之一的“歌”。既然唱念都具有音樂性，就要求譯者格外注意這一效果的傳遞。

首先來看念白。通常念白要求感情濃郁，語言凝練，韻律和諧，節奏鮮明，有音樂性和動作性。它是一種富有音樂性的藝術語言，具有鮮明的旋律性、誇張性和強烈節奏感。其輕重緩急、抑揚頓挫均無腔板可按，無譜籍可查。通常念白的音調起伏、速度快慢，必須與曲調部分的起音協調、速度一致，才能銜接自然。如第一場《漂母飯信》中有一段口白：

聖人有云：天降大任於斯人也，必先苦其心智，勞其筋骨，餓其體膚。(勾羅)

唉吔，講到餓其體膚，我已兩天未曾進食，如今饑餓難抵，只見河邊婦人正在用膳，更令我口角流涎。本欲上前乞討，則怕顏面無光，正是今時欲把饑腸濟，恨難搖尾乞人憐。

As sage Mencius says, when heaven bestows a great mission to a man, it first tests his will, tries his sinews, starves his body...(Gong)

Aiya, talking about starve his body, I have not eaten for two days. I would be grateful for some food except that I cannot bear to beg.

我們對照原文發現譯者有增添的，如“聖人”被譯為“sage Mencius”，這是譯者運用自己的常識而做了添加；至於第二段，譯者只直譯了前面兩句，其他則做了大量的省略，尤其是省去內心的感受（羞恥心、自尊心），並根據英文的文脈和意脈，處理成一個簡單的“I would be grateful for some food except that I cannot bear to beg.”。整個表述非常清晰，且用詞典雅，雖然于原文內容和形象有失落，感染力也不夠強烈，但簡潔自然，適合於演唱。至於將“踟躕介”之說明文字處理成“He hesitates and loiters”，完全是在做字譯。原來的三個字，被拆譯成四個詞，節奏尚在，讀起來與原作相仿佛。

又如第六場《霸王別姬》中有霸王的一段口古：

愛妃不用操心，孤王早有主意，今日一戰，已殺得敵人心驚膽落，孤王回來抖息片時，連夜攻其無備，任他雄獅稱百萬，也難敵孤王江東八千子弟兵。

Do not be troubled, my love. I have made my plans. Today's battle has got them scared. I've come back for a respite. Then I'll attack at night, catch them unprepared. Even if they are a million, they are no match for me and my men.

口古又稱口鼓，是粵劇中一種押韻的念白。其結構分上句和下句，且有上句必須有下句，以收仄聲為上句，收平聲為下句。每句結束都有鑼鼓襯托。這裡原文句式長短不一，但節奏分明。譯文以簡單句和短句為主，稍長的句中運用標點，以此來模仿原文的節奏。同時還使用頭韻法，/m/音不斷重複，並將“江東八千子弟兵”改譯為“my men”，意在與前面的“for me”對應，形成一個頭韻，而且也將霸王“我”與士兵的關係拉近。這段口古讀來節奏分明，很有原文的氣勢。

同樣在第六場《霸王別姬》中還有虞姬目送霸王去後心情不安入帳的一段念白：

大王不知軍心渙散，此戰定然失利，我相告時難不告難，寧不令人急懷。

YU JI (*Watching XIANGYU exit, returns to the tent in trepidation.*) [Speaks]

My lord knows not the low morale in his camp. I fear the battle will not go *well*.  
It was difficult not to tell him or to *tell*, caught in the dilemma it's like torment in *hell*.

原文為念白句，六字句和八字句各兩句，句式整齊。黎譯效仿原文不做分行排列，但句末分別出現了三個詞“well”“tell”和“hell”押韻，句中還出現了頭韻（如“lord”與“low”，“will”與“well”）、重複（如“tell”），最末一句中使用了逗號，這樣整個語段讀來節奏與原文相仿佛，而且大致等時。

當然，在念白中還有不少詩體形式的片段，這些更容易讓讀者將其與詩歌聯繫起來。如第二場《鴻門會中》有范增念白攬，明顯地是以詩的形式排列：

今日慶亡秦，等候劉邦臨席上，  
只許他一君一臣來進見，多來一個劍下亡。

We celebrate the end of Qin,  
We wait for LIU BANG to come in.  
Only one attendant we allow him.  
Come one more and we do him in.

白攬原本是一種有節奏、有腳韻和句逗比較整齊的說白。多為三、五、七字句，句數不限，不分上下句，單數句可不押韻，雙數句必須押韻，平仄自由。它突出的是節奏的作用，即將自由的語言納入固定的節奏規範之中。這裡的原文二四兩句以/soeng/（上）和/mong/（亡）音結尾，押通韻。譯文採取分行形式排列，讀者立刻能明白這是詩體。其中一二四行押/in/韻，另外“him”一詞重複了兩次，儘管從效果上看與原作有差別，但意在模仿原作的韻式。整首詩用詞簡潔，念起來節奏分明。

再來看第六幕《霸王別姬》中霸王與虞姬之間的兩段白攬對白：



霸王 [白欖] 兒女請，英雄性（介）  
你一笑令我振雄風，一顰令我神不定，  
重複霸王威，生死與卿形隨影，  
愛妃騎上烏騅馬，我步行殺敵護娉婷。  
虞姬 [白欖] 大王無馬怎能行，徒步焉能和敵拼，  
大王恩，重拜領，祝王帝業早完成，  
祭壇告與泉台聽。

XIANG YU [Speaks: Baatlaam]

Love for his woman

Becomes a man.

Your smile my spirit charms to life;

I'm troubled if you frown.

I'll pluck up my courage

To take you from this carnage.

You ride my steed, I'll go on foot

To shield you from peril.

YU JI [Speaks: Baatlaam]

The king cannot be without his steed.

On foot you'd be disadvantaged.

Your gracious offer

I decline.

I wish you in victory to bloom,

Tell me the good news at my tomb.

當然，整部戲曲最能體現言語音樂性的還是那些唱段。通常粵劇的句式因曲式不同可分為三類：曲牌體的句式為長短句襯字；板腔體的句式基本為七字句和十字句，上下句結構，可加襯字和活動句；說唱體的句式分起式、正文、收式三部分。正文部分是七字的上下句結構，各句表達一個相對完整的意思，句末必須押韻，且有基本的字數、平仄相間的規定。目前粵劇的句格字數已不受原譜格式限制，但須與音樂旋律結構同步。

詩歌的音韻美是詩歌語言的重要特徵之一，如果我們把詩韻比作人的靈魂，語言就是他的骨幹。一首好的詩首先取決於它的語言。鑒於旋律、節奏與和聲是音樂的三要素，而詩的音樂性之核心便是節奏和旋律，若無二者無以成詩。詩歌的節奏、旋律主要體現在平仄音節和韻腳兩個方面。只要翻譯得法，音節和韻腳也是可以傳譯的。

首先來看節奏的傳譯。如果我們著眼于原唱詞的字數和節奏，可以發現黎翠珍做了一些創格舉措。如第一場《漂母飯信》中有漂母唱七字清下句和滾花句：

[七字清下句]

民不聊生天下亂，老來無靠百愁纏。  
帝主兇殘人命賤，佞臣肆虐榨金錢，  
修築阿房夫命損，愛兒抗暴未歸旋，  
大好家庭遭慘變，空流苦淚落矜前。

[花]

堪憐白髮倚門閭，天涯望斷延殘喘。

WASHERWOMAN [Sings: Chitjiching lower line]

Disasters make life not worth living.  
The old are helpless and besieged by grief.  
The cruel emperor makes lives so cheap,  
And greedy officials from us the last coin extract.  
My man died in forced labor the emperor's palace to build.  
My son joined the rebels has yet to return.  
Such horrors they've brought to my family,  
Leaving me in sorrow, lonely tears to weep.  
[Sings: Gwanfa]  
My sad hoary head against the doorway lean,  
Looking to hope my hopelessness to redeem.

