Ambiguity and Entropy in the Process of Translation and Post-editing

Yuxiang Wei

BSc, MA, MPhil

A Thesis Submitted for the Degree of

Doctor of Philosophy (PhD)

Supervisors:

Prof. Dorothy Kenny

Dr. Joss Moorkens

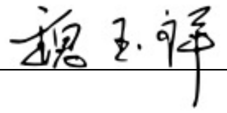
School of Applied Language and Intercultural Studies

Dublin City University

April 2023

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, and that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: 

ID No.: 17212925

Date: 22 April 2023

Acknowledgements

I am greatly indebted to the many people who have supported me in the process of my PhD research, especially my supervisors, whose patience, advice, and encouragement have all been crucial as I complete this thesis amid an unprecedented global pandemic.

My supervisors, Prof. Dorothy Kenny and Dr. Joss Moorkens, have both given me tremendous guidance during my study and research at Dublin City University (DCU), without which this thesis would never have been possible. The research for the thesis had gone through significant changes due to a series of constraints that were far from expected, the most prominent of which were the pandemic restrictions that prohibited any experiment from being carried out for the initial project. The extent to which this research was derailed and reshaped, the uncertainty and anxiety involved in the process, the difficulty therein, and the amount of support, guidance, reassurance, encouragement, and profound academic advice I have received from my supervisors along the way, are all beyond what words can describe.

My sincere gratitude also goes to Prof. Ricard Muñoz Martín and Prof. Sharon O'Brien, who examined my PhD and provided valuable, critical, and constructive feedback on my thesis.

I am grateful to the School of Applied Language and Intercultural Studies at DCU, for providing me with the scholarship, stipend, and various kinds of resources that are needed for this PhD research. The subsequent extension of my funding from DCU, in view of the disruptions to my research due to the pandemic, was no less crucial for the completion of this thesis.

I would also like to thank the organizers of two research boot-camps I have participated, first at the University of Macau in 2018 and then at Kent State University in 2019, and the funding that the organizers have granted me for my travel and accommodation during these month-long, extensive, and fruitful activities. Prof. Michael Carl, with whom I have had frequent communication on conceptual and empirical explorations in translation process research, especially deserves my gratitude.

In the light of the impossibility of generating new data during the pandemic, this thesis makes use of a publicly available database called the CRITT TPR-DB for its empirical investigations. In this regard, I would like to extend my gratitude to the network of researchers who base their studies on the framework of, and actively contribute experimental data to, the CRITT TPR-DB.

I am also grateful to Dr. Áine McGillicuddy, who served as the director of the Centre for Translation and Textual Studies (CTTS) and had regular meetings with postgraduate members including myself. During my PhD years, Dr. McGillicuddy had always been kind and encouraging whenever I had conversations with her, in both formal and casual settings. She also had annual meetings with me and my supervisors as an independent member of my supervisory panel. Through these meetings, the encouragement and motivation I have received had been truly valuable.

The postgraduate members of the CTTS, with whom I served on the organizing committee of the 15th International Postgraduate Conference on Translation and Interpreting, have all left me with precious memories of the time we worked together for the event — Jamie Murphy, Wichaya Pidchamook, Nguyen Nguyen, Boyi Huang, Rana Roshdy, Eddie López Pelén, and Shane Forde.

These years of my PhD study and research at DCU are memorable, and it is impossible to mention all the people who have left wonderful traces in my memory of this journey. My sincerest gratitude goes to all of them.

Last but not least, my parents deserve my most special thanks, as always.

List of publications and presentations from this research

Wei, Y. (2022). Entropy as a measurement of cognitive load in translation. In Proceedings of the 15th Biennial Conference of the Association for Machine Translation in the Americas (Workshop 1: Empirical Translation Process Research) (pp. 75-86): Association for Machine Translation in the Americas.

Wei, Y. (2021). Entropy and eye movement: A micro-analysis of information processing in activity units during the translation process. In M. Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 165-202). Cham, Switzerland: Springer.

Wei, Y. (2021, November). *Entropy-reduction as evidence of horizontal translation.* The 3rd International Conference on Translation, Interpretating & Cognition, University of Bologna, Forlì, Italy, 2-5 November 2021.

Wei, Y. (2020, November). *Post-editing of structurally ambiguous translation: The biasing effect from source text.* The 42nd Translation and the Computer (TC42) conference, online, 18-20 November 2020.

Wei, Y. (2019). Book review — New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB. *Journal of Translation Studies* 翻譯學報 *(New Series)*, 3(2), 122-127.

Wei, Y. (2019, December). *Entropy-based predictions of cognitive effort in translating and post-editing*. Translation Studies and Digital Humanities, The Chinese University of Hong Kong, Hong Kong, 9-11 December 2019 (cancelled).

Wei, Y. (2019, December). *Cognitive effort in translation production: An analysis on entropy and eye movements in the CRITT Translation Process Research Database.* The 2019 Chinese Studies PhD Candidates Workshop, Maynooth University, Maynooth, Ireland, 6 December 2019.

Wei, Y. (2019, November). *Uncertainty and cognitive effort: An investigation into the correlation of word translation entropy with eye movements in translation production*. The 15th International Postgraduate Conference in Translation and Interpreting (IPCITI), Dublin City University, Dublin, Ireland, 22-23 November 2019.

Wei, Y. (2019, August). *Predicting cognitive effort in translation production*. The 17th Machine Translation Summit, Dublin, Ireland, 19-23 August 2019. In M. Carl & S. Hansen-Schirra (Eds.), Proceedings of the Second MEMENTO Workshop on Modelling Parameters of Cognitive Effort in Translation Production (p. 17).

Wei, Y. (2019, August). *Impact of word translation entropy on translation activity units – – An analysis of eye fixations*. The Memento Boot-camp 2019 on Empirical Translation Process Research — Modeling Parameters of Cognitive Efforts in Translation Production, Kent State University, Kent, Ohio, USA, 9 August 2019.

Wei, Y. (2018, November). *Cognitive Effort in Relation to Word Translation Entropy and Syntactic Choice Entropy*. The 5th International Conference on Cognitive Research on Translation and Interpreting, Renmin University of China, Beijing, China, 3-4 November 2018.

do Carmo, F., & Wei, Y. (2018, October). *Studying Editing, Effort and Entropy in TPR-DB, the Translation Process Research Database*. The SALIS Seminar Series: Promoting Excellent Research and Teaching Practices, Dublin City University, Dublin, Ireland, 24 October 2018.

Wei, Y. (2018, July). *Ambiguity, Machine Translation, and TPR*. The Memento Boot-camp 2018 on Empirical Translation Process Research — Modeling Parameters of Cognitive Efforts in Translation Production, University of Macau, Macau, 8-31 July 2018.

Wei, Y. (2017, November). *Customising Machine Translation for Academic Texts: Issues of lexis and syntax*. Paper presented at the 13th International Postgraduate Conference in Translation and Interpreting (IPCITI), Heriot-Watt University, Edinburgh, Scotland, UK, 9-10 November 2017.

Table of Contents

Acknowledgements **2**

Abstract **14**

Chapter 1 Introduction **15**

1.1 Introduction *15*

1.2 Aims, questions, and scope *20*

1.3 Thesis outline *23*

Chapter 2 Ambiguity **27**

2.1 Introduction *27*

2.1.1 Lexical ambiguity 28

2.1.2 Structural ambiguity 29

2.1.3 Temporary ambiguity 30

2.2 Disambiguation in psycholinguistics *32*

2.3 Translation ambiguity in TPR *36*

2.4 Summary *37*

Chapter 3 Entropy **38**

3.1 Introduction *38*

3.2 Information entropy *39*

3.3 Selectional information-content *41*

3.4 Uncertainty and surprisal *42*

3.4.1 Surprisal 43

3.4.2 Surprisal, entropy, uncertainty, and informativity 44

3.5 Entropy in translation *46*

3.5.1 Word translation entropy 46

3.5.2 Syntactic choice entropy 49

3.5.3 Entropy from the sample 50

3.6 Summary *52*

Chapter 4 The process of translating and post-editing **54**

4.1 Introduction *54*

4.2 Models of the translation process *54*

4.2.1 Vertical and horizontal processing 54

4.2.2 Stratification models 57

4.2.3 Unique items hypothesis and the monitor model 60

4.2.4 Recursive model 66

4.2.5 Default translation 70

4.2.6 Other models of the translation process ... 72

4.3 Uncertainty, problem solving, and cognitive load ... 75
4.3.1 Uncertainty in the stratification model ... 75
4.3.2 Means-ends analysis ... 76
4.3.3 Cognitive load ... 79

4.4 Post-editing of Machine Translation ... 82
4.4.1 MT with human post-editing ... 83
4.4.2 Research on post-editing effort ... 84
4.4.3 Ambiguity in translation and post-editing ... 86

4.5 Machine translation approaches ... 87
4.5.1 Symbolic NLP approaches and Rule-based MT ... 87
4.5.2 Empirical NLP approaches and Statistical MT ... 89
4.5.3 Neural MT ... 91

4.6 Towards an entropy-based perspective ... 93

4.7 Summary ... 96

Chapter 5 Ambiguity: statistical approaches ... 97

5.1 Introduction ... 97

5.2 Statistical approaches to psycholinguistics ... 98
5.2.1 Rank-frequency distribution and Zipf's law ... 98
5.2.2 Principle of least effort and lexical ambiguity ... 102
5.2.3 Other theoretical explanations ... 104

5.3 Measuring ambiguity for processing effort ... 104
5.3.1 Surprisal, relative entropy, and cognitive effort ... 106
5.3.2 Surprisal and disambiguation ... 108

5.4 Entropy in the translation process ... 111

Chapter 6 Ambiguity, entropy, and cognitive processes ... 115

6.1 Introduction ... 115

6.2 Entropy and cognitive processes of translation ... 116
6.2.1 Entropy, mental states, and cognitive load ... 116
6.2.2 Surprisal, relative entropy, and ITra ... 121
6.2.3 Decrease of entropy and HTra ... 124
6.2.4 Entropy for the mental state and entropy in the text ... 126
6.2.5 Transition between mental states ... 131
6.2.6 Serial and parallel processing, and entropy ... 133
6.2.7 Systems theory perspective of the translation process ... 135
6.2.8 Surprisal and entropy in translation ... 135
6.2.9 Translation and post-editing ... 137
6.2.10 'Final' choice ... 137
6.2.11 Context and entropy ... 137

6.3 Summary ... 139

Chapter 7 Eye movements ... 141

7.1 Introduction ... 141

7.2 Areas of study using eye-tracking methods ... *141*

7.3 Eye movements ... *142*

7.4 Measuring eye movements ... *144*

7.5 Metrics to represent visual/mental effort ... *147*

7.6 Eye movements in reading and information processing ... *150*
7.6.1 Lexical issues influencing eye movements ... 151
7.6.2 Syntactic issues influencing eye movements ... 154
7.6.3 Eye-mind hypothesis ... 156

7.7 Summary ... *157*

Chapter 8 Lexical analysis ... 158

8.1 Introduction ... *158*

8.2 Dataset ... *161*

8.3 Entropy and fixations in Activity Units ... *163*
8.3.1 Machine Translation Post-editing ... 166
8.3.2 Effect of averaging ... 167
8.3.3 An example of scanpath ... 169

8.4 'Cough up': Analysis on activity units ... *171*
8.4.1 HTra values for AUs containing "cough" ... 172
8.4.2 Progression graph analysis ... 174
8.4.3 Translation tasks into different language ... 178
8.4.4 Dynamic change of HTra within the AU ... 189

8.5 Discussion ... *190*

8.6 Summary ... *194*

Chapter 9 Structural analysis ... 195

9.1 Introduction ... *195*

9.2 Table generation ... *197*
9.2.1 Translog-II and gaze-to-word mapping ... 197
9.2.2 Generation of summary tables ... 200

9.3 Dataset used for analysis ... *200*

9.4 Data inaccuracy and re-processing ... *201*

9.5 Structural ambiguity ... *202*
9.5.1 Sentence from MT ... 202
9.5.2 Structure of the raw MT output ... 206
9.5.3 Structure of the post-edited output ... 211

9.6 Qualitative analysis of the process ... *214*
9.6.1 Keyboard input ... 214
9.6.2 Eye movements ... 217
7.6.3 Detailed fixations on TT ... 219
9.6.4 Discussion ... 221

9.7 Statistical analysis ... *225*
9.7.1 Method of labelling ... 226

9.7.2 Reading time and fixation counts 230
9.7.3 Statistical analysis 235

9.8 Summary *238*

Chapter 10 Concluding remarks 240

10.1 Ambiguity, cognitive models, and statistical approaches *240*

10.2 Entropy *241*

10.3 Lexical ambiguity and entropy *244*

10.4 Structural ambiguity *245*

10.5 Summary *247*

10.6 Limitations and future work *248*

References 251

Appendix A Source text and Machine-translated output A-1

Appendix B Post-edited versions from participants B-1

Appendix C Data inaccuracy and re-processing of JIN15 .... C-1

C.1 Translog replay analysis *C-9*

C.2 Possible causes of the issue *C-21*

C.3 Data re-processing and error correction *C-32*

List of abbreviations

AU	activity unit
BIA+	bilingual interactive activation plus model
CL	controlled language
CLT	cognitive load theory
DeP	*de* (的) phrase/sequence
HCross	syntactic choice entropy
HTra	word translation entropy
IME	input method editor
MT	machine translation
nFix	number of fixations
NLP	natural language processing
NMT	neural machine translation
NP	noun phrase
PP	prepositional phrase
RNN	recurrent neural networks
SL	Source Language
SMT	statistical machine translation
ST	source text
TAP	think-aloud protocol
TL	target language
TM	translation memory
TPR	translation process research
TPR-DB	translation process research database
TT	target text
UAD	user activity data
VP	verb phrase

List of Tables

Table 7.1	Early and later processing measures
Table 8.1	Entropy values for AU 1.1
Table 8.2	Entropy values for AUs 1.2 and 1.3
Table 8.3	Entropy values for AU 2
Table 8.4	Entropy values for AU 3
Table 8.5	Entropy values for AU 4.1
Table 8.6	Entropy values for AU 4.2
Table 8.7	Entropy values for AU 5.1
Table 8.8	Entropy values for AU 5.2
Table 9.1	Frequency of each category of the post-edited result
Table 9.2	Number of participants who had each number of passes for the segment
Table 9.3	Example of TToken and Edit columns
Table 9.4	TToken and Edit columns for P08
Table 9.5	Keystroke data for P08
Table 9.6	Frequency and proportion of different labels for TT tokens
Table 9.7	ANOVA results for fixation data before editing

List of Figures

Figure 4.1	Recursive model of translation
Figure 5.1.	Condon's plot of the rank-frequency distribution
Figure 5.2	Zipf's plot of the number-of-words-frequency distribution
Figure 5.3	Probabilistic context-free grammar
Figure 5.4	Incremental analyses for sentence 6
Figure 6.1	Mental state when entropy is H_{max}
Figure 6.2	Mental state H_1
Figure 6.3	Mental state H_2
Figure 6.4	Entropy values in the text and in the mind when context is not considered
Figure 6.5	Entropy values in the text and in the mind when context is considered
Figure 6.6	Entropy after selection
Figure 6.7	An example of mental states in selection process
Figure 6.8	Illustration of AUs within and between typing bursts
Figure 8.1	Illustration of AUs within and between typing bursts
Figure 8.2	Scatter plot of nFix and average HTra
Figure 8.3	Scatter plot for each AU type
Figure 8.4	Effect of averaging
Figure 8.5	Scanpath 1.1

Figure 8.6 HTra values of AUs containing fixation on *cough*
Figure 8.7 Progression graph of Example 1
Figure 8.8 Progression graph for AU 1.1 and AU 1.2
Figure 8.9 Progression graph for AU 1.3
Figure 8.10 Progression graph for Example 2
Figure 8.11 Progression graph for Example 3
Figure 8.12 Progression graph for AU 4.1
Figure 8.13 Progression graph of the typing bursts after AU 4.1
Figure 8.14 Progression graph for AU 4.2
Figure 8.15 Progression graph for Example 5
Figure 9.1 Eye gaze points in Translog-II file (offline mapping)
Figure 9.2 Result after re-mapping
Figure 9.3 Visualization of offline mapping
Figure 9.4 DeP as adjunct to the head noun
Figure 9.5 Progression graph of a participant who has edited the segment once
Figure 9.6 Progression graph of a participant who has edited the segment three times
Figure 9.7 Keyboard behaviour of P09 at the start of editing on Sentence 1
Figure 9.8.a Progression graph 1 for P09
Figure 9.8.b Progression graph 2 for P09
Figure 9.8.c Translog screenshots for progression graph 1
Figure 9.9.a Number of fixations on TT tokens in terms of STid
Figure 9.9.b Total reading time on TT tokens in terms of STid
Figure 9.10.a Number of fixations on TT tokens in terms of category and STid
Figure 9.10.b Total reading time on TT tokens in terms of category and STid
Figure 9.11 TrtT and nFixT on different categories of TT tokens
Figure C.1 Progression graph 1 for JIN15 dataset
Figure C.2 Vertical drift in fixations in JIN15 dataset
Figure C.3 Progression graphs 2 & 3 for JIN15 dataset
Figure C.4 Progression graph 4 for JIN15 dataset
Figure C.5 Progression graphs 5 for JIN15 dataset
Figure C.6 Progression graphs 6-8 for JIN15 dataset
Figure C.7 Progression graph 9 for JIN15 dataset
Figure C.8 Progression graphs 10 & 11 for JIN15 dataset
Figure C.9 Huang's interpretation of the fixation points
Figure C.10 Screenshot for the replay in Translog-II
Figure C.11 Huang's Screenshot of Translog-II user interface
Figure C.12 Editing on the first sentence
Figure C.13 Fixations during editing of the first sentence
Figure C.14 The displacement of fixations for the first sentence
Figure C.15 Progression graph of P16 for the first sentence
Figure C.16 Editing on the second sentence
Figure C.17 Fixations during editing of the second sentence
Figure C.18 The displacement of fixations for the second sentence

Figure C.19 Progression graph of P16 for the entire experimental session
Figure C.20 Vertical drift in fixations in JIN15 dataset
Figure C.21 Huang's illustration of 'inaccurate data sample'
Figure C.22 Carl's illustration of the error in naïve mapping
Figure C.23 Carl's illustration of the re-computed mapping
Figure C.24 Carl's illustration of naïve mapping vs. re-mapped version
Figure C.25 Fixations when the bar is moved upward
Figure C.26 Huang's screenshot of the replay session
Figure C.27 Data correction for P09
Figure C.28 Data correction for P11
Figure C.29 Data correction for P14
Figure C.30 Data correction for P15
Figure C.31 Data correction for P16
Figure C.32 Data correction for P17
Figure C.33 Data correction for P18
Figure C.34 Data correction for P20
Figure C.35 Data correction for P21
Figure C.36 Data correction for P22

Abstract

Ambiguity and Entropy in the Process of Translation and Post-editing

Yuxiang Wei

This thesis analyses the way in which ambiguity is cognitively processed, in translation in general and post-editing in particular, drawing inferences from psycholinguistics, bilingualism, and entropy-based models of translation cognition. Conceptually, it assumes non-selective activation of both languages (source and target) in the translation process, and explores how entropy and entropy reduction can theoretically describe assumed mental states during disambiguation. Empirically, it uses a product-based metric of word translation entropy (HTra), and eye-movement and keystroke data from the CRITT Translation Process Research Database, to shed light on how the conceptual understanding of lexical and structural ambiguity may be manifested by observable behaviour.

At the lexical level, examination of behavioural data pertaining to a high-HTra item from 217 participants translating/post-editing from English into multiple languages shows that the item tends to result in pauses in production and regression of eye movements, and that the translators'/post-editors' corresponding scrutinization of the source text (ST) tends to involve a visual search for lower-HTra words in the co-text and, accordingly, a decrease in the average entropy of the activity unit. Regarding syntax, a Chinese relative clause in the machine translation output, which can involve a garden-path effect, is examined in terms of eye movements from 18 participants. Results show that, contrary to monolingual reading, disruptions of processing tend to occur not in the later part of the sentence where the wrong parse is disconfirmed, but in the earlier regions where the most quickly-built analysis is semantically inconsistent with the ST. Structural disambiguation and re-analysis seem to be bypassed. This suggests that, on the one hand, reading for post-editing receives a strong biasing effect from the ST, and on the other, argument integration is more appropriately explained from an incremental processing perspective rather than a head-driven approach, as thematic roles seem to be assigned immediately in reading for post-editing. While the lexical analysis supports a parallel disambiguation model, the structural analysis seems to support a serial one. In terms of translation models, both emphasize the impact of cross-linguistic priming and the presence of considerable horizontality in the translation process.

Chapter 1 Introduction

1.1 Introduction

Ambiguity

Language is inherently ambiguous, especially when context is lacking. Every word we use can have multiple meanings and senses, a single sentence can be subject to different analyses of its syntactic structure, and every paragraph or text can be interpreted in various manners depending on the knowledge, experience, cultural background, ideological stance, and many other factors of the reader or listener. The issue becomes increasingly complicated and prominent when pragmatic meanings come into consideration. The speaker or writer's intention, the communicative circumstance, and many other contextual issues all have various impacts on how an utterance or a piece of writing is interpreted and what meaning is made out of it. Meaning-making is almost always dynamic, unstable, and uncertain, leading to the pervasive existence of ambiguity.

In terms of communication across languages, especially the translation process, another aspect of ambiguity comes into play, largely due to a lack of one-to-one correspondence of most linguistic elements between the Source Text (ST) and Target Text (TT), further complicating the issue and increasing the degree of ambiguity involved.

Nevertheless, human beings still seem to manage to communicate via language with fairly high efficiency. The listener or reader, most of the time, intuitively picks a meaning from all these possible interpretations, and assumes this is what is meant by the linguistic symbols. It seems amazing that with language so inevitably and pervasively ambiguous, we still select a meaning very easily, often without being aware in the first place that the utterance or text we are making meaning of is potentially highly ambiguous. How does that happen? What is the cognitive mechanism behind this? How do we, especially in the process of translation, manage to make a selection of the meaning and translation equivalent (so to speak) so quickly, subliminally, and most of the time, with disproportionally little effort, out of so many possibilities not only in terms of the pervasive Source Language (SL) ambiguity itself, but also with respect to the multiple ways of expression that are almost always in existence in the Target Language (TL) for the same SL element? What is it in the mental processing that decides how much cognitive effort is needed to make this happen in the mind of the translator? These are some of the issues that the current thesis seeks to address.

Translation technology

With increasingly popular use of translation technology (i.e., computer-aided translation tools and machine translation systems) in professional translation workflows, the revision or post-editing of either translation memory (TM) suggestions, or machine translation (MT) output, seems to be a typical aspect of the translating routine today (which makes the distinction between such terms as "translator" and "post-editor" much less straightforward than before; see O'Brien, 2022; see also Section 4.3). While this has undoubtedly increased the efficiency of translating to a considerable extent, this increase is largely dependent on the quality of the suggestions from the technology, i.e., the "percentage" of the fuzzy match and the MT output quality. Here, in terms of the cognitive processes, the translation suggestions from the computer can perhaps be considered as additional to what is already activated in the translator's mind (see, e.g., Chapter 6 for details), which can increase the level of translation-ambiguity in the selection process. This computer-generated translation, once inputted into the cognitive mechanism, is likely to join the competition that is already occurring among all the other candidates that are activated by the translator's (or post-editor's) encounter with the ST, together with the information that is also activated by the encounter with the TM/MT suggestion itself. The translator would have to disambiguate, or make a selection, not only among the possible options in his/her mind, but also regarding an additional option provided by the TM or MT, together with the information that is activated as a result of this suggestion. This means that the suggestion from TM or MT might be a factor increasing the cognitive load of the translator, in view of ambiguity resolution.

This is even more so if the ambiguity of the ST results in translation errors, or if the MT engine itself creates a potentially ambiguous output, whether or not the ST was correspondingly potentially ambiguous.

Translation ambiguity

In Translation Process Research (TPR), empirical studies have shown that translation behaviour tends to be affected by the number of alternatives into which individual ST words can be translated (i.e., translation ambiguity). For example, eye-key span, i.e., the time interval between the translator's fixation on an ST word and the keyboard input of its translation, has been shown to be longer for ST words which have many different translation alternatives than those corresponding to only one possible translation (Dragsted, 2010; Dragsted & Hansen, 2008). Such influence of the number of translation alternatives on behaviour is also shown in studies using decontextualized single words (e.g., Tokowicz & Kroll, 2007), and research in psycholinguistics demonstrates that bilinguals often activate lexical information non-selectively from both languages during word recognition (Brysbaert, 1998; De Bruijn et al., 2001; Dijkstra et al., 2000; Marian et al., 2003; Schwartz & Kroll, 2006).

While it can be debated as to how the bilingual mental lexicon is organized and which model best reflects the translating mind (see, e.g., Dijkstra & van Heuven's (1998) Bilingual Interactive Activation Model, or Kroll & Stewart's (1994) Revised Hierarchical Model, etc.), from an empirical perspective the influence of what is sometimes called "translation ambiguity" (i.e., the existence of multiple translations for the same ST word; see, e.g., Prior et al., 2011) on translation behaviour seems evident in experimental studies of the translation process.

Entropy

Entropy generally refers to the degree of uncertainty of the outcome of an event (see Chapter 3). In the context of cognitively-oriented translation process research, an entropy-based predictor variable which mathematically describes the degree of uncertainty in translation choices, namely word translation entropy, has been considered a better measure for the variation of translation alternatives than simply counting the number of these alternatives (Bangalore et al. 2016). In other words, rather than investigating translation behaviour in relation to the *number* of translation alternatives (see studies mentioned above), it is perhaps more indicative of the cognitive aspects of translation if the *entropy* of these alternatives is used as a predictor variable.

Accordingly, word translation entropy, approximated by a metric denoted as HTra (see Chapter 2), has been recently proposed as a statistical measure of the translation product. It is claimed that HTra can represent variance, literality, and translation ambiguity (Carl, Bangalore, & Schaeffer, 2016a; Carl, 2021d), and some empirical investigations have analysed the correlation of HTra with cognitive effort. HTra has also been used to find evidence for the early priming processes, and to discuss ways of quantifying translation difficulty. Importantly, a number of empirical studies show a positive and statistically significant effect of HTra on different measures of effort, including, among others, first fixation duration, total reading time, word production duration, and the probability of a fixation (e.g., Schaeffer, Dragsted, et al., 2016; Carl & Schaeffer, 2017c; Lacruz et al., 2021), and therefore words with higher HTra values have often been considered more difficult to translate (Carl et al. 2019).[1]

However, for such empirical studies, there are two issues that may pose problems in efforts to establish the link between entropy and cognitive processing. First, the approximation of entropy (or sometimes its empirical definition) appears to be more relevant to the variation of the (final) choices made by different experimental participants, rather than the entropy which represents the cognitive environment of a particular individual. Second, entropy is largely used for the technical quantification of a feature of the translation *product*, rather than as a way of conceptually explaining the process. More specifically, in empirical studies of word translation entropy, the concept of entropy seems to be consistently used as a measure of the product, rather than as a representation of specific aspects of assumed mental states during the process or as a way of describing the process of transition from one mental state to another.[2] An exception, however, is the "systems theory perspective" (Carl et al., 2019), where the human translation process is considered "a hierarchy of interacting word and phrase translations systems which organize and integrate as dissipative structures" (p.

[1] Studies of this kind are generally based on an assumption called the eye-mind hypothesis, i.e., the eyes remain on a word as long as it is being processed (see Chapter 7).

[2] The exact meaning of *mental state* is sometimes controversial, with different areas of study using it to refer to different concepts. In this thesis, mental states are similar to the *internal states* in, e.g., the free energy principle and active inference (Friston 2010; Friston et al. 2017) where the cognitive agent constantly updates the probabilistic representation of the external environment. To put it simply, in the context of this thesis these mental states can be understood as different stages of processing during entropy reduction.

211), and where entropy is defined as the internal order of these word translation systems. This definition of entropy is apparently from a systems theory perspective.

In its conceptual investigations of entropy, the current thesis looks at the concept from a different perspective, focussing on the probabilistic nature of its mathematical equation, its representation of cognitive resource allocation in the activation, suppression, competition, and selection of candidates when multiple options are available (i.e., when the ST is translation-ambiguous), and the specific process in which entropy is reduced through the transition between mental states. The process of translation selection is described in close detail through the lens of entropy and entropy reduction, which very few studies have addressed at a conceptual level in the literature on word translation entropy. Following these conceptual explorations, the behavioural manifestations of this process are also explored through observation in a large database.

The view of the translation process as entropy reduction is meaningful, as most studies in the literature use entropy mainly as a product-based measurement of the variation of translation, the number of alternative choices for a particular ST word or word order, the agreement among translators in the experiment, or at most the existence of TL-specific processes. These studies rarely discuss entropy directly as a description of mental states at an abstract and conceptual level. The exception is Carl et al.'s (2019) above-mentioned treatment of the translation process, where entropy refers to the degree of disorder of word or phrase translation systems, drawing on the definition of entropy used in systems theory and in thermodynamics. It seems that very few previous studies have provided a conceptual account of, and a theoretical justification for, the detailed mental processes of lexical selection (or disambiguation) involved in translation/post-editing, using entropy in line with its probabilistic, mathematical definition.[3]

[3] One might wonder why entropy can describe so many different things, or whether it sounds too ambitious for one single concept (or metric) to measure this many aspects. It should be noted that the mathematical concept itself does not specify, and is not confined to, what it is used for. Perhaps a simple analogy is the 'mean', or average, which does correspond to one single equation: $mean = \frac{x_1+x_2+ \dots +x_n}{n}$ but can, in effect, be used to describe anything quantifiable — the mean age of a population, the mean height of certain tables, etc. In the same way, although entropy is defined with a specific equation (see Chapter 3), what it measures can vary in practice. This thesis focuses on the general, mathematical, and abstract

The importance of providing this conceptual account is that the term itself seems to be often vague and ambiguous, and that the entropy value which is calculated from the final translation product, and perhaps more problematically, which is approximated in TPR experimental sessions, does not exactly capture the mental states during the process or the transition from one mental state to another. How this product-based metric relates to the process, and how it can represent cognitive load, are important questions if entropy is to be used for our understanding and conceptualization of the translation process. On the other hand, given empirical investigations confirming the positive statistical correlation of entropy (approximated by HTra) with various measures of effort, and with respect to the conceptual nature of entropy in its definition, a perspective that looks at the cognitive processes of translation through the lens of entropy and entropy reduction seems to bear great potential for conceptual understandings and empirical investigations of translation/post-editing as a cognitive activity. This aspect will be elaborated in later chapters of the thesis.[4]

Relevant to entropy are also a number of concepts that are discussed in detail, including uncertainty and cognitive load. This thesis will show conceptually the way in which entropy represents uncertainty and cognitive load, drawing inferences not only from the definition of entropy in information theory but also from the formulation of resource-allocation processing difficulty in psycholinguistics.

1.2 Aims, questions, and scope

In view of the above, this thesis intends to investigate the way in which ambiguity is cognitively processed, in translation in general and in post-editing in particular, from the lens of entropy. Inferences are drawn from psycholinguistics, bilingualism, translation process research (TPR), and entropy-based models of translation cognition, describing disambiguation in translation/post-editing as a

notion of entropy, and in the conceptual discussion in Chapter 6, entropy will be used to (statistically) describe the mental states, rather than the translated text.

[4] In some parts of this thesis, "translation/post-editing" is used to refer to both translation and post-editing. This usage reflects the fact, on the one hand, that distinguishing between the two processes is not straightforward due to the technological changes in the industry (see also Section 4.4), and on the other, that the perspective of entropy and entropy reduction taken in this thesis is compatible with both translation and post-editing (see Section 6.2.9).

process of resource re-allocation and entropy reduction across different mental states, with cognitive effort expended to re-allocate resources and to reduce entropy. In short, the cognitive processes of disambiguation are explained in terms of entropy and entropy reduction. The investigation involves both conceptual and empirical examinations of the cognitive processes of disambiguation in translation (and post-editing), as well as some of the important cognitive issues that are subject to debate in psycholinguistics and TPR, including parallel vs. serial processing in disambiguation, horizontal vs. vertical views of translation, and incremental processing vs. the head-driven approach in parsing. The thesis also addresses a number of important yet unaddressed issues regarding the conceptualization of entropy as an explanation of mental activities. These include: entropy for the mental state and entropy in the text, relative entropy and decrease of entropy, surprisal and word translation entropy, and the implications of entropy for the mind in view of parallel and serial processing. In so doing, it is hoped that this research can contribute to the shaping of the entropy framework.

Conceptually, it assumes non-selective activation of both SL and TL in the translation process (i.e., both languages are active, which is consistent with the general consensus in bilingualism, see Section 1.1 above), and explores how entropy and entropy reduction can theoretically describe the assumed mental states during disambiguation. While concepts such as word translation entropy have mostly been operationalized as a product-based metric in TPR, this thesis seeks to bring the concept of entropy into the mind of the translator/post-editor, considering entropy as a conceptualization of the mental activities in cognitive processing, rather than merely as an alternative statistical means of counting translation equivalents (which is what it is currently used as). Several important, tricky, and currently unaddressed conceptual issues are discussed in detail in this regard.

Empirically, it uses HTra (word translation entropy as observed in experimental tasks, see Carl et al. 2016 and Chapter 3), eye-movements, and keystroke logs in a publicly-available database (i.e., the CRITT Translation Process Research Database, see, e.g., Carl, Schaeffer, et al., 2016), to shed light on how the conceptual understanding of lexical and structural ambiguity may be manifested by observable behaviour.

The thesis aims to address the following research questions:

RQ 1: How does entropy, other than as a statistical measure of the text, conceptually describe the mental activities in translation/post-editing when it comes to ambiguity?

RQ 2: What are the behavioural manifestations of entropy and entropy reduction in the mental state, as conceptualized here, during translation/post-editing of ambiguous items?

RQ 3: How does this theoretical conceptualization, as well as empirical observation, relate to serial vs. parallel processing of ambiguity and horizontal vs. vertical models of translation/post-editing, at both lexical and syntactic levels?

In addressing these research questions, the thesis focuses on mental processes rather than the larger cognitive environment, considering translation as involving information processing as a major part of the cognitive process. In comparison with, e.g., an embodied, embedded, enactive, extended and affective (4EA) perspective of cognition which emphasizes situatedness (Muñoz, 2016a, 2017; Muñoz & González Fernández, 2021; Risku & Rogl, 2020), this seems to reduce the entire spectrum of cognitive processes to mental activities only. This reduction would largely provide a convenient ground for operationalizing the conceptual and empirical explorations in this thesis.

Nevertheless, the notions of entropy, surprisal, uncertainty, and informativity are consistent with (and also crucial parts of) such neuroscientific theories as free energy principle (e.g., Friston, 2010), active inference (e.g., Parr et al., 2022), Bayesian brain hypothesis (e.g., Knill & Richards, 1996), and predictive coding (e.g., Clark, 2013; Stefanics et al., 2014; Hohwy, 2017). In many of such theories, the embodiment of the brain, the coupling between the cognitive agent and its environment, and the interaction between the internal states in the brain and the external states in the world (through sensation and action) are recognized and incorporated in their entropy-, surprisal-, uncertainty-, and probability-based formulations (see also Friston, 2010; Friston et al., 2013; Parr & Friston, 2017; Parr et al., 2022). Therefore, the discussion in this thesis on entropy in the mental processes of translation can potentially be expanded to a larger framework which incorporates the perspectives of these theories (see Section 3.4.2).

Indeed, entropy itself as a mathematical concept, at an abstract level and regardless of the scope of discussion on cognition in this thesis, does not appear contradictory to the aspects of cognition that are not covered by this thesis — the way in which entropy describes various probabilistic features remains the same. It would only be the specific value of entropy (if we want that value) that would be different as a result of the impact from additional factors (this value, however, is not directly measurable and can only be approximated, since this is within the mind of an individual).[5]

In addition, as the present thesis is focused on ambiguity, the terms and standpoints are often treated specifically in relation to ambiguity (especially from a psycholinguistic view), rather than the larger domain of information processing. This is important in the conceptual discussion of crucial aspects, e.g., serial vs. parallel models of disambiguation.

Regarding contextual effects on ambiguity, this thesis is focused more on co-text than context in its empirical analyses, although context is also reflected in the theoretical discussion in Chapter 6 (when probabilities are conditioned on context). For empirical investigations, eye-movements are analysed, where the effect of co-text would be conveniently operationalized (examined from eye fixations on neighbouring words or phrases). It should also be noted that the translation and post-editing processes that are discussed in the thesis, through the lens of entropy, are processes with respect to written text, rather than spoken language, therefore aspects of intonation, for example, are not within its scope.

1.3 Thesis outline

The rest of the thesis is structured as follows: Chapter 2 explains the theoretical issues, cognitive processing perspectives, and empirical findings in relation to ambiguity, as encountered in three different areas — psycholinguistics, natural language processing (NLP), and TPR. It covers ambiguity of different types, including lexical, structural, accidental, and temporary, and outlines the linguistic features of ambiguity, and the computational and human mental processing issues it raises. In terms of the cognitive activities of disambiguation, it introduces the theoretical

[5] In this regard, HTra can perhaps be considered a metric that imperfectly approximates the entropy in the mind. The reader is once again reminded, however, of the importance of distinguishing between the HTra as calculated in the CRITT TPR-DB and the notion of 'entropy' in the conceptual discussion in this thesis.

assumptions, the metaphors used in studies of monolingual disambiguation, the debatable perspectives of serial vs. parallel processing strategies, and how translation ambiguity is analysed and theorized in TPR. In particular, in its treatment of lexical access and selection, the priming effect, structural ambiguity and opposing views of structural disambiguation (i.e., incremental processing and the head-driven approach), and the use of word translation entropy as a product-based metric to quantify ambiguity, this chapter lays the groundwork for later chapters.

Chapter 3 explains a crucial concept in the thesis, namely, entropy. This concept is defined in line with information theory, and various aspects of its implications are explained, including information, surprisal, and uncertainty. Such definition and explanation provides the basis for the introduction of two entropy-based predictors frequently used in TPR — word translation entropy and syntactic choice entropy. Important to mention is that the entropy defined here as a mathematical concept should be differentiated from the metric frequently used in TPR as an approximation (e.g., HTra in the CRITT database), which, depending on how it is operationalized, may represent different aspects.

This is followed by a general discussion in Chapter 4 on the process of translating and post-editing, introducing some important concepts and models in TPR and particularly discussing another point of debate in the area — horizontal vs. vertical processing, as well as cognitive models based on this distinction. High-level cognitive issues, such as non-representational perspectives of translation, radical enactivism, computationalism, and 4EA cognition, among others, are also outlined and engaged with critically. This also lays the foundation for later chapters where conceptual and empirical investigations on lexical and structural ambiguity are conducted on the CRITT Translation Process Research Database (TPR-DB).

As the thesis uses the entropy framework to describe both translation and post-editing in relation to ambiguity, this chapter also describes approaches to MT, as well as issues related to the temporal, technical, and cognitive effort expended in MT post-editing.

These three chapters, altogether, introduce the three aspects that are crucial to the thesis – ambiguity, entropy, and the process of translating and post-editing. In the subsequent chapters, conceptual and empirical investigations are conducted in view of the research questions.

Chapter 5 reviews some of the statistical approaches to ambiguity in psycholinguistics, as a basis for the conceptual investigations into entropy in mental states during translation. It discusses the metrics that are theoretically formulated and empirically examined in psycholinguistics, and that can be used to quantify ambiguity and cognitive load in the processing of ambiguity. The chapter starts by discussing frequency before proceeding to surprisal, while also explaining the relevant assumptions, theoretical justification, and the contradictions among the different frameworks.

This lays the foundation for the conceptual explorations in Chapter 6, where entropy, mostly used as a product-based metric, is brought into the mental activities of the translator. It addresses, among other issues: how entropy can describe mental states and activities, how entropy can be a theoretically justifiable means of quantification of cognitive load in relation to translation ambiguity, how HTra used in today's TPR studies relates to these conceptualizations, and how entropy that describes mental states differs from entropy that can be observed in the text via statistical means, or the entropy that can be approximated from experimental sessions. Several other important concepts are also disentangled, such as entropy reduction vs. relative entropy as a quantification of cognitive effort in translation/post-editing, and entropy from the serial and parallel processing perspectives of disambiguation. This chapter seeks to answer RQ 1 above.

While Chapter 6 addresses the conceptual aspect of entropy, the following chapters proceed to empirical and behavioural considerations, largely using eye-movement analyses.

Chapter 7 homes in on methodology, and more specifically issues related to eye-tracking. The chapter discusses use of this method in various areas, the various types of eye-tracking equipment, the margin of error in the eye-tracker, etc. Perhaps more important are the findings from empirical studies on ambiguity that are conducted on the basis of eye-movements. Such studies cover lexical and syntactic issues influencing eye-movements, particularly in relation to ambiguity, and are drawn from both psycholinguistics and TPR.

Chapters 8 and 9 investigate respectively the lexical and structural aspects of ambiguity, on the basis of the conceptual ground established in previous chapters, and using the CRITT TPR-DB. The way in which the conceptual understanding is manifested in observable behaviour (i.e., RQ 2) is investigated empirically.

In Chapter 8, behavioural data pertaining to a high-HTra item from 217 participants translating/post-editing from English into multiple languages shows that the item tends to result in pauses in production and regression of eye movements, and that the translator's/post-editor's corresponding scrutinizing of the ST tends to involve a visual search for lower-HTra words in the co-text and, accordingly, a decrease in the average entropy of the Activity Unit (AU). The expenditure of cognitive effort on ST items of high HTra values results in AUs where low-entropy words in the context come into play, thereby decreasing the average entropy of the scanpath in the AU.

In so doing, this chapter addresses RQ 3, both conceptually and empirically, at the lexical level.

Chapter 9 zones in on syntax, examining eye movements captured from 18 participants as they processed a machine translated Chinese relative clause that can involve a garden-path effect. Results show that, contrary to monolingual reading, disruptions of processing tend to occur not in the later part of the sentence where the wrong parse is disconfirmed, but in the earlier regions where the most quickly-built analysis is semantically inconsistent with the ST. Structural disambiguation and re-analysis seem to be bypassed. This suggests that, on the one hand, reading for post-editing receives a strong biasing effect from the ST, and on the other, argument integration is more appropriately explained from an incremental processing (i.e., serial) perspective rather than a head-driven (i.e., parallel) approach, as thematic roles seem to be assigned immediately in reading for post-editing. This chapter addresses RQ 3 at the syntactic level. While the lexical analysis supports a parallel disambiguation model, the structural analysis presented in this Chapter seems to support a serial one. In terms of translation models, both emphasize the impact of cross-linguistic priming and the presence of considerable horizontality in the translation process.

Chapter 10 concludes the thesis, describing its shortcomings and providing suggestions for future work.

Chapter 2 Ambiguity

2.1 Introduction

As can be intuitively expected, natural language is highly ambiguous (Wasow, 2015), where this ambiguity refers to "the potential of a linguistic expression to have more than one meaning" and does not imply vagueness, as the competing interpretations can be "perfectly concrete" (Hogan, 2011 p. 95).

Although many expressions that are ambiguous in isolation would no longer involve much ambiguity when used in a particular context, the mechanism through which ambiguity is processed and resolved has been a crucial aspect of psycholinguistics, and not surprisingly, of computational linguistics. In this chapter, several types of lexical and syntactic ambiguity are briefly described, together with some of the models that describe the disambiguation process in psycholinguistics. A point of debate here, especially at the syntactic level, is the question of parallelism, i.e., whether multiple interpretations of the sentence are entertained simultaneously or whether only one single interpretation is attended to at a time.[6] It is this difference between serial and parallel models of disambiguation that gives rise to different predictions that can be made on the basis of two proposals of how syntactic arguments are integrated into verbal predicates (to form predicate-argument structures) and when (i.e., how early) these arguments are assigned their thematic roles as they are encountered in a sentence — namely the incremental processing and head-driven approaches — when it comes to head-final relative clauses in languages such as Chinese.[7] A serial model predicts a garden-path effect (see Section 2.2 below; see also 7.6 and 9.5 in later chapters), while a parallel approach suggests that the clause would

[6] Note that the terms here (parallel vs. serial) are used in line with their usage in psycholinguistic research on ambiguity, referring specifically to whether multiple interpretations are entertained simultaneously. In information processing, parallel and serial processing are defined according to whether or not two or more sequences of operations are carried out simultaneously by independent processors.

[7] In this thesis, "argument integration" refers to the integration process at the syntactic level, while "thematic role assignment" refers to the semantic level.

result in extra processing cost in the ambiguous region. At the lexical level, this is relevant to the activation of multiple meanings or one single meaning. In this chapter, the above aspects are outlined, which provides a basis for the discussion in subsequent chapters. The chapter starts with a brief description of different categories of ambiguity — lexical, structural, and temporary, and proceeds to the theoretical perspectives of disambiguation in psycholinguistics (Section 2.2). In addition, researchers in TPR have also studied ambiguity, showing that translation ambiguity increases cognitive effort. This is also briefly described in the present chapter (Section 2.3).

2.1.1 Lexical ambiguity

Ambiguity at the lexical level can generally be divided into categorial ambiguity and homograph/polysemy. Categorial ambiguity refers to very commonly seen instances where a word is assigned to more than one grammatical or syntactic category, e.g., *show* as a noun and as a verb, and *round* as a noun, verb, adjective, preposition, particle, and adverb. This type of ambiguity tends to result in syntactic (i.e., structural) ambiguity as well, and since compound nouns are prevalent in English, this kind of structural ambiguity is in fact "very common" (Hutchins & Somers, 1992, p. 91), especially when each of the nouns within the compound can also function as verbs or other categories in many cases.

Other than categorial ambiguity, words can have different meanings, which results in polysemy (sometimes this also includes homographs and homophones). A classic example in this regard is *bank*, which can refer to either a riverside or a financial institution. Some words can have multiple categories and multiple meanings at the same time, e.g., *duck* as a noun and as a verb with different meanings. Here, the sentence *They saw her duck*, in addition to lexical ambiguity, is also structurally ambiguous (example from Hogan, 2011 p. 95).

Regarding translation specifically, another kind of ambiguity is often relevant — the one-to-many correspondence of a word between the source and target languages. Hutchins & Somers (1992) use "transfer ambiguity" and "translational ambiguity" to refer to this type, while it is also sometimes called "translation ambiguity" (e.g. Prior et al., 2011). In this thesis, discussion of lexical ambiguity is focused on this type, which may often overlap with polysemy and homograph.

2.1.2 Structural ambiguity

Structural ambiguity, or ambiguity at the syntactic level, arises when there are multiple ways of parsing a structure or sentence, due to "phenomena such as prepositional phrase (PP) attachment, conjunction, ellipsis of articles, embedded clauses and adjective scope" (O'Brien, 1993 p. 65).

A classic example of the structural ambiguity arising from PP attachment is below (see, e.g., Hutchins & Somers, 1992, pp. 88-89):

The man saw the girl with the telescope. (1)

In the sentence, it is unclear as to whether the attachment site of the PP (*with the telescope*) is the preceding noun phrase (*the girl*), or the main verb (*saw*), and in either case the meaning would be very different from the other. When attached to the noun, it would be interpreted as meaning that the girl whom the man saw had a telescope; when attached to the verb it would be interpreted as meaning that the man used the telescope to see the girl:

The man [$_{VP}$ saw [$_{NP}$ the girl [$_{PP}$ with the telescope]]]
The man [$_{VP}$ saw [$_{NP}$ the girl] [$_{PP}$ with the telescope]]

This issue becomes more complicated when there are two prepositional phrases at the end, where there would be as many as five possible parses (Wasow, 2015, p. 34):

We saw a man in a strange hat with a telescope. (2)

In such sentences, there can even be more PPs put at the end, which theoretically means that the number of possible parses "explodes" (ibid) as the number of PPs increases.

Regarding structural ambiguity, Hutchins & Somers (1992) distinguish "real" and "accidental" ambiguities, in terms of whether the sentence is intrinsically ambiguous for the human reader, or merely ambiguous for NLP (including MT). The latter occurs essentially due to "an accidental combination of words having category ambiguities, due to alternative grammatical uses for syntactic constituents, or due to different possible combinations of syntactic constituents" (p. 89).

For both real and accidental ambiguities, many issues are also related to categorial ambiguity, i.e., words that can serve different grammatical functions, as mentioned above (see the example of *duck* in 2.1.1). In such cases, when categorial ambiguity is resolved the structural ambiguity would be eliminated accordingly. Wei (2017), for example, shows how lexical customization (where entries in the lexicon incorporate information on parts-of-speech) of MT can lead to considerable improvement in the syntax of the TT in terms of the ST main-verb identification, subordinate clauses, modifier attachment, conjunction, etc., which often involve accidental structural ambiguities for NLP.

2.1.3 Temporary ambiguity

A special kind of structural ambiguity is temporary, where the ambiguity occurs in a certain region of the sentence but is resolved in a later region. This can be, among others, the type of ambiguity caused by a verb that can either lead a sentential complement or be directly followed by an object noun (*I believe Sandy…*), or the inflection of certain verbs being in the same form for the past tense and the past participle (*The policeman questioned…*). See examples (3) and (4) (taken from Wasow, 2015, p.36):

I believe Sandy… (3)

… because Sandy is honest.

… is lying.

The policeman questioned… (4)

… the suspect for two hours.

… by the defense attorney changed his story.

In the case of such temporary ambiguity (which can only be resolved at a later part of the sentence), Wasow (2015) maintains that

> … the listener, hearing the sentence one word at a time, either must keep multiple possible (partial) interpretations in mind until the disambiguating material is encountered, or must guess which of the possible interpretations will turn out to be the correct one. (p. 36)

These two perspectives (i.e., whether multiple interpretations are entertained or a single one is committed to at a time), as will be described below, represent one of the points of considerable debate in psycholinguistics, and correspond to parallel and serial processing strategies, respectively. They also have to do with whether syntactic argument integration and thematic role assignment are immediate as online sentence processing proceeds. In later parts of this thesis, this will also be a crucial point in the discussion of structural ambiguity.

Temporary ambiguity can often involve a favoured interpretation in contrast to a less-favoured one, although the less-favoured interpretation can turn out to be the correct parse as confirmed in the later regions of the sentence. This results in a garden-path effect, which has been used as a tool for analysis of the two processing strategies mentioned above, i.e., regarding single or multiple interpretations being maintained in the ambiguous region. Typical examples of the garden-path sentences are as follows:

The horse raced past the barn fell. (Bever, 1970) (5)

When the dog scratched the vet and his new assistant removed the muzzle. (Staub, 2007) (6)

For the above two sentences, a garden-path effect would usually occur during reading, as the temporary ambiguity strongly favours an interpretation that would be inconsistent with, and disconfirmed by, a later part of the sentence. In terms of eye-movement, disruptions of processing would be observed in the region that disconfirms the wrong parse, which is represented by lengthened gaze durations and increased frequency of regressions (see Chapter 7 for a more detailed description). In examples (5) and (6), the disruptions would be expected to be observed on the words *fell* and *removed* respectively.

In this regard, such garden-path findings have often been used as evidence to support a serial, single-analysis model for sentence processing (e.g., Frazier & Rayner, 1982), claiming that readers provisionally accept the very first structure that is built along the sentence, which is generally the simplest possible structure, until this is inconsistent syntactically (or semantically) with the material later.

2.2 Disambiguation in psycholinguistics

The garden path effect, which has been observed in numerous empirical studies, may serve as strong evidence for serial processing. This is in contrast to a parallel perspective, which considers the parser as pursuing multiple interpretations simultaneously, as in the Parallel Processing Hypothesis (Fodor et al., 1974), or delaying making a decision about temporary ambiguity until sufficient information is received for disambiguation, as in the Minimal Commitment Hypothesis (Marcus, 1980).

In the unrestricted race model (Traxler & Pickering, 1998), multiple analyses are activated, although it is only the most probable structure that completes the race. This multiple activation of syntactic alternative analyses is also assumed in constraint-based competition models (McRae et al., 1998; Spivey & Tanenhaus, 1998), where activated alternatives compete for selection. A ranked parallel model (Gibson, 1991; Gorrel, 1987) describes these syntactic alternatives as ordered by various constraints (context, syntactic complexity, semantic information, etc.), and as maintained by the parser at varying levels of activation throughout the ambiguous region.

In this regard, consider the following sentences (a modified version of the sentences used in Frazier & Rayner, 1982):

The wife will claim the inheritance. (7a)

The wife will claim the inheritance belongs to her. (7b)

A serial processing framework (e.g., minimal attachment principle) would predict that in the latter sentence, the reading would show disruptions of processing on the word *belongs*, which is indeed observed in Frazier and Rayner (1982). Here, the issue regarding this example is whether the verb *claim* precedes a sentential complement or a noun phrase. For Garnsey et al. (1997), statistical preference of the verb would be utilized by the parser, so that one version (sentential complement or noun phrase) is preferred rather than the other in sentence processing. For example, as shown in sentences (8) and (9) below (ibid), a sentential complement (i.e., version b) is expected by the readers when they read *argue*, while a noun phrase (version a) is expected for the verb *discovered*:

The divorce lawyer argued the issue… (8)

a. The divorce lawyer argued the issue.

b. The divorce lawyer argued the issue was irrelevant to the case.

The scuba diver discovered the wreck… (9)

a. The scuba diver discovered the wreck.

b. The scuba diver discovered the wreck was caused by a collision.

This means that the parser exploits information contained in the verb phrasal head (Yamashita et al., 2011, p. viii), involving a head-driven parsing approach. In constructing dependent relations between linguistic units, the head has a crucial role "as it specifies subcategorising and adjunctive relations with other elements within the same phrase and gets integrated with linguistic units outside of the phrase" (Lin & Bever, 2010, p. 277). For this integration process, there are two main perspectives regarding sentence processing (ibid) — an *incremental* perspective where arguments are assigned their thematic roles immediately as they are encountered (Bader & Lasser, 1994; Miyamoto, 2002), and a *head-driven* approach (Abney, 1989; Pritchett, 1988, 1991, 1992), based on which thematic roles are assigned only after the head is reached.

For English, such phrases are head-initial structures, as in the verb phrases in examples (7), (8), and (9) above, and a head-driven approach seems consistent with incremental processing to a large extent, since the head (i.e., the verb in this case) appears early on, making the onset of the phrase and "providing selectional restriction and probabilistic information" on, in the case here, "the type of complements the verb may take" (Yamashita et al., 2011, pp. vii-viii). For languages such as Japanese, Korean, and Chinese,[8] however, phrases are frequently head-final, which means that these two strategies — incremental processing and head-driven approach — would be sharply different and would make distinct predictions.

Lin & Bever (2010) hold that, when it comes to head-final relative clauses, incremental processing would predict a garden-path effect, with the relative clause prior to the head being preferably analysed as a main clause, whereas based on the head-driven approach, "the temporary ambiguity prior to the head makes processing

[8] Chinese is different from Japanese and Korean in this regard, because it can involve mixed-headed structures (e.g., relative clauses), with the VP being head-initial and the NP being head-final.

laborious since multiple possibilities need to be maintained before the head is reached" (pp. 277-278).

Regarding sentence (7), i.e., *The wife will claim the inheritance*, when there is a modification for the noun *inheritance*, such as *the inheritance that the husband has left*, a head-final relative clause in Chinese would place the modification before the head noun (i.e., the relativized gap before the filler):

妻子 将会 领取 丈夫 留下 的 遗产。 (10a)

*qizi jianghui lingqu zhangfu liuxia*_$_{\text{GAP}}$ *de yichan*$_{\text{FILLER}}$

The wife will claim the husband left _$_{\text{GAP}}$ relativizer inheritance$_{\text{FILLER}}$

relative clause ––– relativizer ––– head noun

gap >>>>>>>>>>>>> filler

"The wife will claim the inheritance that the husband has left."

In contrast, a head-initial relative clause (in English) produces distinct filler-gap relations, with the head (filler) being prior to the relativized gap:

The wife will claim the inheritance$_{\text{FILLER}}$ that the husband has left_$_{\text{GAP}}$ (10b)

head noun ––– relativizer ––– relative clause

filler >>>>>>>>>>>>>>>>>>>>>>>>>>> gap

Regarding such a head-initial structure, sentence processing theories such as the Active Filler Strategy (Frazier & d'Arcais, 1989) stipulate that when the filler (i.e., the head) is encountered in (10b), a gap-searching process is initiated, where the parser actively postulates a gap in the upcoming words (which are linearized to the right side of the head). In (10a), however, the head and the relativizer are both positioned *after* the relative clause, with the gap preceding the filler. In this regard, detecting the gap prior to the head noun "becomes a challenge for the parser" (Lin & Bever, 2010, p. 278), as there is no overt marking for relative clause. While the left edge of a head-initial relative clause (e.g., as in English, see 10b) is "unambiguously marked by the relativizer" (ibid, p. 279), the left edge of a head-final relative clause (e.g., as in

Chinese, see 10a), on the other hand, is not "overtly indicated" and "both a main-clause analysis and a relative-clause analysis are initially possible" (ibid, p. 279).

Such structures as (10a) can therefore easily create temporary ambiguity prior to the head noun, as Chinese verb phrases are head-initial (i.e., the sentences follow a SVO structure). In this example, the modifier for the head noun is an object-extracted relative clause in the form of NP_1 *Verb de* NP_2 where NP_2 is the head (filler) and *de* (的) is the relativizer. This structure functions as the object of the main verb (i.e., *claim*): NP_3 *Verb* NP_1 *Verb de* NP_2. It can be seen that the first part of the sequence creates a convenient NVN structure (i.e., NP_3 *Verb* NP_1), where NP_1 can be considered part of the main clause (as the object of the main verb), rather than the subject of the relative clause. In addition, if the main verb can lead to either a noun phrase or a sentential complement, then this adds another layer to the ambiguity. Chapter 9 analyses a sentence of this kind.[9]

Here, with respect to the predictions made by the two processing perspectives, incremental processing would mean that argument integration and thematic role assignment are immediate, adopting such strategies as minimal attachment and committing to the simplest parse. This creates a garden path, indicating observable disruptions of processing in the word or region which disconfirms the wrong parse. In most cases, this disruption would be on the relativizer and/or the head noun regions (indicated in eye-tracking data as increase of fixation numbers, reading time, or regression of eye movements in the regions of the relativizer and/or the head noun).

For the head-driven approach, this means that thematic roles are not assigned until the head is reached, and that multiple interpretations are maintained until the head, resulting in extra processing effort in the ambiguous region (i.e., the relative clause). This would indicate increased levels of fixation (as captured in eye-tracking data) and reading time in the region of the relative clause.

[9] It is important to note that the temporary ambiguity caused by head-final relative clauses is not confined to Chinese, although the way in which the sentences are ambiguous may be different for other languages. For example, both Japanese and Korean follow strictly head-final structures, meaning that not only nominal structures, but also verb phrases, are head-final (i.e., SOV sentences). Regarding the garden-path effect caused by subject-extracted and object-extracted relative clauses in Japanese, Miyamoto & Nakamura (2003, p.343) provide a good example: 年寄のおばあさんをバス停まで見送った女の子 / 年寄りのおばあさんがバス停まで見送った女の子

As will be seen in Chapter 9, such hypotheses are examined in view of the reading process in post-editing.

As indicated above, incremental processing is characterized here as serial, while head-driven processing is parallel. Although these two perspectives are somewhat controversial in sentence processing, at the lexical level there seems to be a general consensus that the multiple meanings of ambiguous words are simultaneously activated, although this activation may occur to varying extents for different meanings.

For translation ambiguity, models of the mental lexicon in psycholinguistics seem to mostly endorse activation of both languages (source and target), and this seems to be supported by evidence from many studies in both bilingualism and TPR (Balling, 2014; Bangalore et al., 2016; Carl, Tonge, & Lacruz, 2019; Grosjean, 1997; Macizo & Bajo, 2006; Ruiz et al., 2008; Schaeffer, Dragsted, et al., 2016; Schwartz & Kroll, 2006; Wu & Thierry, 2012), although the *extent* of this activation is often relative to language proficiency, particularly in the case of non-dominant languages (e.g., experiments also show that semantic priming is stronger for those with higher proficiency, see Favreau & Segalowitz, 1983).

2.3 Translation ambiguity in TPR

As mentioned in Chapter 1, studies in empirical TPR have shown a positive and statistically significant impact of word translation entropy (and syntactic choice entropy) on different aspects of behaviour in translation. This follows a line of research that analyses the cognitive effort of translation when a certain ST word corresponds to multiple translations (see Chapter 1), i.e., translation ambiguity.

In such studies, the entropy-based metrics approximated from the text are used as a predictor variable for various measures of effort in the translation process. On the one hand, this reflects a crucial aspect of psycholinguistic research on ambiguity – statistical approaches to its measurement and to the associated cognitive load, an issue to which the thesis will return in Chapter 5. On the other hand, the conceptual aspects of entropy in these metrics are indicative of core elements of information theory, including surprisal, uncertainty, and freedom of choice. Tracing the concept back to the fundamental premises of information theory would shed light on its core value when the concept is used for translation processes.

Despite all the shortcomings of the way in which HTra and HCross (syntactic choice entropy, see 3.5.2) are defined, approximated, operationalized, and interpreted in current empirical TPR, the conceptual, mathematical meaning of entropy at the abstract level perhaps has far better value than the technical use of the metrics.

In the following chapter, entropy as encountered in information theory (and thus also often referred to as *information entropy*) is defined and its implications in relation to information, uncertainty, and surprisal are outlined. Word translation entropy and syntactic choice entropy are then explained in mathematical detail.

2.4 Summary

This chapter has outlined several important aspects of ambiguity and disambiguation, which largely lays the foundation for the analysis and discussion in later chapters. The ambiguity introduced above, as well as such theoretical perspectives as entropy, serial vs. parallel processing, and incremental vs. head-driven approaches, will be a recurring theme in the later parts of the thesis.

In the following chapters, the conceptualization of entropy as a description of mental processes is explored, and the dynamic change of entropy during the process of the resolution of translation ambiguity is both theoretically analysed and empirical examined on the basis of the user activity data (UAD) from the CRITT TPR-DB, on lexical and structural levels.

Chapter 3 Entropy

3.1 Introduction

This chapter explains a crucial concept in the thesis — entropy. While the definition of entropy, as well as its implications, can vary depending on the area of study, the mathematical expression of the concept remains largely consistent across different fields. The definition of this concept in this thesis sticks firmly to its mathematical equation in information theory, and its probabilistic nature in this mathematical definition. This is the basis on which the implications of entropy in relation to ambiguity, as well as the cognitive processing of ambiguity in translation, will be investigated in later chapters.

In this chapter, a definition of entropy is provided in line with information theory, followed by a discussion of its implications in terms of information, uncertainty, surprise, and freedom of choice. While these related terms (i.e., information, uncertainty, surprise, and freedom of choice) can bear different meanings in different contexts (which is a vivid example of the lexical ambiguity pervasively in existence in language), it is important to mention that they are used in this thesis in a *probabilistic* sense. For example, *uncertainty* can, in some contexts, refer to a cognitive state of indecision which an individual consciously experiences (as in, e.g., Angelone's (2010) description of uncertainty management in translation), while the *freedom of choice* might be understood as an individual's autonomy to make a choice unconstrained by external parties. However, these concepts, in the context of the current chapter, should be construed from a probabilistic point of view. This means, e.g., that uncertainty refers to the extent to which the value of a random variable (or the outcome of an event) is not certain, where being 'certain' entails a probability of 1 for a particular value or outcome (and thus 0 for all other values or outcomes).[10] The freedom of choice indicates the extent to which the probabilistic selection of a value for a stochastic

[10] Note that the *value* here may not necessarily be numerical. The value of a stochastic variable can also be, among many others, categorical (e.g., a gender, a word, a political party, a blood type, etc.).

variable is deterministic. In the same way, the other concepts mentioned here are all used in line with probability. These concepts, and their relationship with entropy, are explained in more detail in this chapter.

After a definition of entropy and explanation of its implications and related concepts, the chapter proceeds to the two entropy-based metrics frequently used in empirical TPR — word translation entropy and syntactic choice entropy. These metrics measure lexical and syntactic aspects respectively, on the basis of entropy, which is essentially consistent with information theory.

3.2 Information entropy

The information theory famously formulated in Shannon (1948), Wiener (1948, 1954), and Shannon and Weaver (1949) is a theory which studies, mostly by mathematical means, "the role of information in communication" (see, e.g., Chan, 2004, p. 105), and over decades it has been influential for a variety of areas of study (Bazzanella, 2011) ranging from telecommunication to, among others, computer data storage, physics, biology, physiology, neuroscience, linguistics, and more importantly, translation studies where the translation of texts is sometimes discussed from the perspective of transcoding the ST content into the target form (especially in the equivalence paradigm). This theory models the communication process as an information channel consisting of three basic components: the channel encoder, the channel, and the channel decoder, respectively defined as the following:

> The channel encoder is the source of the message. It sends the coded signal through the channel, the communicative medium. The channel may be noisy (adding distortion) or clean (without distortion). Finally, the channel decoder attempts to decipher the original message sent by the channel encoder. (Collins, 2014, p. 652)

This theory has been applied in translation studies (e.g., Nida & Taber, 1969; Wilss, 1982) because the process of translating can be viewed in a similar fashion — translators open "an information channel between senders and receivers" and "recreate a linguistic surface" from which readers will retrieve the information in the source

(Chan, 2004). The relevant mathematical expressions, however, lying at the core of Shannon's theory of communication, seem much less discussed in translation studies.

Perhaps far more important than the general idea of modelling communication as an information channel is that the properties of the channel can be represented in mathematical terms, e.g., signal-to-noise ratio, redundancy, information load, etc. (Marcus, 1963; Pierce, 1980; Shannon, 1948; Shannon & Weaver, 1949). Among these formulated expressions, a largely fundamental and profoundly meaningful concept would be Shannon's mathematical definition of *information* by probabilistic means, using the calculation of entropy which in many respects also measures uncertainty, surprisal, and the freedom of choice. If a source of a message generates a discrete stochastic variable X, i.e., a variable which can take discrete values $x_1, \ldots, x_n$ with probability $p_1, \ldots, p_n$, respectively, then the amount of information for this message is represented by its entropy, $H(X)$, described mathematically via

$$H(X) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{1}$$

where the value of $H(X)$ would be maximum if all probabilities $p_1, \ldots, p_n$ are equal (Shannon, 1948).

Mathematically, the value of entropy is indicative of the number of informational units which are needed for message encoding,[11] and this is considerably useful for the quantification of information regarding data storage and transmission. The specific entropy value of the source, as a function of $p_1, \ldots, p_n$, refers to the expectation of the minimum number of bits that are required in order to encode this message (when the base of the logarithm is 2).

As will be described below, the entropy of a random variable is the average level of "information", "surprise", or "uncertainty" inherent to the variable's possible outcomes.

[11] Specifically, the value of entropy is the *expectation* of the bits of information for the message.

3.3 Selectional information-content

It is obvious that this mathematical definition of information does not take into account any aspect of meaning, the concept of which is relatively qualitative in comparison to the measurements used in information theory, but which is nevertheless important in terms of actual communication. Given this shortcoming in Shannon's theory, MacKay (1969) establishes a more general theory of information which seeks to incorporate meaning — one broadly regarding the "processes by which representations come into being, together with the theory of those abstract features which are common to a representation and that which it represents" (ibid, p. 80). In this theory, a distinction is made between the common sense of information and what he calls *information-content* — the amount of information which is contained in a message and measured by certain technical means. For MacKay, Shannon's original definition of information is considered as one of the three basic types of information-content — selectional, structural, and metrical.

Selectional information-content refers to Shannon's definition and measurement, for which the main point of interest is the "relative improbability of a message given an ensemble of messages" (Kockelman, 2013, p. 116), and this type of information-content can be used to understand the replication of representations, a process regarded by both MacKay and Shannon as the central function of communication (ibid). The other two kinds of information-content, structural and metrical, turn respectively on "the frame of relevance that is used to construct a mapping between messages and referents" and "the degree of resolution such a mapping is capable of capturing", and can be used to understand the production of representations (ibid, 117). Meaning, as the interpretation of representations, is defined as the effect of a representation on a receiver (MacKay, 1969). The meaning of a message refers to "its selective function on the range of the recipient's states of conditional readiness for goal-directed action" (ibid, p. 24).

Although MacKay's theory categorizes information-content into three different types, the structural and metrical information-content are understood to be presupposed by selectional information-content and these two types have been argued to have already been built into selective information-content in his theory (Brillouin, 1962). In this regard, while MacKay's theory of information has been useful, important, and profound in many aspects (see, e.g., Kockelman, 2013), and while it does complement Shannon's theory in incorporating meaning in addition to the account of information, it seems that the original measure in Shannon's mathematical theory might be adequate as a practical means for measurement in most cases.

In addition, many models of semantic analysis are in fact consistent with Shannon's definition, at least in terms of the "Inverse Relationship Principle" (see Floridi, 2010, pp. 53-54) by which the probability of a proposition, sentence, event, or situation (in terms of occurrence) is in a reverse relation to the amount of semantic information conveyed. The probabilistic approach to semantic information defines the information in such proposition, sentence, event or situation in terms of the inverse relation between information and probability (ibid, p. 54).

Indeed, among the three interrelated processes with which MacKay's theory is concerned — the production, replication, and interpretation of representations, the appropriate measures for both replication and interpretation (if the states of conditional readiness of the receiver *can* be observed) are the same as Shannon's measure of information. As Kockelman (2013) rightly comments,

> … (the interpretation of a representation) would have a measure identical to selective information-content: they just turned on different ensembles, the relation between a message and an ensemble of possible messages (given some code); or the relation between a state of readiness and an ensemble of possible states of readiness (given some mind). While the numerical values might be different (because the ensembles are different), MacKay thought Shannon's measure could still be applied here. (p.118)

As can be seen in Chapters 5 and 6, this concept of information entropy in Shannon's theory can be easily borrowed to quantify or predict the cognitive effort of language processing, and more importantly to model the choice of translation in an ensemble of possible TT items (given an ST) when the translator is working on a certain part of the text.

3.4 Uncertainty and surprisal

In addition to the above definition's reference to the number of informational units (i.e., bits) in encoding (see Section 3.2), Shannon and Weaver (1949) succinctly explain that the term information measures the *uncertainty* involved in estimating the value X can take (from a probabilistic perspective), which means, in other words, its "freedom of choice" when a message is chosen.[12] This can be reflected by the mathematical expression of *H(s)* in equation (1) and is closely relevant to the issue of uncertainty (again, from a probabilistic perspective) regarding the translation choice which one has to make in the process of translation and post-editing. If translation and post-editing can be considered a cognitive process which involves frequent choice-making at different levels (e.g., Angelone, 2010; see Chapter 4 for more detailed discussion),[13] the extent of uncertainty concerning the translation choice at a particular point of the text, and thus the cognitive effort required to resolve this uncertainty, could perhaps be modelled and measured via the entropy value in view of all the alternatives concerning this choice (see 3.5.1 below). Important to mention is that the use of 'uncertainty' in Shannon's information theory, as well as the use of it here in the thesis, is from a probabilistic perspective (i.e., referring to the uncertainty in the probabilistic selection of the value of the variable), rather than the uncertainty (i.e., cognitive state of indecision) consciously experienced by an individual.

3.4.1 Surprisal

Here, the entropy as a concept in information theory is also crucially relevant to surprisal — or to use e.g., Kockelman's (2013) term, surprise-value — which not only reflects the intuition that "unpredictable items should carry a large amount of information while predictable items should not" (Collins, 2014, p. 652) but can also be used as an important quantification of cognitive effort in psychology (Attneave, 1959; see also Levy, 2013; Levy & Gibson, 2013).

The surprisal of an event, which is sometimes called *Shannon information content* or *self-information* in the information theory literature, is defined simply as the negative logarithm of its probability, or in an equal manner, as the logarithm of the

[12] Note that these terms (i.e., uncertainty and freedom of choice) are used in their technical sense in a mathematical context, despite their psychological connotations.

[13] The first paragraph in Section 4.2.1 briefly summarizes Angelone's description. The same section also discusses the issue of uncertainty management, where the use of the term 'uncertainty' seems to be from a different perspective than the way it is used here. As the thesis will argue, a probabilistic view of uncertainty provides a convenient means for both conceptual and empirical investigations into ambiguity.

inverse of the probability of the event. In the above case of a message with discrete values $x_1, \ldots, x_n$, the surprisal (s) of x_i in the message would thus be:

$$s(x_i) = -\log p(x_i) \tag{2}$$

A lower probability, $p(x_i)$, consequently leads to a higher surprisal, as can be seen from the equation. This means that an outcome which is less probable indicates more surprise (if that outcome is observed).

3.4.2 Surprisal, entropy, uncertainty, and informativity

Comparing equation (2) with equation (1) above, it is not hard to find that the entropy (i.e., information) of the entire message is mathematically equal to the weighted sum of the surprisal of each individual option in an ensemble of messages.[14] In other words, surprisal is associated with a particular value of the stochastic variable (or a particular outcome of an event), while entropy refers to the probability-weighted sum of all values (or outcomes).

In probability theory, the probability-weighted sum of all possible outcomes of a random variable is the expected value of the variable. Therefore, entropy of a random variable X is the *expected* surprisal based on the variable's probability distribution over all possible outcomes (see equations 1 and 2).

Surprisal quantifies the uncertainty in this random variable X taking a certain value x_i based on its probability of occurrence $p(x_i)$. Entropy measures the uncertainty in estimating the value of X.

A higher entropy, which indicates a larger amount of (average) information in the entire message (as in the case of message encoding), is equivalent to a higher level of probability-weighted sum of surprisal over all outcomes of the stochastic variable, which results from generally lower probabilities (i.e., higher surprisals) regarding each individual outcome. For the probabilities to be "generally lower", these probabilities need to follow more of an even distribution (i.e., without strikingly high or low probabilities).

---

[14] Note that the surprisal equation here has one single element, x_i, rather than the sum over i between 1 and n (see equation 1).

Regarding surprisal, the less probable an item is (i.e., higher surprisal), the larger amount of information would be associated with that item in the message. Information is positively related to the extent to which the item is surprising.

This seems to be largely the basis for the conceptualization of *informativity* in text linguistics (de Beaugrande & Dressler, 1981), which designates "the extent to which a presentation is new or unexpected for the receivers" (p. 139), as well as for similar notions in translation regarding the recreation of "a linguistic surface with which readers will retrieve the informativity" in the source (Chan, 2004, p. 107). Although this purely mathematical notion of information contains no semantic content (in a similar manner to Zipf's law, see Zipf (1935, 1949) and Chapter 5), the statistical, probabilistic approach to meaning and informativity in general has nevertheless been widely accepted in various disciplines related to language. In situations where meaning is involved, this definition of information is still consistent with many models of semantic analysis, where the same means of measurement can still be used for quantitative analyses (see Section 3.3 above).

In short, the entropy of a random variable is the average level of "information", "surprise", or "uncertainty" inherent to the variable's possible outcomes.