所謂七字清，是每句七字，分兩頓，每頓六拍。第一頓四字，第二頓三字。每句均由漏底板後起唱（叮位起唱），板位收，無過序。七字清的節奏通常為 2/4 拍，即一板一叮。而滾花接近於滾調，長短快慢可以自由變化。原唱詞平仄分明，句末押韻。黎譯每行以五音步抑揚格為主體，而五步抑揚格（iambic pentameter）又是英詩中最常用的格律形式。從文類的角度著眼，中國戲曲最接近的西方文類形式便是戲劇。由於黎翠珍長期從事莎劇教學和翻譯工作，她在翻譯那些長篇詩體作品中就常借用莎劇的一些表現手段。眾所周知，莎劇均為詩劇，絕大部分是用素體詩（blank verse）寫成。這種詩不用韻腳，不用韻但有韻律，每行以五步抑揚格為主體。黎翠珍在翻譯粵劇《霸王別姬》時再次嘗試借用這種表現形式規範，其用詞相當典雅，形象生動。譯詩不押韻，很有莎翁素體詩的風格，但各行以/m/ /n/ /ŋ/，或以含長母音、雙母音的詞結尾，這些音對應于漢語中的平聲韻，吟誦起來悠揚婉轉，極具音樂效果。而且，這種詩行唱起來大致與原詩等時。

又如第一場還有韓信反線中板下句唱段：

自負經緯才，胸懷天地志，  
誰願長貧潦倒，株守在家園。  
空有抗暴心，怎奈賞識無人，  
雖效當年，毛遂來自薦。  
家道歎飄零，飽受饑寒交襲，  
未忘求上進，習武讀書篇。  
壯志未能伸，蹇滯未逢時，  
淪落窮途，才華難外現，  
雖未至討飯為生，也弄到捱饑度活，  
捱盡人間苦，積恨苦難宣。  
[花] 惟有韜光養晦待時機，豈願長淪貧與賤。

I pride myself on varied talents,  
My aspirations high as the sky.  
Hard it is to stay in dire poverty.  
In the humble homestead.  
I long in vain to fight tyranny,  
But find no opportunity to forge ahead.  
So I venture forth,  
In hope to volunteer.  
My family is scattered,  
In adversity I strive  
In hope of improvement  
In learning and martial arts.  
My time has yet to come,  
Misfortunes dog me still.  
Bedraggled and in hunger,  
My talents cannot be revealed.  
Though fate has not beggared me,  
I live in hunger and cold.  
Hardened so with suffering,  
Bitterness in my heart congeals.

[Sings: Gwanfa]

I can but bide my time.

To shed the curse of poverty.

粵劇中所謂中板，其音樂節拍為二拍子，即“一板一叮”。在整個唱腔及鑼鼓體系內，板式控制速度，並且唱腔收結最後的一個字一定在板的位置。叮又稱眼，是控制板與板之間過渡的拍子。“一板一叮”相當於西方音樂的兩拍子，第一拍為板，第二拍為叮，即板叮相隔迴圈組合。原唱詞以五字句主體，間用四字句、七字句或九字句，讀來節奏分明。譯詩不押韻，但以四音步為主體，間用六音步，這樣的節奏基本對應於原作，抑揚頓挫，音韻感強，而且讀來每行基本與原詩等時。

第六場《霸王別姬》有唱反線《花底琴音》一段則試驗了另一種快速、激昂的節奏：

再添杯，應飲勝，建邦基，千秋永，

盼早晚，亂世戰禍平，施恩德政萬民敬。

One more cup

Please you to drink.

Start a dynasty,

Forever reign.

Soon or late,

Peace to bring,

Enjoy the praises the people sing.

漢語中有所謂“句短則意無窮”之說。連續的短句並用，語速自然加快，能夠表現很強的氣勢。這點在英文中同樣如此。不過適當地換用一個較長句，又能起到變奏的效果。在此，黎譯與其說是在以音節對譯平仄，毋寧說是在字對字譯。整段譯詩除了三個詞為雙音節或三音節外，其他都是單音節，節奏完全對應於原唱詞，各行吟誦起來基本與原文等時，很有氣勢，完全可以配以粵劇音樂演唱。

其次來看尾韻的復現。從音樂學角度著眼，押韻在音樂上是長主音。一般而言，音樂的旋律是從主音開始，到主音結束。聽到一個長的母音，聽眾會感覺到旋律又回到起點或基準線，所以會有一個樂句結束的感覺。從語言的角度來說，押韻就是把一個韻的聲音拖長放大，讓人體會到這個聲音的含意及內在情緒。韻的長度相當於句子的四分之一到二分之一，因為押韻拖韻，讀者會關注這個不同，並體會韻的含意。而在詞和曲中，押韻規定是出於音樂旋律的要求，樂句要求舒緩，便押平聲韻；樂句要求激越，便押仄聲韻；要求表現不同的感情色彩，便讓平韻與仄韻互換。無論一韻到底，還是數韻互換，詞和曲都要求押韻的句子起碼有兩個樂句，其收束的語音、韻母相同。總之，無論是詞、曲還是近體詩，在叶韻方面遵循和諧與變化交錯，但始終以

和諧為主的原則（黃天驥 2003，37-38）。這些押韻方式（rhyming scheme）能否在英譯詩中復現呢？

總的來說，黎翠珍在翻譯長詩時很少押韻，只是通過輕重音的組合來形成節奏。不過她在翻譯《霸王別姬》時偶爾也嘗試押韻，並試驗了不同的押韻方式。

如第二幕《鴻門宴》有項羽唱的一段花下句：

莫在人前多作態，巧言令色語荒唐。

居心叵測再難瞞，待我直言你三罪狀。

Feign not courtesy in pretense;

False words ring hollow and *untrue*.

Your evil design is too apparent.

Let me name three charges against *you*.

長花下句即長句滾花，其特點是在基本句間插入若干個七字的活動句而構成，有上句、下句之分。這裡的兩段唱詞均為七字句，可分為兩頓，即前四字一頓，後三字一頓，各占六拍，節奏為二二二一，且句尾押韻。譯詩每行以四音步為主體，與原詩相仿佛，且模仿原作採取二、四行押韻，只是原唱詞是以響亮的/oeng/和/ong/音押通韻，被換成了雙母音/ei/和長母音/ju:/，其聯想義自然發生了一些變化。

又如項羽唱催爽七字清：

俺有容忍滄海亮，是非黑白察端詳。

沛公果然真好漢，痛陳肺腑在當堂。

I am magnanimous as the *sea*,

Evil and good I can perceive.