Concerning information, a simple example is that if someone is told something they already know (i.e., unsurprising, or expected message, where the expected surprisal over all possible possibilities are low, i.e., the entropy *H(X)* is low), the information obtained would be very small. This information would have a very low level of entropy.

On the contrary, if someone is told about something which is barely known to them, new information would be obtained and this information would be valuable. Such information is unpredicted (i.e., probability $p(x_i)$ is low), surprising (i.e., surprisal $s(x_i)$ is high), and informative in the mind of the message receiver.

From the perspective of cognitive processes, this view of information in terms of probability, surprisal, and unexpectedness is also meaningful. Neuroscience has proposed that the brain prevents sensory overload by *predictive coding* (e.g., Stefanics et al., 2014; Hohwy, 2016), where predictions of the sensory inputs are constantly generated and where expected inputs are unsurprising, uninformative, suppressed, and likely not fully processed. On the contrary, unexpected (i.e., surprising) inputs are informative and are fed forward to update the mental model of the world. Note that such notions of expectedness and prediction are in probabilistic terms, in exactly the same way as what has been explained above using $p(x_i)$.[15] The description of surprise and informativeness in association with probability is in line with the discussion above.

In addition, neuroscientific theories that draw extensive inferences from mathematics, such as the *free energy principle* and its corollary *active inference* (e.g., Friston, 2010), tend to see perception as being supported by neural processes optimizing probabilistic representations of the causes of sensory inputs. From this perspective, "any sentient creature must minimize the entropy of its sensory exchanges with the world" (Friston et al., 2017, p. 2636), where this entropy indicates the uncertainty or expected surprise of the sensory input, and is an important part of the mathematical equations describing the internal states of the cognitive agent (e.g., variational free energy, for which the mathematical expressions consists of surprisal and entropy). The meaning of entropy, uncertainty, and surprisal used in these theories are exactly the same as what has been outlined above (and discussed in later chapters).

3.5 Entropy in translation

3.5.1 Word translation entropy

As mentioned in Chapter 1, the word translation entropy (approximated by a metric called HTra) has been used as a predictor variable in many studies of TPR — particularly ones which make use of the CRITT TPR-DB (Carl, Bangalore, et al., 2016a) — to describe the lexical translation choices that a translator has to make at a given point of the source text.[16] As mentioned in Chapter 1, the use of this mathematical expression for modelling translation choice has been considered a perhaps "better reflection of the cognitive environment" of the translation process, than simply counting the *number* of possible translation alternatives for a given word in the source (Bangalore et al., 2016, p. 214). Its conceptualization and calculation are closely linked to the concept of word translation perplexity, largely due to the mathematical relationship between perplexity (PP) and entropy (H) as an exponential function (see Carl, Schaeffer, et al., 2016, p. 31):

[15] In probability theory, as mentioned, the expected value of a random variable is a probability-weighted sum of all possible outcomes.

[16] For a full list of the studies on the basis of this database, which is updated from time to time, see https://sites.google.com/site/centretranslationinnovation/tpr-db-publications?authuser=0

$$PP(s)=2^{H(s)} \quad (3)$$

The word translation perplexity can be considered an indication of the difficulty of the translation choice at a certain point of the source text, based on the idea that the more similar the likelihoods of the alternative choices are, the more difficult it is to make a decision at this point.

Assuming the choice of translations for the ST follows a distribution of probabilities ***p***, then the probability of a certain choice of word translation t_i in the target text over all possible translations $t_1 \ldots t_n$ of a particular ST word ***s*** can be represented by $p(s\text{->}t_i)$, where $s\text{->}t_i$ represents word-level alignments, and this probability can be estimated via a parallel corpus of aligned translations (ibid). In terms of computation, this probability would be the ratio of the alignments $s\text{->}t_i$ over the total number of translations for ***s***. It indicates the expectation of the translation choice, i.e., the choice of a certain TT word to be used for the translation of a given ST word.

These word translation *probabilities* can be directly combined with Shannon's information concept (i.e., entropy, see Section 3.2 above) to indicate the minimum number of informational units (i.e., bits) with which the expectation of word translation choices can be encoded (Carl, Schaeffer, et al., 2016). The expectation of this information, in turn, can be represented by the entropy value across the probability distributions:[17]

$$H(s) = -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) = -\sum_{i=1}^{n} p(s \to t_i)\log_2 p(s \to t_i) \quad (4)$$

The ***H(s)*** here is called *word translation entropy*, defined in Carl et al. (2016) as "the sum over all observed word translation probabilities (i.e., expectations) of a given ST word ***s*** into TT words $t_{i \ldots n}$ multiplied with their information content" (p. 31). The "information content" here refers to Shannon information content, which, as described in Section 3.4, is equivalent to surprisal — in this case the surprisal of the word translation choice at a particular point of the ST, i.e., $\log_2 p(s \to t_i)$. In a similar manner to Shannon's information entropy discussed in previous sections, the word

[17] This is the probability-weighted sum of surprisal, in the same way as in equation (1).

translation entropy here can be used to represent information, uncertainty, surprisal (marginalizing over all possible items), and unpredictability.

As intrinsically a mathematical expression of information entropy, it represents "the average amount of information contained in a translation choice" (ibid). When in a given context the ST word has only one translation option in the target language, its word translation probability would be 1, the Shannon information content would accordingly be 0, and this leads to a minimum level of entropy. In this case the translation options are concentrated on one single choice with very low uncertainty and unpredictability. If on the other hand a very large number of translation alternatives are equally probable, the entropy would accordingly be very high, as can be seen from the equation above. The more evenly distributed the individual probabilities are for all the translation alternatives, the higher the entropy value would be; the more different translations there are for a source word, the higher its word translation entropy would be. In other words, this value indicates not only the number of translation alternatives for a certain source (as a sum from i to n in the equation), but also the probability distribution of these TT word alternatives (as a probability-weighted sum in the equation).

This also means that the entropy here is closely similar, although not identical, to what is meant by variance (e.g., Miller, 1956), in addition to its reference to the unpredictability and uncertainty of the content which is conveyed by a message as implied by the concept of entropy in information theory. Therefore, a higher entropy value not only indicates a higher level of uncertainty and unpredictability, but also represents higher variance. When used in the context of the translation process, the way in which this value is calculated means that the more likely translation choices and less likely ones are weighted accordingly, capturing the distribution of probabilities for each translation option (Bangalore et al., 2016). This is argued in Bangalore et al. (2016) as what makes entropy a better metric to reflect the cognitive environment of a translator, in comparison to the simple counting of the number of translation alternatives.

While Bangalore et al. (2016) are right that entropy can be a better metric to measure what they call 'variance' — largely because of the probability weight given to each option, and that entropy can reflect the cognitive aspects of the translator, the argument here seems rather weak. From a different perspective, this thesis focuses on the probabilistic nature of the concept (in the present chapter) and discusses the dynamic *change* of probability distribution during the translator's decision-making (see Chapters 5 and 6), directly investigating how entropy can reflect and quantify cognition from a translation process perspective.[18] This provides a perhaps better conceptual basis for entropy as a statistical description of the mental processes.

In short, the word translation entropy can statistically describe "a set of co-activated translation possibilities that are equally good choices for the translation of a source text item" (Bangalore et al., 2016, p. 213), as well as "the degree of uncertainty regarding which lexical TT item(s) are chosen given the sample of alternative translations for a single ST word" (Schaeffer, Dragsted, et al., 2016, p. 191). Its features can be seen from the succinct description below:

> … if the probabilities are distributed equally over a large number of items, the word translation entropy is high and there is a large degree of uncertainty regarding the outcome of the translation process. If, however, the probability distribution falls unto just one or a few items, entropy is low and the certainty of the TT item(s) to be chosen is high.

3.5.2 Syntactic choice entropy

In a similar manner as word translation entropy, the entropy value can be calculated for another variable which represents the ST-TT word alignment information "as a local cross-lingual distortion" (Carl, Schaeffer, et al., 2016, p. 26). This variable is called *Cross* in CRITT TPR-DB, and is related to the difference in the order of words between the ST and the TT. The larger this difference, the larger the value of *Cross*. When transformed into entropy (HCross), it consequently models the variance of the position of the corresponding TT word within the sentence, regarding a particular ST word in question. At the same time, HCross also represents the surprisal and unpredictability of this word position in the TT sentence, as well as the degree of uncertainty involved in the translator's decision of choosing a certain order of words at a particular point of the ST.

[18] That is, in contrast to a translation product perspective.

This measurement captures certain aspects of syntactic choice in the process of translating, using word order as a feature of syntax. While studies (e.g., Bangalore et al., 2016; Behren, 2016) have resulted in important findings on this basis, compared to the surprisal framework which will be outlined in Chapter 5, the syntactic choice entropy here may seem somewhat oversimplified in view of structural disambiguation.

3.5.3 Entropy from the sample

The above descriptions of entropy, surprisal, word translation entropy, and syntactic choice entropy are all grounded in probabilities and probability distributions. An important aspect here is that the probabilities are associated with the values that the stochastic variable X can take, i.e., $x_1 \ldots x_n$.[19] This means, *what* values of the variable are taken into consideration (as x_i) largely determines what is entailed by the resulting entropy.[20]

In practice, word translation entropy and syntactic choice entropy are approximated in studies of TPR via a sample of translations produced in an experimental session (using metrics of HTra and HCross respectively in the CRITT TPR-DB). For a particular ST item, the translation options provided by all participants are entered into equation (4) for calculating entropy for that item. In this way, entropy would be more appropriately understood as a measure of how varied the participant's translations are.

As an example, consider the following explanation of HTra in the CRITT TPR-DB, from Ogawa et al. (2021, p. 143):

> … if a group of 12 Japanese-English translators were given the word 猩々紅冠鳥 to translate, and all 12 of them translated it as "cardinal," then the HTra would be 0. Now, if they all translated it differently—no matter how slight the differences, e.g., "cardinal" and "cardinals" would be counted as non-identical strings — then they would achieve the maximum entropy value for

[19] That is, in the case of equation (1), i.e., entropy *H(X)*. In equation (4), these values are $t_1 \ldots t_n$. See also next note.

[20] As mentioned, the *value* here may not necessarily be numerical. The value of a stochastic variable can also be, among many others, categorical (e.g., a gender, a word, a political party, a blood type, etc.).

any group of 12 strings: 3.58. HTra also takes into account the frequency (or probability) with which any given string occurs. For example, if 11 translators rendered 猩々紅冠鳥 as "cardinal" while 1 translator wrote "redbird," the HTra value would be 0.41. If, however, 6 translators wrote "cardinal," and the other 6 wrote "redbird," then the HTra value would be 1.0.

This succinctly illustrates a case of how the mathematical value of equation (4) is usually applied in practice in current TPR. It appears to reflect the choices that all individuals have made (i.e., the extent to which the experimental participants agree with one another), rather than the choices that are available for the translator him/herself. In this respect, the link between this metric and the cognitive processes would be somewhat weakened. This can be problematic when establishing the conceptual connection between the entropy of the translation choice and the cognitive load/effort involved in that choice.

At an abstract level and in connection to the cognitive processes, however, the conceptual formulation of entropy (as outlined in previous sections) does not seem to be impacted by the way in which it is practically approximated, or the way in which the metrics are used and interpreted in empirical studies. What this entropy means, and whether it extrapolates from the observed performance of many individuals to the potential performance of one individual, are dependent on what is included in the probability distribution of X (i.e., what the $x_1 \ldots x_n$ include). If the probabilities of items included in the equation refer to the items that are available for a particular participant (i.e., the items in his/her mental lexicon), then this entropy would describe the participant's cognitive environment regarding a translation choice.

That being said, the way in which HTra and HCross are operationalized in almost all TPR studies is based on the translated texts produced by the participants in an experimental session (in the same way as explained in the example above), which seems to be largely different from the cognitive environment of a particular individual. Perhaps this can be understood as more social than cognitive, as the resulting entropy is intrinsically relevant to the choices made by a group of individuals. On the one hand, the choices which other translators have made may not necessarily be available for the individual in question, and on the other, the choices observed in the sample of text produced in the experimental session may not include all the options that are available to a particular individual. This is an issue of approximation and operationalization, which does not mean that the conceptual formulations are wrong or inappropriate, but is also an issue that needs to be addressed if entropy is to be used to indicate the cognitive processes.

To contest the link between entropy and cognitive load in translation, it is important to focus on the fundamental, mathematical, and conceptual underpinnings of entropy (in connection to resource-allocation, see Chapters 5 and 6), rather than seeing metrics of HTra or HCross as the essential meaning of entropy in the context of translation. In this regard, the discussion in the sections in this chapter is vital for further investigations in the thesis.

In this thesis, entropy is used in accordance with its mathematical, probabilistic definition, rather than to refer to the specific way of calculation in the CRITT TPR-DB. This is an important difference between the concept discussed in this thesis, and what most studies in empirical TPR would entail when it comes to translation entropy (which, as seen from the example above, is more related to the extent to which the translation choices vary among different experimental participants).

3.6 Summary

This chapter has explained the concept of entropy from the perspective of information theory, and discussed how it relates to a number of other important concepts, including information, surprisal, and uncertainty. This discussion is largely grounded in probabilities, and the concepts are all introduced with reference to probability or probability distribution.

In terms of how the equations are applied in practice, the chapter points out that the way entropy is approximated in empirical TPR is from the translation options observed in the experimental session, which is more relevant to the differences between the (final) choices of different translators rather than to the options that are available in a particular individual's cognitive environment before that final choice comes into being. While the two may be hypothesized to be related to each other, the link seems rather uncontested.

Another point to note is that in this chapter, the discussion on entropy in translation is based on the text, and approximated from the text. It describes a feature of the text, with respect to how many translation options there are in the observed text (either from the experimental session, or otherwise). This feature shows a statistically significant impact on many aspects of translation behaviour (as mentioned above), and seems to correlate with effort based on those empirical studies. However, it does not directly describe the mind, or the translation process in itself, although the basic concepts introduced in this chapter can be applied to the cognitive process (if the probability distribution is not for the items in the text, but in the mind, see also Section 3.3).

The notion of entropy (and surprisal), however, has potential to directly model the process of translation choice, rather than serving merely as a metric for quantifying the translation alternatives in the observed translation. This will be elaborated in the following chapters of the thesis.

Chapter 4 The process of translating and post-editing

4.1 Introduction

This chapter delves into the detail of various processing models in the domain of Translation Process Research, covering vertical and horizontal processing, stratificational processing models, literal translation, the monitor model, etc. In so doing, counterarguments and controversies are also discussed, while high-level cognitive issues are outlined and engaged with critically. The notion of translation as uncertainty and problem solving is subsequently introduced, and since the thesis uses the entropy framework to describe both translation and post-editing in relation to ambiguity, this chapter also introduces post-editing research and MT approaches. This lays the foundation for later chapters where conceptual and empirical investigations on lexical and structural ambiguity in translation and post-editing, from the lens of entropy, are conducted.

Importantly, Chapter 2 has mentioned parallelism in the disambiguation models that have been proposed in psycholinguistics, explaining that parallel models characterize the disambiguation process as one where multiple interpretations are processed at the same time, while serial models endorse the view that only one interpretation is attended to at one point of time. In the present chapter, a similar issue is also relevant — whether SL comprehension and TL reformulation can occur simultaneously. These two processes occurring at the same time would indicate parallelism as well. This will be explained in more detail in the following section.

4.2 Models of the translation process

4.2.1 Vertical and horizontal processing

In TPR, the mental processes in translation have often been considered to involve two strategies, respectively described as vertical and horizontal (de Groot, 1997; de Groot & Christoffels, 2006; Macizo & Bajo, 2006; Paradis et al., 1982; Ruiz et al., 2008). This arises from the fact that many psycholinguistically-motivated theories of translation differ in terms of their assumptions on how comprehension and code-switching processes take place. The former strategy, i.e., vertical translation, is conceptually mediated, where the process is assumed to involve "giving lexical expression to the meaning extracted after comprehension" (Macizo & Bajo, 2006 p. 1). From this perspective, comprehension and recoding occur in a serial and sequential manner, with "no direct links between the SL and the TL lexical/syntactic levels of analysis" (ibid, p. 24). As the SL message is decoded, a conceptual representation is activated through "phonological, morphological and semantic analysis", followed by an activation of its lexical representation in the TL (Chmiel, 2021 p. 229), which thereby encodes this conceptual representation as the outgoing text. Both decoding and encoding are "semantically and conceptually computed" (He, 2007 p. 79), and this produces a "content-oriented target text" (ibid) which loses the linguistic forms of the ST during the translation process. In de Groot's (1997) description of the vertical strategy, comprehension of the ST is "full", meaning that the construction of meaning in the comprehension component includes the pragmatic intention of the SL text/discourse, therefore the subcomponents of this comprehension would also include pragmatic analysis (together with perceptual analysis, word recognition, and syntactic and semantic analyses). For the same reason, the "conceptual mediation" in vertical processing is sometimes also described, perhaps more accurately, as "conceptual-intentional mediation" (e.g., He, 2009). This vertical view tends to assume that the comprehension and production components in translation are largely characterized by the same processes as their counterparts in within-language tasks (i.e., monolingual comprehension and production), therefore the process of translation can be understood by investigating the details of L1 comprehension, L2 comprehension, L1 production, and L2 production individually.

In contrast, an opposing view which de Groot calls "horizontal" construes translation as "trans-coding", where the target text is produced by direct, on-line, and memory-based searches for links between linguistic entries in the two languages involved, with no conceptual mediation invoked between the ST and the TT. From this perspective, there would be a difference between within-language reading and reading for translation, since partial reformulation processes already occur as the translator reads the ST (Macizo & Bajo, 2006). In terms of the translation product, horizontal processing is considered to generate a "form-focused target text" (He, 2007). Experimental studies of the translation process have produced much evidence in support of the horizontal view (e.g., Jakobsen & Jensen, 2008; Ruiz et al., 2008; Schaeffer et al., 2017), and de Groot (1997) explains that even for strong advocates of the other (i.e., vertical) perspective, the presence of some horizontal processing is still acknowledged, albeit as of inferior status, in both translation and interpretation. Since the translation process does not seem to be vertical only, the above opinion that the cognitive processing of translation can be inferred from an understanding of comprehension and production in L1 and L2 respectively, "would be obviously flawed" (de Groot, 1997 p. 31).

In addition, another reason which de Groot — among others — uses to argue against the opinion that reading for translation is largely the same as within-language reading, is that translation and interpretation have a unique set of requirements "to comprehend and produce language material simultaneously or in continual alternation and in different languages", which can "modulate the various processing components" and lead to differences from the same components in monolingual language use (ibid, p. 31).

The division between vertical and horizontal views of the translation process lies indeed in "whether translation is a sequential process or whether and to what extent comprehension and production activities may occur in parallel" (Schaeffer, Dragsted, et al., 2016 p. 184). The vertical perspective supports a sequential processing strategy, considering ST reading a phase distinct from the reformulation phase, while horizontal translation maintains that TL reformulation already commences during ST comprehension, i.e., it uses a parallel processing strategy (ibid). Accordingly, this division — sequential vs. parallel strategy — is closely related to the issue of whether the processes involved in reading are characterized in the same way between interlingual translation and monolingual comprehension, as described above. A vertical/sequential view considers the two as largely the same in terms of the cognitive processes, while a horizontal/parallel view holds that the two processes are different.[21]

[21] Here, the terms "parallel" and "sequential" are used specifically in the context of whether comprehension and production activities occur in parallel, although there may be other ways of understanding the meaning of these processing strategies.

As will be mentioned below, the processes involved in reading for post-editing, especially reading of the raw MT output, are largely assumed in some studies as cognitively the same as reading for comprehension: "…despite its defects, the machine translation is initially nothing else than a *text* to be read and understood" (Krings, 2001 p.361), while empirical and detailed examination of the behavioural data in Chapter 7 provides evidence which seems to show otherwise.

4.2.2 Stratification models

Despite the existence of much empirical evidence in support of horizontal translation from different aspects — e.g., direct links between the SL and TL which are independent of meaning (Ruiz et al., 2008), differences in reading processes between comprehension and translation (Jakobsen & Jensen, 2008), the impact of cross-linguistic lexical ambiguity on translation recognition speed (Eddington & Tokowicz, 2012), the co-activation of both SL and TL as examined by reaction times in word recognition (van Heuven, Dijkstra, & Grainger, 1998) as well as by brain activity during lexical access (Wu & Thierry, 2010) — it seems that the issue of vertical vs. horizontal processing is still, to a considerable extent, a subject of debate in present-day Translation Studies.

While there seems to be a general consensus regarding co-activation of SL and TL items in lexical access in bilingualism studies (see below), "many (if not most)" of the models of the translation process "imply some sort of stratificational processes in cycles of comprehension-transfer-production" (Carl et al., 2019 p. 213). Examples of this are not hard to find in the literature. In Gile's (2009) seminal book, the translation process is described as one in which an ST is read, a "meaning hypothesis" is formulated based on linguistic and extra-linguistic knowledge (i.e., a meaning is assigned to the segment), this meaning hypothesis is then checked for plausibility, and upon completion of the processes involved in ST understanding, the translator formulates the meaning hypothesis in the TL drawing on general world knowledge as well as knowledge of the TL. Similar to the ST meaning hypothesis, this newly-formulated meaning hypothesis is also subsequently checked, this time for fidelity and general acceptability, while the TT is continuously revised to arrive at a satisfactory version.

A stratificational, three-stage process of translation is also explicit in Angelone (2010), where the process is considered as cycles of comprehension, transfer, and production, and the "uncertainties" encountered by the translator can be attributed to any of the phases (i.e., comprehension, transfer, and production). A similar perspective can also be found in many other models, e.g., the cognitive model in Shreve & Lacruz (2017) which considers reading, writing, and transfer as "constituent activities". All these models are reflective of a vertical/sequential view of the translation process, and the assumption of complete ST comprehension as the first stage in translation seems to be apparent in many descriptions of the human translation process (e.g., Craciunescu et al., 2004).

Co-existence of horizontal and vertical processing

As already mentioned, there has been evidence from empirical investigations of the translation process that some reformulation processes occur during ST reading, which poses a challenge to the stratificational model (van Heuven et al., 1998; Mossop, 2003; Jakobsen & Jensen, 2008; Ruiz et al., 2008; Wu & Thierry, 2010; Eddington & Tokowicz, 2012; Carl & Dragsted, 2017; Carl et al., 2019). What can be seen from these empirical studies is that the translation processes "can also be based on a *shallow* understanding", and that "ST understanding and TT production can occur in *parallel*" (Carl & Dragsted, 2017 p. 6).

It is therefore not surprising for, e.g., Lin and Xiang (2021), to find that in terms of both theoretical models and empirical investigations, there seem to be more papers in the literature (from 1970 to 2020) endorsing parallel processing than otherwise (i.e., sequential processing), with more recent ones showing an apparent inclination to a "hybrid" view suggesting co-existence of both parallel (horizontal) and sequential (vertical) processing in translation. In particular, the studies that aim to directly verify whether the cognitive processing of language in translation is horizontal or vertical, via robust psycholinguistic experimental methods, are mostly supportive of a horizontal/parallel view (e.g., Macizo & Bajo, 2004, 2006; Ruiz et al., 2008). Some of these studies involve very influential experiments providing convincing evidence that horizontal/parallel processing does in fact occur in the mind of the translator even if vertical processes are also present in the meantime (see also Dong, 2010). In the

Chinese context, the cognitive models proposed by researchers in TPR (e.g., Dong, 2010; Dong & Lin, 2013; Dong & Wang, 2013; Liu & Xu, 2017; Wang & Mei, 2017; Wang & Liu, 2020) seem to consistently endorse the perspective that both processing strategies exist in translation (see also Lin & Xiang, 2021).

In addition to challenging the (exclusively) vertical/sequential view in stratification models, the experimental findings in relation to the (simultaneous) co-occurrence of ST reading and TL reformulation processes are, as Carl & Dragsted (2017) rightly argue, “obviously in contrast with the eye-mind hypothesis when assuming a stratification model”, as the eyes remain on the ST in *both* understanding and reformulation processes, making it fairly difficult to “disentangle” which fixations are associated with understanding and which are “due to pre-translation” (pp. 6-7). Therefore, either the eye-mind hypothesis “has to be weakened” (ibid), or if it is justifiable in its current form (as generally assumed in most eye-movement studies), the stratification model would not be an accurate description of the cognitive processes and “has to be reconsidered” (ibid, p. 7), given the empirical findings from eye-tracking and key-logging experiments.

In this regard, Carl & Dragsted’s argument seems to suggest both (i.e., weakening the eye-mind hypothesis and reconsidering the stratification model).

They argue that during the translation task, translators switch between the two modes, i.e., shallow/parallel and deep/sequential processes, with the former being more frequent than the latter. This is consistent with the standpoint of Dong (2010), who argues that horizontal processing is the normal mode, although this parallel processing is “asymmetrical” and “limited”, and with the “two-stage account” in Dong & Wang (2013), which stipulates that SL comprehension is “accompanied by a parallel processing of TL that merges with SL deverbalization through the function of incremental processing” (p. 126).

Carl & Dragsted’s account of the translation process is built upon the assumption that literal translation proceeds as a default procedure until a monitor interrupts this procedure due to a problem in the outcome, where conscious decision-making is triggered for problem-solving (Ivir, 1981; Toury, 1995; Tirkkonen-Condit, 2005). This literal default procedure is interpreted in Carl & Dragsted (ibid, p. 7) as implying “parallel, tightly interconnected text production and comprehension processes" — "while the mind is engaged in the production of a piece of text, the eyes search for relevant textual places to gather the required information needed to continue the text production flow". Here, the suggestion that literal translation is a default procedure is consistent with what was initially proposed by Tirkkonen-Condit (ibid) as a process-oriented explanation for one of her corpus-based observations in the translation product — the unique items hypothesis. This hypothesis, and the models arising from corresponding perspectives, are described in the next and subsequent sections of the present chapter.

4.2.3 Unique items hypothesis and the monitor model

In the context of an investigation into the target-side universals of translation (T-universals, see e.g., Chesterman, 2004b), i.e., features of translated language that contrast with those in originally written material in the same language, Tirkkonen-Condit (2002) hypothesizes that highly TL-specific items would be underrepresented in translations:

> "… translated texts would manifest lower frequencies of linguistic elements that lack linguistic counterparts in the source languages such that these could also be used as translation equivalents" (p. 209).

This has been named the unique items hypothesis, and considered, although arguably, a potential candidate for translation universals in many descriptive studies. Here, the "linguistic elements" which are deemed "unique" in the above sense, according to Tirkkonen-Condit's description, "may appear in lexis, phraseology, syntax, textual organization, collocational patterns, or pragmatic habits, in other words at any linguistic level" (ibid, p. 216).

Following this hypothesis, as well as her conclusion (among others) that "the role of the unique items is worthy of further research" (ibid, p. 216), a later investigation in Tirkkonen-Condit (2004) includes a corpus-based analysis of the frequency of a few Finnish verbs whose lexical domain does not have direct lexicalized translation equivalents in English, and two clitic particles for which the translation in English depends on their pragmatic function. The relative frequency of those items (i.e., frequency per 1000 words) in four sub-corpora of the Corpus of Translated Finnish — namely, original fiction, translated fiction, original academic, and translated academic — was extracted and compared between original and translated texts in each genre. As Tirkkonen-Condit argues, the results from her investigation support the hypothesis "very strongly" (p. 179), while it is suggested that the explanation for this observation "should be sought in the translation process itself" (p. 177; see also pp. 182-183).

Tirkkonen-Condit's (2004) explanation, largely from a process perspective, is that since these "unique" items are not obvious equivalents for any particular items in the ST, they do not "suggest themselves as first choices for translators, even when they would fit the context very well" (p. 182). Instead, the translator tends to adopt more of a literal translation for the item in question, resorting to an expression which resembles the ST rather than adopting what is typical in the TL. If this (so-called) literal option "makes perfect sense and does not violate the target language norms, there is no immediate reason to discard it" (p. 182). In other words, the observed underrepresentation of those verbs and particles in translated Finnish — from which Tirkkonen-Condit generalizes to TL "unique items" — is due to the "potentially universal" tendency of the translation process to proceed "to a certain extent" in a literal manner, with the translator picking out lexical, syntactic, and idiomatic expressions from the bilingual mental lexicon (p. 183).

This process-oriented explanation of the (product-based) observation is further elaborated in another paper (Tirkkonen-Condit, 2005) which revisits, and provides a proposal in favour of, the "monitor model" of translation — a model which is in effect not new but can be traced back decades (see Toury, 1995, p. 184, pp. 191-192). Tirkkonen-Condit (2005) argues that literal translation seeking formal correspondence is the default procedure regardless of the translator's level of expertise, and that this literal procedure "goes on until it is interrupted by a monitor that alerts about a problem in the outcome" (p. 408), with the monitor triggering off conscious decision-making to find a solution to this problem. Attempts to translate in a literal manner are argued to be present on a lexical as well as syntactic level, and to receive evidence from psycholinguistic experiments, TPR studies using key-logging methods, think-aloud records of translation, repetition and repair in simultaneous interpreting, and product-based analyses regarding what Tirkkonen-Condit calls "unique items" (ibid).

In other words, Tirkkonen-Condit (2002) hypothesizes, largely on the basis that translators do not seem to fully exploit the resources in the TL (Reiss, 1971) and that the translation process results in a tendency to resort to SL-resembling expressions instead of typical TL expressions in a similar context (Toury, 1995), that the linguistic elements unique to the TL in relation to the SL would be under-represented in translation, which, in terms of observed phenomena in the corpus (i.e., translation *product*), is manifested in a lower frequency of occurrence of these "unique items". This hypothesis is elaborated in Tirkkonen-Condit (2004), who also provides empirical evidence for the under-representation of "unique items" from a corpus-based investigation on Finnish, and an explanation for same from the perspective of the translation *process*, while also emphasising that attempts to translate literally (seeking formal correspondence) constitute the default mode of translation, and endorsing the monitor model. The monitor model, which is essentially a model of the translation process and closely relevant to the cognitive activities in the mental processing of language, is revisited in Tirkkonen-Condit (2005) where further evidence is discussed regarding, among others, typical methods of TPR including think-aloud and key-logging. Here, attempts to "translate through formally corresponding material" are described as manifesting themselves "in key stroke logged data as false starts or misprints", although on "close analysis", they "often turn out as attempts to translate literally, or with a formally corresponding item" (ibid, p. 408). Such phenomena are, according to Tirkkonen-Condit, "very frequent" (ibid, p. 411) in Translog files.

Controversy of the hypothesis

Regarding the unique items hypothesis, although it appears intuitively reasonable that certain linguistic elements in translated texts would be systematically under-represented in comparison to non-translations, the hypothesis itself has nevertheless been fairly controversial in corpus-based translation studies. It has indeed received some support from several corpus-based studies on different languages — Finnish (Mauranen, 2000; Eskola, 2004), French (Cappelle, 2012), Spanish (Vilinsky, 2012), and even Chinese (Xiao & Dai, 2014), while there is also considerable criticism and scepticism both conceptually and empirically.

Chesterman (2007), importantly, raises substantial doubts regarding the concept of "unique items", including its vagueness and problems in the identification of these items. Perhaps it is the vagueness of the concept, as well as in practical ways of identifying instances of unique items, that have led to a number of findings which are in sharp contrast to those investigations that are supportive of the hypothesis. For example, Kenny and Satthachai's (2019) analysis of passive markers in Thai shows evidence "strongly against" the unique items hypothesis, in obvious contradiction to Xiao and Dai's (2014) conclusion on Chinese which supports the hypothesis, although the passive markers discussed in both studies show quite some similarities. While it is reasonable to attribute this contradiction to the intrinsic linguistic differences between Chinese and Thai (which therefore refute the notion of "universal" in this regard), it is also not hard to see that the way in which unique items are defined and identified is fundamentally different between these two (and many other) studies. What is meant by "unique items", therefore, seems to be in lack of a proper definition which is straightforward, unambiguous, and more importantly, operationalizable across languages, despite attempts (e.g., Cappelle, 2012, p. 181) to reformulate the hypothesis in different ways.

The fact that the unique items hypothesis depends on the relationship between SL and TL also "contradicts Baker's understanding of a translation universal as arising from the translation process itself and, by implication, as therefore not having to do with the relationship between the languages or textual systems involved" (Malmkjær, 2007, p. 57). In terms of whether it is a translation-inherent feature, Kolehmainen (2013, p. 92) argues that "this kind of under-representation can be triggered by cross-linguistic influence in many different kinds of contact settings, not only in translations", and that the hypothesis therefore "cannot be considered as a universal feature of translation" but instead "a characteristic of multilingual communication in general". Malmkjær (2007), however, is in strong favour of the hypothesis in terms of its potential in this regard as "an excellent candidate for the status of a universal" (p. 56), largely because it receives a cognitive explanation.[22]

Cognitive perspectives

[22] Specifically, Malmkjær describes the unique items hypothesis as the only potential "universal" that receives a cognitive explanation (p. 56):
"It seems to me that of the candidates for universal-hood proposed by Baker (1993), listed by Chesterman (2004) and discussed by Mauranen and Kujamäki (2004), very few qualify for the status of cognitively determined universals. One that does qualify, though, is identified by Tirkkonen-Condit (2004)."

Here, what is important and relevant in this thesis is not the unique items hypothesis itself, but the cognitive perspective on how translation is produced. No matter whether the actual translated product manifests over- or under-representation of certain features, or in whatever way these features are defined and operationalized, perhaps the suggestion that literal translation (or a translation based on "shallow" understanding) is the default mode and proceeds until it is interrupted by a monitoring process, with support from empirical and theoretical investigations of the cognitive aspects of translation, is far more meaningful than the frequency of occurrence of those so-called "unique items", or "literal translations", or items of "formal correspondence".

Indeed, when it comes to universal features of translation (as a product), especially in corpus-based approaches, the issue of causality is of vital importance in the understanding of the nature of translation, although it does not seem to be addressed in much depth in most empirical, corpus-based studies. As Chesterman (2004a) rightly points out,

> To claim that a given linguistic feature is universal is one thing. But we would also like to know its cause or causes. Here, we can currently do little more than speculate as rationally as possible. (p. 44)

It is far from unexpected that theoretical discussion on translation universals has inevitably involved, albeit in a somewhat speculative manner, some aspects of the *process* of translation. Blum-Kulka (1986) holds that "the process of translation necessarily entails shifts", and that explicitation is "inherent in the process of translation" (pp. 17-19). Similarly, Baker (1993) considers the "universal" features as "linked to the nature of the translation process itself" (p. 243) and "a product of constraints which are inherent in the translation process itself" (p. 246). These accounts show "a strong suggestion…that translation universals are cognitive phenomena, since the processes of translation that they inhere in are certainly cognitive processes" (Malmkjaer, 2011). These cognitive processes, however, tend to be either neglected, or addressed in a very speculative manner, in many corpus-based studies, regardless of how "rational" their speculations are.

In this regard, perhaps Carl et al. (2016) are right when they argue that although corpus-based studies have been prolific in generating hypotheses regarding the translation product, drawing inferences on the translation process from those studies has not been easy. Xiao (2010, p. 5) also describes corpus-based Descriptive Translation Studies as a venue which has "primarily been concerned with describing translation as a product". This is perhaps where the discussion, analysis, and suggestion in Tirkkonen-Condit (2002, 2004, 2005), providing insights into the mental processes, especially subliminal processes in the mind, are truly meaningful.

In terms of the cognitive explanations, the hypothesis proposes that these items are under-represented due to the nature of the cognitive mechanism in translation. Their uniqueness means that there are no readily available counterparts in the other language, and thus no stimuli in the ST that can trigger their use. In psycholinguistic terms, these unique items can perhaps be considered as less quickly activated in the bilingual mental lexicon during translation. A neurolinguistic theory of bilingualism (Paradis, 2004) would perhaps explain this in terms of "the activation threshold hypothesis", i.e., the threshold for these items is gradually elevated to a level higher than that for alternative options due to a lack of stimulation, whereas the positive neural impulses associated with the alternative (less unique) options keep their threshold at a low level. What is activated more easily and quickly, according to this explanation, would be the alternative lexical and syntactic options which are not "unique" vis-à-vis the ST, and to Tirkkonen-Condit, this usually means a version that bears much resemblance to the ST, or perhaps more of a literal option.

Tirkkonen-Condit uses the "monitor model" to describe translation processes in which the translator tends to proceed in a literal manner until the literal choice of translation creates an apparent problem and leads to a monitoring process. If a literal choice of translation "makes perfect sense and does not violate target language norms, there is no immediate reason to discard it" (Tirkkonen-Condit, 2004, p. 182). In other words, the unique items do not suggest themselves as first choices for the translator and end up being under-represented.

It is also important to note that in Tirkkonen-Condit's explanation, the literal translation procedure as a default is a direct search in memory (e.g., in the mental lexicon) for "lexical items, syntactic patterns and idiomatic expressions" (Tirkkonen-Condit, 2004, p. 183), which obviously endorses the horizontal/parallel view of translation mentioned above. It seems safe to assume, meanwhile, that the monitoring process would involve a certain extent of conceptual-intentional mediation, which means vertical/sequential processing, although this is not explicitly mentioned in Tirkkonen-Condit's work.

In the following chapters, this process (although not focusing on the "uniqueness" or "literality" of the items) is described as a process of lexical activation and selection at a micro-level, where the TL lexical items are assumed to be subliminally activated upon encounter of the ST in the process of translating, with certain items (regardless of the debatable issue as to whether they are "unique" or "literal") being activated faster (or to a greater extent, or with more cognitive resources allocated) than others, resulting in an entropy that can describe the mental state in this regard. These items (or candidates) come into a competition (or race, to use another metaphor which is common in psycholinguistics), during which the mental state transitions from one to another, re-allocating cognitive resources and reducing entropy, all with the expenditure of cognitive effort. The process will be explained in detail in Chapter 6, where the conceptualization of this process assumes a horizontal view as the default before the monitor interrupts, thus endorsing the monitor model and the general perspective of the cognitive processes in Tirkkonen-Condit's explanation.

4.2.4 Recursive model

The above section has mentioned that the cognitive explanation provided by Tirkkonen-Condit can quite reasonably be described, in psycholinguistic or neurolinguistic terms, as involving a lack or latency of activation, or an elevated activation threshold, for certain items in the TL. To Tirkkonen-Condit, these items are "unique", "non-literal" translations, and ones that do not conform to "formal correspondence" (therefore, her cognitive explanation is sometimes also referred to as the "literal translation hypothesis"). While the so-called non-unique or literal items in the translation product seem to be interpreted in her explanation as evidence for the relevant cognitive processes which are inclined to being "automatic" and "default" (and indeed horizontal), some (e.g., Englund Dimitrova, 2005) suggest that it is also possible to find literal translations which come from non-automatic processes. Therefore, the literal translation product does not necessarily, or strictly, correspond to automatic translation processes, and comparison between ST and TT does not provide a solid basis for determining the cognitive activities in the translation process (Schaeffer & Carl, 2013).

In this regard, the literal translation hypothesis and monitor model are revised by Schaeffer & Carl (2013) on the basis of empirical evidence from priming experiments, where the authors argue that it is the activation of shared lexico-semantic and syntactic representations between source and target language items that facilitates "automated" processing due to a priming effect, and in turn they propose a recursive model of translation (see Figure 4.1 below.) Here, "automated" processing, i.e., the cognitive activity which Tirkkonen-Condit argues results in literal translations (but which was vaguely defined), is understood by Schaeffer & Carl in a way that is consistent with the neurolinguistic theory of bilingualism (see Paradis, 2004). To Schaeffer & Carl, this cognitive activity refers to "procedural, implicit mechanisms and memory which are not accessible to consciousness." Such cognitive activity is also disentangled from "literal translations" or "formal correspondence" — both of which are used by Tirkkonen-Condit in her description of the monitor model, but both of which focus on the translation product only.

In Schaeffer & Carl's (2013) model, the identification of shared (cognitive) representations between SL and TL is automatic, and as an implicit procedural process, once initiated it is not subject to "conscious control over how source and target are aligned cognitively"; neither can this process be verbalized (p. 173), which is in line with Paradis' (2004 pp. 38-45) neurolinguistic description of the controlled and automatic processing of language. The automatic processes, where shared representations are accessed, are horizontal and occur very early on, and while they can be stopped and replaced by controlled and conscious processing on the basis of the output of the automatic processes, switching to a vertical process occurs later. The basis of the horizontal translation process is understood to be formed by priming effects, rather than explicit metalinguistic knowledge regarding translation equivalents (which was assumed to be the basis in Paradis, 1994). In this regard, Schaeffer & Carl explain that their understanding is consistent with Pickering & Ferreira's (2008) argument that priming reflects an implicit learning mechanism.

On the basis of the empirical findings from bilingual priming studies, where results seem to consistently suggest cross-language priming effects both semantically and syntactically, as well as an experiment on priming in translation, Schaeffer & Carl (ibid) describe the literal translation hypothesis in terms of shared representations and automaticity rather than the literality of the TT, and propose that concurrent reading and writing during translation indicate automatic processes and shared representations. The translations produced from automatic processes may or may not be "literal" or instances of "formal correspondence". The horizontal processes are understood as priming processes, while both horizontal and vertical models are integrated into one model accounting for both automatic processes and problem-solving activity.

This recursive model therefore consists of early priming processes in which shared bilingual representations are activated, and vertical monitoring processes, which are later, more conscious, and essentially monolingual (see also Carl & Schaeffer, 2017a, p. 62).

The shared bilingual representations involve both syntactic and semantic aspects, defined respectively in accordance with the shared syntax account (Hartsuiker et al., 2004) and in terms of the Distributed Feature Model (de Groot, 1992). Horizontal priming, which occurs early during ST reading, functions as a default mode of translation, with the primed (and activated) representations serving as "a basis for regeneration in the target language" (Schaeffer & Carl, 2013, p. 185). In translation, these representations are activated twice, as the TT can also result in a prime for the aspects of representations which are shared with the ST (ibid, p. 183). On the basis of Potter & Lombardi's (1998) argument on syntactic priming and sentence recall (i.e., Potter-Lombardi Regeneration hypothesis), Schaeffer & Carl's model seems to stipulate that instead of "*reproducing* a verbatim representation" of the text, translators "*regenerate*" it "on the basis of primed representations" (p. 175, their emphasis). It is perhaps for this reason that in Carl & Schaeffer's (2017a p. 62) description of the model, early horizontal priming processes during ST reading involve regeneration of the ST (based on the primed representations in these processes), and the ST regeneration is distorted by monitoring processes which introduces "translation variance on a lexical and syntactic level" (ibid, p. 62). Here, monitoring processes are vertical. The vertical, monitoring processes depend on context (as well as target norms) which becomes available later, and these processes "access the output from the automatic default procedure recursively in both the source and the target language and monitor consistency as the context during translation production increases" (p. 186). It is worthy of mention that the "output" from the automatic default procedure does not necessarily refer to a translation draft that is already produced (or partially produced) but can also be (perhaps more commonly) an interim (and roughly reformulated) TT that is kept in working memory. In this model, Schaeffer & Carl argue that during decoding, “both horizontal and vertical processes are always active at the same time” (ibid, p. 185).

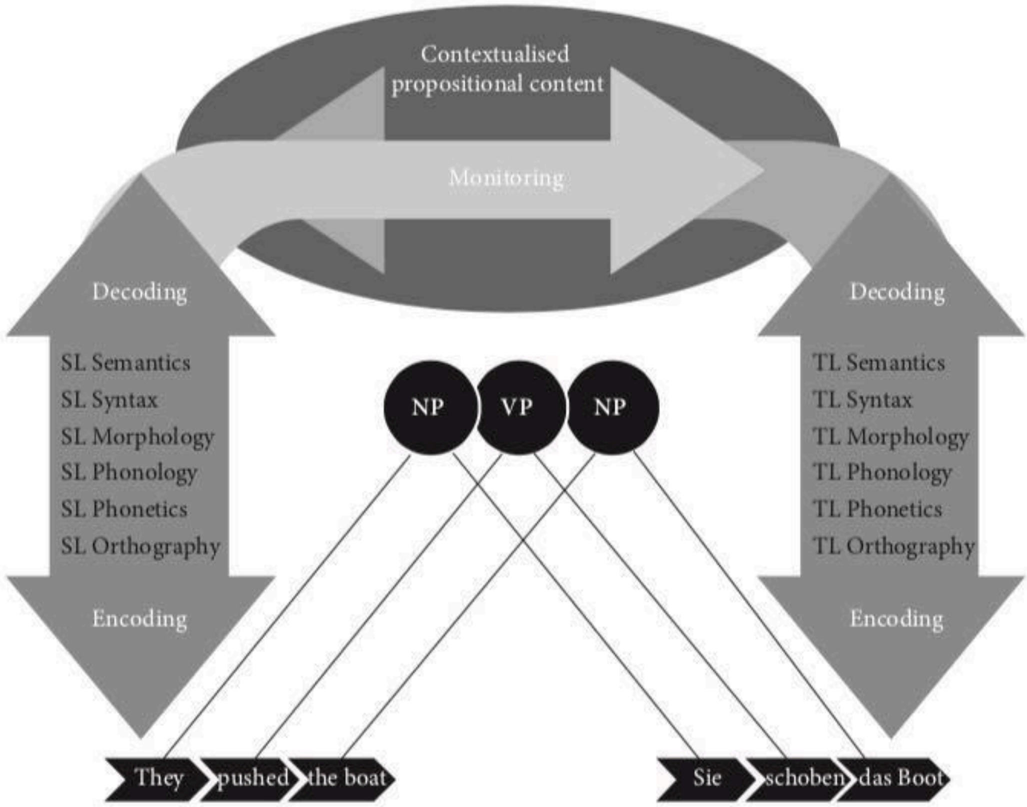


Figure 4.1 Recursive model of translation (from Schaeffer & Carl, 2013, p. 182)

The investigation in Chapter 6 into patterns of gaze behaviour — as well as the conceptual explorations in that chapter — adopts a perspective that shares considerable similarity with this model, assuming subliminal, automatic, and early activation of lexical entries in the TL upon encounter of the ST, but whereas the recursive model holds that the monitoring processes which interrupt the default, early, automatic, and horizontal priming processes are vertical, the observations in that chapter seem to suggest that these processes might still involve much horizontality to a large extent.

4.2.5 Default translation

The above description of the default mode of translating — which to Tirkkonen-Condit means "literal translation" and to Schaeffer & Carl means automatic, early and horizontal priming — shares much similarity with Halverson's (2019) construct of "default translation", although she considers this as a construct within an opposing paradigm to TPR (i.e., cognitive translatology, see Muñoz, 2016a, 2017).

To Halverson (2019), this is a phase where "translators demonstrate stretches of uninterrupted production", drawing on "easily accessible, routinized knowledge, including bilingual linguistic knowledge, metalinguistic knowledge (including knowledge of communication norms), and knowledge of the specific task" (ibid, p. 190). She explains that her construct of default translation rests "clearly on an embodied cognition philosophy and a theory of language that is consonant with it (cognitive grammar)", and that these foundations "are central to the content of the construct and the hypotheses that may be derived from it" (p. 189). Despite the strong emphasis on the dichotomy, difference, and "mutual exclusiveness" (Muñoz, 2016a, 2017) between what is called the *computational approach* (referring to traditional TPR, and described as a view of translation as a problem-solving activity) and the opposing *cognitive translatology* (which draws on the perspective that cognition is embodied, embedded, enacted, extended, and affective, i.e., 4EA cognition, see also Risku & Rogl, 2020), it seems that the methods of empirical investigation into the default mode of translation, via behavioural data of the translation process, are largely consistent with the typical methods used in TPR. Conceptually, Halverson puts considerable emphasis on bilingual co-activation during translation, on cross-linguistic priming effects, and on "a control mechanism to suppress potential targets in the process of selecting among many candidates and for some, inhibitory links in addition to excitatory ones" (ibid, pp. 192-194), which are perhaps all consistent with the basic assumptions in the recursive model, with the general psycholinguistic perspective of lexical access and disambiguation, and with what will be explored in the following chapters. In addition, the description that this default translation proceeds "in a relatively routine, uninterrupted manner" where "translational choices seem to be made without obvious deliberation" (p. 188), appears to share the same view as Schaeffer & Carl's recursive model in terms of the automaticity of the procedure as well as the procedure's being not consciously controlled. With regards to the parallel processing of ST comprehension and TL reformulation, Halverson (2019) stresses that the task of translation "affects ST reading quite early, suggesting that translation-relevant cognition is involved also in ST reading while at task", which is apparently consistent with the recursive model in the claim that horizontal processes occur early in translation.

The default mode of translation is summarized in Halverson (2019) as involving "the generation of fast and easy translations which are based either on coactivation patterns that have been established through bilingual language use (without translation) or that have been established through prior translation activity and subsequently reused for that purpose more or less frequently, or both" (p. 196). The fact that this perspective draws heavily on the patterns of joint co-activation across the SL and TL, on priming effects, and on the competition among the co-activated candidates as well as the selection process in which some potential targets are suppressed by a control mechanism, is also consistent with the discussion in the present thesis, especially the conceptual explorations in Chapter 6.

In addition, Halverson also seems to closely relate the default mode of translation to issues of ambiguity, stating that:

> … default translation can be found where the translator is translating a particular stretch of language for the first time, if the linguistic material used is familiar and the cross-linguistic relationships relatively unambiguous. Conversely, co-activation patterns in non-translational activity, for example, may also result in there being more than one translation candidate, which would most likely prompt a transition to deliberate consideration and choice, and a transition into a non-default phase. (p. 196)

The specific, micro-level processes of "deliberate consideration and choice", as well as behavioural manifestations of these processes, are analysed in Chapter 6 through the lens of bilingual activation and entropy reduction, drawing inferences from resource-allocation processing difficulty in psycholinguistics, and on the basis of large-scale empirical data involving eye-movements.

4.2.6 Other models of the translation process

Here, the illustration of the cognitive models of translation has focused primarily on horizontal vs. vertical views, and on the models that adopt non-selective co-activation of both SL and TL (based on a broad consensus in psycholinguistics and bilingualism). These models are also proposed on the basis of empirical evidence from experiments, using either product- or process-based methods on large-scale data (e.g., corpus, eye-movement, and key-logs). While these seem to be the most relevant to what this thesis will be investigating in the coming chapters, there are nevertheless numerous other models in the literature, either of a somewhat speculative nature or summarized from a bottom-up approach, that are equally important in terms of explaining the complex cognitive activities in translation.

Halverson (2003) explains the cognitive basis for the "universals" proposed in corpus-based studies and argues that various translation universals "designate essentially the same thing, and represent the effects of gravitational pull exerted by category prototypes" (p. 221; see also Halverson, 2017). Therefore, she proposes what has been called the *gravitational pull hypothesis*, integrating cognitive grammar and the Distributed Feature Model of bilingual representation into her psycholinguistic model of translation (see also Carl & Schaeffer, 2017a, pp. 60-62). The "gravitational pull" model suggests that due to the existence of asymmetries in the cognitive organization of semantic information, highly salient linguistic items are more prone to be chosen, thus leading to their over-representation in the product data (i.e., corpora of translated language).

Cognitive activities in the translation process have also been modelled from a relevance-theoretic perspective. Alves & Gonçalves (2007) integrate relevance theory (Sperber & Wilson, 1986; Gutt, 1991) into a connectionist framework (Elman et al., 1996), and describe translation as implying "the progressive and recursive generation of contextual effects from two counterpart translation units and the search for an optimal overlapping between these two sets of effects" (p. 44). Such a psycholinguistically oriented, relevance-theoretic model (see also Alves & Jakobsen, 2022, p. 38) assumes that the principle of relevance guides translators in the processes of problem-solving and decision-making, and that the driving force and the ultimate goal of the translation process is seeking for interpretive resemblance. In Alves & Jakobsen's (2022) description of such a model, a translation unit is "first processed automatically in the Adhoc Block as a default procedure" and "moves into the Rest Block only when a solution cannot be found", while in the Rest Block "mechanisms of external and internal support interact recursively" (p. 38).

More recently, a post-cognitivist perspective of the translation process has been developed on the basis of the dynamic systems approach of the radical embodied theory of cognition (Carl, 2021c), where the notion of translation affordances is introduced and extended with a probabilistic recursive layer. In this account, translation affordances are viewed as basic units of cognition in translating, and translation units are conceptualized as cycles of perception-action in translation production.

Some other models of the translation process (Muñoz, 2016a, 2017; Muñoz & González Fernández, 2021; Risku & Rogl, 2020) place a strong focus on 4EA cognition, i.e., the perspective that cognition is embodied, embedded, enacted, extended, and affective, aiming to become an alternative to the traditional approach. Such models are often described as belonging to cognitive translatology, and tend to criticize TPR studies for adopting a problem-solving, information-processing, and "computational" perspective, although the term "computation" here seems to be very vaguely defined and somehow taken to be the same as "information processing". While the situatedness of cognition is without doubt an important aspect in translation, it has been argued that computation and information processing should not be used interchangeably, that theorists need to be careful in choosing notions of computation and information, and that the assimilation of such concepts as computation, information processing, and the manipulation of representations "does more harm than good" (Piccinini & Scarantino, 2010, p. 237). In Carl (2021a), various notions of computation, and the typologies of representation, are discussed in view of this terminological confusion, while TPR is argued to be compatible in many aspects with 4EA cognition perspectives (i.e., cognitive translatology). Similarly, Alves & Jakobsen (2021) state that the paradigmatic and methodological differences between computational information processing and the situated-action paradigm "are best addressed from a complementary, integrative approach" (p. 546), although their account also provides no definition of "computation". Despite the controversy, drawing on the 4EA perspective of cognition, cognitive translatology understands the cognitive aspects of translation by reference to more context-sensitive, embodied frameworks (see Muñoz, 2016b).

In this regard, the above description regarding priming, automaticity, monitoring, bilingual co-activation, competition between various activated items, etc. seems to be consistent with the definition that translating as a cognitive activity encompasses controlled and uncontrolled processes and requires problem solving, decision making, and the use of translation strategies and tactics (Alves & Hurtado Albir, 2010). Alves & Jakobsen (2022, p. 48) argue that most of the empirical results from TPR support this definition. This thesis does not assume a dichotomy of mutually exclusive views between the computational approach and cognitive translatology, nor an affiliation to either, but follows the broad consensus in psycholinguistics regarding non-selective activation of both SL and TL, which has been supported by evidence from both bilingualism and translation studies, and considers this consensus as the assumption behind theoretical explorations. Attempts are also made to analyse large-scale eye-tracking and key-logging data to shed light on the behavioural manifestations of the conceptual explorations (and to validate the theoretical discussion).

4.3 Uncertainty, problem solving, and cognitive load

4.3.1 Uncertainty in the stratification model

From a cognitive perspective, the process of translating involves decision-making as well as uncertainty management. Angelone (2010, p. 17), for example, succinctly describes the task of translation as essentially "a chain of decision-making activities relying on multiple, interconnected sequences of problem solving behaviour for successful task completion". Based on this view, translation is considered as a progression of "problem solving task sequences" which are "interspersed with unproblematic translation sequences where comprehension, transfer, and production occur unimpeded by any difficulties" (ibid). Once difficulties of any kind arise during this progression, there will likely be some extent of *uncertainty* in the translator's decision-making, i.e., a cognitive state of indecision (ibid), which would in turn require additional cognitive effort from the translator.

In Angelone's description, such uncertainty can be marked by an "observable interruption" in the natural flow of the translating process, which may be empirically studied through certain diagnostic behaviours including keyboard and interface behaviours such as "extended pauses in target text generation, deletions and/or revisions, cursor repositioning, and information retrieval behaviour (e.g., dictionary look-ups, terminology look-ups, Internet searches, to name just a few)", as well as physiological indicators such as "eye movements, changes in pupil size, increases in certain kinds of brain activity, or even changes in physiological parameters such as galvanic skin response" (p. 18).

Uncertainty is considered to be associated with the general translation-oriented processes in the form of comprehension uncertainty, transfer uncertainty, and production uncertainty, of which the specific behavioural indicators are illustrated in Angelone (2010, p. 21).

However, as Angelone admits, "non-articulated indicators, such as pauses and eye-fixations, give us no real clue as to how and where to allocate the uncertainty" (p. 23). Instead, Angelone suggests using verbal protocols as a way of helping to distinguish comprehension-oriented phenomena from transfer- and production-oriented phenomena (ibid, p. 23). As is generally believed, though, Think-Aloud Protocols (TAP) can only reflect the processes that are conscious, while the subliminal, subconscious mental activities are not indicated. Together with the fact that the experimental participant may not be objectively aware of his/her complete cognitive processes, it seems that the allocation of the uncertainty to one of the processes is largely infeasible. In addition, as described above, these processes "do not normally exist independently in the translator's mind" (Carl & Dragsted, 2017, p. 8), making it even more difficult for a clear-cut allocation of such kind.

As explained in Chapter 3, the notion of "uncertainty" associated with entropy is from a slightly different perspective. This notion, especially based on bilingual lexical activation, seems better suited to the translation processes where both parallel (horizontal) and sequential (vertical) processes are involved.

In the meantime, a probabilistic view of uncertainty provides a convenient means for both conceptual and empirical investigations into ambiguity, as this can be mathematically formulated. Such mathematical formulations, as mentioned in Sections 1.2 and 3.4.2, are entirely consistent with neuroscientific theories such as the free energy principle, where this (probabilistic) notion of uncertainty describes the internal state of the cognitive agent. A change in the uncertainty (and entropy), from that view, is an update of belief in the mental state.

Regarding the issue of ambiguity, disambiguation in translation and post-editing can perhaps be considered, from a cognitive perspective, as a phenomenon in relation to instances of decision-making on translation choices which may result in some extent of uncertainty. This can involve, in the case of translating, a choice among the different translation alternatives for an ST lexical, phraseological, or syntactic item, while in post-editing it can also involve the additional choice-making between what is presented to the post-editor (i.e., MT output) and what is stored in the post-editor's mental lexicon, in relation to the same ST item.

4.3.2 Means-ends analysis

In the literature of cognitive science and educational psychology, a general problem-solving strategy called means-ends analysis (e.g., Newell & Simon, 1972; Sweller, 1988) seems to be widely considered as a typical — if not the "most important" or "best-known" — cognitive procedure for finding solutions to problems (quoted from Sweller et al., 2011, p. 9). Means-ends analysis is considered an essential skill which "all normal humans have evolved to use" (ibid), relying on biologically primary knowledge (see Geary, 2007, 2008) which is acquired easily, automatically, unconsciously, and without tuition.

According to Sweller et al. (2011), means-ends analysis is adopted whenever a problem is encountered to which the solver does not have domain-specific knowledge of a solution. Irrespective of the problem domain, it includes the following steps — simultaneous consideration of the current problem state and the goal state, search for differences between the two states, search for legitimate problem-solving operators to reduce these differences, and selection of a problem-solving operator which can be applied to the current problem state. This final step of applying the problem-solving operator leads to the creation of a new state which is subsequently checked by the problem solver as to whether it corresponds to the goal state. If it does, the problem is solved; otherwise, "the procedure is used recursively until the goal is attained" (p. 9).

This strategy, however, would become less prominent as the problem solver increases his/her exposure to the domain and accumulates domain-specific knowledge of solutions, thereby recognizing the relevant problem states and detecting the appropriate solution move for each state without searching for moves (ibid). This change occurs only after "considerable learning" and "only applies to particular, domain-specific problems" (ibid).

In other words, the difference in problem-solving when means-ends analysis is used versus when previously-learned domain-specific knowledge is used, can be safely regarded as a crucial difference between novices and experts in the relevant domain. When solving problems in a particular domain, novices tend to use means-end analysis while experts are more inclined to domain-specific knowledge. This is in fact a pattern which has been supported by evidence from a substantial amount of literature on expert-novice differences in problem-solving (e.g., Larkin et al., 1980a, 1980b; Sweller et al., 1983). On the basis of these experimental findings in the literature, Sweller et al. (2011) comment that the awareness of some knowledge indicating a solution of a problem will lead to routine use of that knowledge, and that in the absence of the knowledge of appropriate problem-solving moves, a means-ends strategy will be used (p.10).

The above findings have implications for TPR. Domain specialists translating/revising/post-editing texts in their own domains can be assumed to be capable of effortlessly, intuitively, and sometimes automatically associating particular domain-specific terms with the corresponding concepts (or even with the equivalent target-language terms). Any ambiguity involved in the relevant lexical items would not incur much difficulty or uncertainty for them, leading to an absence of problem-solving strategies such as means-ends analysis but instead resulting in a direct recognition of the problem state and the appropriate solution move. This is the consequence of enhanced, domain-specific learning into which these specialists have devoted significant amounts of years in their education and career. On the other hand, when it comes to issues which are more relevant to linguistics and translation in general (e.g., syntactic ambiguity) than to domain-specific knowledge, it may be assumed that more problems, together with a higher level of difficulty and uncertainty associated with these problems, would be prominent, and that in the process of generating moves that are appropriate to their solutions, there could be an inclination to general problem-solving strategies such as means-ends analysis rather than to direct moves on the basis of domain-specific knowledge. This is why it is important to differentiate between novices and experts, on the one hand, and non-specialists and domain specialists, on the other hand, when analysing the kind of user activity data collected in contemporary TPR. In the empirical analysis in Chapter 8, consideration is also given to different levels of expertise, where the observed pattern seems to be consistent across different levels (see Section 8.4). This is also an issue to which Chapter 10 returns, in terms of the conceptualization of entropy and entropy reduction in the mental states during lexical selection for different individuals. With respect to this conceptualization, note that the approximation of entropy, especially in the CRITT method (calculated from the translation text produced by all participants in a particular experimental session), extrapolates from the observed performance of many individuals to the potential performance of one individual, as mentioned in Section 3.5.3. The choices which other translator/post-editors have made may not necessarily be available for the individual in question, and all choices might not be available to any one individual translator/post-editor either. This will be addressed in Chapter 10, and similar issues are also touched upon in Chapter 7 (Section 7.5.1).

4.3.3 Cognitive load

In educational psychology, the concept of cognitive load is often used to refer to the amount of working memory resources which need to be allocated to process the information in instructional materials, i.e., the load imposed on working memory by the instructional information (e.g., Leahy & Sweller, 2011; Sweller et al., 2011; Sweller, 2016). Although this was originally intended as an inquiry into the most effective means to teach, the concept of cognitive load as a psychological construct (e.g., Moreno & Park, 2010) can be usefully applied to the broader scope of problem-solving.[23] Cognitive load theory (CLT), first proposed by Sweller (1988) and comprehensively explained in, e.g., Plass et al. (2010) and Sweller (2011), describes how this load can be divided into several categories on the basis of its function (see also Sweller et al., 1998; Paas et al., 2003, 2004; van Merriënboer & Sweller, 2005), frequently including such types as intrinsic, extraneous, and germane cognitive load.[24]

Intrinsic cognitive load refers to the working memory load that is imposed by "the intrinsic nature of the information" (Sweller et al., 2011, p. 57), i.e., by the basic structure of information (or task, problem, etc.) rather than by the manner in which this information (or task, problem, etc.) is presented.

A contrast to this is *extraneous cognitive load* — the type of cognitive load which is imposed on the working memory as a result of the presentation of the material (or, in the context of education, the instructional procedures) and which can be "unnecessary" in many situations (ibid).

Germane cognitive load, although frequently used in an "inappropriate" (ibid) manner to refer to germane resources in the working memory, was actually first described in Sweller et al. (1998) as a reference to the working memory resources that are devoted to creating a permanent store of knowledge, generally in the form of schema construction and automation. This load arises from the processes that are directly relevant to learning. Unlike the previous two types of cognitive load which are imposed by either the nature, or the structure, of materials, germane cognitive load

---

[23] According to Moreno & Park (2010), *cognitive load* is one of the main constructs of interest in cognitive load theory (the other being *learning*).

[24] CLT is an instructional theory on the optimization of teaching, which explains how instructional design affects two psychological constructs — cognitive load and learning. It provides guidance as to how extraneous cognitive load can be diverted to germane cognitive load, thereby maximising learning in the form of schema construction, storage, and automation.

is not imposed by the materials themselves, but rather, by information which is "relevant or germane to learning" (Sweller et al., 2011, p. 57), i.e., the cognitive processes of constructing schemas as acquisition of competence and expertise.

Since the present thesis is focused on the effort which is expended *during* the translation and post-editing tasks, instead of a longitudinal study on the subjects' acquisition or development of translation or post-editing skills, perhaps germane cognitive load bears very little relevance to the research questions. In a similar manner, the fact that the focus in this research are linguistic issues related to the ST and the TT — the textual materials on which a translator/post-editor works, rather than the environment in which the translation/post-editing is conducted, or the interface/functions of the software used — means that the cognitive effort to be analysed in relation to the linguistic features (i.e., lexical and structural ambiguity) would be *intrinsic* rather than extraneous. Therefore, the cognitive effort investigated in this thesis is equivalent to *intrinsic cognitive load* from the perspective of Sweller's CLT.

This view is consistent with some other theoretical discussions on the effort involved in post-editing MT output. For example, Vieira (2016) draws a similar parallel between post-editing and the intrinsic vs. extraneous classification of cognitive load in Sweller's theory:

> … the complexity/difficulty of the ST and the quality of the MT output could be seen as representing intrinsic cognitive load. Extraneous cognitive load would be posed by external factors such as the editing tool and any extra technical difficulty it might pose, for example. (p.12)

Although in this respect, Vieira's discussion seems limited in that it has entirely neglected germane cognitive load — an indispensable category of cognitive load which would be vital for translation/post-editing in terms of competence acquisition, it does provide a meaningful, succinct, and accurate analogy between the cognitive load theory and the task of translation/post-editing.

In addition to this analogy, he gives a concise but comprehensive analysis on the (potentially confusing) terminology which is employed in cognitive psychology and cognitive load theory, particularly regarding *mental* as opposed to *cognitive* as well as *effort* as opposed to *load*, *capacity,* or *attention*. Vieira adopts the term *cognitive effort*, as in his study the demands of the task are estimated independently of effort, and as he regards *cognitive* effort and *mental* effort as interchangeable terms for the same concept (ibid, p. 12).

Here, some theories tend to differentiate these terms (e.g., *mental load* v.s. *mental effort*; *cognitive load* v.s. *mental load*), often using them to refer to concepts that are relevant but not exactly the same (according to those theories), especially regarding the variables involved in cognitive load.

For example, Kirschner (2002) maintains that cognitive load is affected by both causal and assessment factors, where these types of factors are explained as follows:

> Causal factors can be characteristics of the subject (e.g., cognitive abilities), the task (e.g., task complexity), the environment (e.g., noise), and their mutual relations. Assessment factors include mental load, mental effort, and performance as the three measurable dimensions of CL.[25] (pp. 3-4)

Here, it can be seen that mental *load* and mental *effort* are not considered as one and the same concept, but two different dimensions of cognitive load. In the same manner, *mental* load is not considered the same as *cognitive* load, as the former is one of the factors effecting the latter.

In a more specific manner, Kirschner provides a concise description of the three assessment factors:

> Mental load is the portion of CL that is imposed exclusively by the task and environmental demands. Mental effort refers to the cognitive capacity actually allocated to the task. The subject's performance, finally, is a reflection of mental load, mental effort, and the aforementioned causal factors. (p. 4)

In fact, the concept of *mental load* as a psychological construct — and perhaps the idea of cognitive load in general — was not new, even at the time when CLT was developed (Moreno & Park, 2010). In the domain of human factors psychology, for example, Moray (1979) defines mental load (or mental workload) as the difference between the demands of the task and the ability of the human to handle these demands.

---

[25] In Kirschner's description, cognitive load is abbreviated as CL. The same is for cognitive load theory, which is abbreviated as CLT.

Many other psychological theories define it in a way that recognizes its "multidimensional nature of the mental load construct" (Moreno & Park, 2010, p. 10).

Therefore, the use of such terms can sometimes be confusing (with different theories defining it in different manners). In this thesis, however, consideration of their detailed conceptual differences, although important indeed, does not seem necessary in view of the research questions related to entropy. This thesis considers *cognitive load* as the difficulty (or demand) that is posed by a task or process (i.e., the required amount of cognitive effort), and considers *cognitive effort* as the actual effort expended in the process or task (where this effort is realized by optimizing the allocation of limited cognitive resources, see Chapters 5 and 6).

In the discussion in the following chapters, the thesis adopts the term *cognitive load* to refer to the load that is imposed by the ambiguity involved at a particular point of the text, and does not differentiate cognitive load and mental load, similar to Vieira (2016). This load is theoretically discussed in terms of the shift of resource allocation and in turn via entropy, on the basis of bilingual lexical activation (see Chapters 5 and 6).

4.4 Post-editing of Machine Translation

Computers have brought radical changes to the way translation is produced today. One such change is the use of translation memory (TM), i.e., previously translated segments aligned at sentence level, which can be searched and suggested to the translator as a rough translation for a source sentence that bears similarity to the segment they are currently translating. Along with TM, MT output is increasingly used as part of the workflow in the industry, largely due to the significant increase in its translation quality in recent years. As a consequence, the translator in a professional setting is perhaps becoming a post-editor working on the rough translation suggestions either from the TM of a computer-aided translation system, or from an MT engine. Today's translation involves a considerable extent of human-computer interaction (O'Brien, 2012), and professional translation is increasingly becoming a form of activity in which translators "switch between fuzzy match editing and MT post-editing as they move from one segment to the next within their TM tools" (O'Brien &

Moorkens, 2014, p. 2). The distinction between such terms as "translator" and "post-editor" seems to be becoming vague (see also O'Brien, 2022), as translators are extensively involved in the revision of either TM suggestions or MT output in typical professional translation workflows. In view of the discussion on entropy, the perspective of entropy and entropy reduction seems compatible with both translation and post-editing (the only difference being the presence of MT output as an additional factor impacting the entropy value, if we want that value).

Along with this change, there is growing interest in studying the cognitive processes of the human-computer interactive production of translation. The investigation of the translation process in this thesis incorporates both translation and post-editing.

4.4.1 MT with human post-editing

Although Fully Automatic High Quality Machine Translation largely remains unfulfilled to the present day, the use of MT in different scenarios has always been recognized since the beginning of MT development, for both dissemination and assimilation purposes (e.g., Hutchins, 2003). In recent decades, the popularity of MT has further increased dramatically (e.g., Doherty & O'Brien, 2012; Graham et al., 2014), together with an earlier change in paradigm from the rule-based approach towards the corpus-driven, statistical approach (i.e., SMT, see, e.g., Koehn, 2010), and then a recent shift towards a deep-learning approach which is based on neural networks (i.e., NMT, see, e.g., Kalchbrenner & Blunsom, 2013; Cho et al., 2014; Bahdanau et al., 2015; Kohn, 2020; Pérez-Ortiz et al., 2022). This has resulted in many successful examples of the practical use of the raw output from MT, and of combining MT with human post-editing to produce publishable translation — to the extent that the use of MT seems to be becoming a de facto standard in the translation industry.

With the pervasive popularity of MT both in commercial settings and among the general public, it seems well-accepted that MT nowadays is not confined to information gisting purposes, but is largely and increasingly incorporated into the process of creating texts of publishable quality.

Nevertheless, human post-editing is still undeniably indispensable if the output from MT is to be used for serious application. In this regard, a question is often raised:

how effective is MT, in terms of facilitating the production of translations of publishable quality?

4.4.2 Research on post-editing effort

Not surprisingly, much recent research on MT output has focused on the effort involved in the post-editing of the raw output from MT, typically following Krings' (2001) seminal categorization of post-editing effort into three aspects — temporal, technical, and cognitive. Most, if not all, of the results from these studies are positive about the effectiveness of MT post-editing compared with translating without this aid (Garcia, 2011; Groves & Wicklow, 2008; Guerberof, 2014; Kirchhoff et al., 2011; Turner et al., 2014), even though there is still "significant translator resistance to the task" (O'Brien & Simard, 2014, p. 159). In general, such research reveals that MT with human post-editing has been very useful in the translation workflow, providing rough draft translations for the segments whose fuzzy match to the TM is below a threshold (e.g., 75% similarity).

In more specific directions, much research is now focusing on investigation of the variables which influence one or more aspects of post-editing effort, particularly regarding ST features, MT output quality, and the characteristics of post-editors.

Krings (2001) provides some preliminary indications regarding the influence of ST features on the speed of post-editing, the verbalization effort in Think-Aloud Protocol (TAP), the frequency of cognitive processes and attentional shifts in post-editing. While Krings' research is useful, important, and profound, the unit of his analysis is the text, and consideration of ST features is confined to text and sentence length. It has been pointed out that a more precise approach would be to analyse linguistic features at the sentence level (Tatsumi, 2010) and that more sophisticated linguistic aspects (e.g., lexical frequency, Coh-Metrix indices, etc.) are needed (Vieira, 2016).

In this regard, O'Brien's (2004) analysis of negative translatability indicators (see, e.g., Underwood & Jongejan, 2001) in relation to post-editing speed is particularly meaningful, highlighting the impact of different negative translatability indicators on post-editing. Similar research also considers the relationship between post-editing and controlled language (CL), regarding the technical (Aikawa et al., 2007) and temporal (O'Brien, 2006a) effort of post-editing, while Tatsumi (2010) and Tatsumi and Roturier (2010), among others, correlate automatic scores of ST complexity (or quality) to post-editing time. Gerlach (2015) analyses how pre-editing the ST impacts post-editing speed in the context of SMT for English-to-French translation, and found that the pre-edits result in faster post-editing.

For the correlation between ST features and post-editing effort, studies are often confined to the particular type of MT technology at the time, and findings in older-generation systems may not be transferrable to newer systems. This is also consistent to findings of pre-editing studies, as the linguistic features that cause issues for the MT system would be different between, e.g., NMT and SMT or RBMT. For example, Miyata and Fujita (2021 p.1547) point out that, in the context of NMT, traditionally recommended pre-editing methods (e.g., making sentences shorter and simpler) are not as important as some other aspects of the ST (e.g., clarity of the content, syntactic relations, and word senses) for better output. Similarly, Marzouk and Hansen-Schirra (2019) found that although pre-editing the ST leads to improvement in the output from an RBMT, an SMT, and a hybrid system, this improvement is not shown in the context of NMT (their study tested technical translation from English into German).

Regarding correlations between MT quality measures and post-editing effort, there seem to be fewer studies to date. Some of such work includes Krings (2001), using human ratings of MT quality at the sentence level, and O'Brien (2011), who uses GTM and TER evaluation metrics. Regarding other MT evaluation metrics, Gaspari, et al. (2014) investigated how BLEU, TER and Meteor are potentially correlated to post-editing time. Some other investigations of post-editing effort are based on error typologies, e.g., Temnikova (2010), Lacruz et al. (2014), and Zaretskaya et al. (2016). A succinct overview of such research can be found in O'Brien (2022).

The characteristics of post-editors have also been investigated in connection with post-editing effort, regarding post-editing professional experience (e.g., Guerberof, 2014), hourly payment (Green et al., 2013), perceived post-editing effort (Moorkens et al., 2015), etc.