LIU BANG a true hero *be*,

And he bares his soul to *me*.

原唱詞同樣為七字句，句尾一韻到底。譯詩每行四音步，除第二行外，亦一韻到底。這裡原文向上揚的響亮音/oeng6/（亮、詳）、/on3/（漢）、/ong4/（堂）押通韻，再度被換成了幽遠的長母音/i:/，聯想義亦隨之改變。

第三場《追賢》蕭何唱合尺花下句：

我主求才若渴，你莫生疑慮在心間。

難得天降奇才，當寄付乾坤重擔。

My master's wish is sincere,

Leave your doubts *behind*.

Talent such as yours

A leading place will *find*.

合尺滾花又稱二黃滾花，唱詞基本格式為七字句，有嚴格的上下句結構。基本拍為散板，其音樂節拍自由，即“無板無叮”。在粵劇中，以合尺為骨幹音的旋律結構，具有色彩華麗、格調清晰、音本色調纏綿委婉等特點。這裡的唱詞以七字句為主體，間用六字句和八字句，且句末/gan1/（間）、/danm3/（擔）押韻。譯文以三音步為主體，間用四音步，意在模仿其參差不齊的句式；單行不押韻，二、四行押韻，亦與原唱詞相仿佛。

又如第四場《九裡山》有虞姬唱長花下句：

一發系千鈞，為防留破綻，  
 估道大王人猛勇，誰知才略亦非凡。  
 斬草再難留根患，深冀一戰功成指顧問。  
 拈杯含笑祝升平，莫負萬民同切盼。  
 Much hangs on this, t'is best to move with care.  
 I know my lord has valor, and now stratagem as well.  
 The roots of evil to cut in one blow fell,  
 One brave battle to destroy the enemy's lair.  
 I drink to my lord's enterprise.  
 The peace that our people aspire.

原唱詞以七字為主體，間用五字、九字，系典型的長短句形式，句末/zan6/（綻）、/fan4/（凡）、/men6/（問）、/pan3/（盼）押通韻。譯詩亦效仿原唱詞長短不齊的句式，但節奏分明。在尾韻上，譯詩改為第一四行、二、三行押韻，五、六行押中韻。本來在詩中換韻就意味著情緒的轉換，原詩以開口的/aan6/音結尾，譯詩尾韻換成了雙母音/eə/、短母音加舌折音/el/以及中韻/ai/，其效果同樣也發生了變化，不過這種韻式倒是為西方讀者所熟悉，容易被接受。

又如霸王唱反線《百花亭鬪酒》一段：

酒中見肝膽，痛飲休怠慢，  
 他朝九裡山，滅漢於一旦。  
 Courage shines through in the **wine**,  
 Gladly I quaff this cup of **mine**.  
 And then on top of Jiuli **Hill**,  
 The King of Han I will **kill**.

所謂反線就是在正線的基礎上改弦變調，以改變調高。這裡的詞每句五字，且一韻到底。譯者則改成每兩行押韻，每行為四音步，這又與原詩每句的三音組大致相仿佛。只是原本響亮的/danm2/（膽）、/man6/（慢）、/san1/（山）、/dan3/（旦）結尾，分

別被換成了開口的雙母音/ai:/和短母音加舌折音/iI/。從聽覺效果著眼，如果說前者尚系勉強接近原作，後者的效果就稍打折扣了。可見，在翻譯中為了押韻而押韻，固然在形式上看來比較順眼，但在詩學效果和接受效果方面就有缺陷。

## 五、氣勢與情感之傳譯

正如詩歌是“形”(form)與“質”(substance)緊密結合的整體。詩的“形”就是詩中的音節和詞句的構造，詩的“質”就是詩人的感想情緒。詩的“質”是情。“質”是詩的核心，也是詩的靈魂。詩的“形”始終都得服務於意蘊的傳遞。同樣，在戲曲表演藝術中，氣勢或情感無疑屬於“質”的方面，這些又是可以借助於“形”即詞句和聲音效果來營構的。受她早年翻譯氣魄雄渾的莎劇《李爾王》之影響，黎翠珍無論是在欣賞話劇、京劇還是粵劇時，最看重的還是劇中的那些對話節奏、威力和氣勢。這點首先可以從第二場《鴻門會》項羽責備劉邦和劉邦回話中領略一二：

項羽 [古口] 哼，你休要狡辯，我來問你，你先進關中，不放諸侯進入，其罪一也，緊封倉庫，其罪二也，子嬰來降不殺，其罪三也，如今事實證明，你還敢呼冤叫枉。

XIANG Yu [Speaks] Huh! No sophistry! Let me tell you: you went to Guanzhong and no other generals admitted. That's the first charge. You locked up the granary. That's the second charge. When Ziyang came to surrender, you did not have him executed. That's the third. These are facts you cannot deny. I wrong you not.

劉邦 [古口] 臣不放諸侯入關，只為提防搶掠，替將軍保境安民，其德一也，緊封倉庫，待將軍來獻，其德二也，體將軍廢秦苛政，揚將軍除暴之心，其德三也，使百姓知道沛軍尚且如此，上將軍一定倍施仁政，是揚將軍美名，其德四也，子嬰降而不殺，是等候將軍到來，判決其罪，劉邦不敢僭越，其德五也，試問我何來有罪，望將軍仔細參詳。

LIU BANG [Speaks] I did not let other generals in, to prevent pillage and to keep the city's peace, to your honor and merit. Two, I locked up the granary to offer to your triumphant entry and to your merit. I freed the people from Qin rule in your name. That's three. To let people know that even your vassal gave them peace, you, sir, would most certainly surpass this. Honoring your name and merit, that's four. I did not execute Ziyang when he surrendered was to await your pleasure for I would not bypass your authority. That fifth merit. Pray, tell me, what guilt do I bear.

項羽責備話中連續使用了“其罪一也”“其罪二也”“其罪三也”來推進，對應的英文分別為“*That's the first charge*”，“*That's the second charge*”和“*That's the third*”三個結構相同的簡單句。其他各句也以簡單句為主，這種短句念來很有力，且句中以使用單

音節的詞為主，其次才是雙音節詞，三音節的詞甚少，這樣的語篇讀來節奏分明，很有氣勢。

劉邦的回答中連續使用了“其德一也”“其德二也”“其德三也”“其德四也”“其五德也”來推進語篇，黎譯的對應英文分別是“to your honor and merit”，“Two”，“That’s three”，“that’s four”，“The fifth merit”，這樣的表達中有變化，但列舉數字明確；且其他各句亦為簡單句，相對於項羽的語句稍長，但在一些句子中運用插入語如“sir”和“tell me”，其意在增強語句的節奏感，讀起來很有氣勢。

如果說項羽和劉邦的表達均很有氣勢，代表的是男性話語的特點，那麼在傳達女性話語中的複雜的情感時，譯者則運用了其他策略來分別對待。如在第六場《霸王別姬》中虞姬聽到四面楚歌之前有一段獨自幽怨的歌唱：

虞姬 [歎氣浪白] 戰雲彌漫，秋雨淋鈴，春山愁鎖，秋水恨凝。

[唱反線《婢女淚》]

一片風聲與鶴唳，山川皆血腥，  
楚漢爭，苦戰為停，冒險進軍一著差，  
垓下困，皆因貪勝造成，  
大王太自持授人以柄，  
漢軍早佈陣形，韓信敗裡取勝，  
痛被圍困，楚王一怒揮軍反撲敵營，  
金風卷起悲歌聲，恨去愁來何時寧，  
未見王歸倍心驚，（歎氣）

[頹頭花上句]

只見暮雲四合夜蒼茫，  
但覺坐甯不安時行不定。（彷徨焦躁介）

YU JI [Sighs and speaks]

Clouds of battle loom dark, and autumn rain drizzles. The hills are shrouded in sorrow and water does not flow.

[Sings: Faansin “The Maid’s Tears” ]

The wind walls and birds fly screeching.

The heels and rivers are bleeding.

This struggle between Chu and Han

Has yet to end.

One rash a trap,

The price of pride overweening.



The king's confidence turns into a flaw.  
The Han army has set a trap  
For Han Xin to reverse his fate.  
Angry at being snared,  
Our King now turns on their base.  
The autumn wind brings sounds of weeping,  
When will sorrow and hatred be ending,  
Anxious I await the King's Returning. (*Sigh*)  
[Sings: Gwanfa upper line without prelude]  
As evening clouds gather in darkness,  
I wait in anxious yearning. (*Agitated*)

在翻譯開唱前的景物描寫的時候，其用詞非常準確。如將“彌漫”譯為“loom dark”，“loom”一詞非常生動形象，有四面楚歌，如履薄冰的戲劇感和畫面感；將“愁鎖”譯為“shrouded in sorrow”，“shrouded”被動的用法很妙，讓人有被哀愁包裹的感覺。再來看整個唱段各行不押韻，各行長短也不整齊，但其節奏分明，意在模仿原詩的節奏；其次在用詞方面，“風聲鶴唳，山川血腥”中的動詞譯得更是傳神，其中“screeching”一詞本身發音偏尖細，讓人不寒而慄，如芒刺背，符合原文危如累卵的景象；“bleeding”一詞對譯“山川染血”，觀眾的緊張感和刺激感油然而生。至於虞姬所唱“反調”的部分則含蓄地表達出霸王此戰已是有去無回，哀怨與憤懣，溢於言表。其唱段表現的是幽怨之情，結尾字基本上是以-ing 為韻母，分別是：鈴（ling4）、凝（jing4）、腥（sing1）、爭（zang1）、停（ting4）、成（sing4）、柄（bing3）、形（jing4）、勝（sing3）、營（ying4）、聲（sing1）、寧（ning4）、驚（ging1）、定（ding6），這種鼻腔共鳴的發音可以直抒胸臆，表達鬱結於心的怨恨。針對這些，譯者儘量採用了一連串的-ing 結尾的字，意在使聲音和意義都變得微弱和猶疑，因為在漢語中 an、en、in、un、ing、eng、ong 等鼻音韻，音色聽起來就比較冷淡，而且這些鼻音韻拖腔來唱，能將原文那種幽怨之情充分地再現出來。而且譯文也以採用單音節或雙音節詞為主，為的是與原詩行取得等時的效果。

總之，由於粵劇厚載著中華文化，這中間的許多因素是難以傳譯的，然而黎翠珍卻通過反復聆聽粵劇原唱腔，嘗試各種變通和協調措施，在形式上儘量接近原作，力圖再現原作的音樂性。然而，畢竟中文系表意性文字，英文系表音文字，試圖將英文中的節奏和韻式直接照搬過來傳譯絕無可能，但求其大致相仿佛還是可能的，至少在形式的傳譯上是如此。不管怎麼說，上述種種變通手段也是譯者在協調中西兩種詩學體系之後做出的抉擇。

## 六、結語

文化走出去是當今中國社會探頭打量的時尚，也是主流意識形態重點打造的文化工程。如何傳播中國聲音，講好中國故事，展示中國形象，提升中華文化軟實力，這些都是重要的命題。粵劇作為中國文化藝術的一個瑰寶，已被列入聯合國教科文組織“人類非物質文化遺產代表作”名錄，並在世界範圍內擁有一批忠實的追隨者。如何將粵劇藝術更好地推向國外，也成了新時代翻譯家的重要使命。而今黎翠珍英譯粵劇《霸王別姬》做出了一個表率。其譯本原是供粵劇團做字幕之用，不是直接用於演唱的，但是黎翠珍卻能在翻譯中嘗試協調中英兩套詩學規範，充分傳遞了原作的信息和音樂效果，同時也照顧到西方讀者的閱讀和欣賞習慣，其譯本一方面可以作為劇本來閱讀，另一方面經過打磨後是可以作為腳本來排演的。她的譯本具有很強的可讀性、可表演性和可接受性。她的成功經驗無疑是值得借鑒的。最後需要指出的是，所謂“學術乃天下之公器”，藝術無國界，更何況是厚載著文化的粵劇藝術對外傳播。在這方面，香港粵劇發展基金做出了表率。近十幾年來，該發展基金致力於粵劇藝術的推廣和資助，先後贊助和翻譯數量眾多的粵劇作品，這些作品均可以在公開網站上免費閱讀。這樣無疑有助於粵劇藝術的推廣和傳播，有助於提高國家文化軟實力，這種做法值得我們借鑒和效仿。

## 參考文獻

- Bassnett, Susan. 2002. *Translation Studies*. 3rd ed. London: Routledge.
- Hatim, Basil and Ian Mason .1997. *The Translator as Communicator*. London & New York: Routledge.
- Toury, Gideon.1980. *In Search of a Theory of Translation*. Tel Aviv: The Porter Institute for Poetics and Semiotics.
- 曹廣濤、彭可欣、張派豪。2018。英語粵劇翻譯及其發展走向簡論。《英語廣場》5: 29-30。
- 陳宇。2020。非物質文化遺產嶺南粵劇英譯研究現狀分析。《英語廣場》18: 29-30。
- 黃天驥。2003。《詩詞創作發凡》。廣州：廣東人民出版社。
- 蘭青編。2014。《中國傳統音樂概論—戲曲與說唱音樂》。瀋陽：東北大學出版社。
- 黎翠珍、張佩瑤。2005。〈黎翠珍與張佩瑤翻譯對談〉。黎翠珍，《搖搖一生》，19-83。香港：香港浸會大學翻譯學研究中心、國際演藝評論家協會（香港分會）聯合出版。
- 李燕霞、曾衍文。2018。跨文化傳播視域下嶺南粵劇的譯介與流變—以粵劇經典〈帝女花·香夭〉英譯為例。《四川戲劇》9: 18-23。
- 潘福麟。2000。粵劇的外文翻譯及其它。《南國紅豆》6: 40-41。
- 吳鳳平、陳鈞潤編。2013。《葉紹德粵劇劇本精選》。香港：香港大學教育學院中文教育研究中心。

張鼎程 張旭

楊浩然、張映先。2021。粵劇英譯與文化傳承——以〈牡丹亭·驚夢〉譯本為例。海外  
英語 7: 172-173+175。

# Compositional Translatology: Unveiling New Perspectives on Translating Chinese Classics into English

*Xiangyu Yu    Yifan Zhu*

**Address:** <sup>1,2</sup>School of Foreign Languages, Shanghai Jiao Tong University, China

**E-mail:** <sup>1</sup>yuxiangyu113@sjtu.edu.cn; <sup>2</sup>aliciazhu@sjtu.edu.cn

**Correspondence:** Yifan Zhu

**Citation:** Yu, Xiangyu, and Yifan Zhu. 2022. “Compositional Translatology: Unveiling New Perspectives on Translating Chinese Classics into English.” *Translation Quarterly* 105: 131-135.