4.4.3 Ambiguity in translation and post-editing

This thesis can be considered as following the discussion on the relationship between ST features and translation/post-editing effort, but more importantly, the mental processes therein, seeking a step further by homing in on lexical and syntactic ambiguity. In addition to being difficult for MT (especially those instances which are usually termed "accidental ambiguity" (see, e.g., Hutchins & Somers, 1992; see also Section 2.1.2), the types of ambiguity related to the human perception of language (i.e., real ambiguity) are also important for translation and post-editing, as the translator/post-editor has to cognitively resolve the ambiguity both in the ST and, where appropriate, in the machine-translated output. In the context of human sentence processing, psycholinguists have often argued that ambiguity tends to be a hindrance to efficiency in the cognitive processes (which can in turn be reflected in behavioural observations such as eye movement). Studies in TPR have also found similar phenomena in this regard, both theoretically and experimentally, as mentioned in Section 2.3. In MT post-editing research, however, this seems to be little investigated.

"Temporary ambiguities", or ambiguous regions of sentences that can be resolved by the following parts of the sentence (Winkler, 2015), are also interesting in terms of the cognitive effort of the translation/post-editing process, where the ambiguity can be either lexical (e.g., Tanenhaus et al., 1995), or more commonly, structural (e.g., "garden path sentences" discovered by Bever, 1970, see Chapter 2).

Statistical measurements of ambiguity including word translation entropy (Schaeffer, Dragsted, et al., 2016) and syntactic choice entropy (Carl, Schaeffer, & Bangalore, 2016) have been introduced in Chapter 3. Other lexical variables that can predict translation probability (Prior et al., 2011) and the cognitive effort of language processing (Attneave, 1959; Hale, 2001; Levy, 2013; Levy & Gibson, 2013) on the basis of frequency (e.g., Zipf, 1949; Rayner et al., 1996) and surprisal (e.g., Levy, 2008, 2013) will be reviewed and analysed systematically in Chapter 5.

The conceptual and empirical investigations in Chapters 6 and 8, largely from the perspective of bilingual lexical access and entropy reduction, apply to both translation and post-editing, as can be seen from the examples in Chapter 8. The analysis on structural ambiguity in Chapter 9 focuses specifically on post-editing. These chapters involve detailed analyses on the cognitive processes involved, on the basis of empirical data from eye-tracking and key-logging studies.

4.5 Machine translation approaches

Since approaches in NLP are often categorized into symbolic approaches and empirical (statistical/stochastic) approaches (Indurkhya & Damerau, 2010), MT systems have correspondingly been distinguished by the type of NLP techniques they adopt in the computational design. The difference between these two kinds of approaches reflects the difference between stratificational (sequential) views of translation and the horizontal processing based on shallow understanding (see Section 4.2 above). In Carl & Schaeffer's (2017a) account of the historical development of translation models in both computational linguistics (for MT engines) and translation studies (regarding human translation), it is clearly shown that although "there does not seem to be a large overlap between the two research communities, there are a number of commonalities in the features of the models they developed" (p. 50).

With respect to different types of MT systems (rule-based, statistical, or neural, see following sections), the empirical investigations in Chapters 8 and 9, where post-editing is involved, are mostly based on statistical MT. This will also be described in those chapters.

4.5.1 Symbolic NLP approaches and Rule-based MT

As illustrated in Kwong (2015, p. 565), symbolic approaches in NLP use "manually crafted" procedural/declarative knowledge, and mostly follow the AI tradition, having later been developed into knowledge-based or rule-based approaches where *knowledge* may refer to rules, semantic lexicons, ontologies, etc. This is largely the basis for rule-based MT (RBMT) systems or knowledge-based MT (KBMT) systems, which rely on "morphological, syntactic, semantic, and contextual knowledge" regarding the source and the target languages in the process of translating (Yu & Bai, 2015, p. 186).

While early proposals of MT were largely based on a direct, word-to-word replacement of the ST with equivalents in a bilingual glossary, it was soon found to be inadequate, oversimplistic, and perhaps naïve (e.g. Arnold, et al., 1994). For an MT system to really function in a satisfactory manner, an adequate degree of syntactic and semantic analysis is a necessity, and in the ideal situation for multilingual systems, there would be an abstract and somehow language-independent representation (i.e., interlingua) that can be defined to bridge the source sentence analysis and target sentence generation, so that the translation would be based purely on semantic and rhetorical information. Such systems try to find a structured representation of the ST meaning in a universal formal language, which may be a logical expression, a semantic network, a knowledge representation, etc., and to express this meaning using the lexical units and syntactic constructions of the target language. This "interlingual" approach adds "deep" semantic analyses to the process other than grammatical or syntactic rules, and it is also called knowledge-based approach when a knowledge representation is used as the interlingua (Nirenburg, 1989). Although there have been systems adopting the interlingual approach, the techniques proved to be far more challenging than practical, while practices in some large-scale projects showed that it may encounter uncontrollable complexity when many languages are involved (Nagao, 1989; Patel-Schneider, 1989).

Consequently, traditional research in MT has resulted in a third, perhaps more viable approach, using intermediary levels of abstraction from the source sentence, usually with much of the syntactic information and independent of the target text. The system analyses the source sentence according to the source language syntactic representation and, by use of the contrastive lexical and syntactic differences between the source and target languages, converts this representation into the corresponding to target language representation. Based on this target-language representation, the target sentence is then generated accordingly.

In the processes of analysis, transfer, and generation, regarding issues ranging from morphology, syntax, and semantics, rules are used in all components and are crucially important. These rules are encoded by linguistic experts, therefore the rule-based paradigm provided a mechanism for linguists and computer scientists to collaborate in the development of MT systems (Liu & Zhang, 2015, p. 111). However, development of such systems is time-consuming and labour-intensive, to the extent that the effort involved in crafting the rules often “confine the resulting systems to toy systems” (Kwong, 2015, p. 565) and that it may take years to develop an RBMT that can be put to commercial use (Liu & Zhang, 2015). In addition, human-encoded rules are largely inadequate in terms of the coverage of linguistic phenomena, while there may be conflicts between rules when the system is translating large-scale, real-life texts (ibid).

For such approaches, ambiguity is of crucial importance, and was a heavily discussed topic when rule-based MT was state-of-the-art, especially with respect to the types of ambiguity and how one could disambiguate (see, e.g., Hutchins & Somers, 1992). For syntax, as these systems had to assign a syntactic structure to an input, they necessarily had to deal with competing syntactic analyses. However, with the technological shift towards data-driven systems (see next section), it seems that ambiguity has receded into the background in the context of MT development, as the systems no longer had to parse the sentence based on grammatical rules.

4.5.2 Empirical NLP approaches and Statistical MT

Given the problems of rule-based paradigm, MT has been continuously shifting towards corpus-based methods and strategies since the end of the 1980s, in the context of important progress in NLP based on statistical approaches.

Empirical and statistical approaches — also called stochastic approaches — are data-driven, and based on a statistical analysis of large corpora (i.e., training data) for “estimating the probabilities of various linguistic units or phenomena with respect to particular statistical language models" (Kwong, 2015, p. 565). These approaches in their purest form note statistical regularities in a corpus (Charniak, 1993), evaluating different outcomes probabilistically and selecting the one with the highest value. As Kwong (2015) describes, many algorithms in statistical models are based on machine learning, which attempts to learn patterns of linguistic features that are extracted from the corpora, in either supervised (where the data is annotated) or unsupervised (where the data is not, or is only minimally, annotated) mode. With increasing accessibility of large text corpora, machine learning has now become the dominant approach for many NLP tasks, often supplemented with rule-based post-processing of the results, i.e., hybrid methods (ibid). Although statistical methods do not necessarily model human cognitive processes, they overcome the "severe limitation" of rule-based systems, largely because of the methods' "general coverage regardless of the frequency or rarity of individual linguistic phenomena" (ibid).

The application of empirical and statistical approaches to NLP for the translation of languages, accordingly, has been known as statistical MT (SMT). Whereas the traditional wisdom of MT depicted different levels of intermediary representation between the source and the target texts (see above), in the 1990s, SMT offered a new perspective to view the issue of automatic translation.

Based on an aligned parallel corpus, an SMT system mathematically models the probability of a target sentence being the translation of a given source sentence. Automatic translation is thus achieved by finding the translation which gives the highest probability P(T|S), where T is the target and S is the source. This is called a translation model. In the first SMT system developed by Brown et al. (1990), this is fitted into the noisy-channel model and becomes equivalent to finding the optimal target sentence which gives the highest probability P(S|T)P(T), thereby introducing to the system a so-called language model. With the translation model attempting to ensure that the source and target sentences have the same meaning, the language model helps in producing an output that is relatively fluent. What seems particularly meaningful in this regard is that these two models analogously quantify two of the important translation criteria — faithfulness and fluency, respectively (Kwong, 2015).

This not only pioneered research on SMT, proposing the word-based SMT approach (Brown et al., 1990; Brown et al., 1993), but also provided the theoretical basis to word alignment which is "fundamental to all other SMT methods" (Liu & Zhang, 2015, p. 109). This word-based SMT model was subsequently developed into phrase-based models (Koehn et al., 2003; Marcu & Wong, 2002; Och & Ney, 2002, 2004; Och et al., 1999; Och & Weber, 1998), syntax-based translation models such as inversion transduction grammar (Wu, 1995, 1997), hierarchical models (Chiang, 2005), string-to-tree models (Galley et al., 2006; Galley et al., 2004), and tree-to-string models (Huang et al., 2006; Y. Liu et al., 2006), as well as SMT models involving predicate-argument structures (Zhai et al., 2012).

Despite the "apparent superiority and dominance" of SMT, "traces of transfer models" could still be found in SMT research into the mid-2010s (Kwong, 2015, p.

571). As statistical approaches to NLP were often supplemented by rule-based methods (see above), there was "a clear trend toward hybrid MT", with systems incorporating more linguistic knowledge into their statistical MT models (ibid).

4.5.3 Neural MT

Since the first SMT system developed by Brown et al. (1990), SMT systems have been under development at increasingly large scales both in academia and in industry. The availability of large corpora has not only brought the statistical, data-driven approach to the forefront of NLP and MT research, but also, as Liu and Zhang (2015, p. 109) describe, encouraged a number of studies on parallel computation, neural networks, and connectionism. Particularly worthy of mention is the emergence in the mid-2010s of neural MT (NMT) systems (see, e.g., Kalchbrenner & Blunsom, 2013; Cho, Van Merriënboer, Bahdanau, et al., 2014; Bahdanau et al., 2015; Koehn, 2020; Pérez-Ortiz et al., 2022), together with the "enthusiastic reception" (do Carmo, 2017, p. 88) of this new technology in the translation industry. This indicated a trend towards deep-learning approach which is based purely on neural networks, to the extent that NMT displacing "its corpus-based predecessor" (i.e., SMT) (Forcada, 2017) has become a reality today.

NMT is an end-to-end learning approach for MT, which is said have the potential to overcome many of the weakness of conventional SMT. As demonstrated above, statistical approaches to MT consider translation as equivalent to finding a target sentence T to maximize the conditional probability of T given the source sentence S (i.e., the translation model). In NMT, this translation model is parameterized and the conditional distribution of the target sentence is directly learned from a parallel training corpus using neural networks (Forcada & Ñeco, 1997; Kalchbrenner & Blunsom, 2013; Cho, Van Merriënboer, Bahdanau, et al., 2014; Cho, Van Merriënboer, Gulcehre, et al., 2014; Sutskever, Vinyals, & Le, 2014), consisting of two components — encoder of the source sentence and decoder of the target sentence. Typically (e.g., Cho, Van Merriënboer, Gulcehre, et al., 2014; Sutskever et al., 2014), two recurrent neural networks (RNN) are used to encode the source sentence (whose length is variable) into a fixed-length vector and to decode this vector into a variable-length sentence. Nowadays, transformers are more commonly used for this encoding and decoding processes, since the training times are shorter and the quality is higher for transformers than for RNN (e.g., Pérez-Ortiz et al., 2022). With large-enough training data, powerful computational devices can implement a large number of layers for training, with each layer comprising multiple heads. Once the distribution of the conditional probability P(T|S) is learned, the translation of an input can be then produced accordingly — by calculating argmax_T P(T|S). In addition, there is often an attention mechanism (Bahdanau et al., 2015) in NMT for handling long sequences in the input, and for incorporating contextual aspects in word embeddings (i.e., vectors that represent each word) using the query, key, and value vectors (Vaswani et al., 2017; see also Pérez-Ortiz et al., 2022).

The advantage of NMT, therefore, is its ability to learn directly and in an end-to-end manner from the parallel training corpus, in order to give the optimal output from the translation model.

In fact, neural networks had already been used as a component in SMT systems prior to the advent of direct NMT, e.g., the joint language model for learning phrase representations in Devlin et al. (2014), and the concept of end-to-end learning for MT had been attempted with "limited success" (Wu et al., 2016) in the past, e.g., Chrisman (1991). However, it is perhaps only following such seminal papers as Sutskever et al. (2014) and Bahdanau et al. (2015) that the quality of NMT began to show promising results in comparison with phrase-based SMT (e.g., Luong et al., 2014). Many techniques have also been proposed since then to further improve NMT, bringing considerable progress in the area.

Although NMT is a relatively new technology, online translation systems such as Google Translate already deploy this approach. Translation services are also beginning to be offered using NMT (e.g., Shterionov et al., 2017), and deployment of this technology can even be found in the internal operation of some major international corporations (e.g., Levin et al., 2017).

The technology itself is not without shortcomings, however. Wu et al. (2016) succinctly describe the inherent weaknesses of NMT — slower training and inference speed, ineffective processing of rare words, and occasional failure to translate all words in the source sentence. The processing of "rare words" is particularly important in the context of specialized translation, as terminologies are dense and can be considered rare in view of the general language at large.

Based on the success of the "Zero-Shot Translation" (Johnson et al., 2016) in Google's NMT system, Schuster et al. (2016) raise the question as to whether the system is learning a common representation "in which sentences with the same meaning are represented in similar ways regardless of language" (para. 6), i.e. the so-called "interlingua". To answer this, Schuster et al. (ibid) use a 3-dimensional representation for internal network data for analysis, concluding that the network is encoding a certain amount of the semantics of the sentence "rather than simply memorizing phrase-to-phrase translations" (para. 7). To them, this suggests an existence of an interlingua in the network (ibid), although the actual form of representation within the NMT seems somewhat difficult to observe in an accurate manner.

4.6 Towards an entropy-based perspective

The sections in this chapter have reviewed models of the translation process, largely focusing on a point of debate in the area of TPR regarding horizontal (parallel) vs. vertical (sequential) processing. Here, the vertical processing is described as sequential, which is consistent with much literature in the area, but it is important to note that some also prefer to use "serial" rather than "sequential", adopting a term which is used in the broader perspective of psycholinguistics and cognitive psychology. The difference between these two views of processing bears resemblance to the views of parallel vs. serial processing that have been discussed in Chapter 2 in the context of disambiguation. Indeed, the parallel and serial views reflect the general perspectives in cognitive psychology and psycholinguistics in terms of problem-solving and information-processing strategies, which seems to be a point of debate in itself.

Here, while the aspects of disambiguation (in psycholinguistics) and translation (in TPR) discussed in this thesis in view of serial vs. parallel processing are focused on whether two or multiple processes occur at the same time (regarding the processing of multiple lexical/structural interpretations, or the processes of SL comprehension and TL reformulation), it should be noted that the notion of serial vs. parallel processing might be understood differently if one adopts a different perspective.

For example, serial processing can also be associated with a chain of linguistic units being processed in a sequential manner, or different types of processing — phonological, morphological, lexical, syntactic, semantic and pragmatic — being processed sequentially. Parallel processing, accordingly, would be understood as those units, or types of processes, occurring at the same time independently. From that perspective, perhaps the *type* of processing in, e.g., parallel disambiguation models at the lexical level, would be the same (i.e., lexical), and thus understood as serial even if multiple interpretations are processed at the same time. In the same manner, horizontal translation (where ST comprehension and TL reformulation can occur simultaneously) would be understood as serial rather than parallel, as the type of processing is only lexical. Vertical translation would be construed as parallel due to conceptual and lexical processing occurring at the same time (see Section 4.2.1 for description of horizontal and vertical models).

Here, the understanding of serial vs. parallel processing in this thesis is not focused on whether there are multiple *types* of processing, but whether two or multiple processes themselves occur simultaneously (and thus independently, as one process does not necessarily depend on the outcome of another process). Therefore, when multiple interpretations are processed at the same time, rather than one after another, it is understood as parallel processing; in terms of translation, if SL comprehension and TL reformulation processes occur simultaneously rather than sequentially, the process is understood as parallel processing (see Section 4.2.1 for details).

The discussion in this regard has argued that recent TPR perspectives tend not to consider the two processing strategies as opposing each other, but to recognize their complementarity and to adopt a hybrid view endorsing the co-existence of both horizontal and vertical processes. While concurring with this position, this chapter has also put an emphasis on the effect of cross-linguistic priming which occurs both on the lexical and syntactic levels. In the following investigations, especially in Chapters 6, 8, and 9, this emphasis is also apparent, together with a psycholinguistic view of lexical access and selection, and a recognition of considerable horizontality in the translation process.

In the meantime, the opposition vertical-horizontal processing entails a view of the mental lexicon as conceptual and composed of representations liked to a unitary, stable, self-contained concept. Sub-symbolic models of the mental lexicon, however, adopts a different perspective, in which links are established between concept's meaning traces, rather than on whole concepts. As will be described below, a connectionist perspective (i.e., BIA+ model, see Chapter 6) is also mentioned to explain the process of entropy reduction, and in this respect, the conceptualization of entropy for the mental states seems to be more relevant to a connectionist model of the mental lexicon, although the same perspective can also be compatible with a conceptual model. Here, the model of mental lexicon in the discussion of vertical vs. horizontal processing in this chapter, in view of the aim of the thesis, is not as important as parallelism in this discussion. In other words, the thesis looks at whether SL comprehension and TL reformulation processes occur in parallel, rather than whether the mental lexicon model is conceptual (or the extent to which translation is conceptually mediated).

Another perhaps debatable aspect is the view of translation as an activity involving, among others, information processing, largely consistent with the broad view taken in psycholinguistics of many human cognitive processes and supported by most empirical TPR studies on the basis of behavioural data. While this thesis does not consider itself affiliated to computationalism, it does focus mostly on mental activities rather than the larger cognitive environment in which these activities are situated. As mentioned in Chapter 1, this provides a convenient ground for analysing the process of translation and post-editing through the lens of entropy, and also makes empirical investigation (based on eye-tracking data) more easily operationalizable. In the meantime, as the notions of entropy, uncertainty, and surprisal are consistent with the mathematical formulations in neuroscientific theories that emphasize the situatedness of cognition (e.g., active inference, see Chapters 1 and 3), the discussion in this thesis on mental activities can be a good starting point for further elaboration incorporating more aspects of cognition.

The view of the translation process as entropy reduction will be outlined in Chapters 5 and 6, providing a conceptual account on the basis of the probabilistic, mathematical definition of entropy, and drawing inferences from the formulation of resource-allocation processing difficulty in psycholinguistic studies of sentence processing. As mentioned in Chapter 1, the entropy value which is calculated from the final translation product, and perhaps more problematically, which is approximated in TPR experimental sessions, does not exactly capture the mental states during the process or the transition from one mental state to another. How this product-based metric relates to the process, and how it can represent cognitive load, are important questions if entropy is to be used for our better understanding and conceptualization of the translation process. A perspective that looks at the cognitive processes of translation through the lens of entropy and entropy reduction seems to bear great potential for the frameworks, conceptual understandings, and empirical investigations of translation as a cognitive activity.

4.7 Summary

In summary, this chapter has outlined the models of the translation process in Section 4.2, highlighting horizontal vs. vertical perspectives which resembles the serial vs. parallel views discussed in Chapter 2. Additional issues related to the cognitive activities in translation are also touched upon, including uncertainty and cognitive load. As these concepts are frequently discussed in TPR (and educational psychology) from a slightly different perspective, Section 4.3 reviews these concepts as well.

Since the conceptual and empirical investigations in this thesis involve both translating and post-editing, Sections 4.4 and 4.5 has reviewed post-editing as well, including the status of MT as an important tool in modern translation profession, research on post-editing effort, and MT approaches in NLP.

The chapter ends with a preliminary discussion on the entropy-based perspective of the mental processes in translation and post-editing, which will be elaborated in the following chapters.

Chapter 5 Ambiguity: statistical approaches

5.1 Introduction

This chapter analyses the statistical means used in psycholinguistics in search of a method to quantify ST features in relation to ambiguity, which is subsequently built upon in the conceptual explorations in Chapter 6. This largely helps to provide theoretical justifications of the entropy-based perspective of mental states in translation and post-editing.

Starting from pre-Chomskyan theories and findings regarding the frequency and meaning of words, the chapter proceeds to relatively more sophisticated analyses of entropy and surprisal, together with the theoretical justifications for these metrics of measurement. In doing so the chapter outlines several methods of quantification which are worthy of investigating with regards to the research question in this thesis. As will be shown below, metrics to be studied for lexical issues include the square root of ST word frequency, the surprisal of this frequency marginalizing over all possible alternatives (i.e., the corresponding entropy), and the word translation entropy which relates the ST to the TT. Those for structural issues include the syntactic choice entropy and the surprisal framework.

These metrics are not without theoretical justifications and empirical evidence, from the perspectives of comprehension, communication, and translation. Some of the research in the literature which is described in Chapters 2 to 4 has already touched upon, one way or another, the influence of these statistics on reading, translating, and interpreting, but this chapter provides a systematic analysis on the statistical approaches.

5.2 Statistical approaches to psycholinguistics

The statistical means to study language is without doubt an indispensable, if not crucial, part of modern psycholinguistics in many ways. However, studies on such statistics as word frequency as well as the compilation of word lists based on their frequency of occurrence are by no means the outcome of modern-day technology only. As Levelt (2014) rightly asserts in his historical account of the pre-Chomskyan era of psycholinguistics, the statistical tools which are now taken for granted in modern language studies, including not only extensive lexical and textual databases, but also phonological, morphological and lexical frequency data for various languages, were largely the result of pioneering work which traces all the way back to the late nineteenth century (p. 449).

Initial studies of such kinds were conducted in an explicitly practical context, with stenographers being perhaps the first to adopt this statistical approach, in their investigations of the frequency of words in various languages such as German (e.g., Kaeding, 1898), French (e.g., Estoup, 1902), and English (e.g., Dewey, 1923), to name just a few. In addition, language education in schools at the time was another typical context for collecting data on the frequency of words (Levelt, 2014), with numerous published compilations of word lists based on occurrence counts (e.g., Eldridge, 1911; Thorndike, 1921, 1932; Thorndike & Lorge, 1944). As a matter of fact, using word frequency data for educational purposes still continues to be a popular area of research in today's quantitative applied linguistics, where frequency data and other statistical features of languages are collected, studied, and applied to practical approaches to pedagogy in various domains.

5.2.1 Rank-frequency distribution and Zipf's law

Although these early, pioneering efforts in the collection and analysis of frequency data have been important, profound, and crucially foundational, two truly influential works in this regard, standing upon the shoulders of the above pioneers, are particularly meaningful for the quantitative studies of language, and more importantly for the present thesis — the rank-frequency distribution discovered by Condon (1928), and the number-of-words-frequency distribution (i.e., Zipf's law) discovered by Zipf (1935, 1949).

On the basis of published word frequency data from Ayres (1915) and Dewey (1923), Condon (1928) discovers what he calls "a rather interesting functional relationship" between the words' frequency of occurrence and their corresponding ranks in the frequency list (see Figure 5.1 below). According to Condon, if the words in "a large representative sample of written English" are arranged in order of decreasing frequency, the n^{th} word in the list "will then occur with an observed frequency which is a function of *n*", i.e., *f(n)*, and this frequency can be given by the following formula:

$$f(n) = \frac{k}{n}$$

where *k* is a constant and can be readily determined on the basis of the number of different words in the sample (i.e., *m* in the following equation):

$$k \sum_{n=1}^{m} \frac{1}{n} = 1$$

Here, what is particularly interesting is that the first equation above indicates that the product of rank and frequency of a word is constant, as the equation can be alternatively written as $nf(n) = k$. What it also indicates in the meantime is that the frequency of a word, *f(n)*, is inversely proportional to the word's rank in the frequency list, *n*.

Although this relationship between word rank and frequency was only briefly described by Condon's (1928) short article (perhaps excessively so), with very limited explanation of the principles underlying this phenomenon,[26] the rank-frequency

[26] Condon mentions explicitly that the explanation of this relationship is "admittedly incomplete" because the subject is one which he "can not pursue" (p. 300). However, the article does explain at the end that this phenomenon is "perhaps a quantitative appearance in language of the Weber-Frechner law of psychology". It adds that "the marginal increase in idea-transmitting power which can be accomplished by the addition of another word to the vocabulary is smaller the greater the value of n, according to the same law which governs the relation between the psychological increase in sensation accompanying an increase in the total intensity of the physical stimulus".

distribution has nevertheless been a useful inspiration for numerous investigations in the study of language and beyond.[27]

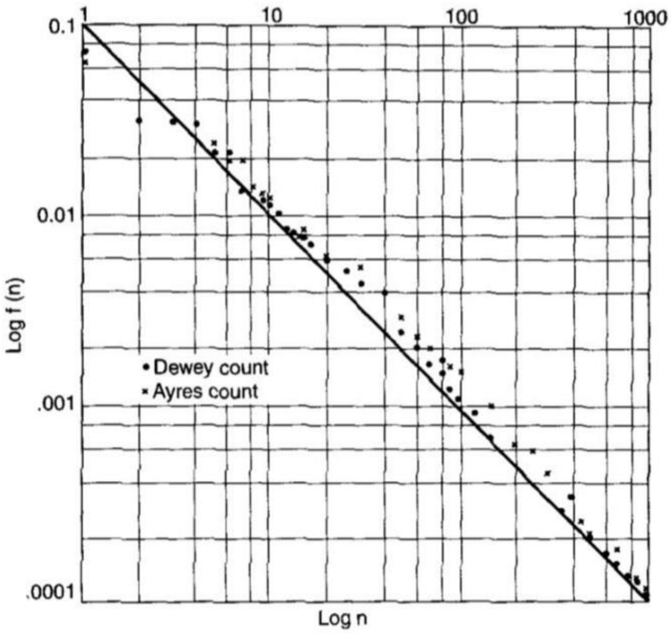


Figure 5.1 Condon's plot of the rank-frequency distribution — the logarithm of the observed frequency of the n^{th} word against the logarithm of n, based on Dewey's and Ayres' counts

Not surprisingly, the same rank-frequency distribution pattern is also found by Zipf (1935) using English and Latin texts as data, where he introduces the notion of *lexical wavelength* — the average length of interval (measured by the number of words) between a word and its reoccurrence in the text, calculated as the length of the text divided by the word's frequency. Zipf discovers, in addition to the rank-frequency relationship, that a word's wavelength equals ten times its rank, a phenomenon which he calls the *harmonic series* relation.

Regarding the systematic frequency distribution of words which had been coincidentally discovered and incompletely explained by Condon, Zipf's description of the same phenomenon is as follows:

[27] This phenomenon discovered by Condon is in fact not confined to language studies. For example, the same distribution has also been found in city sizes in relation to population (Auerbach, 1913; Zipf, 1941). Similar analysis based on the work of Condon and Zipf can also be found in numerous studies of DNA sequences in biology (Mantegna et al., 1994), the size distribution of firms in business (Stanley et al., 1995), etc.

The r^{th} most frequent word occurs with a frequency *f(r)* which scales according to $f(r) \propto \frac{1}{r^{\alpha}}$ for $\alpha \approx 1$ (Zipf, 1935, 1949).

Here, *r* is referred to as *frequency rank* of a word, while *f(r)* is its frequency in a natural corpus. This rather simple mathematical expression has been known as *Zipf's law*, a power law which human natural language approximately obeys in terms of frequency distribution.

In a similar manner to Condon's study, Zipf (1935) takes a step further and uses the word frequency data collected by Eldridge (1911) to draw another, slightly different curve in terms of the logarithm of word frequency, here not against the rank of the word, but against the logarithm of the number of different words with the corresponding frequency (see Figure 5.2 below). An important observation from this is that only a small number of words are highly frequent and that there is a very large number of words with very low frequency — even occurring once only.

In addition, another finding which is similar to Condon's discovery is particularly interesting — the linear relationship in the double-logarithmic plot.

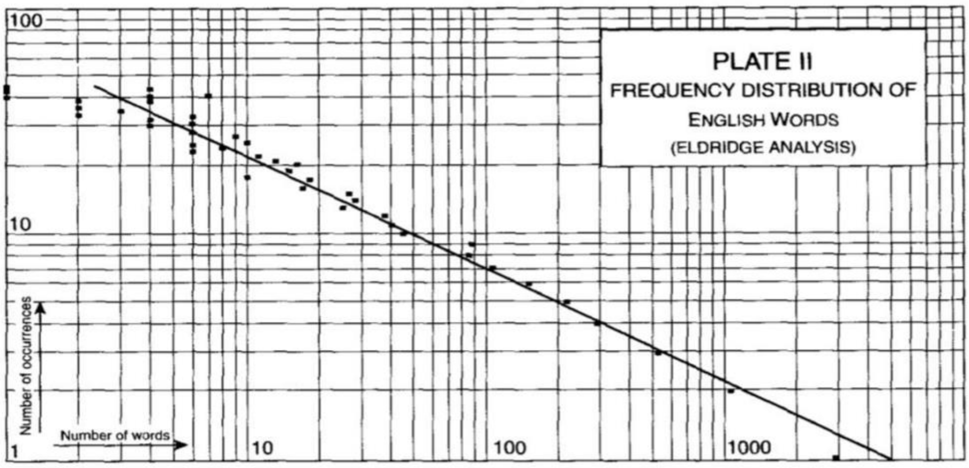


Figure 5.2 Zipf's plot of the number-of-words-frequency distribution — the logarithm of frequency against the logarithm of the number of words sharing that frequency

In this plot, one can see a linear pattern similar to the one in the frequency-rank plot (Figure 5.1), but with a different slope. If the frequency of occurrence is denoted by *f*, and the number of words sharing this frequency is *n'*, then the function describing Zipf's plot would be: $n' * f^2 = k$. In other words (to quote his own description), The product of the number of words of given occurrence, when multiplied by the square of their occurrences, remains constant for the great majority of the different words of the vocabulary in use, though not for those of highest frequency" (Zipf, 1935, pp. 41-42).

What this also means is that the number of words of a particular frequency is inversely proportional to the square of this frequency. This can be viewed as a mathematical expression of the observation that there are a very small number of words occurring highly frequently, and very large numbers of words with rather low frequency of occurrence.

5.2.2 Principle of least effort and lexical ambiguity

The regularities mentioned above have been interesting and meaningful, not least because of the peculiarity of Zipf's law. The fact that words which vary in frequency would follow such a precise mathematical rule, a rule which does not make reference to "any aspect of each word's meaning" (Piantadosi, 2014, p. 1113), is particularly puzzling, given that the actual use of language for communication is closely associated with meaning-making where utterances "obey much more complex systems of syntactic, lexical and semantic regularity" (ibid, p. 1113). Questions as to why the intricate processes of language production would conform to a perhaps "unreasonably" simple (Wigner, 1990) mathematical frequency distribution, not surprisingly, necessitate theoretical explanations for this law.

According to Zipf's (1949) own explanation, these regularities in fact go beyond purely observational phenomena or stochastic effects. They are more profoundly the result of the property of the human mind, particularly regarding what is often called the *principle of least effort* (c.f. Schultze, 1880). In other words, these observed statistical phenomena seem to be not merely a stochastic necessity, but rather a law which expresses a deep principle of mental economy.

From the perspective of the principle of least effort, Zipf (1949) explains two aspects of latent economy when communication happens — 1). the speaker's economy, as a *force of unification* which aims at minimizing the effort to select words to convey meaning, thereby pushing to decrease the number of different words and resulting in a small range of high-frequency words; and 2). the hearer's economy, as a *force of* *diversification* when detecting the intended meaning with minimal effort, pushing to maximize the number of different words to be associated with every meaning. These two economies are obviously in conflict, which eventually results in a compromise between them, where the joint expenditure of effort from both sides of communication is minimized.

As a consequence, the compromise between the forces of unification and diversification leads to the characteristics shown in the above plot (see Figure 5.2) — a very small number of very high-frequency words which are associated with multiple meanings, and a very large number of low-frequency words with very few meanings associated with each of them.

This is clearly an important aspect of lexical ambiguity. High-frequency words tend to have more meanings — and are thus more ambiguous — than low-frequency words, which benefits the speaker's interest. On the other hand, the low-frequency, rather unambiguous words come in very large numbers, benefiting the hearer's interest. The principle of least effort on both sides of communication creates forces of two conflicting directions, which then compromise each other and eventually result in a pattern in relation to word frequency, the number of words corresponding to a particular frequency, and the degree of lexical ambiguity associated with these words.

More importantly, based on the count of word meanings (e.g., Thorndike, 1941), this number of meanings can be plotted in the same manner as above, against the ranks of each word, where a linear relationship has also been found with a slope of -0.5 (Zipf, 1935), indicating a proportionality between the square root of a word's frequency and the number of meanings it has — in other words, the degree of lexical ambiguity of a word seems to be proportional to the square root of its frequency of occurrence.

This proportional relationship between the number of different meanings of a word and the square root of its frequency is the result of an equilibrium of the two forces mentioned above. Meanwhile, this relationship is perhaps a crucial aspect of the lexical ambiguity in this research, particularly regarding the statistical means to measure ambiguity and in turn to predict cognitive effort.

5.2.3 Other theoretical explanations

The psychological explanation of the law in Zipf's analysis, however, is not without criticism and has met with scepticism. Thorndike (1937), for example, explicitly states that the phenomenon is unlikely to be "evidence of any uniform and ubiquitous tendency toward a certain equilibrium between frequency and variety", but rather, "in some measure a statistical effect". Meanwhile, Levelt's (2014, pp. 454-455) description of Zipf's explanation is not without negative comments either, stating that while his psychology "may be correct", it "needs a different kind of empirical support".

In fact, for the past seven decades the question mentioned above as to the peculiarity of Zipf's law has been a "central concern" in statistical theories of language (e.g., Piantadosi, 2014), with subsequent mathematical formalization, generalization and modification to the rank-frequency relation (e.g., Mandelbrot, 1953, 1962), as well as different explanations for the distribution with many formal frameworks and various sets of assumptions. Among the wide range of theoretical explanations for this phenomenon — in studies of language as well as in many other scientific domains in which the same law can be observed, particularly useful ones include optimization of entropy (Mandelbrot, 1953) or Fisher information (Hernando et al., 2009), mixtures of exponential distributions (Farmer & Geanakoplos, 2008), semantic organization (Guiraud, 1968; Manin, 2008), preferential reuse of certain forms (Simon, 1955; Yule, 2014), and communicative optimization similar to Zipf's (1935, 1949) own explanation mentioned above (Mandelbrot, 1962; Ferrer i Cancho & Solé, 2003; Ferrer i Cancho, 2005a, 2005b; Salge et al., 2015). While these frameworks are not completely consistent with one another in terms of theoretical explanations and standpoints, they seem to be in no position to contradict the phenomena discovered by Zipf.

5.3 Measuring ambiguity for processing effort

Although the theoretical explanations for the mathematical relationship regarding word frequency have been diverse and sometimes controversial, the phenomenon is quite straightforward. More importantly, its relevance to the cognitive effort of translation and post-editing with regards to lexical ambiguity is also a crucial aspect of the statistical measures involved in the current research. If the degree of ambiguity of a word is somewhat related to its frequency of occurrence in a corpus, this might be an indication of a possible means to quantify lexical ambiguity and thus to predict the additional post-editing effort arising from this ambiguity.

In other words, if the square root of frequency for a word is proportional to its degree of ambiguity, and if the word's ambiguity is in some way positively related to the cognitive effort required to process it, one might expect a positive relation between the square root of frequency and the statistical quantification of effort in post-editing.

On the other hand, this seems to indicate at the same time that more frequent words might correspond to a larger amount of effort required, which might be in some ways contradictory to the findings of many investigations into eye movements in reading. As will be discussed in Chapter 7 (see Section 7.6.1), experimental studies on the influences of word frequency on reading have shown that more frequent words are processed faster — words with high frequency receive shorter fixation durations.

Regarding this contradiction, both sides seem reasonable, each focusing on a different aspect of processing. The former assumption which positively relates frequency with effort (if Zipf's explanation stands) can be described as emphasizing the number of meanings for each word, i.e., lexical disambiguation, while the latter has to do with the expectedness (or in an equal sense, unexpectedness) of a particular word appearing and being encountered at a certain point of the sentence during online processing. Rather than looking at this from a black-and-white perspective, perhaps a better standpoint would be to recognize the role that both play in the cognitive processes.

Despite the theoretical analyses above, the metric of word frequency (or the square root of frequency) alone seems perhaps somewhat oversimplified in the context of the actual processes of language comprehension, translation, and MT post-editing. Another set of measurements which are more sophisticated and focus on entropy and surprisal, have already been introduced in Chapter 3. Such entropy-based measurements are beyond simply counting the *number* of options,[28] but largely represents some more sophisticated aspects, including uncertainty and probability

[28] The 'frequency of words' discussed in the above sections in this chapter are, in essence, a simple count of the number of occurrence.

distributions (see Chapter 3). They have their roots in the mathematical theory of information, a theory which has been influential for — among a variety of areas — translation studies, where translation, especially in the "equivalence" paradigm,[29] has often been considered a process of transcoding the ST content into the target form (e.g., Nida & Taber, 1969). The relevant mathematical expressions, however, lying at the core of Shannon's theory of communication, seem much less discussed in translation studies.