**Wenguo Pan.** *A Comprehensive Course for Rendering Chinese Classics into English*. 2021. 867pp. ISBN: 978-7-5760-2028-1. Shanghai: East China Normal University Press.

This monograph, authored by Pan Wenguo, features an eclectic and panoramic approach to the translation of Chinese classics. It is richly imbued with a profound cultural consciousness while prioritizing a rigorous comparative viewpoint. Based on the author’s teaching material for the postgraduate course “Chinese Classics in English Text” at the Academy of Sinology, the University of Wales Trinity Saint David, this book primarily represents his ideas on “Wenzhang Fanyi Xue” (文章翻譯學, Compositional Translatology) derived from Chinese traditional discourse of translation and the accompanying view on how to provide a good translation of Chinese classics in practice.

Dexterously divided into two parts, the book comprises a total of fourteen chapters. The first six chapters are mainly devoted to the theoretical foundation and the essence of “Compositional Translatology,” while the remaining eight chapters focus on how “Compositional Translatology” informs the translation of Chinese Classics in practice.

The first part, entitled “Lilun Pian” (理論篇, Theory Module), contains six chapters. Chapter 1 mainly answers the questions of what the core of Chinese culture is and what should be translated through a historical comparison of the genealogies of knowledge in European and Chinese traditions. In contrast to the Dewey Decimal Classification (DDC), Library of Congress Classification (LCC), and Chinese Library Classification, the author expresses his preference for the traditional fourfold classification of Chinese classical works: “Jing” (經,

classics), “Shi” (史, historiography), “Zi” (子, schools of thought and miscellaneous arts), and “Ji” (集, collections). This bibliographic product of Chinese civilization highlights coherence and consistency with the evolution of traditional Chinese culture (p. 12). Pan emphasizes that the Six Canons: *Shi* (詩, The Canon of Poetry), *Shu* (書, The Canon of Ancient Documents), *Li* (禮, The Canon of Rites), *Yi* (易, The Canon of Change), and *Chunqiu* (春秋, Spring and Autumn Chronicle), as well as *Yue* (樂, The Canon of Music), which together make up “the six King” (Legge 1861, xvi), are recognized as the cornerstone of the intellectual structure of top authority in China (29) and reflect the core feature of Chinese culture—“Zhidao” (治道, Tao of National Governance) (26). Pan has also made an effort to sort out the voluminous ancient Chinese classics by compiling a table that further categorizes “Jing,” “Shi,” “Zi,” and “Ji” into 25 classes, including 96 sections. Within this framework, the author draws up a long list of 260 (and a concise version of 60) recommendable classical books for reading and translation.

In Chapter 2, Pan defends and justifies the role of Chinese translators in spreading Chinese works globally while proposing a refreshingly systematic principle for rendering Chinese classics into English. This principle includes contextualizing interpretation within traditional Chinese culture, distinguishing the root and branch of the Chinese intellectual structure, examining the genuine original text, and clarifying the terminology of specific cultural terms (69). The author warns against the tendency of “Geyi” (格義, naturalizing) Chinese notions with modern Western terminology (77). Thus, he encourages representing the cultural “otherness” by creating “Zhengming” (正名, nomenclature) for Chinese culture in English, along with an explanation of its original meaning in a Chinese context. The author demonstrates this idea by presenting and discussing his own translation version as an example (Pan 2019, 79-109).

In the next three chapters, Pan provides readers with briefings on how translation studies have historically and presently grown and thrived in both the Anglo-European context and China. Pan takes a comparative perspective to examine the two different translation traditions with the intention of proposing the theory of “Compositional Translatology” in Chapter 6. In Chapter 3, the author critically analyzes the foundation and developments of Western translation studies and its five major development stages. Pan, taking a progressive perspective, focuses his discussion on key figures of each stage and their contributions to the field.

In Chapters 4 and 5, Pan shifts his focus to Chinese translation activities and discourse. He first divides the history of Chinese translation into two stages, with the year 1898 as a watershed moment when Yan Fu famously presented the three challenges or principles of translation in his groundbreaking work “Yili Yan” (譯例言, Self-Note on Translation Strategy) during his Chinese translation of Huxley’s *Evolution and Ethics* — “Tianyan Lun” (天演論, On the Natural Evolution). This event is renowned for marking the beginning of modern Chinese translation theories (195). Specifically, in Chapter 4, Pan details the lineage and characteristics of Chinese practice in translating Sanskrit Buddhist sutras and reviews the efforts of pioneers in the Ming and Qing Dynasties to introduce China to Western scientific achievements. More

importantly, from a Chinese exegetical approach, Pan dispels the persistent myth surrounding Yan Fu's three translational principles of "Xin" (信, sincerity), "Da" (達, communicability), and "Ya" (雅, norm) (p. 226). He acknowledges that Yan's thoughts on translation epitomize the long-established tradition of translation in China and lay a solid foundation for "Compositional Translatology." Chapter 5 begins with robust debates such as "literal translation versus liberal translation," caused by indiscriminately modeling Chinese translation practice on Western analogous concepts before addressing the issue of Europeanized Chinese. The chapter also provides insight into how translation as a field of study has developed through representative scholarly writings over the past century. Pan reviews attempts to acculturate foreign terms or names into the Chinese language and addresses the topic of how poems should be translated and what theories have been advanced in that regard.

Built on the extensive survey of Western and Chinese translation studies that preceded it, in the last chapter of Part One, Pan finally introduces "Compositional Translatology" (340) as a unique theory system that has grown out of the Chinese translation tradition and practice. This theory system possesses three quintessential Chinese characteristics: "Li Gaobiao" (立高標, aiming high) <sup>[1]</sup>, "Zheng Yiming" (正譯名, scrutinizing nomenclature) <sup>[2]</sup>, and "Zhong Wencai" (重文采, highlighting literary grace). According to Pan, "Compositional Translation" has its roots in the "Chinese academic tradition of language and poetics" (Lin 2015, 96), which emphasizes the translators' sincerity before they commence their work and maintains that "all translation needs a careful post-process of the text to suit the compositional requirements of the target language" (Pan 2019, 94). Within the framework of "Compositional Translatology," the author further proposes "Yiwen san he: yi ti qi" (譯文三合義體氣, mapping translation onto the source text in meaning, style, and rhythmic harmony) (348), a set of principles for translation practice and translation criticism.