The definitions, implications, and mathematical expressions of entropy and surprisal have been introduced in Chapter 3 from an information theory perspective, but entropy and surprisal have in fact been used in psycholinguistics as a measurement of the cognitive effort that is required to process a particular point of the text during (monolingual) reading. Such a psycholinguistic perspective of language processing will be explained in the following sections.

5.3.1 Surprisal, relative entropy, and cognitive effort

As mentioned in Chapter 3, Shannon information content (i.e., surprisal) indicates the uncertainty involved in a random variable taking a particular value based on its probability of occurrence (see Section 3.4). This is the implication of surprisal from mathematical, probabilistic, and information-theoretic perspectives. From another perspective, the use of surprisal in psycholinguistics to investigate the cognitive processing of language is by no means infrequent.

For example, on the basis of Attneave's (1959) application of Information Theory to psychology, Hale (2001) suggests that the surprisal of a word in its context can be used to quantify the cognitive effort which is required to process the word in the sentence. In his view, incremental sentence comprehension is a step-by-step disconfirmation of possible phrase-structural analyses for the sentence, and cognitive load can thus be interpreted as the combined difficulty of "disconfirming all disconfirmable structures at a given word" (ibid, p. 4). It would be easy to assume that more probable words require more effort to disconfirm — i.e., the processing difficulty

[29] In translation studies, the "equivalence" paradigm refers to a set of (somewhat marginalized) theories that assume translations can and should be "equivalent" to their original, although the definition of equivalence may vary. These theories consider language, communication, and translation as static, where the meaning can be transferred from the ST to the TT without being changed.

of a word would be proportional to its surprisal in the corresponding context. In this regard, surprisal can be a reasonable metric of word-by-word cognitive load.

This quantification of cognitive effort, more importantly, also raises "a unified treatment of structural ambiguity resolution and prediction-derived processing benefits" (Levy, 2013). Both Hale (2001) and Levy (2008) illustrate much successful use of the surprisal framework to adequately analyse a variety of psycholinguistic phenomena, many of which are closely relevant to this thesis including the disambiguation of garden-path sentences (i.e., temporary ambiguity). Surprisal, especially when marginalizing over all the possible partial structural descriptions, seems to account perfectly for empirically observed syntactic-processing effects which involve disambiguation. In the following Section 5.3.2, specific ways of using surprisal for quantitative analysis of structural ambiguity, especially regarding garden-path sentences, will be described.

In addition to the above-mentioned rationale for using surprisal as a metric for cognitive processing difficulty, theoretical justifications in this regard have not been lacking, especially within the frameworks of rational cognition (e.g., Shepard, 1987; Tenenbaum & Griffiths, 2001) where expectation-based models[30] making use of surprisal to quantify the on-line processing difficulty can be explained as "the consequence of optimization in comprehension" (Levy, 2013, p. 179). As demonstrated in Chapters 2 and 3 (see above), difficulty, or measurable disruption, in real-time sentence processing can arise either from an overload in memory — the cognitive resources for the storage and retrieval of the representational units which are used to analyse the linguistic input, or from a sufficiently unexpected input which causes a shift in resource allocation "to various alternatives in the face of uncertainty" (Levy, 2013, p. 144). Although theories based on the former (i.e., resource-limitation theories) have been a dominant paradigm for studies of differential processing difficulty, the latter (i.e., resource-allocation approach) has been a line of investigation which largely has ambiguity resolution as a primary concern (Levy, 2008).

In the latter approach, the size of the shift in resource allocation which is induced by a word *w* is indicative of the difficulty of processing this word, and is equivalent to the *change* in the conditional probability distribution over all

[30] This is in contrast to the memory-based models. In Levy (2013), two broad approaches to understanding real-time language comprehension are outlined — memory-based and expectation-based, and the use of surprisal falls into the expectation-based approach.

interpretations (Levy, 2013) which in mathematics would be measured in terms of entropy (e.g., Cover & Thomas, 1991) — specifically, the *relative entropy* between the conditional distributions before and after encountering the word *w*. This is consistent with the idea of using word translation entropy and syntactic choice entropy to measure the difficulty of a translation choice in the face of uncertainty (see below), where the difficulty is measured by the conditional probability distribution over TT alternatives. In terms of monolingual comprehension of the ST, the TT, and the MT output (in translation and/or post-editing), the use of this measurement can also be largely useful for the discussion on cognitive processing effort.

Of particular note here is that in sentence comprehension, the relative entropy mentioned above has been shown to be equivalent to the surprisal of *w* (Levy, 2008, pp. 1131-1132), which Levy views as the reranking cost in incremental disambiguation where cognitive resources are allocated to the possible analyses of the sentence: "Empirically observed processing difficulty after local ambiguity resolution is informally ascribed to either a *reranking* of the favoured analysis, or *competition* between closely-ranked analyses."

In short, the close relevance of the mathematical definition of information to such concepts as uncertainty and surprisal is of profound importance in the investigation of human language processing, particularly regarding translation, post-editing, and the processing of ambiguity. The uncertainty involved in translation where a chain of decision-making activities take place in view of a number of translation alternatives can be modelled using entropy, e.g., word translation entropy and syntactic choice entropy. Surprisal (or surprise-value) can be a reasonable measurement of cognitive load in comprehension and particularly in disambiguation — based on, among others, Hale's intuition that processing difficulty is the work related to disconfirming possible interpretations or analyses, as well as Levy's closely related proposition that the difficulty refers to the cost of the reranking of possible interpretations or analyses for reallocating cognitive resources.

5.3.2 Surprisal and disambiguation

As mentioned above, surprisal can be a reasonable measurement for disambiguation and for cognitive effort in language processing. Levy (2013) holds that surprisal can "simultaneously account for both empirically observed syntactic-processing effects which involve disambiguation and effects which do not" (p. 159), and gives very explicit illustration of combining surprisal with probabilistic grammatical analysis using the garden-path sentence which is mentioned in Section 2.1.3 of Chapter 2 (i.e., sentence 6):

When the dog scratched the vet and his new assistant removed the muzzle.

In this sentence, a clear garden-path ambiguity is involved regarding *the vet and his new assistant*, as to whether it is the object of *scratched* or the subject of *removed*. An increase of cognitive processing effort occurs when the processor encounters the second verb — *removed*. To calculate the surprisal of the word *removed*, the probabilistic context-free grammar in Figure 5.3 can be used together with the incremental analyses (Figure 5.4) of the sentence before the ambiguity, to obtain the probability of the tree in which *the vet and his new assistant* is the object of *scratched* (0.826), the probability of the tree in the other analysis (0.174), as well as the conditional probability of *removed* to be the next word respectively corresponding to these two analyses (0 and 0.2). The probability of *removed* would be the weighted average, or computed via marginalizing over the possible structural analyses of the sentence up to this point:

$$P(w_{11}=\textit{removed}|w_{1...10}) = 0.2*0.174 + 0*0.826 = 0.0348$$

This corresponds to a surprisal of $\log_2 0.0348 = 4.85$ bits.

Here, a comparison can be made with the following sentence to describe how the value of surprisal corresponds to garden-path effect and processing difficulty:

When the dog scratched its owner the vet and his new assistant removed the muzzle.

The surprisal for this sentence, which is unambiguous regarding the instance above, would be $\log_2 0.2=2.32$ bits. Since this sentence does not lead to a garden-path effect, the processing difficulty arising from the temporary ambiguity in the previous sentence is absent, and the surprisal is accordingly lower, reflecting this difficulty in terms of mathematical value.

Rule		Prob.	Rule		Prob.	Rule		Prob.
S	→ SBAR S	0.3	Conj	→ and	1	Adj	→ new	1
S	→ NP VP	0.7	Det	→ the	0.8	VP	→ V NP	0.5
SBAR	→ COMPL S	0.3	Det	→ its	0.1	VP	→ V	0.5
SBAR	→ COMPL S COMMA	0.7	Det	→ his	0.1	V	→ scratched	0.25
COMPL	→ When	1	N	→ dog	0.2	V	→ removed	0.25
NP	→ Det N	0.6	N	→ vet	0.2	V	→ arrived	0.5
NP	→ Det Adj N	0.2	N	→ assistant	0.2	COMMA	→ ,	1
NP	→ NP Conj NP	0.2	N	→ muzzle	0.2			
			N	→ owner	0.2			

Figure 5.3 Probabilistic context-free grammar

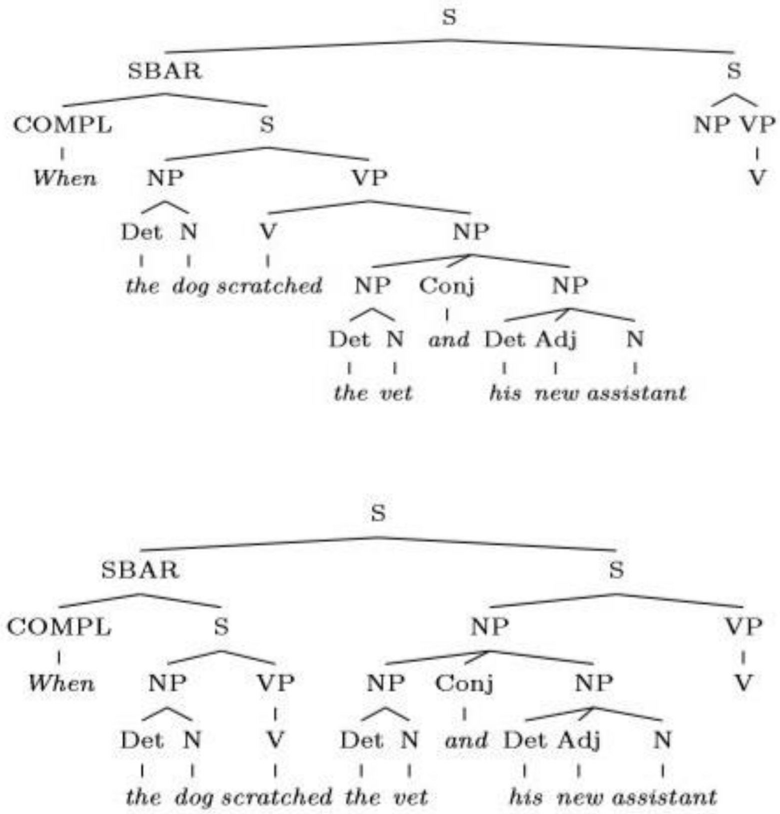


Figure 5.4 Incremental analyses for sentence 6

5.4 Entropy in the translation process

Section 5.3 has explained the theoretical underpinnings, and the practical use, of surprisal (and relative entropy) in psycholinguistics as a quantification of cognitive load in reading, drawing on Levy's formulations of resource-allocation processing difficulty. This provides the basis for the discussion in chapter 6, where the concept of entropy as defined in the mathematical expressions is analysed in more detail in the context of the translation process, as well as from a psycholinguistic perspective of language processing, to explore how entropy can conceptually describe mental states and to provide theoretical justifications for entropy reduction as a representation of cognitive load.

So far, the thesis has explained the concept of entropy, and its related concepts, in view of both information theory (Chapter 3) and psycholinguistics (Chapter 5). In such explanation, the probabilistic nature of the concept is emphasized, and analyses are firmly grounded in the relevant mathematical expressions.

Regarding information theory, although its definition of information does not seem to accurately or thoroughly reflect meaning-making (or "information") in human communication, the mathematical expression of entropy as used in that theory shows several important aspects of language processing through a probabilistic and statistical lens (e.g., frequency of occurrence, uncertainty, unpredictability, and freedom of choice in selection among an ensemble of candidates, as mentioned in Chapter 3). While the theory itself adopts a somewhat static view of language and communication, the probabilistic perspective and statistical approach to the psychological reality of language (as frequently observed in empirical investigations) seem to bear important implications for the study of ambiguity and disambiguation in the translation and post-editing processes.[31]

To reiterate, information theory defines the information (i.e., surprisal, or Shannon information content) of an item as the negative logarithm of its probability (see Section 3.4.1), which reflects the intuition that unpredictable items should be more informative than predictable items. Specifically, according to this theory, the amount

[31] One might realize that in this thesis, the probabilistic and statistic approaches to psycholinguistics is a recurring theme, and is also the underlying perspective of the conceptualization based on entropy.

of information for a message which is generated as a discrete stochastic variable is represented by the weighted sum of the surprisal for each individual value that the variable can take. This weighted sum is mathematically defined as entropy (see equation 1 in Section 3.2), and the entropy would be maximum if all probabilities are equal.

Although this definition is initially intended as a technical means for indicating the number of informational units needed for message encoding, when it is used to refer to the broader sense of communication such a definition is considered the "selectional information-content" (MacKay, 1969), for which the main point of interest is the "*relative improbability* of a message given an ensemble of messages" (Kockelman, 2013, p. 116, my emphasis), as explained in Chapter 3. In addition to the number of informational units in encoding, entropy also measures the *uncertainty* involved in estimating the value which a variable can take, i.e., the freedom of choice when a message is chosen (see Section 5.3.3).

For the process of translation — a task frequently described as essentially "a chain of decision-making activities" (Angelone, 2010, p. 17, as mentioned in Section 3.4; see also Hervey et al., 1995; Levý, 1967; Tirkkonen-Condit, 1993) — the *relative improbability* of a translation option for a particular ST item, given an ensemble of translation options for the same item, can perhaps be defined in a similar manner. In this regard, Section 3.5 has explained an entropy-based metric for word translations, namely, HTra, as a description of the word translation choices at a given point of the ST, and syntactic choice entropy (HCross), as a representation of "local cross-lingual distortion" (Carl et al., 2016). Word translation entropy represents the "degree of uncertainty" involved in the choice of lexical TT items "given the sample of alternative translations for a single ST word" (Schaeffer, Dragsted, et al., 2016, p. 191), and this entropy is also indicative of the probability distribution over all the alternative choices in the sample. With probabilities equally distributed over a large number of choices, the entropy would be high, and the outcome of lexical selection would involve considerable uncertainty. On the other hand, probabilities being concentrated on very few items would lead to a low entropy and a high level of certainty (see Section 3.5 for details).

These points, as well as the psycholinguistic formulations of processing difficulty mentioned in the previous section, will be an important for the next chapter.

5.5 Summary

This section has outlined several statistical means to study ambiguity which are particularly interesting and worthy of investigation in the context of the current research. The simple metric of word frequency (or the square root of this frequency) in the ST seems to be related to both lexical ambiguity and expectedness, and both these two aspects can be assumed to play a part in the cognitive processing of the ST in translation and post-editing.

Given that relying solely on word frequency data would perhaps oversimplify predictions of processing difficulty, the other metrics outlined above are also analysed.

As probability indicates the expectedness of an event, the entropy value (or the surprisal marginalizing over all possible events) would represent the frequency distribution of all the possible events, i.e., the variance of these alternatives. Regarding the word frequency here in sentence comprehension, a calculation of the entropy in view of all the possible words at a certain point of the sentence would provide more information as to the uncertainty and unpredictability involved in the word, and thus reflect the cognitive effort accordingly. Theoretically, the frequency data models the probability with which a word would be predicted correctly by a participant, while the entropy indicates *whether* the participant would make a prediction at this point, based on how certain he/she is in this regard. It seems the value of entropy can facilitate deeper understanding of the cognitive process. It is worthy of mention that the calculation here would not involve syntax but is a transformation of the individual word frequency data into entropy over all possible words in the corpus to appear in the same position, as well as all possible meanings in the corpus for the same word.

These analyses are focused on reading and comprehension at the lexical level, which is an indispensable part of the translation/post-editing process regarding both the ST and the MT output.

More specific to the cognitive processing of translating and post-editing, word translation entropy can be a further indication of the uncertainty, surprisal, unpredictability, variance, and difficulty of lexical translation in view of all the possible TT options for a particular ST word. The correlation of this metric with cognitive processing in translation and post-editing can also be a valuable contribution to the search of means to predict and measure effort, as well as to the conceptual understanding of the mental activities themselves.

In terms of syntax, the syntactic choice entropy describes the extent to which different word orders are possible when translating a given ST word, and in a way models the uncertainty involved in determining TT syntactic structures in translating and post-editing. Correlation of this metric to effort is also useful in the search for statistical means to predict effort, and it is interesting to see how this reflects the structural ambiguities involved in the ST.

While syntactic choice entropy touches upon syntax and might be related to ambiguity, a statistical method which directly models structural ambiguity resolution is the surprisal framework. Calculation of the values of surprisal regarding structural ambiguity, especially when garden-path sentences are involved, can perhaps also be hypothesized as a predictor of effort in disambiguation.

In the following chapters, both conceptual and empirical aspects are systematically examined, building upon the basis formed by the present and the previous chapters, and addressing the research questions posed in Section 1.2.

Chapter 6 Ambiguity, entropy, and cognitive processes

6.1 Introduction

The previous chapter has given a detailed description, explanation, and theoretical justification of some statistical measures, including entropy and surprisal, that are conceptually formulated or empirically observed in psycholinguistics to quantify the ambiguity and cognitive load in monolingual processing. This chapter further explores the conceptual basis on which entropy (and the reduction of entropy) can describe the cognitive activities in translating and post-editing, and in the meantime, represent the cognitive load therein. This brings the concept of entropy into the mental processes of translating, considering it a (probabilistic) description of mental states during cognitive activities, rather than merely a statistical measurement of the number of translation equivalents in the translation product.

The issues that are addressed in this chapter include, among others, how entropy can (statistically) describe mental states and activities, how entropy can be a theoretically justifiable means of quantification of cognitive load in relation to translation ambiguity, how HTra used in today's TPR studies relates to these conceptualizations, and how entropy that describes mental states differs from entropy that can be observed in the text via statistical means, or the entropy that can be approximated from experimental sessions. Several other important concepts are also disentangled, such as entropy reduction vs. relative entropy as a quantification of cognitive effort in translation and post-editing, entropy from the serial and parallel processing perspectives of disambiguation, etc. On the basis of the discussions in the previous chapter (i.e., Chapter 5), this chapter intends to answer RQ 1 (see Chapter 1), i.e., how entropy, other than as a statistical measure of the text, conceptually describe the mental activities in translation/post-editing when it comes to ambiguity.

6.2 Entropy and cognitive processes of translation

As mentioned in Chapter 5, the thesis has so far discussed entropy (and surprisal) in view of both information theory and psycholinguistics, where the discussion in both places puts an emphasis on the mathematical, probabilistic properties of entropy (and surprisal). While the notion of entropy in empirical TPR studies refers to a feature of the text, the discussion in Chapter 5 starts to connect this feature with the cognitive processes by introducing the surprisal framework in psycholinguistic studies of monolingual reading. In this section, the concept of entropy as defined in the mathematical expressions is analysed in more detail in the context of the translation process, as well as from a psycholinguistic perspective of language processing, to explore how entropy can conceptually describe mental states and to provide theoretical justifications for entropy reduction as a representation of cognitive load.

In empirical studies, HTra has been used as a predictor variable for cognitive effort (see also, e.g., Lacruz et al., 2021), and the value of word translation entropy seems to show a statistically significant impact on many aspects of translation behaviour, e.g., word production duration, first fixation duration, the probability of a fixation, and total reading time (Carl & Schaeffer, 2017c; Schaeffer, Dragsted, et al., 2016). Such observations are meaningful, but theoretical justifications of the connection between entropy and cognitive load is lacking.[32] The following sections, on the basis of the discussion provided in Chapters 3-5, will address this issue.

6.2.1 Entropy, mental states, and cognitive load

Given its definition, and results from empirical studies, it is easy to hypothesize that some aspects of this entropy-based metric could be an indicator of cognitive load in translating. Bangalore et al. (2016), for example, propose using entropy as a measure of the translator's cognitive effort which is expended in making choices during translation. While their argument is that entropy captures the *weight* of each alternative to describe probability distribution (without detailed discussion of the mental

[32] Importantly, note that the entropy here refers to the general, abstract, mathematical, and probabilistic concept, rather than the metric of HTra used in TPR studies (i.e., the metric used in the CRITT TPR-DB).

processes), the discussion in the present chapter is from a slightly different perspective, focusing on the dynamic *change* of probability distribution during the translator's decision-making.

And rather than considering entropy as a mere statistical means, or an alternative way of counting the possible translation options available to the translator (see, e.g., Bangalore et al., 2016), the discussion here brings the concept into the assumed mental states at a micro level, using entropy to describe the specific cognitive activities involved when mental states transition from one to another during the lexical activation and selection processes in translation (or post-editing).

Entropy in psycholinguistics

In Chapter 5, Section 5.3 has already described some typical (and influential) psycholinguistic theoretical frameworks that are based on entropy (or surprisal). For example, the surprisal (i.e., Shannon information content, or self-information) of a word in its context has been argued to be a useful quantification of the cognitive effort required to process this word during online sentence processing (e.g., Hale, 2001). This is because, from that view, incremental sentence comprehension is a step-by-step disconfirmation of possible phrase-structural analyses for the sentence, which means that cognitive load can be interpreted as the combined difficulty of disconfirming the disconfirmable structures at a particular point of the sentence.

In expectation-based models of real-time sentence comprehension (in terms of parsing), processing difficulty (or measurable disruption) can arise from a sufficiently unexpected input which causes a shift in resource allocation "to various alternatives in the face of uncertainty" (Levy, 2013, p. 144). The size of this shift in resource allocation, in, e.g., Levy's (2008) analysis, can be measured by the *change* (or update, to use Levy's word) of the probability distribution over possible interpretations after the current word is processed (see Section 5.3 for details). In mathematical terms, this difference in probability distribution would be expressed by the *relative entropy* (also known as Kullback-Leibler divergence, see Kullback, 1959) of the updated distribution with respect to the old distribution.

Entropy in the process of translating

Regarding the process of translating, perhaps a similar perspective can be adopted for the translation of each ST item. Upon encounter of a particular ST item, possible translations for this item are likely to be subliminally co-activated (Balling, 2014; Bangalore et al., 2016; Carl et al., 2019; Grosjean, 1997; Macizo & Bajo, 2006; Ruiz et al., 2008; Schaeffer, Dragsted, et al., 2016; Schwartz & Kroll, 2006; Yan Jing Wu & Thierry, 2012),[33] and a higher entropy value for the activation associated with this (ST) item indicates a higher level of uncertainty in choosing from these translation alternatives. This increased level of uncertainty may be the result of a larger number of translation alternatives which are activated, or a lack of highly likely choices from these alternatives, or both (see also Sections 3.5 and 5.4).

In this respect, the activation within the bilingual lexicon is often assumed, e.g., in the Bilingual Interactive Activation Plus model (BIA+, see Dijkstra & van Heuven, 2002), to be directly affected by surrounding linguistic context which provides lexical, syntactic, and semantic information. This assumption seems to be supported by experimental studies in bilingualism (e.g., Schwartz & Kroll, 2006). Similarly, in the re-ordered access model (see, e.g., Duffy et al., 2001), the relative frequency of the alternative meanings of an ambiguous word (which often correspond to different translation alternatives) determines the order (or relative speed) in which these meanings are activated and compete for selection, while this activation can be re-ordered by a strong biasing context. For structural building framework (Gernsbacher, 1990, 1997), the mental processes involve a combination of enhancement and suppression, in which the activation of contextually relevant information is enhanced, whereas a suppression effect reduces the activation of information that is irrelevant to the context. Although there does not seem to be "a uniform theoretical account" as to "how sentence context exerts its influence on bilingual lexical access" (Schwartz & Kroll, 2006, p. 209), the fact that linguistic context does aid the interpretation of

[33] As mentioned in Chapter 2 (see 2.2), this is suggested by evidence from many studies in both TPR and bilingualism, although the *extent* of this activation is often relevant to language proficiency, particularly regarding a non-dominant language (e.g., experiments also show that semantic priming is stronger for those with higher proficiency; see Favreau & Segalowitz, 1983).

ambiguous words, and also reduces the number of appropriate translations for an ST word, is perhaps without much disagreement.[34]

In terms of probabilities which are observed in the *text*, this results in a distribution where the different translation choices for a given ST word are not equally probable (i.e., not equally appropriate for the context). This distribution of observed probabilities can be described by its entropy value, and approximated from the translation choices made by different translators regarding the ST word (i.e., HTra value). As essentially a statistical feature of the translation product, a higher entropy (i.e., HTra) in the observed translations indicates a higher level of uncertainty in the selection among the possible TT items in the sample.[35]

In terms of *mental processes*, the translator can perhaps be assumed to engage in an activation pattern where the activated items receive different degrees of priority for resource allocation (or according to the re-ordered access model, a pattern where the items are activated in a certain order in terms of relative time course). This pattern is largely affected by linguistic context,[36] and can appear as a distribution of probabilities which are observed in the produced translation. During the subsequent selection process, the pattern would be dynamically updated to arrive at a translation choice (see below), while the distribution of probabilities which can describe this new pattern is updated accordingly. In a similar manner to the probability distributions observed in the textual material, the activation pattern in a particular mental state can also be represented by entropy (i.e., the distribution of resource allocation to various alternatives which are activated, or, perhaps equivalently, the distribution of temporary probabilities with which the activated candidates are to be selected), in turn indicating the (temporary) uncertainty level in the mental state regarding the selection among the activated candidates.

---

[34] The disagreement in this regard is around *when* (i.e., how early), rather than *whether*, sentence context exerts its effect. See also the next footnote below, where more detail is described.

[35] Note that this is the entropy in the *text*, rather than in the mind.

[36] It can be debated as to how early this effect is exerted — some context-dependent accounts in monolingual processing argue that the conceptual representations which are built along the sentence have an early effect on lexical access, while in some context-independent accounts, sentence context influences not the initial activation, but the subsequent selection process after the word has been accessed. Despite this disagreement, however, it seems that the effect of context concerning the pattern here, which is subsequently updated in the mental processes, can be safely assumed.

As the translator attempts to resolve the uncertainty arising from a high-entropy ST item (e.g., high HTra as observed in the translation) — typically by making use of additional contextual information — a change (i.e., update) occurs in the pattern of the co-activated possible (TT) translations for this ST item, as new input is received from the context (e.g., information which is provided by the surrounding items in the textual material). The extent to which each candidate is activated would then be further enhanced or suppressed in accordance with its relevance to the contextual cues which are increasingly inputted as cognitive activities proceed. Some choices would become more probable while others would become less so, as the translator proposes and evaluates solutions on the basis of contextual information received from scrutinizing the items preceding or following the current high-entropy item. If further contextual information is inputted as a further step to arrive at a decision, the pattern continues to change.

Consequently, as the translator proceeds with the selection process for a suitable TT item, the allocation of cognitive resources becomes less evenly distributed over the various alternatives (or in other words, the probabilities which describe the updated activation pattern become less evenly distributed), concentrating on the items which are more likely than others to be chosen by the translator. Since the value of entropy represents — among other aspects — the extent to which probabilities are evenly distributed (see Chapters 3 and 5), this means that the updated entropy (in the assumed mental state) decreases, together with a decrease in the uncertainty involved, and the choice of a TT option for the ST item becomes more straightforward. When the resulting entropy decreases to zero, the choice would be restricted to only one option (i.e., the choice made by the translator).[37]

During this process, there is continual shift in the allocation of cognitive resources from one mental state to another, while the entropy level is continually reduced towards the end of the selection. The process can therefore be described as a process of resource allocation shift and entropy reduction, where cognitive effort needs to be expended and can be quantified accordingly by these two means — shift of resource allocation and reduction of entropy.

[37] Mathematically, the entropy value for a distribution that is concentrated on one single item equals 0.

If Levy's formulation (see Section 6.2.1; see also 5.3 for details) can be adopted, the size of the shift of resource allocation between two certain points during this process would be represented by the *relative entropy* of the updated pattern with respect to the previous pattern.[38] In this view, the cognitive effort which is expended in the entire translation selection process can therefore be quantified via the relative entropy of the pattern when the translation choice is made at the end of the process, with respect to the beginning of the process when the items are activated. The following section will show that mathematically, the value of this relative entropy would be equal to the surprisal of the TT item that is chosen by the translator.

From a different perspective, the above process can also be viewed as continual reduction of entropy levels in the mental states, where cognitive effort expended can reasonably be quantified by the extent to which entropy is reduced. In this regard, if the decrease of entropy value is used as the measurement of cognitive effort in this selection process, then after the choice is made (i.e., after the entropy decreases to zero), this decrease would equal the initial entropy when all the TT candidates are activated given the ST item (i.e., the entropy in the mental state between activation and selection).

The following sections will describe these two in more detail, and show that the former equals to surprisal, while the latter is perhaps equivalent to word translation entropy described in Chapter 3. The relationships between the entropy for the mental states and the entropy that can be observed in the text via statistical means are also explained, with the two being equal to each other under the assumption of lexical activation being modulated by frequency of occurrence and contextual effects.

6.2.2 Surprisal, relative entropy, and ITra

Here, between these two ways of representing cognitive load via entropy — relative entropy and the decrease of entropy value, the relative entropy of the activation pattern at the end of the process is described as being equal to the surprisal of the TT item eventually chosen by the translator. This is because at the end of the selection process (i.e., when the mental processing has arrived at a decision as to which

---

[38] From an information theory perspective, this relative entropy indicates the penalty incurred from encoding the new pattern with the old one.

particular target item is to be selected), the distribution of cognitive resources in the mental state can be reasonably assumed to have, after a series of continuous update (or shift) along with the expenditure of cognitive effort, eventually concentrated on one single item (i.e., the item chosen by the translator) whose probability therefore equals 1 given this mental state. According to the definition of Kullback-Leibler divergence (i.e., relative entropy), the divergence of the updated distribution Q(x) from the original distribution P(x) equals the expectation of the logarithmic difference between Q(x) and P(x), with the expectation taken using Q(x). Suppose there are n possible items in the mental lexicon (i.e., n values for x in $x \in \chi$), among which the item chosen by the translator is W, then the above description would mean that $Q(W)=1$, that $Q(x)=0$ when $x \neq W$, and that P(x) represents the probabilities in the initial activation pattern for both $x=W$ and $x \neq W$. In this case, the divergence $D_{KL}(Q \parallel P)$ would be:

$$D_{KL}(Q \parallel P) = \sum_{x \in \chi} Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$= \sum_{i=1}^{n} Q(x_i) \log\left(\frac{Q(x_i)}{P(x_i)}\right)$$

$$= Q(W) \log\left(\frac{Q(W)}{P(W)}\right) + (n-1) \lim_{Q(x)\to 0^+} Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$= \log\frac{1}{P(W)} + (n-1) \lim_{Q(x)\to 0^+} Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$= -\log P(W) + (n-1) \lim_{Q(x)\to 0^+} Q(x) \log\left(\frac{Q(x)}{P(x)}\right)$$

$$= -\log P(W) + (n-1) \lim_{Q(x)\to 0^+} [\, Q(x) \log Q(x) - Q(x) \log P(x) \,]$$

$$= -\log P(W) + (n-1) [\, \lim_{Q(x)\to 0^+} Q(x) \log Q(x) - \lim_{Q(x)\to 0^+} Q(x) \log P(x) \,]$$

As $\lim_{Q(x)\to 0^+} Q(x) \log P(x) = 0$ and

$$\lim_{Q(x)\to 0^+} Q(x)\log Q(x) = \lim_{Q(x)\to 0^+} \frac{\log Q(x)}{\frac{1}{Q(x)}}$$

$$= \lim_{Q(x)\to 0^+} \frac{\frac{d}{d(Q(x))}\log Q(x)}{\frac{d}{d(Q(x))}\frac{1}{Q(x)}}$$

$$= \lim_{Q(x)\to 0^+} \frac{\frac{1}{Q(x)\ \ln 10}}{-\frac{1}{[Q(x)]^2}}$$

$$= -\lim_{Q(x)\to 0^+} \frac{Q(x)}{\ln 10}$$

$$= 0$$

it then follows that

$$D_{KL}(Q \parallel P) = -\log P(W)$$

In other words, the Kullback-Leibler divergence of these two distributions (i.e., the relative entropy between initial activation and final selection) equals the surprisal of the item that is eventually chosen by the translator, i.e., -log$P(W)$.

Regarding the concept of surprisal, Chapter 3 and Section 6.2.1 have also mentioned that this concept is also known — in different contexts — as information, self-information, or Shannon information content, all referring to essentially the same mathematical equation (i.e., the negative logarithm of probability, see equation 2 in Chapter 3). In some recent papers, the surprisal regarding a particular translation item is specifically called “word translation information”, and approximated by ITra (see, e.g., Carl, 2021b; Heilmann & Llorca-Bofí, 2021). These terms, although focusing on quite different aspects, are in fact mathematically expressed in the same manner as the surprisal discussed here.

As the $P(W)$ in the surprisal above represents the probability of W in the initial activation pattern (i.e., when W is first activated together with all other items), this surprisal should in theory refer to the surprisal in the corresponding mental state at the initial stage, rather than the surprisal of the item in the textual material.

However, if the activation of lexical items is (to a considerable extent indeed) modulated by context and the frequency of the different meanings/translations (e.g., in the re-ordered access model), the P(x) which describes the mental state of activation would be the same as the probabilities that can be observed in the text. This means that the initial surprisal for this item W in the mental state, i.e., $-logP(W)$, is equivalent to its surprisal in the text. This is explained in more detail in the following sections.

In this manner, the relative entropy with respect to the above mental process would be, albeit arguably, equal to the corresponding surprisal in the text. Cognitive effort can thus be represented by this surprisal (consistent with Levy's formulation), and in turn approximated by corpus-based analyses on the text. As mentioned, this surprisal is the same as word translation information (an approximation of which is ITra).[39]

6.2.3 Decrease of entropy and HTra

From the same perspective, the initial entropy value in the mental state would be equal to the entropy value that is observed in the text (e.g., HTra). If the *decrease* of entropy value, i.e., the absolute difference between the two respective entropy values regarding the initial and final mental states, is used as a measurement of cognitive effort in the selection process, then at the point when the translation choice is made, this decrease would equal the initial entropy when all the TT candidates are activated given the ST item (i.e., the entropy in the mental state between activation and selection), and in turn equal the word translation entropy.

Specifically, when the choice is made, the entropy in the mental state refers to the entropy for distribution Q(x), which equals zero:

[39] Note that the CRITT TPR-DB estimates this value on the basis of the translation choices made by all participants in each experiment. However, the surprisal here can in fact be approximated in other ways as well, using different corpus data, and would result in different ITra values than those in the CRITT TPR-DB. This is the same for HTra.

$$
\begin{aligned}
H_1(x) &= -\sum_{x\in\chi} Q(x)\log Q(x) \\
&= -\sum_{i=1}^{n} Q(x_i)\log Q(x_i) \\
&= -Q(W)\log Q(W) - (n-1)\lim_{Q(x)\to 0^+} Q(x)\log Q(x) \\
&= -Q(W)\log Q(W) \\
&= 0
\end{aligned}
$$

The initial entropy associated with the pattern of activated lexical items, i.e., the entropy in the initial mental state, is as follows:

$$H_0(x) = -\sum_{x\in\chi} P(x)\log P(x) = -\sum_{i=1}^{n} P(x_i)\log P(x_i)$$

where $P(x_i)$ refers to the conditional probability with which x_i is to be selected, given the mental state at the initial stage of activation.

Accordingly, the decrease of entropy value between these two points, i.e., from *P*(x) to *Q*(x), or from initial activation to final selection, would be simply:

$$H(x) = H_0(x) - H_1(x) = H_0(x) = -\sum_{i=1}^{n} P(x_i)\log P(x_i)$$

Here, if the activation of lexical items is modulated by context and the frequency of meanings/translations, as mentioned above, the *P*(x) in this equation can be considered equal to the probabilities observed in the text. This means that the *H*(x) here would be the same as word translation entropy explained in Chapter 3, i.e., that which is calculated from the probabilities in the text and approximated from the sample. In other words, the decrease of entropy value in the mental state can perhaps be approximated by the HTra value (i.e., the formulation in Carl et al. 2016).[40]

[40] See previous note.

6.2.4 Entropy for the mental state and entropy in the text

In the above description, the activation of lexical items is considered to be modulated by context and the frequency of meanings/translations, which is directly related to the value of entropy in the initial mental state. This is also directly relevant to the relationship between the entropy value which can describe the mental state, and the entropy value which is observed in the text. Here, more specific analysis is provided in this regard.

Strictly speaking, if the activation is completely non-selective, the activated entropy would be genuinely maximum — everything in the mental lexicon is activated. However, this seems a rather extreme scenario, and it is perhaps more appropriate to assume that only the multiple meanings (translations) relevant to the word are activated, and that these meanings are activated completely non-selectively. This means that the entropy in the mind at the initial stage of activation is maximum for the meanings/translations (H_{max}, i.e., the entropy when all probabilities of the meanings/translations are equal, see Figure 6.1 below for illustration of a simplified example for four items). For this scenario, the cognitive resources in the mental state are perhaps equally distributed among all the possible items (assuming parallel processing in disambiguation, see below), or alternatively, equally likely to be allocated to each activated item (assuming serial processing, see below).

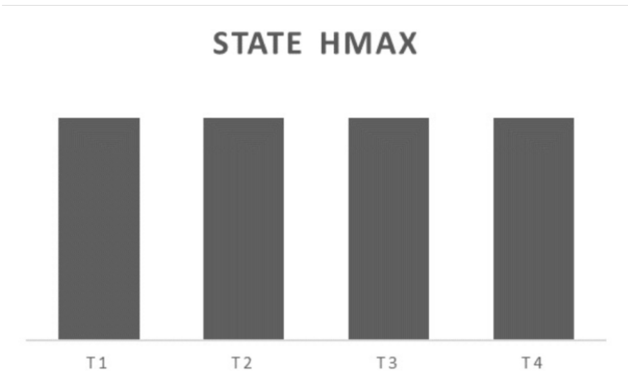


Figure 6.1 Mental state when entropy is H_{max}

The entropy observed in the actual meanings/translations of the word, however, is the distribution of the relevant frequencies (H_1) of these meanings/translations, which can be obtained from a well-balanced corpus of general language when context is *not* considered. This is shown in Figure 6.2, where the probabilities of the items are not equally distributed, and where the entropy value H_1 would be lower than H_{max}.

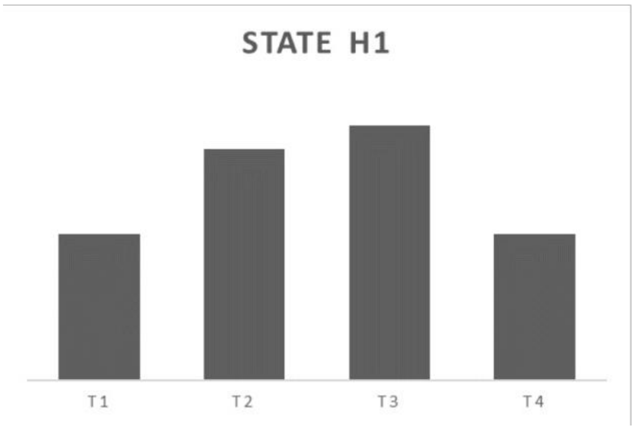


Figure 6.2 Mental state H_1

When context is considered, on the other hand, the distribution of the frequencies of meanings *within* context corresponds to a smaller entropy (H_2), which can be illustrated by Figure 6.3. Regarding the four items in the example, Figures 6.2 and 6.3 illustrate that while item T3 could be the most frequently used (and hence most probable) translation for the ST word in question, the usage of the word in its context may result in very low frequency for T3 (as well as T4) but a strong inclination towards T2. While T1 could be a relatively rare translation, in the specific context it may become a very frequent (and thus likely) translation for the ST word. In general, it seems safe to assume that when the context is taken into consideration the items would be less evenly-distributed, or the translation options are restricted to fewer possibilities. Therefore, H_2 would be smaller than H_1.

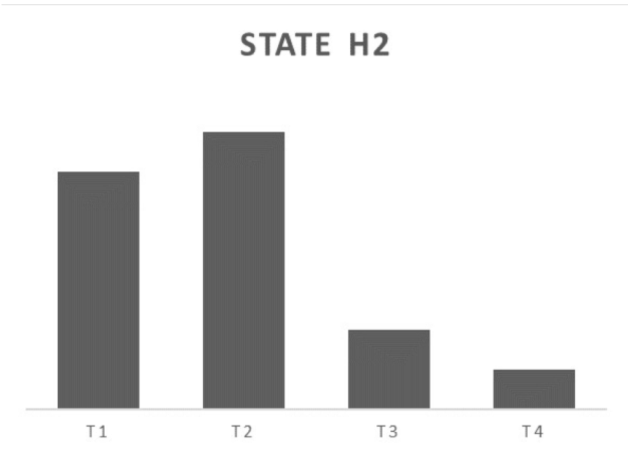


Figure 6.3 Mental state H_2

Depending on the method of data processing, HTra can be H_1 or H_2 (this has to do with token-level alignment and the type of text data used for calculation), but both reflect the observed entropy in the text and both are smaller than H_{max}. At the end of the selection process, the entropy in the mind is reasonably low and can be assumed to be zero.

Regarding lexical activation, all above mental states are possible, depending on what theoretical assumptions are made. Therefore, the entropy in the mind for this mental state can be one of the following: H_{max} (completely non-selective), H_1 (modulated by frequency of translation), or H_2 (modulated by frequency of translation in context).

In terms of the observed entropy in the text, context reduces the entropy from H_1 to H_2. As mentioned above, while the impact of context on bilingual lexical access is generally agreed upon, it is still largely debatable as to *when* (i.e., how early) this context exerts its influence. Some context-dependent accounts (in monolingual processing) argue that the conceptual representations which are built along the sentence have an early effect on lexical access, while in context-independent accounts, sentence context influences not the initial activation, but the subsequent selection process after the word has been accessed.

Here, if the effect of context in the cognitive processes is assumed to be exerted early on, the entropy in the mind would perhaps be H_2. On the other hand, if a context-independent account is adopted (i.e., it influences not activation but selection), the entropy in the mind would be H_1. If the activation is completely non-selective for these meanings, the entropy would be H_{max}.

This can be demonstrated by Figures 6.4, 6.5, and 6.6 below. In terms of words *out of context*, Figure 6.4 shows the entropy values that can be observed in the text, together with the entropy for the mental state at the initial stage of activation. As the words are decontextualized in this case, the entropy in the text would be associated with the frequency of occurrence for the corresponding meanings/translations regardless of its context, which is H_1 (see also Figure 6.2 above). For the entropy that describes the mental state, if the items are activated completely non-selectively without any bias or influence of their frequencies or contextual cues, this entropy would be H_{max}, i.e., the corresponding entropy value when the probabilities for all items are equal to one another. This is shown in the chart on the left side. If, however, the lexical activation is modulated by frequency of occurrence (and free of contextual impact), i.e., the more frequent items are activated to a greater extent than less frequent ones, then the entropy in the mental state would be equal to the entropy that is based on (only) the frequency of occurrence, i.e., H_1 (see also mental state H_1 above).

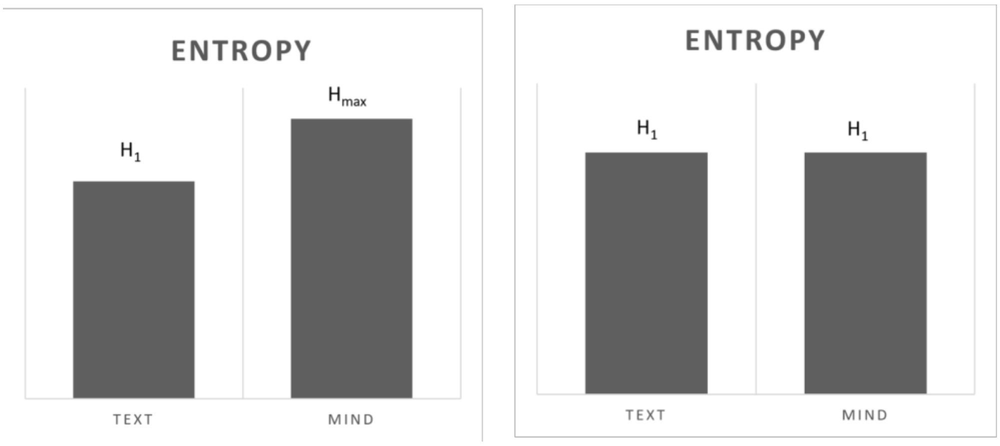


Figure 6.4 Entropy values in the text and in the mind when context is not considered

When context is taken into consideration, the entropy in the text would be reduced to H_2 (see above). For the entropy in the mental state, the same as above can also be hypothesized. If lexical activation is, again, completely non-selective and non-biased, meaning that all items are activated to an equal extent, this entropy would be H_{max} (see left chart in Figure 6.5). If this activation is modulated by frequency of occurrence, then this entropy would be the entropy value that can be calculated from these frequencies.

Here, there would be two scenarios for the latter case, depending on when, or how early, context exerts its influence in the mental processing (i.e., lexical access). For context-independent accounts of lexical access, the impact from the context comes only at a later stage after the items are activated (i.e., the context influences the subsequent selection, rather than the initial activation, of lexical items), and with the modulation of the frequency of occurrence, the entropy in the mental state would be the one modulated by frequencies out of context, i.e., H_1. On the other hand, adopting a context-dependent account, the contextual effect would be already exerted in the lexical activation process, and this means the entropy would be H_2, which is quite reasonably lower than H_1. These two are illustrated in Figure 6.5 as the charts in the middle and the one on the right side respectively.

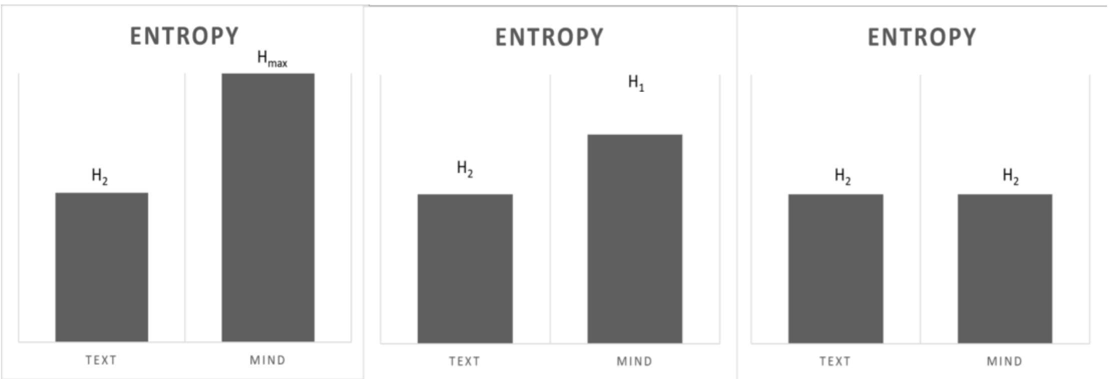


Figure 6.5 Entropy values in the text and in the mind when context is considered

These are possible mental states at the initial stage of cognitive processing in terms of the lexical choice for a particular ST word. From these states, the mental processing proceeds to a selection process whereby entropy is continuously reduced along with the expenditure of cognitive effort, and at the end of the process the entropy for the mind is reasonably low (and can be assumed to be zero). The entropy in the text, however, stays unchanged and becomes increasingly higher than the entropy in the mind. Figure 6.6 shows the entropy values after the selection process.

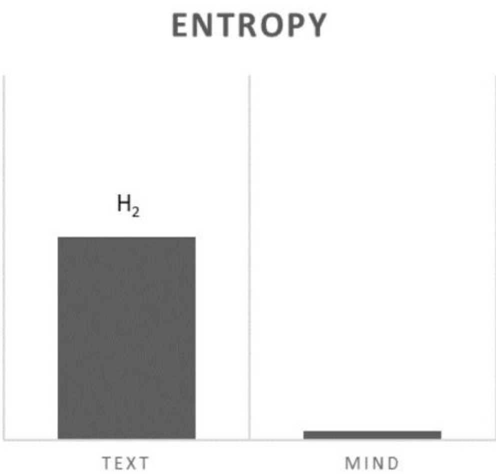


Figure 6.6 Entropy after selection

The process of making lexical choices in translation regarding a particular ST word can therefore be described as a process of entropy reduction, and the cognitive effort would be quantified via this change of entropy between the initial and final mental states.

When measuring the cognitive effort in the process, the *decrease* of entropy can be H_{max}, H_1, or H_2, depending on which of the above-mentioned mental state is most likely at the initial stage, while the *relative entropy* equals the surprisal of the selected item for these respective mental states (i.e., the item's surprisal corresponding respectively to H_{max}, H_1, and H_2).

6.2.5 Transition between mental states

Specifically, the above process can be illustrated by the following charts (see Figure 6.7). Suppose the initial mental state is H1 (with the effect of context exerted at a later stage *after* the lexical item has been accessed), where all the activated items receive proportional cognitive resources consistent with their frequency of occurrence in an ideally well-balanced, general-language bilingual corpus. Here, T3 is the most likely option, as can be seen in the chart, receiving the most cognitive resources and corresponding to a very high probability of being selected given this mental state.

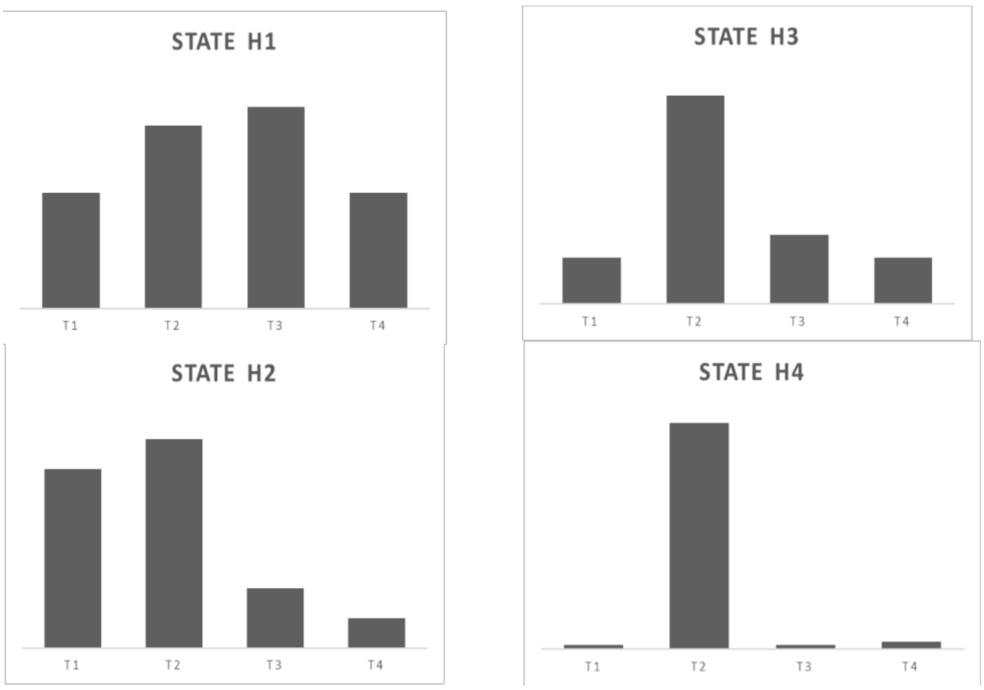


Figure 6.7 An example of mental states in selection process

As the mental processing proceeds, the effect of context comes into play, making T3 much less likely than T1 and T2. The translator recognizes that in this specific context, the meaning/translation for the ST word is most likely T1 or T2. Accordingly, the cognitive resources are now concentrated on these two, with the other two options (T3 and T4) receiving very low levels of attention. From the previous mental state, a shift of cognitive resources has occurred which can be quantified by the relative entropy between H_1 and H_2, according to Levy's formulation. This indicates the cognitive effort expended between these two mental states. On the other hand, as entropy is reduced between these two mental states, cognitive effort can also be indicated by the decrease of entropy value, and this decrease would be the absolute difference of the entropy values: H_1-H_2.

Then the mental processing continues to receive (and integrate) further contextual cues, leading to further adjustment in the allocation of cognitive resources — based on each candidate's appropriateness for (i.e., relevance to) the contextual information that has been inputted, processed, and integrated at this point. This can be manifested in behaviour by scrutinization of the text preceding and/or following the ST word in question (and a decrease of entropy in the text scrutinized, which will be described in the empirical analyses below). The mental state now transitions to H3, as the translator becomes increasingly aware that T1 may be much less likely than T2 to be appropriate (see Figure 6.7) regarding the translation of the ST word. The distribution of mental resources (or the distribution of the probabilities of being selected) among these items becomes less even, increasingly concentrating on T2. The corresponding entropy for this mental state, accordingly, is lower than that for the previous states.

After H3, the mental processing further evaluates the translation options, perhaps represented by more scrutinization of the ST/TT in terms of eye-movement behaviour, and continues to shift resource allocation, expend cognitive effort, and reduce the entropy level, until arriving at a selection which, in the above example, is T2 (see Figure 6.7). The resources are now concentrated on T2, while the probability of being selected, given this mental state, is concentrated on T2 as well. The corresponding entropy, accordingly, is zero.

Here, the above four mental states are of course very simplified illustrations of the process, and there can be a number of mental states between any two of them, gradually transitioning towards H4.[41] In view of the entire process, the cognitive effort expended can perhaps be quantified by the relative entropy, or decrease of entropy, between H1 and H4, if the process starts with H1. As mentioned above, the entropy in the initial mental state can possibly be any of the following: H_{max}, H_1 and H_2,

[41] In the meantime, the illustration is also simplified in terms of the number of possible choices (this example shows only four possibilities), although the general conceptualization is not confined by the number of available choices. If one considers an infinite number of choices, the entropy equation can be adjusted to reflect the integral, rather than the sum, of surprisals (using probability density function), but the discussion would still be consistent.

depending on when (or perhaps to what extent) context exerts its influence on lexical access.

6.2.6 Serial and parallel processing, and entropy

The above has given considerable conceptual attention to entropy, mental states, and cognitive load. More importantly, a distinction has been made between the entropy that describes the mental state, and the entropy that can be observed in the text. However, although the discussion is largely dependent on an essential equation involving probabilities, namely P(x) and/or Q(x), what exactly this probability means seems to merit further clarification.

Here, to disentangle the specific meaning of P(x) with respect to a particular mental state, it seems necessary to also make a distinction between serial and parallel processing of lexical ambiguity (see Chapter 2).

As described above, the translator's mental processes can perhaps be assumed to engage in an activation pattern, with the activated items receiving different degrees of priority for resource allocation. This probability can be interpreted as the probability with which the cognitive resources are allocated to a particular item, and from this perspective, it seems to assume a serial processing view, as the resources are allocated to one item at a time. The mental activities can be understood as attending to this or that item, among all the items that can be accessed in the mental lexicon. Entropy, accordingly, refers to the distribution of such probabilities.

On the other hand, a parallel processing perspective can assume that all items are accessed and processed at the same time, and that all of them receive cognitive resources simultaneously. From this perspective, it is perhaps reasonable to consider P(x) as the relative amount (i.e., the proportion) of cognitive resources allocated to x, as some of the activated items can receive more resources than others. This results in a curve distribution of resource allocation among them, which is represented by entropy. A change of this distribution of P(x), accordingly, corresponds to a shift of resource allocation.

Regarding lexical activation, if one adopts the re-ordered access model, the alternative meanings/translations of the word are activated at different points of time, with more frequent ones activated earlier (faster) than others. From this perspective, the P(x) can perhaps represent the relative speed in which x is activated and compete for selection (or perhaps equivalently, the probability for x to be activated first). In this model, the activation can be re-ordered by a strong biasing context, which is consistent with the discussion above (H_{max}, H_1, and H_2).

In terms of the structural building framework, the mental processes described above would be the effect of enhancement and suppression combined, with the activation of contextually relevant information being enhanced and irrelevant information suppressed. From this perspective, the P(x) can perhaps represent the extent to which an item is enhanced relative to the most suppressed ones (as a bottom line). Each item is enhanced/suppressed to a different extent regarding a particular mental state, with this extent being influenced by contextual cues received. This corresponds to a distribution of P(x) among the items, represented by entropy. As more context is processed, this distribution changes as a result of further enhancement and suppression, in turn indicating a change in entropy.

From a much simpler probabilistic view, the P(x) above can represent the conditional probability with which x is to be selected given a certain mental state, which continuously changes as mental states transition from one to another. This results in a probability distribution represented by entropy. As the mental state approaches the end of the selection process, the conditional probability of one item to be chosen, given the corresponding mental state, approaches 1, while that of all other items approaches 0.

For all the ways of understanding probability as considered in this section, entropy and entropy reduction appear to allow a reasonable description, compatible with different perspectives, of the mental processes involved.

Concerning the relationship between entropy and the serial vs. parallel models of lexical access, it seems that the conceptualization here is compatible with both. However, the assumption of the entropy framework is based on bilingual co-activation, and in this respect, it endorses parallelism (in view of whether ST comprehension and TL reformulation occurring simultaneously).

6.2.7 Systems theory perspective of the translation process

The above description of the dynamic change in probability distribution, borrowing Levy's (2008) formulation of resource-allocation processing difficulty, seems consistent with many aspects of a systems theory perspective on the translation process, a framework proposed in Carl et al. (2019), where the use of entropy as a description of the translation process is more inclined to the way in which entropy is defined in systems theory (or thermodynamics) rather than in information theory. From that perspective, entropy refers to the amount of disorder in a system, and the process of translating is considered "a hierarchy of interacting word and phrase translation systems which organize and integrate as dissipative structures" (p. 211). The expenditure of cognitive effort, or "average energy", to arrive at a translation solution decreases the internal entropy (i.e., disorder) of the system.

If an ST item together with all the possible translation alternatives of this item can be considered a word or phrase translation system, the empirical investigations in Chapter 8 will show that for a highly entropic system, the decrease of its internal entropy tends to involve surrounding low-entropy systems with which the current system interacts.

6.2.8 Surprisal and entropy in translation

In Chapter 3, the difference between surprisal and entropy is explained in terms of their mathematical definitions and their implications in information theory. Here, the two are analysed in view of their quantification of the effort required in cognitive processing (i.e., cognitive load), and using Levy's formulation, *surprisal* of the chosen item (i.e., word translation information) is argued to be equal to relative entropy between the final and the initial mental states, which represents the shift of resource allocation. The *entropy* generalizing over all alternative options (i.e., word translation entropy) is argued to be equal to the decrease of entropy in the mental states.[42]

These two perspectives (i.e., shift of resource allocation and entropy reduction) can therefore be represented by surprisal and entropy respectively, as quantification of cognitive load. In the context of translation, these two would be in the form of word

[42] Note that the options here refer to options in the mind, rather than in the text.

translation information (i.e., surprisal) and word translation entropy (i.e., entropy). If considered from a *product* perspective, the difference between them seems straightforward — word translation entropy generalizes over different translation items while word translation information indicates the unpredictability of a specific translation item given a particular ST word (see equations 1 and 2 in Chapter 3, for surprisal and entropy). From a *process* perspective, the above sections have shown that between the mental state of initial activation and that of final selection, word translation entropy represents the reduction of entropy while word translation information indicates the size of the shift in cognitive resource allocation. Levy's formulation for quantification of cognitive load, however, would appear to be consistent with word translation information, rather than word translation entropy.

In terms of their mathematical expression, word translation entropy represents the initial $P(x)$ distribution when alternative options are activated, whereas word translation information indicates the surprisal of the final choice. Word translation entropy is equivalent to the *absolute difference* of entropy between the two mental states, while word translation information is equivalent to the *relative* entropy of the final mental state with respect to the initial mental state.

Both of the two concepts, i.e., word translation information (i.e., surprisal) and word translation entropy (i.e., entropy), can be approximated from the text through corpus-based means.

When approximated from the experimental session (rather than a carefully-selected, large-size corpus), perhaps the metric of ITra in the CRITT TPR-DB would be indicative of shift of resource allocation, while HTra is relevant to entropy reduction. In this regard, the entropy-based metric frequently used in empirical TPR, calculated from the CRITT TPR-DB (Carl et al., 2016), seems to be an approximation of the entropy reduction (i.e., decrease of entropy in the mental states) discussed in this chapter, rather than the size of shift in resource allocation (i.e., relative entropy between two mental states).[43]

[43] Again, the HTra in the CRITT TPR-DB is merely *one* of the ways to approximate this entropy, and does not seem to be a way that exactly reflect the entropy in the mind.

6.2.9 Translation and post-editing

Regarding translation and post-editing, it seems that the perspective explained in this chapter is compatible with both. From the lens of entropy and entropy reduction, the two activities do not seem to be fundamentally different, although in post-editing the input of the raw MT output provides another source of activation, resulting in a perhaps high level of entropy in the mental state. The process of entropy reduction in the transition between one mental state and another, as explained in the above sections, would seem to remain consistent. Therefore, the discussion here is not really confined to translation, but also applies to post-editing. Although the cognitive processes are not exactly the same between the two, the probabilistic description of the mental states is largely consistent.

6.2.10 'Final' choice

In the description of the mental processing above, the stage where a decision is arrived at in view of the translation selection is considered an endpoint in the decision making and translation process. However, in reality it is entirely possible that a decision is made, looks final, but is then revised either immediately or at a later time during self-revision. In this regard, the actual process would involve several rounds of entropy increase and decrease, rather than a simple process of reduction in entropy. At the initial stage of activation, the entropy for the mental state would be high, and towards a decision, the entropy would be gradually reduced to a reasonably low level as mental states transition from one to another (i.e., the process described above). At a later point of time, the translator/post-editor would perhaps process more contextual information and activate more options in the mind, thereby increasing the entropy for the mental state. By the time the initial choice is reconsidered, another round of entropy reduction would occur, which may or may not result in the same option being selected in the mind of the translator/post-editor. The 'final' in the description above should therefore be relative.

6.2.11 Context and entropy

In this chapter, there is much discussion on the impact of context on entropy (both in the text and in the mind). Although the discussion in the present thesis (especially the empirical investigations in the following chapters) has not considered those aspects of language beyond lexis and syntax, or the situatedness of cognition, the description of mental processes via entropy is not limited to the scope of this thesis. The impact of context discussed here, for example, is not confined to co-text, although the discussion in Chapters 8 and 9 considers co-text only, for the convenience of operationalization. The theoretical account here, however, is not constrained by this aspect.

Specifically, the impact of context can be incorporated in the probability distribution that describe the mental state, as context can condition these probabilities. What was processed prior to the encounter of the current word, phrase, or sentence, can also condition these probabilities. The way in which the text is presented to the translator can also condition the probabilities. There would be numerous factors can impact these probabilities. It should be noted that the discussion of entropy in this chapter is focused on the entropy for the mental states in the mind (the value of which, if we want that value, would be different from individual to individual), rather than a statistical feature extracted from the text and used to describe every individual.

In the meantime, the entropy (for the mental states) under discussion here is conceptual and abstract, rather than referring to a specific, concrete, and constant result (in the form of a number) that is calculated from a metric that operationalizes it (i.e., the approximation of it in, e.g., the CRITT TPR-DB, which extrapolates the performance of a group of translators to the behaviour of a particular individual). Nor is this entropy, in the present chapter, a feature of the text but a (probabilistic) feature of the mind. This entropy describes the mental states and its exact value can only be obtained if one had, somewhat miraculously, access to the items in the mental lexicon of a particular translator, as well as the specific lexical information that is activated in his/her mind, the extent to which this information is active, and the change of this extent through enhancement and suppression. These are impossible to obtain.

That being said, it is possible to approximate the entropy value via different means, one of which can be the calculation of HTra as operationalized in Carl et al. (2016), a metric that may not accurately capture the cognitive aspects of a particular individual. Using corpus of different kinds also results in various approximated values of this entropy, but the conceptualization provided in this chapter remains the same.

Another important issue in this regard is that there are always individual differences from person to person, while this approximated value of entropy, operationalized via whatever means, would be one single number that seemingly describes every individual. Here, it should be emphasized that, on the one hand, the conceptualization of the mental states, processes, and entropy as abstract concepts would not change from person to person — it would only be the specific values of this entropy that are different. On the other hand, the approximation of the value of this entropy can be different depending on what data is used for calculation. This by analogy would be similar to the relationship between the frequency of occurrence for lexical items and the familiarity of the item for an experimental participant. If the frequency is obtained in a corpus that is maximally similar to the collection of linguistic input that this participant has received all his/her life, it might be reasonable to consider this frequency as a dependable indication of familiarity. Similarly, the approximated value of entropy would be dependent on the specific corpus used, if one considers the genuine level of entropy for an individual. In the meantime, subjective measure of frequency, i.e., rating of familiarity, can also be transformed into entropy values and in turn indicate the actual entropy for a given population of readers.

Regardless of the way in which entropy or surprisal is approximated or operationalized, however, the specific value of this approximation does not impact the conceptualization provided in this chapter.

6.3 Summary

The above sections have explored the theoretical issues of entropy as a statistical, probabilistic description of the mental processes of translation selection, drawing inferences from both the notion of entropy as defined in Chapter 3, and the notion of entropy as used in psycholinguistics outlined in Chapter 5. The way in which entropy can quantify cognitive load, regarding both relative entropy and entropy reduction, is described in detail, and mathematical poof is also provided as to its relationship with word translation information and word translation entropy. In addition, issues such as entropy for the mental state vs. entropy in the text, approximation of relative entropy and entropy reduction, as well as the transition between mental states, are also analysed. Importantly, the chapter also explains, on the basis of such analyses, that the entropy-based, product-oriented metric calculated from the CRITT TPR-DB (Carl et al., 2016) and frequently used in empirical TPR studies is an approximation of the decrease of entropy in the mental states, whereas the shift of resource allocation in Levy's formulation, represented by relative entropy, would be equal to word translation information, for which an approximation can be ITra.

The conceptualization in this chapter has addressed RQ 1. But in the meantime, Section 6.2.6 has explained the relationship between entropy and the serial vs. parallel models of lexical access, and shows that it seems to be compatible with different models. That being said, the assumption of the entropy framework is based on bilingual co-activation, and in this respect, it endorses parallelism (in view of whether ST comprehension and TL reformulation occurring simultaneously).

In the following chapters, empirical investigations will be conducted on the basis of the CRITT TPR-DB.

Chapter 7 Eye movements

7.1 Introduction

The previous chapters have provided a theoretical account of entropy and entropy reduction in view of the research questions in this thesis. In later chapters, the focus will shift to the empirical aspects, largely addressing the behavioural manifestations of the processes explained in Chapter 6. This is based on eye-movements, as well as an assumption widely known as the eye-mind hypothesis. Therefore, before delving into the empirical part of the thesis, the present chapter provides the methodological ground on which subsequent chapters proceed. It discusses the use of the eye-tracking method in various areas, the types of eye-tracking equipment, and some of the findings from empirical studies on ambiguity that are conducted on the basis of eye-movements. Such studies cover lexical and syntactic issues influencing eye-movements, particularly in relation to ambiguity, and are followed by a critical discussion on the eye-mind hypothesis.

7.2 Areas of study using eye-tracking methods

For decades, analysing the movements of the eye has been used as an important method to study human behaviour, attention, and cognition in various disciplines. In TPR, O'Brien (2006b) is one of the first examples of using eye tracking, where pupil size measurements are used as an indicator of cognitive load and where this load is analysed in relation to TM matches. A follow-up study in O'Brien (2008) uses fixation duration as a measurement of cognitive load, also discussing TM matches. Teixeira (2014) analyses the impact of metadata on translation performance via a combination of eye-tracking, key-logging, and screen and face recording. In interpreting and sight translation, the use of the eye-tracking method is also not hard to find (e.g., Chang, 2009; Chmiel & Mazur, 2013; Dragsted & Hansen, 2009; Dragsted & Hansen, 2007; Hyönä et al., 1995; Seeber, 2015; Seeber & Kerzel, 2011; Shreve et al., 2010). Some studies in TPR use eye-movements to analyse reading behaviour in translation (Alves et al., 2011; Jakobsen & Jensen, 2008; Schaeffer et al., 2017). For measuring cognitive effort, eye-tracking has become a "particularly popular method" (Moorkens, 2018, p. 56) in translation studies, while in the context of MT post-editing, it is also used in addition to pause analyses.

These are just a few examples of the areas in TPR where eye-movement is used as a method to investigate how translation is produced, and this thesis does not aim to provide an exhaustive review of the literature in this regard. However, it is important to note here that most empirical studies related to word translation entropy and syntactic choice entropy involve analyses of eye-tracking data. For example, such studies tend to show that the value of word translation entropy has a statistically significant impact on first fixation duration, the probability of a fixation, and total reading time (Carl & Schaeffer, 2017c; Schaeffer, Dragsted, et al., 2016).

The CRITT Translation Process Database (see, e.g., Carl, Schaeffer, et al., 2016), a publicly-available, large-scale database, includes a considerable amount of eye-tracking data as well. The database systematically integrates eye-tracking and key-logging aspects of each experimental session, therefore the majority of studies using this database seem to adopt at least some of the eye-movement data. The fact that the size of the database is remarkable, and that different parts of the data are systematically integrated into a whole set of tables capturing various aspects of the translation or post-editing process, makes it possible to conduct rigorous statistical investigations on translation or post-editing behaviour across different participants, tasks, and languages, so that theories, models, and hypotheses can be verified, confirmed, or falsified more easily. In Chapters 6 and 7, the empirical investigations into lexical and structural aspects of ambiguity are also based on this database.

The following sections outline methods of measuring eye movements, as well as the lexical and syntactic issues that influence the movements of the eye.

7.3 Eye movements

Issues regarding how the eyes move have long been important for studies in many areas — psychology, biology, neuroscience, medicine, computer science, to name just a few, as the movements of our eyes are commonly considered "a vital part of our interaction with the world" (Wade & Tatler, 2005, p. 1), an interaction which plays a crucial and pivotal role in our perception, cognition, and education. The movement of the eyes is perhaps "the commonest potentially intentional human behaviour", occurring on average three times per second and indicating in a precise manner "what information is being received by the subject at the given time" (Soluch & Tarnowski, 2013, p. 90). The control of such movements is often seen as "a model example of cooperation between automatic and controlled mechanisms both in the aspect of perception as well as action" (ibid, p. 85). Perhaps for these reasons, and particularly because of its close relationship with many aspects of human cognitive processing, eye movement as a methodology is now "firmly established as a core tool in the experimentalist's armoury" (Liversedge et al., 2011, p. v), for the majority of research in psychology and a rapidly increasing number of studies in other disciplines (e.g., computer science, industrial engineering, linguistics, marketing/advertising, translation studies, etc.).

On the other hand, although the measurement of eye movements for different topics is increasing at a remarkable rate in recent decades, such research on the movement of the human eye is in fact not new at all. Even before modern recording devices became available, substantial strides had already been made in the experimental study of eyes and their movements, first by physicians and subsequently by psychologists, over one or two centuries' time, converging within contemporary neuroscience (Wade & Tatler, 2005, p. 1).

Laying the foundation of all subsequent studies regarding eye movements is perhaps the measurement of the kinematics of such movements, such as the velocity and duration of saccades (i.e., fast eye movements that redirect gaze),[44] together with their relationships with the amplitude of the eye movement (e.g., Carpenter, 1988; Westheimer, 1973). These studies reveal the characteristics of various kinds of eye movements, largely paving the way for the application of eye-movement research.

[44] It is generally believed that humans use a "saccade and fixate" strategy when viewing the world, where information is gathered during stabilized fixations, and where gaze direction is shifted by saccades as rapidly as possible. Saccades are one of the four types of eye movements, which also include stabilizing movements, smooth pursuit, and vergence movements. (see, e.g., Land, 2011; Gilchrist, 2011)

Subsequent investigations into the neural and muscular mechanisms generating saccades and other categories of eye movements have allowed researchers to infer the processes involved in the programming of these movements, including the factors impacting the selection of targets for fixation (Findlay, 1981; Findlay & Walker, 1999; Itti & Koch, 2000; Karn & Hayhoe, 2000; Krieger et al., 2000; Rao et al., 1996; Reinagel & Zador, 1999; Theeuwes & Burger, 1998).

More importantly, studies of issues related to vision during the movements of the eyes have not only revealed the mechanisms of vision suppression during saccades, but also raised crucial questions as to "how our smooth and continuous perception arises from the disjointed input to the perceptual system that is a necessary consequence of our saccade-and-fixate behaviour" (Wade & Tatler, 2005, p. 14). It is perhaps such questions that led to the prominence of eye-movement investigation in studies of many information processing tasks such as reading, scene viewing and perception, visual memory, and of course human translation processes as well as the post-editing of MT, which will be illustrated below in more detail.

7.4 Measuring eye movements

The techniques used for the study of how the eyes move have a long history of development, dating back centuries. For example, the historical review by Wade & Tatler (2005) illustrates vividly and succinctly how ancient as well as modern techniques to measure the movements of eyes have been employed over the centuries, starting from early investigations into how the eyes operate as an optical instrument (by scholars such as Kepler, Platter, Scheiner, Descartes, etc.). As these early studies led to increasingly thorough and accurate understandings of image formation and ocular anatomy, "attention could be directed to the functions served by the structures of the eye — both internal and external" (ibid, p. 5). The issue of accommodation, i.e., the ways in which eyes focus on objects at variant distances, was one of the initial issues investigated regarding the "functions" of the eyes' structures, where competing theories have resulted in a diversion of concern from other aspects of vision and into directions including how the eyes move.

In this respect, Descartes was probably the first to provide a "speculative mechanistic account" of the reciprocal action of the muscles in the human eye, after which considerable interest emerged in how the eyes move as well as the consequences of such movements for vision, to the extent that issues regarding the eye's movements actually "took a central role" in vision analysis at that time (ibid, p. 5).

In order to approach these issues, early researchers have used different techniques to monitor the movements of the eyes, from the subjective methods of direct observation (James, 1890; Erdmann & Dodge, 1898; Newhall, 1928; Peckham, 1934; Ogle et al., 1949; Fantz, 1961; Salapatek & Kessen, 1966; Hainline, 1998; Atkinson, 2000) to the more reliable and objective use of "afterimage" (Wells, 1792, 1794a, 1794b; Darwin 1786), making a number of meaningful discoveries on how the eyes move and how vision is affected during their movements, without use of modern sophisticated eye-tracking devices. In contemporary research, devices for studying eye-movements are far more advanced, allowing easier collection of data and more accurate tracking of the eyes' movement and position.

Modern devices

Just as afterimages at the early stages of research provided an important tool for obtaining knowledge on how the eyes move, modern studies have used sophisticated equipment to measure the movements and positions of the eyes in a more reliable and accurate manner. These tools are diverse, and can be classified in different ways. For example, Wade & Tatler (2005) group them in accordance with the principle they employ for monitoring eye position, into attachment devices, optical devices, and electro-oculographic devices. Duchowski (2007), on the other hand, classifies them into four generations in terms of the technological taxonomy:

1). First generation: eye-in-head measurement of the eye consisting of techniques such as scleral contact lens/search coil, electro-oculography;

2). Second generation: photo- and video-oculography;

3). Third generation: analog video-based combined pupil/corneal reflection;

4). Fourth generation: digital video-based combined pupil/corneal reflection, augmented by computer vision techniques and Digital Signal Processors.