Entitled "Yinyong Pian" (應用篇, Application Module), "Part Two" focuses on how to apply "Compositional Translation," which was examined in the previous part, to translation practice, especially with a pedagogical purpose. Starting with "intra-lingual translation" (Jakobson 1959, 233), which involves translating classical Chinese texts into vernacular Chinese, an essential step before translating Chinese classics into English (p. 383), Pan first acquaints readers with basic knowledge of locution, sentential construction, and modal particles, among other elements, in traditional Chinese. Then he discusses "interlingual translation," covering the translation of English prose and Chinese philosophical excerpts. For the former, Pan suggests a general guideline of "Meiwen Meiyi" (美文美譯, graceful translation for graceful letters) (Pan 2019, 94) to capture the flavor of the original prose. For the latter, he emphasizes the importance of "Mingbai" (明白, Understandability), "Tongchang" (通暢, Readability), and "Jianjie" (簡潔, Succinctness) (441) in creating a successful translation of the profound wisdom of Chinese ancient sages.

In the second chapter, Pan delves into translation criticism in philosophical classics, epit-

omized by *Laozi* (老子), and in poems, with a focus on the singular and plural forms of nouns. Remarkably, the author ingeniously compares four translation versions of *Laozi* (469-71), illustrating the challenges translators face in nomenclature. He also explores ways to analyze and handle Chinese-specific cultural terms, using “Jiangnan” (江南, South of the Yangtze River) as an example.

In Chapter 3, Pan demonstrates how to teach a translation class with Chinese classics as the material, drawing from his personal anecdote. The author provides detailed, incisive commentary on students’ versions of the first ten chapters in *Suishu jingjizhi zongxu* (隋書·經籍志·總序, General Preface, Canons and Classics: A Bibliographical Record, Book of Sui) (515).

In the subsequent chapter, as a means to help improve students’ proficiency in both traditional and vernacular Chinese, the author analyzes six popular Chinese classic essays and prose before justifying and criticizing each of his own rewordings in turn. Chapter 5 contains four literary works by William Hazlitt, often referred to as “England’s greatest essayist” (598). In his translations, Pan reiterates his principle of “graceful translation for graceful letters” and demonstrates the subtlety of traditional Chinese in English-Chinese translation practice.

The next two chapters are adaptations of Pan’s previous translated work—Philosophical Maxims of 2000 Years Ago (Zhao 1998), with many translations modified for the purpose of this book. Chapter 6 includes extracts from the Five Canons and the Four Books—*Daxue* (大學, The Great Learning), *Zhongyong* (中庸, The Mean), *Lunyu* (論語, Analects), and *Mengzi* (孟子, Mencius), while Chapter 7 mainly contains selections from the magnum opuses of Confucianism, Daoism, Legalism, and Mohism, such as *Yanzi chunqiu* (晏子春秋, Yan Zi’s Spring and Autumn Annals), *Zhuangzi* (莊子), *Guanzi* (管子), *Mozi* (墨子), etc., along with other well-known works of Pre-Qin Period (Before 221 BCE) literature, like *Sunzi* (孫子) and *Chuci* (楚辭, Songs of Chu). The author provides brief but clear explanations of all excerpts’ original texts and their translations.

The last chapter of the volume is dedicated to poem translation, where the author chiefly presents English translations of forty-one well-chosen Chinese poetic masterpieces, including one by the author himself (865). At the end of the chapter, Pan also shares his version of two poems by William Shakespeare and William Wordsworth.

As a potential textbook for courses, this encyclopedic volume covers both theoretical considerations in translation studies and perceptive case studies of translation practice. Written in lucid and expressive language, it avoids many abstract scholarly concepts, making it easily understandable even for translation beginners. In the book, the author unreservedly recommends a list of Chinese classics worth reading and translating, with a view to providing purposeful classroom activities. Overall, Pan’s well-researched and well-argued work represents an invaluable contribution to the translation studies of Chinese classics and will inspire more focused studies on the materials and topics touched on in the book. It is highly recom-

mendable not only for novice or expert translators but also for a larger readership interested in the translation tradition in China.

While the book's name suggests it is designed for Chinese-English translation, not the opposite, the author appears to disproportionately allocate too many pages in Part Two (the second section of Chapter 1, the entire Chapter 5, and the last section of the last chapter) to the translation of English essays or poems, which may confound readers' expectations.

In the book, the author has introduced "Compositional Translatology" and explained what it entails and how it functions in translation practice and criticism. However, it would have been more beneficial and applicable if more evaluation parameters or indicators had been specified. Besides, the framework of the "three-mapping" principles does not seem as "workable" as argued by Lin (2015, 75). For instance, according to the author, "qi" (氣) in the theory, on one hand, clearly refers to rhythmic harmony and balanced sentence length throughout the text (367); on the other hand, "qi" also connotes "shen" (神, spirit or animation) (368) and "wei" (味, flavor) (370), making it somewhat opaque or elusive for translators to adhere to.

## Notes

- [1] In Chinese translation tradition, translators, as intellectuals, who typically show a sense of mission, regard translation as a way to partake in government administration.
- [2] The problem of incommensurability is especially salient between European languages and Chinese, partly due to the lack of contact between European and Chinese cultures up to 19th Century, and also owing to the different textual tradition of these languages. Thus, the scrutiny of nomenclature was and still is key to translation involving Chinese and phonogramic languages.

## References

- Jakobson, Roman. 1959. "On Linguistic Aspects of Translation." In *On Translation*, edited by Reuben A. Brower, 232-239. Cambridge: Harvard University Press.
- Legge, James. 1861. *The Chinese classics: with a translation, critical and exegetical notes, prolegomena, and copious indexes*. London: TRÜBNER & CO.
- Lin, Yuanbiao 林元彪. 2015. "The Continuation of Chinese Traditional Approach Towards Translation Studies—On Pan Wenguo's 'three Matching' Theory of Translating and Translation." *Foreign Language Learning Theory and Practice* (02), 73-79+96.
- Pan, Wenguo 潘文國. 2019. "Name and Nature of Compositional Translatology." *Shanghai Journal of Translators* (01), 1-5+24+94.
- Zhao, Changping 趙昌平 eds. 1998. *Er qian nian qian de zheyen 二千年前的哲言 [Philosophical maxims of 2000 years ago]*, translated by Pan Wenguo 潘文國. Shanghai: Shanghai Chinese Classics Publishing House.





## 關於宗教翻譯的幾點思考： 翻譯的宗教性和宗教的翻譯性

羅選民

各位好！歡迎參加第四屆宗教經典翻譯理論與實踐研討會！這是中國海洋大學第二次承辦該會議，感謝海大外國語學院為宗教翻譯研究所做的貢獻！

2013年，我們在這裡，就在這個會議室裡，舉行了第二屆宗教經典翻譯理論與實踐研討會。今天我們又匯聚一堂齊研共討宗教翻譯。諸位隸屬不同的院校、分屬不同的民族、面向不同的學科，涉及不同的語言、不同的文化、不同的宗教，採用不同的視角、不同的方法、不同的文本，這充分說明了宗教文化的多元性和翻譯研究的跨學科性，也充分體現了我們會議的宗旨：“開放、包容、平等、學術”。我記得上一屆會議以宗教界人士為多，而本次參會的主要是翻譯界學者。除了宗教學界和翻譯學界應盡量多切磋琢磨之外，我們還可以邀請更多具有相異學科背景的專家來參加討論，吸納眾多學界人士關心和關注宗教翻譯。因為，宗教關乎人類精神的依託，翻譯關乎人類精神的交流，兩者干係人類文明的發展，其交融和結合無疑具有極為重要的學術價值和社會意義。

宗教翻譯大有可為，是翻譯學為數不多有待開拓的疆土之一。人類翻譯的歷史悠久，宗教典籍翻譯構成了人類歷史上最為古老、歷時極為漫長的翻譯浪潮。我們傳統的翻譯研究往往著眼於一般意義上的翻譯原則和標準的討論，聚焦於翻譯中文本風格傳達和語言轉換技巧等話題，鮮有涉及宗教翻譯的宗教性和社會性。現在，翻譯已成為了一門具有獨立身份的學科，其跨域發展迅速，方法和視角不斷豐富多元，分支學科逐漸增多，從結構和規模上來看，翻譯學已儼然大矣。就目前翻譯學所涉及的方向領域而言，宗教翻譯研究可謂一塊大的亟待開墾的學術疆域。

根據蓋洛普國際調查聯盟2012年的“全球宗教信仰和無神論指數”調查資料，全球有59%的人信教；美國《國際傳教研究公報》2009年發佈的資料則顯示，信教人口甚至佔到了全球人口的85%（加潤國2017）。雖然這兩份資料有一定的出入，但足以說明宗教在當今世界各國、各民族生活中的重要性。我們本次會議的召集人之一、中國宗教學會會長卓新平教授從事了三十多年的宗教研究工作，他對此有著深切的體察。他曾在一次採訪中說，現在宗教領域日漸以正常面目重回中國人的日常生活與思考視域，我們對之應當“脫敏”，不要把它看作敏感的領域，而要把它看作正常的精神現象。宗教始終是我們必須妥善處理的重大問題，對於人類如此重要的精神活動，翻譯學理應參與其考察和研究，從譯學角度認識宗教、認識宗教翻譯、認識人類自身，以更好地推進世界文明的發展。

英語中有一個習語“An elephant in the room”（房間裡的大象），指顯而易見又被忽

略的事實。宗教翻譯就是這樣的一頭“大象”，它對於全人類精神世界的形成和發展都起著十分重要的作用，我們的語言、文化、思維方式無不深受其影響，但由於某些原因，我們對宗教翻譯的研究卻未予以相應的重視。我們首先應正視這頭“大象”的存在，其次要盡量了解它。單一個體從某個角度所摸到的可能只是這頭“大象”的一個局部碎片，但我們和不同學科的研究者一起努力，所謂“橫看成嶺側成峰”，從各個角度來探索，就有可能拼湊出這頭“大象”的輪廓全貌。

所謂宗教翻譯研究大有可為，主要基於宗教與翻譯緊密相關的事實：宗教的翻譯性和翻譯的宗教性。宗教的翻譯性指宗教作為人類精神的寄託所具有的跨民族、跨語言、跨文化傳播的特性。世界三大宗教的形成、各種宗教經典的反覆翻譯和廣泛傳播已充分地說明了這一點。人類有史以來翻譯得最多、傳播得最廣的文本就是宗教文本，有資料顯示，《聖經》全本已翻譯成約四百種語言，而部分節段甚至已被譯成了兩千多種語言，基督教一直藉由翻譯在傳播，其他宗教也莫不如是。

翻譯的宗教性指人類翻譯活動中宗教因素的重要性，宗教典籍在很長的一段時間裡都是人類翻譯活動的主流翻譯要件。翻譯文學發端於宗教典籍的譯介，我國唐詩、宋詞、元曲、明清小說，無一不受到漢譯佛典的影響，而在西方，各種《聖經》譯本也總是被奉為文學典範。現在我們進入了翻譯的職業化時代，翻譯也離不開宗教因素，後者已經滲入到了我們的靈魂深處，融入到了我們的日思夜想，即使是無意識的驚呼“天啊”、“My God”等，都與宗教密切相關。因而，宗教翻譯不僅包括對宗教性文本翻譯的研究，也包括對非宗教性文本翻譯中所涉及的宗教因素的討論。後者賦予了宗教翻譯研究更為廣闊的學術空間。近年來在文學翻譯研究領域，宗教因素開始受到越來越多的關注。

我曾在一次翻譯研討會上講過典籍外譯中的文化自覺，其中談到了典籍翻譯選材應遵循的幾個原則：普適性、契合性、思想性、工具性。普適性指外譯選材時要找那些具有人類精神普適意義、可以在世界範圍內得以推廣的典籍；契合性指外譯選材要發現自己喜歡，同時也契合譯入語讀者興趣的作品；思想性指外譯選材要優先選取具有豐富思想內涵以及能啟迪讀者思想的作品；工具性指外譯選材應為現實服務、為滿足社會需求服務。宗教經典文本就符合上述所有原則的要求。通常而言，一個國家越是強大的時候，也越是其宗教翻譯興盛發達的時候。現在我國正處於努力實現中華民族偉大復興的歷史時期，物質變得越來越豐富，人們在精神層面的需求也將越來越強烈，相應地對宗教翻譯的需求也必然會越來越迫切。

宗教的翻譯性意味著多元宗教文化之間具有可交流性，也意味著信教者和非信教者之間的可理解性，還意味著信徒和“神（上帝）”之間的可溝通性。文化間的交互和人之間的理解都要基於相互的尊重，而人與神之間的溝通則要基於信仰和敬畏。從事宗教翻譯要懷有敬畏之心，以嚴謹的態度對待宗教經典；要從元角度看待宗教性文本，做到知行合一，追求事物的本源。譯文文字講究通曉易懂，保持原文本的觀點和風格，避免用詞過於絢麗，慎用帶有強烈情感傾向的詞語，以保證和維護宗教翻譯的宗教性。

“歷史表明，社會大變革的時代，一定是哲學社會科學大發展的時代。當代中國

正經歷著我國歷史上最為廣泛而深刻的社會變革，也正在進行著人類歷史上最為宏大而獨特的實踐創新。這種前無古人的偉大實踐，必將給理論創造、學術繁榮提供強大動力和廣闊空間”（習近平 2016）。在“一帶一路”和“走出去”戰略的實施過程中，翻譯將起到十分關鍵的作用，翻譯過程中如何處理所涉的宗教因素則是我們翻譯工作者必須面對的重要問題，翻譯的宗教性也就必然成為了我們翻譯研究的重大課題。關懷現實是學術研究的本義所在，也是理論創新的動力源泉，翻譯的宗教性和宗教的翻譯性無疑將成為翻譯學理論創新的雙核動力，也自然會為宗教研究帶來新的主題、視角和維度。

今天，我們的宗教翻譯理論與實踐研討會已經舉辦到了第四屆。我相信，宗教翻譯將吸引越來越多的學者予以關注，我們的宗教翻譯實踐者和研究者的隊伍將越來越壯大，隨著這一領域的發展，我們還要成立學會、創辦刊物，將我們的宗教翻譯研究建構成“宗教翻譯學”。

**銘謝** 本文為國家社科基金西部項目《翻譯理論在世界文學研究中的價值凸顯研究》（項目編號：20XWW005）的部分成果。<sup>[1]</sup>

## 注釋

[1] 該文為第四屆宗教經典翻譯理論與實踐研討會的致辭。

## 參考文獻

- 加潤國。2017。〈全球信教人口有多少〉。中國社會科學院馬克思主義研究院編，《馬克思主義無神論研究（第5輯·2015）》。北京：中國社會科學出版社。
- 習近平。2016。《在哲學社會科學工作座談會上的講話（全文）[EB/OL]》。[http://news.xinhuanet.com/politics/2016-05/18/c\\_1118891128.htm](http://news.xinhuanet.com/politics/2016-05/18/c_1118891128.htm).



## Translation Quarterly Editorial Policy

1. *Translation Quarterly* is a major international scholarly journal published by the Hong Kong Translation Society. We welcome academic articles, translated works and book reviews in all areas of translation studies written in Chinese and English, as well as research materials involving other language combinations.
2. All contributions should be hitherto unpublished and not being considered for publication elsewhere. Authors take responsibility for their views.
3. Translated works should be submitted with a copy of the source text and a brief introduction to the source-text author. It is the translator's responsibility to obtain written permission to translate.
4. All submissions are **peer reviewed by the Editorial Board members and double-blind peer reviewed** by referees before acceptance for publication. Care should be taken by authors to avoid identifying themselves or acknowledging others in the submission, including the main text, notes and page headers; such information should preferably be supplied when the submission is accepted for publication.
5. Once a contribution is published, its copyright is transferred to the publisher; any reproduction, distribution or display thereof is subject to the written consent of the publisher.

### Guidelines for Contributors

1. Manuscripts, except for book reviews, should preferably be **no less** than 8,000 words (about 16 pages).
2. **Manuscripts should be accompanied by an abstract in English of 200–300 words**, with no more than **five** keywords, and a title page containing the author's name, affiliation, correspondence address, email address and telephone number.
3. Manuscripts written in a language which is not the author's mother-tongue should preferably be checked by a native speaker before submission.
4. Manuscripts should be submitted in word-processed electronic files to the Chief Editor at [translationquarterly@gmail.com](mailto:translationquarterly@gmail.com).

### Manuscript Stylesheet

\*Manuscripts should be consistent in their use of language, spelling and format. Once your paper is accepted for publication, it is essential that the final submission be formatted to

the specifications given in these guidelines to avoid publication delay. This journal uses the “Author-Date” style as described in the latest edition of *The Chicago Manual of Style*.

### Overview

1. Manuscripts should be typed in **12** point-font, in Times New Roman for English and PMingLiU for Chinese traditional font. **Allow double-spacing throughout.**
2. Use **10** point-font for figures and tables, which should be numbered consecutively in Arabic numerals and provided with appropriate captions, e.g., Table 1: Interview subjects.
3. Subheadings within the main text should be aligned expressly according to the following order: 1. / 1.1 / 1.1.1
4. Start each new paragraph with a single tab.
5. Quotations longer than **five** lines should be set off from the surrounding text in block form, indented from the left margin with a single tab. **Use font size 10.**
6. Notes, to be provided in the form of endnotes, should be kept to a minimum. Endnotes should be numbered consecutively using Arabic numerals within square brackets in superscript, e.g., [1]. Note indicators in the main text should appear after punctuation marks.
7. Proper nouns and personal names, in English or otherwise, should appear in their full original form where they first appear in the main text. Chinese names and book titles in the text should be romanised according to the Hanyu Pinyin, “modified” Wade-Giles or other pertinent systems, and then, where they first appear, followed immediately by the Chinese characters and translations **in parentheses. Translations of Chinese terms obvious to the readers (like wenzue), however, are not necessary.**
8. Book and periodic titles should be italicised, while titles of articles, chapters, poem and songs, etc., should be placed within double quotes (“ ”). The same applies for titles within titles.
9. Double quotes (“ ”) are used for quotations and special terms; single quotes (‘ ’) are only used for quotations within quotations.
10. Ellipses come in three dots, with a space between the dots, e.g., “Jack and Jill went up the hill . . .” and “Jack and Jill went up the hill . . . And Jill came tumbling after.”
11. Please use **American** spellings and punctuation, including spellings like -ze, -or, etc.
12. Dates should be of the form “1 July 2022.”

### References in the text

In-text citations in English and other languages should be as precise as possible, giving the author-date and, where applicable, page references. Here are some examples: (Bassnett 2014, 28-32); as in Chesterman et al. (2003, 198); (see Kurland and Lerner 1987, chap. 10, doc. 19); (Toury 1995/2012, 10).

### References section

All in-text citations in English and other languages should be matched by items in the references section, instead of listed in note form. All and only references cited in the text must be listed. References should be listed first alphabetically and then chronologically. For Chinese names and titles, please start a separate list or use Hanyu Pinyin, Wade-Giles transcription, the translated title as published or the title in literal translation where appropriate. Below are some examples:

#### Book

Bassnett, Susan. 2014. *Translation Studies*. 4th ed. London & New York: Routledge.

Chan, Tak-hung Leo, ed. 2003. *One into many: Translation and the dissemination of classical Chinese literature*. Amsterdam and New York: Rodopi.

Chan, Tak-hung Leo. 2004. *Twentieth Century Chinese Translation Theory: Modes, Issues and Debates*. Amsterdam and Philadelphia: John Benjamins Publishing Company.

Toury, Gideon. 1995/2012. *Descriptive Translation Studies – and beyond*. Rev. ed. Amsterdam and Philadelphia: John Benjamins.

Yang, Hsien-yi 楊憲益, and Gladys Yang 戴乃迭, trans. 1956. *Selected Works of Lu Xun*. Vol. 1. Peking: Foreign Languages Press.

#### Article (in book)

Fawcett, Peter, and Jeremy Munday. 2013. “Ideology.” In *Routledge Encyclopedia of Translation Studies*, edited by Mona Baker, and Gabriela Saldanha, 137–141. London: Routledge.

Liang, Qichao 梁啟超. 1984. “Fanyi wenxue yu fodian 翻譯文學與佛典 [Translated Literature and Buddhist Sutras].” In *Fanyi lunji 翻譯論集* [An Anthology of essays on translation], edited by Luo Xinzhang 羅新璋, 52-67. Beijing: Commercial Press.

#### Article (in journal)

Even-Zohar, Itamar. 1990. “The position of translated literature within the literary polysystem.” *Poetics Today* 11 (1), 45-51.

Keng, Shao-Hsun, Chun-Hung Lin, and Peter F. Orazem. 2017. “Expanding College Access in Taiwan, 1978–2014: Effects on Graduate Quality and Income Inequality.” *Journal of Human Capital* 11, no. 1 (Spring): 1–34. <https://doi.org/10.1086/690235>.



Other materials

Liu, Zhengyan 劉正琰, Mingkai Gao 高名凱, Yongqian Mai 麥永乾, and Youwei Shi 史有為, eds. 1984. *Hanyu wailaici cidian* 漢語外來辭辭典 [Chinese dictionary of foreign-imported words]. Shanghai: Shanghai cishu chubanshe.

Zhu, Xi 朱熹. 1270/2017. *Zhuyi yulei* 朱子語類 [Classified discussions of Master Zhu], compiled by Li Jingde 黎靖德. Chinese Text Project. <http://ctext.org/zhuzi-yulei/zh>.

\*For further guidelines, please consult the latest edition of *The Chicago Manual of Style*.