An earlier categorization by Young & Sheena (1975) seems to complement the above taxonomy, which critically and comprehensively illustrates eye-movement techniques including: 1). electro-oculography, 2). corneal reflection, 3). limbus, pupil, and eyelid tracking, 4). contact lens method, 5). point of regard measurement, which tracks corneal reflection centre with respect to the pupil centre, 6). measurement of eye rotation by the double purkinje image method, 7) measurement of fixation point by determining the rotation of a plane attached to the eye, and 8). head movement measurement to obtain point of regard.

While these and many other classifications are systematic, comprehensive, and very thoroughly researched, they tend to be either historical, or inclined to aspects of technical details of the devices or techniques. From the perspective of the eye-tracker users, such classification might seem impractical for selection of an appropriate device for his/her study at hand.

In a much simpler and perhaps more practical manner, modern eye-trackers have also been divided into two main categories — intrusive eye-trackers and non-intrusive eye-trackers (e.g., Sharafi et al., 2015).

Intrusive devices for eye-tracking require participants to wear a padded headband with three miniature cameras used respectively for capturing the two eyes' movements and for tracking head movements. Eye-movements are captured via infrared lights which are reflected on the participants' pupils. Such devices include, for example, *Eye-link II,* a 500 Hz binocular eye tracking device with 0.5º average accuracy and 0.01º resolution, developed by SR Research Ltd.[45] This type of device has some limitations including stability issues, as head movements can lead to inaccurate eye-movement coordinates, and the "heavy goggle with several wires" is likely to "worry participants" (ibid), which could have an impact on the data obtained from the experiments (Sharafi et al., 2015).

Non-intrusive eye-trackers fall into two generations in terms of their development. Older devices typically rely on reflections of the light beams which are projected to the participant's eyes, and their resolution and precision are relatively low (ibid). Those of the newer generation are video-based, consisting of a computer, two miniature cameras, and an infrared pad. Eye movements are detected by tracking the

[45] See https://www.sr-research.com/products/eyelink-ii/ (retrieved 10 June 2018).

participant's head using eye-brows, noses, and lips, while also making use of corneal reflection and pupil centre for differentiating head- and eye-movements (ibid). Such newer-generation devices include *FaceLab*[46] and *Tobii 1750* (ibid).

7.5 Metrics to represent visual/mental effort

The metrics used in eye-movement studies to measure and calculate visual effort are perhaps as diverse as the various types of devices for eye-tracking. They can be generally divided into four categories according to the parameters on which they are based (e.g., Sharafi et al., 2015), and these parameters include: 1). number of fixations; 2). duration of fixations; 3). saccades; and 4). scan-paths.

Metrics based on the number of fixations calculate fixation count, fixation rate, fixation spatial density, or convex hull area to represent visual effort. Those based on the duration of fixations compute such values as average fixation duration, fixation time, proportional fixation time, average duration of relevant fixations, or normalized rate of relevant fixations. The metrics based on saccades are relatively straightforward, using the number of saccades, saccade duration, saccade amplitude or regression as an indication of visual effort and mental workload. The ones based on scan-paths compute such metrics as transitional matrix, attention switching frequency, scan-path precision, scan-path recall, and scan-path f-measure as mathematical representations of the effort expended by the participant.

In addition, some other techniques can also be used as similar metrics to scan-paths, including sequential pattern mining and ScanMatch.

For investigations into the ongoing processing in reading, especially regarding word recognition, psycholinguistic and TPR research tend to use first-pass measures to indicate early, subliminal processes, and a combination of early and later processing measures to indicate cognitive effort. The first-pass measures, i.e., early processing measures, refer to those measures that are prior to any regressions back to the word in question, and "are particularly informative for examining variables that are likely to have an impact on the initial access to representations of a word's orthography, phonology, or meaning" (Juhasz & Pollatsek, 2011 p. 874). Such measures include

[46] Developed by Seeing Machines. See http://www.seeingmachines.com/technology/

first fixation duration, single fixation duration, and gaze duration, and are computed only when the word is not skipped in the initial reading (i.e., when the word is fixated on the first pass). Other than these measures, another important and informative early measure is whether or not the word is initially skipped during the first-pass, i.e., whether the first fixation duration, single fixation duration, or gaze duration is zero in each trial, which is measured by skipping rates. The definitions of these measures are shown in Table 7.1 (from Juhasz & Pollatsek, 2011, p. 874).

Later processing measures include spillover duration, regression rates, second pass duration, and total fixation duration. These are indicative of the impact of the independent variable on the later cognitive processes regarding the word of interest. These are also shown in Table 4.1 below.

Early processing measures	**Definition**
First fixation duration	The duration of the very first fixation on the target word during the first pass, irrespective of number of fixations
Single fixation duration	The duration of the first fixation on the target word if it only received one fixation during the first pass
Gaze duration	The sum of all first pass fixations on the target word
Skipping rates	The percentage of cases in which the target word is not fixated on the first pass
Later processing measures	
Spillover duration	The duration of the next fixation after a reader moves their eyes off a target word (usually excluding regressions from the target word)
Regression rates	The percentage of regressions into a target word (**regressions in**) or out of a target word (**regressions out**)
Second pass duration	The amount of time spent re-reading a target word after first pass reading
Total fixation duration	The total time spent reading a target word (a sum of gaze duration and second pass duration)

Table 7.1 Early and later processing measures (from Juhasz & Pollatsek, 2011, p. 874)

As mentioned in Section 7.2, TPR studies on word translation entropy and syntactic choice entropy have used eye-movement data to shed light on the impact of entropy on translation behaviour as well as different aspects of cognitive processes, and in this regard, the study in Schaeffer et al. (2016) is particularly worth mention in terms of the early and later processing measures mentioned above. In their study,[47]

[47] The study is based on nine datasets in the CRITT TPR-DB: ACS08, BD08, BD13, BML12, HLR13, KTHJ08, MS12, NJ12, and SG12, consisting of experiments on both from-scratch translation and post-editing, in six languages: German, Danish, Spanish, Estonian, Chinese,

first fixation duration, defined in the same way as above, is used as one of the indicators of early priming processes, and is shown to be "significantly" influenced by both word translation entropy and the word order difference between the ST and TT. This is interpreted as a suggestion of TL-specific aspects playing a role at the earliest stages of ST reading, i.e., it suggests very early co-activation of both SL and TL upon visual encounter with the ST, as well as aspects of lexical and structural priming (for example, words with lower word translation entropy are more likely to prime TT equivalents, and structural priming is existent during early ST reading). While their interpretation is open to criticism, the use of first fixation duration as an indication of early cognitive processes in relation to the word of interest is consistent with Table 7.1 above. In terms of methodology, first fixation duration as an indication of early cognitive processes (under the eye-mind assumption, see Section 7.6.3) is commonly used in eye-tracking studies.

As mentioned above, another indicator of early processing which is equally informative is whether a word is fixated in the first pass. Here, whether a word is fixated at all in the entire experimental session, regardless of first or second or later pass, is perhaps more useful in TPR studies. More specifically, researchers are interested in the *probability* of a fixation on the word in question, i.e., the likelihood that the word is fixated. This probability is analysed in Schaeffer et al.'s study, which is perhaps not dissimilar to the skipping rates above (see Table 4.1), when such rates are used to refer to the entire experimental session rather than to the first pass. For Schaeffer et al. (ibid), reasonably, this is an indication of how predictable a word is or how well it can be pre-processed, as some words are never fixated in the translation process "because they can be guessed from the context and/or they can be pre-processed during the fixation(s) on the prior word" (p. 202).

Similarly, the probability of a regression, which seems another form of "regression rate" in the table above, is used in Schaeffer et al. (ibid), to indicate integration issues in reading for translation. Total fixation duration is, in their study, termed *total reading time* and refers to all fixations on the word irrespective of when the fixations have taken place. It is considered a measure that applies to the very late cognitive processes in translation (which to the authors means "later regeneration of

---

and Hindi. Altogether there are 295 recordings of UAD (see p. 195; see also CRITT website for details of the datasets).

the ST in the TL and … later monitoring processes", p. 205). Some other measures used in Schaeffer et al.'s study are not included in Table 4.1 above, and are specific to the metrics formulated by the CRITT method, e.g., regression path duration.

In the CRITT database, various measures of processing based on eye-movements, as well as on the coordination between the eye and the hand (e.g., eye-key span, referring to the interval between the first fixation of a word and the start of keyboard typing for its translation production), are included, making it very convenient to systematically investigate each stage of the cognitive processes from different angles, statistically verify or reject hypotheses in a rigorous manner, and shed light on the behavioural manifestations of conceptual, theoretical explorations of the translation process.

In the empirical investigations in Chapters 6 and 7, scan-paths, the number of fixations and fixation durations are analysed both qualitatively and quantitatively, but not adopting excessively sophisticated metrics beyond the scope of the CRITT data.

7.6 Eye movements in reading and information processing

Early observations regarding the role of eye movements in reading, which discovered many phenomena that are known today as the basic facts about eye movements, covering such issues as saccadic suppression, saccade latency, and the size of the perceptual span (Rayner, 1998), largely laid the foundation for the following investigations.

Under the influence of the behaviourist movement in experimental psychology, these studies "tended to have a more applied focus" (Rayner, 1998), involving issues such as improving reading by training eye-movement patterns (Rayner, 1978). The studies in an early era were largely confined to education, with an emphasis on applied issues, because the perceptual and cognitive aspects have become "taboo" dur to the popularity of the behaviourist position in psychology (ibid). This resulted in a lack of theory-building research in this period, and more importantly a lack of research that infers cognitive processes from the movements of the eyes. Rayner argues that despite the presence of two "classic" works on reading (Tinker, 1946) and on scene perception

(Buswell, 1935), respectively, the majority of the eye-movement research that appeared during this early period “seems to have focused on eye movements per se”, i.e., on “surface aspects of the task being investigated” (Rayner, 1998, p. 372).

It is since the mid-1970s that eye-movement research has advanced substantially, primarily because of the significant improvements in recording systems which made it considerably easier to obtain eye-movement data, thereby enlarging the amount of data that can be collected and analysed, and which also allowed more accurate measurements of the eye’s movements (ibid). In addition to the works associated with the methods of analysing eye-movement data, the characteristics of various eye-tracking systems, and the so-called *eye-contingent display change* paradigm (e.g., McConkie, 1997; McConkie & Rayner, 1975), perhaps more important is the development of general theories regarding how language is processed in the human mind, which has allowed critical examination of the cognitive processes in reading by analysing eye movements (Rayner, 1998).

7.6.1 Lexical issues influencing eye movements

In cognitive psychology, the mental processes of reading in relation to lexical issues including visual word recognition, for example, have been rigorously investigated leading to various sophisticated models of processing, e.g., the Dual-Route Cascaded model (Coltheart et al., 2001), the family of triangle parallel distributed processing models (e.g., Plaut et al., 1996; Harm & Seidenberg, 2004), the E-Z Reader model (Reichle et al., 1998), etc. Although the general architectures of these models are different from one another, they are largely common in the belief that representation of words in the reader’s mind involves three different levels: orthographic, phonological, and semantic (Juhasz & Pollatsek, 2011), each of which has been shown in experimental studies to influence eye movements. In addition to these properties of words at different levels, frequency of individual words in the language have also been shown to have an impact on the movements of the eyes.

Word frequency

In eye-tracking studies of reading, it is commonly found that more frequent words are processed faster than words that occur less frequently. The eye-movement analyses in this regard have indicated that high-frequency words receive shorter fixation durations (e.g., Inhoff & Rayner, 1986; Rayner & Duffy, 1986; Kennison & Clifton, 1995; Rayner et al., 1996; Rayner & Fischer, 1996; Schilling et al., 1998; Juhasz & Rayner, 2003; Kliegl et al., 2004; Rayner et al., 2004; Juhasz & Rayner, 2006; Kliegl et al., 2006; Yan et al., 2006; Juhasz et al., 2018). This effect from frequency is considered, as Juhasz & Pollatsek (2011, p. 875) succinct describe,

> … a general word recognition variable that is likely to influence access to many levels of a word's representation. How often a reader encounters a given word form can potentially influence how quickly the orthographic or phonological representations are activated once the word form is encountered in text. This, in turn, will affect the speed of access to semantic representations. In fact, the relative frequency of the two meanings of an ambiguous word has a large impact on fixation durations.

The effect of frequency of the meanings of an ambiguous word, as will be shown later, is largely relevant to the explorations in Chapter 6, where entropy can be understood as representing the frequency distribution of the alternative meanings (alternative TT options, in the case of word translation entropy).

In relevant studies, frequency is typically understood as the frequency of occurrence of a unit of analysis in a corpus, where such a unit can be, e.g., a word form, a lemma, a sense, a grammatical category, a lexical bundle, or a syntactic structure, etc. Frequency can also be treated as *current frequency* and *cumulative lifespan frequenc*y, referring respectively to how often the word is encountered as an adult, and how often it has been encountered since the word was initially learned. As the word frequency in corpora "may not adequately measure actual frequency of occurrence for a given individual", an alternative method would be a subjective measure of frequency, i.e., a rating of familiarity, which is "a good index of the actual frequency for a given population of readers" (Juhasz & Pollatsek, 2011 p. 875).

Semantic ambiguity

In addition to the effect of word frequency, orthographic, phonological, morphological and semantic aspects are also shown to influence eye movement in reading (see, e.g., Juhasz & Pollatsek, 2011). In terms of semantic influence, lexical ambiguity has been a fruitful line of research where fixation durations on words are often used in empirical investigation and in the validation of theoretical models.

In this regard, two important variables are often considered regarding lexical ambiguity — the context prior to the ambiguous word (neutral vs. biased) and the relative frequency of the two meanings of the word (balanced vs. biased), primarily focusing on cases where a single word has two different meanings. In a neutral context, both meanings are plausible, while a biased context disambiguates the word in favour of one of its meanings. A balanced ambiguous word means the meanings of the word are equally frequent, and a biased ambiguous word has a dominant (i.e., more frequent) meaning and a subordinate (i.e., less frequent) meaning.

Relevant studies (e.g., Rayner & Duffy, 1986; Duffy et al., 1988) have shown that in a neutral context, fixation durations tend to be longer on balanced ambiguous words than words with single meanings (i.e., unambiguous words), but biased ambiguous words result in approximately the same fixation durations as unambiguous words. The former can be explained as arising from the competition between the two meanings (because of the balance in their relative frequency), while the latter indicates that the dominant meaning rather than the subordinate meaning is accessed when prior context is neutral (Juhasz & Pollatsek, 2011, p. 883).

When the context favours the subordinate meaning, however, the fixation durations on biased ambiguous words become significantly longer than unambiguous words, while balanced ambiguous words are, to the contrary, processed faster than when the context is neutral (ibid). This means that for balanced ambiguous words, the bias in context facilitates disambiguation, while for the biased ambiguous words, the fact that prior context favours a less frequently-used meaning increases the effort in cognitive processing.

These aspects of relative frequency of alternative meanings of an ambiguous word, as well as the effect from context, are also somewhat reflected in the conceptual explorations in Chapter 6, where the case of ambiguity is enxtended to more than two alternative meanings/translations, using probability distribution, surprisal, and entropy to indicate the bias or balance of these alternative options. It can be understood that the relative frequency of the meanings as mentioned above indicates, to a large extent, similar aspects of the surprisal for the activated candidates (see Chapter 6), and that the bias or balance of this relative frequency represents similar aspects of entropy (when it refers to probability distribution). The bias introduced by context (i.e., the change of entropy due to context) is also reflected in the discussion in Chapter 6.

Regarding the studies on the above-mentioned variables (i.e., neutral vs. biased context, and balanced vs. biased ambiguous words), Juhasz & Pollatsek (2011 p. 883) emphasize that the eye-movement pattern after the word "is striking" — when the prior context is neutral but posterior context biases the subordinate meaning, subsequent disruption in reading would occur, with many more regressions back to the ambiguous word, and this disruption is more prominent for biased ambiguous words. In other words, when the context is neutral, the ambiguous word tends to be interpreted in terms of its dominant meaning, and when posterior context indicates that this meaning is wrong, the misinterpretation needs to be corrected (ibid).

Juhasz & Pollatsek (2011 p. 883) explain such findings in terms of the reordered access model (Duffy et al., 1988; see also Chapter 6), but it seems that these findings can also be interpreted via other models, e.g., the Bilingual Interactive Activation Plus model (BIA+), or structural building framework, etc. (see Chapter 6)

7.6.2 Syntactic issues influencing eye movements

Similar to lexical issues that influence eye movements, various syntactic aspects have also been shown to impact the movement of eyes in reading, including syntactic prediction, syntactic complexity, memory effects, etc. Of particular note is syntactic ambiguity, especially garden-pathing, which is relevant to the investigations in Chapter 7 below.

As mentioned, much attention in the psycholinguistic study of structural ambiguity is given to the issue of parallel processing, i.e., whether a reader entertains multiple interpretations simultaneously, and if so, whether this incurs additional processing cost. Regarding multiple vs. single interpretations of ambiguity, there is considerable controversy and debate theoretically, but "truly convincing evidence is lacking" (Clifton & Staub, 2011, p. 987). However, Clifton & Staub (2011) state that from analyses of eye movements, there is clear evidence about the issue of whether there is increased processing effort when ambiguous syntactic structures are encountered — and this is a very surprising conclusion in sharp contrast to findings for lexical ambiguity — "there does not seem to be an extra processing cost in the ambiguous region when material is syntactically ambiguous", as "there is essentially zero evidence that reading a syntactically ambiguous phrase is slower than reading an unambiguous one, when other factors are controlled" (p. 987).

While the extra processing cost on lexical ambiguity that has been consistently observed from eye-tracking studies (see above) can be reasonably explained as due to the competition between different meanings of the word, when it comes to the syntactic aspect of ambiguity, it seems that the competitive multiple-analysis models may not be convincing, judging from the findings of empirical studies. However, this does not convincingly refute multiple-interpretation perspectives of structural ambiguity (i.e., it does not mean for certain that only one interpretation is entertained at a time), since "it is also possible that multiple analyses are considered in a non-competitive, cost-free manner, perhaps in a 'race' to finding a single acceptable interpretation" (ibid, p. 987).

In addition, eye movements have been particularly useful in investigations into garden-pathing. For example, eye tracking data shows that when there is temporary ambiguity, disruptions of processing occur when material later in the sentence is syntactically (Frazier & Rayner, 1982) or semantically (Rayner et al., 1983) inconsistent with the simplest and most quickly built analysis. The word or region providing disconfirmation of the wrong analysis would correspond to lengthened gaze durations and increased frequency of regressions, and if it is a single word disconfirming the analysis, this disruption is also represented by first fixation on that word.

The garden-path effect of Chinese head-final relative clauses is also demonstrated via eye-tracking data in Lin & Bever (2010), where increased reading time on the relativizer and the head-noun regions is taken as evidence that those clauses are susceptible to garden-path readings.

In the empirical investigations in Chapter 7, a similar perspective is adopted, particularly focusing on the disruption of processing, which is used as evidence in the debate on incremental processing vs. head-driven approach.

7.6.3 Eye-mind hypothesis

Eye movements can reveal much of the cognitive processes, as can be seen from the description in this chapter. However, in most studies based on eye movements, there is an important assumption about the correlation between where one is looking and what one is cognitively processing. This is known as the eye-mind hypothesis (Just & Carpenter, 1980), and originates from a model of cognitive architecture which is used to "simulate and understand the perceptual, cognitive, and motor processes that mediate human performance in a variety of tasks" (Reichle, 2011, p. 768). Just & Carpenter's model (called the Reader model) has a broad theoretical scope, and its core assumptions can be described by two simple principles, namely, the immediacy assumption and the eye-mind assumption.

The eye-mind assumption has commonly been the basis of data interpretation in studies using eye-tracking methods. This assumption hypothesizes that the eyes remain on a word as long as it is being processed (or more broadly, the eyes fixate on the entity with which the mind is engaged). The other assumption, i.e., immediacy assumption, states that "the reader tries to interpret each content word of a text as it is encountered" (Just & Carpenter, 1980, p. 330).

The "extreme" version of Just & Carpenter's theory can assume that "words that are not fixated are not processed" (Underwood et al., 2004, p 154), or that "at any point in time, processing is completely restricted to the word being fixated" (Reichle, 2011, p. 769), with both hypotheses strongly assumed (i.e., eye-mind assumption and immediacy assumption). This would be problematic, as attention and the current fixation location may not be strictly consistent with each other (see, e.g., reading models in Reichle et al., 1998; Reichle, 2011; Salvucci, 2001). The strong versions of both hypotheses are thus questionable.

In Chapter 4, the thesis has mentioned that when the eyes remain on the ST, *both* SL understanding and TL reformulation processes can occur, so the eye-mind hypothesis is suggested to be weakened in TPR (see Section 4.2.2). In reading, the limitations of the eye-mind hypothesis have also been discussed to a considerable extent in Underwood & Everatt (1992).

Underwood et al. (2004), however, are in support of the standpoint that "fixations provide an 'on-line' indication of reading difficulty that also involves moment-to-moment control of the dynamics of reading", and that "when fixating a relatively important part of the field, our eyes will remain stationary for a duration that is indicative of the increased amount of processing that is being performed" (p. 154). This seems to be based on a somewhat weakened version of the eye-mind hypothesis.

Here, it should be noted that the conceptual discussion on entropy in this thesis does not depend on the eye-mind hypothesis, nor is the approximation of entropy (if one wants a value) based on eye-movement data. Entropy itself, as a mathematical concept and viewed in terms of information theory (see Chapter 3), psycholingusitics (see Chapter 5), and translation as described in this thesis (see Chapter 6), are not dependent on the eye-mind hypothesis.

For the approximation of such entropy, specific metrics or operationalized methods would be based on a corpus of text, rather than on eye-tracking records, including the HTra used in Carl et al. (2016) (although HTra does not seem the best way for this approximation and may not accurately reflect the cognitive processes, as mentioned in Chapters 1, 3, and 6).

The conceptual discussion above, in all previous chapters, are therefore not impacted on the controversial issues regarding the eye-mind hypothesis.

In the empirical investigation in the following chapters, however, observations are from eye-movement data, and therefore the eye-mind hypothesis is assumed.

7.7 Summary

This chapter focuses on the methodology of eye-tracking, reviewing some of the ways of measuring eye movements, as well as the metrics used in eye-tracking studies. Importantly, the chapter discusses the lexical and syntactic issues that have been shown to influence eye movements in reading, which will be revisited in the investigations in Chapters 8 and 9 and which forms the basis for the analysis of the empirical observations in those chapters.

Chapter 8 Lexical analysis

8.1 Introduction

In this chapter, empirical analyses are conducted on the CRITT TPR-DB at the lexical level. The chapter starts with an explanation of the way in which analyses are conducted, as well as the dataset used, before proceeding to detailed investigations of the behavioural manifestations of what has been discussed in previous chapters. The empirical analyses here are based on eye-movements, and the approximation of entropy via the HTra metric (or perhaps more accurately, the approximation of entropy reduction, see Sections 6.2.3 and 6.2.8). A pattern of HTra decrease in Activity Units (see Section 8.1 for definition) is observed quantitatively in each experimental session of the dataset, and the details of this pattern is then qualitatively explained in detail in view of the scanpaths in eye-movement, via typical examples of HTra decrease (which are selected for illustration purposes). Following such analyses, the chapter proceeds to a discussion of the observed pattern of eye-movements in terms of the conceptual perspectives described in previous chapters.

Activity Units in translation

In TPR, translation behaviour is often analysed in terms of small segments of reading or typing activities. These segments of behaviour tend to reflect a cognitive definition of a crucial and much-debated concept in translation studies — the "unit of translation" (Alves & Vale, 2009; Barkhudarov, 1993; Bennett, 1994; Carl & Kay, 2011; Kondo, 2007; Malmkjær, 2006; Rabadán, 1991; Swadesh, 1960).

Since the process of translating involves a behavioural pattern where "sudden bursts of production are followed by shorter or longer intervals with no typing activity while the source text (ST) is scrutinized" (Jakobsen, 2019, p. 71), these intervals (i.e., the pauses between typing bursts) have been regarded as indicators of the boundaries between different production units (Dragsted, 2010) and the cognitive processes concerning the change of attentional state (Schilperoord, 1996).

In this regard, empirical studies have used different approaches to fragment UAD to investigate the translator's cognitive effort and cognitive rhythm on the basis of typing pauses and gaze behaviour (see Carl & Schaeffer, 2017b), which include the production unit mentioned here. Another perhaps more detailed fragmentation is attention unit (Hvelplund, 2016), a unit consisting of uninterrupted processing activity allocated to either the ST or the TT, or to the ST with concurrent typing. Similarly, in the CRITT TPR-DB (Carl, Schaeffer, et al., 2016), this is represented by "Activity Units" which are categorized into the following types:

Type 1: ST reading.
Type 2: TT reading.
Type 4: Typing activity.
Type 5: ST reading and typing.
Type 6: TT reading and typing.
Type 8: No activity recorded.[48]

Along the process of translating, transitions between one type of AU and another, accordingly, indicate shifts in activity.

Among all the AU types, four of them involve reading: 1, 2, 5, and 6, as can be seen above. This means that these AUs contain corresponding scanpaths where each individual word fixated is associated with an HTra value. These values can be used to calculate the mean of all fixated words to represent the HTra of the AU.

As the AU types 4 and 8 do not contain fixation data, this chapter is focused on AU types 1, 2, 5, and 6.

An example of how these AU types are categorized is illustrated in the following progression graph (picture taken from Schaeffer, Carl, et al., 2016).

[48] Note that there is no type 3 or 7 in this categorization. In an earlier version of the database, there is a type 7: translation typing while reading the source and the target text (see Carl et al., 2016, p. 39). The numbers for these types of AUs are designated in such a way that those for concurrent reading and typing behaviours (AU types 5, 6, and 7) reflect the combination of type 4 (i.e., translation typing) and a type of reading either on the source (type 1) or the target text (type 2). For example, AU type 5 is a combination of typing and ST reading, therefore the number 5 reflects the sum of 1 and 4.

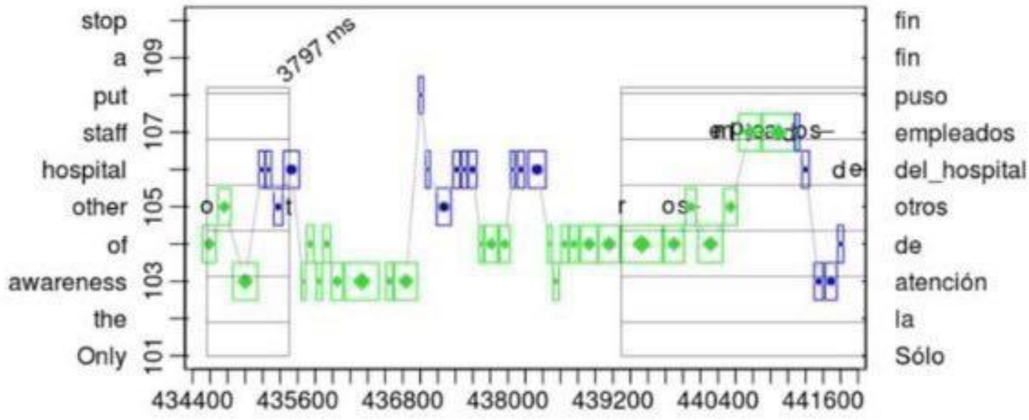


Figure 8.1 Illustration of AUs within and between typing bursts (taken from Schaeffer, Carl, et al., 2016)

Here, the ST tokens are presented on the left side from the bottom to the top in sequential order, with the aligned TT tokens on the right side corresponding to the order of the ST. The horizontal axis indicates time, in milliseconds, during the experimental session. The blue dots indicate fixations on the ST, while the green diamonds refer to fixations on the TT.

The progression graph in Figure 8.1 shows two typing bursts, "*ot*" and "*ros empleados*", as well as the movement of the eyes in this process. In the first typing burst (i.e., "*ot*"), the eyes fixate on a TT item (i.e., *de*) while typing "*o*", which is categorized as AU Type 6 (TT reading and typing); before typing "*t*", the eyes fixate on another two TT items (AU Type 2 — TT reading) and move to the ST side (AU Type 1 — ST reading). When "*t*" is typed, the eyes fixate on the ST word *hospital* (AU Type 5 — ST reading and typing).

After this typing burst, the eyes move back and forth between the ST and the TT, starting from a few fixations on the target side (AU Type 2 — TT reading), then switching to the source window (AU Type 1 — ST reading), and then coming back to the target words for a shorter period of time (AU Type 2 — TT reading), then to the source (AU Type 1), and then to the target (AU Type 2). While the eyes continue to fixate on the target item *de*, the typing activity resumes (AU Type 6 — TT reading and typing), and this AU comes to an end when the eyes switch to the ST window while typing "*empleados*", marking the start of the subsequent AU which contains concurrent typing and ST reading (AU Type 5).

These AUs within and between typing bursts, which can be analysed via the progression graph in Figure 8.1, are indicative of the cognitive processes when the translator is producing *otros empleados*.

In the following sections, the HTra values of each fixated word within such AUs, as well as the overall HTra for each AU, are discussed to gauge the cognitive processes in the translation of high-HTra (i.e., highly translation-ambiguous) items which tend to cause processing difficulty, on the basis of the CRITT TPR-DB.

8.2 Dataset

While the CRITT TPR-DB keeps being updated with new studies and annotations, for the time being it contains "more than 40 translation (and other text processing) studies"[49] which are recorded with data acquisition software such as Translog, Translog-II and with the CASMACAT workbench, where keystrokes and eye-movements are logged, integrated with the ST and TT tokens, and segmented into various units through systematic processing (including the activity units mentioned above). At an earlier stage,[50] Carl et. al (2016) explain that the studies include translation, post-editing, revision, authoring and copying tasks, involving more than 300 experimental participants, over 500 hours of text production, and over 600,000 translated words in more than ten different languages. Needless to say, the size of the database is remarkably large, and this makes it not only possible, but also convenient, to conduct rigorous statistical analyses on translation behaviour across different translators, tasks, and languages. Other than the processed data (in the format of summary tables), the raw logging files are also available and can be downloaded from the CRITT website.[51] In this chapter, a dataset called multiLing from the CRITT TPR-DB is used to empirically analyse the entropy values within AUs.[52]

---

[49] Quoted from the website of CRITT TPR-DB, as of 14 September 2020: https://sites.google.com/site/centretranslationinnovation/tpr-db

[50] That is, when the database contained 30 studies.

[51] See the "Public Studies" page: https://sites.google.com/site/centretranslationinnovation/tpr-db/public-studies

[52] A detailed description of the multiLing dataset is available here: https://sites.google.com/site/ centretranslationinnovation/tpr-db/public-studies?authuser=0

The multiLing dataset includes multiple studies of different tasks of translation production into various languages from the same English STs and is therefore convenient for comprehensive analysis. This dataset is chosen for analyses also because all experiments are on the same set of source texts (in English). Among the studies in this dataset, ten are used for the present study, incorporating all the languages in multiLing: Arabic, Chinese, Danish, German, Hindi, Japanese, and Spanish (AR19, BML12, ENJA15, KTHJ08, MS12, NJ12, RUC17, SG12, STC17, STML18; see below for details).[53]

Arabic

AR19 from-scratch translating, post-editing, and sight translation

Chinese

RUC17 from-scratch translating and post-editing, 21 MTI students

STC17 from-scratch translation, post-editing, sight translation, 16 MTI students

STML18 sight translation (ST), 9 professional interpreters

MS12 post-editing and editing

Danish

KTHJ08 from-scratch translation, 12 students and 12 professionals

Spanish

BML12 from-scratch translating, post-editing and editing

Japanese

ENJA15 translation dictation, post-editing, translation, participants having different levels of translation experience (between 0 and 22)

Hindi

NJ12 from-scratch translating, post-editing and editing by professional translators

German

SG12 from-scratch translating, post-editing and editing, 31 translation students

---

53 Here, more than one study is on Chinese— RUC17, STC17, STML18. For other languages, there is one study for each language. Among the studies in multiLing, ten are selected because others in multiLing are either monolingual reading/copying experiments on the same ST (e.g., SJM16), or experiments on the parts of ST that do not contain the example of *cough up* (e.g., SPC15). It is also important to note that the CRITT TPR-DB is constantly being updated with new studies, so updated versions of the multiLing dataset may incorporate more suitable studies than these ten studies selected.

For these ten studies, analyses are conducted independently for each of them, rather than with all the experimental sessions merged together. As will be illustrated, the pattern of entropy values is observed for all these parts of the data, although the languages, translation tasks, and participants are different across different studies.

8.3 Entropy and fixations in Activity Units

Based on the discussion in previous chapters, it seems reasonable that a higher level of entropy (in the mind) would correspond to more effort involved and therefore lead to a larger value for fixation-based measurements of cognitive effort. If this entropy is approximated by HTra, some studies seem to have empirically observed a statistical correlation between the two, as mentioned in previous chapters.

For the AUs, one might easily assume that the same correspondence can also be found between entropy and cognitive effort — if more effort is expended in an AU, the corresponding words which are processed in this AU should be more likely associated with high HTra values, so the scanpath would result in a higher level of average HTra. This would mean a visible positive relationship between HTra values and the number of fixations (nFix) within the AU.

Interestingly, in a preliminary study in Wei (2018), three scanpath measures (number of fixations, number of different words fixated, and duration) of all AUs which involve reading activity on either the ST or the TT (i.e., AU Types 1, 2, 5, or 6) are analysed in relation to the corresponding HTra values for these AUs (where the HTra value for an AU is calculated as the mean of all fixated words in the AU; see Section 6.2.9), and the relationship is found not to be a simply positive one. Based on a small dataset in an earlier version of the CRITT TPR-DB,[54] a pattern shown in Figure 8.2 is discovered for all languages (Danish, German, Spanish, Hindi, Japanese, and Chinese), tasks (from-scratch translating, MT post-editing, sight translation, editing,

[54] This includes the following studies in the multiLing dataset: BML12, ENJA15, KTHJ08, NJ12, RUC17, SG12, STC17.
Further description of the multiLing dataset in the CRITT TPR-DB can be found in 6.4 and on this webpage: https://sites.google.com/site/centretranslationinnovation/tpr-db/public-studies?authuser=0

and translation dictation), and AU types (1, 2, 5, and 6) in the data. As can be seen from the plots, this relationship does not follow a simply positive trend; instead, there seems to be a certain point of entropy as a threshold before which the number of fixations (nFix) tends to increase as entropy increases, and after this point, the number of fixations begins to show a trend of decreasing. The AUs with maximum number of fixations correspond not to the maximum of entropy, but to its medium level.

When the outliers are removed by 2.5 standard deviations per subject, the pattern becomes somewhat more apparent, as can be seen from the plot on the right side in Figure 8.2.

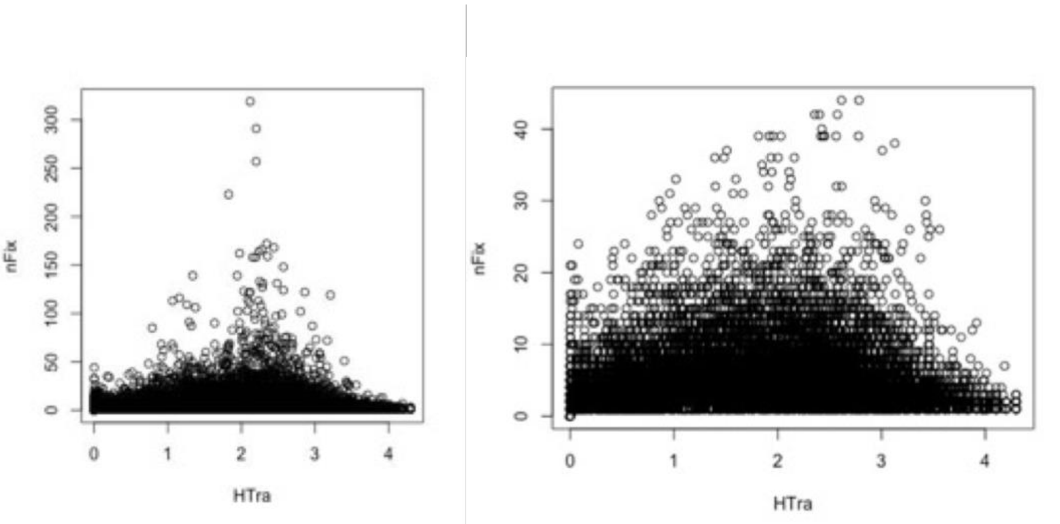


Figure 8.2 Scatter plot of nFix and average HTra
(Left: scatter plot with original data. Right: scatter plot when outliers are removed)

Specifically, for each AU type, this general pattern is also consistent (see Figure 8.3).

As a preliminary study, Wei's (2018) discussion of this pattern is fairly limited, other than illustrating the scatter plots and describing the entropy values in a single example of scanpath (i.e., the scanpath in Section 8.3.3).

In addition to the number of fixations, the same pattern is also found regarding another two features of the AU: the number of different words fixated (DFix) and the duration (Dur) of the AUs. This consistency is in fact not surprising, given that there is a "relative high" correlation among nFix, DFix, and Dur (Schaeffer, Carl, et al., 2016, p. 339):

… the longer a coherent reading activity is (Dur), the more likely it is that more different words are fixated (DFix), resulting in a larger ScSpan. It is also more likely that progressions and regressions occur (higher Turn).

Since the two entropy-based predictors — HTra and HCross — correlate with each other for all the languages in the CRITT TPR-DB (see e.g., Carl et al., 2019), it is not surprising either that this pattern is found to be consistent for both HTra and HCross.

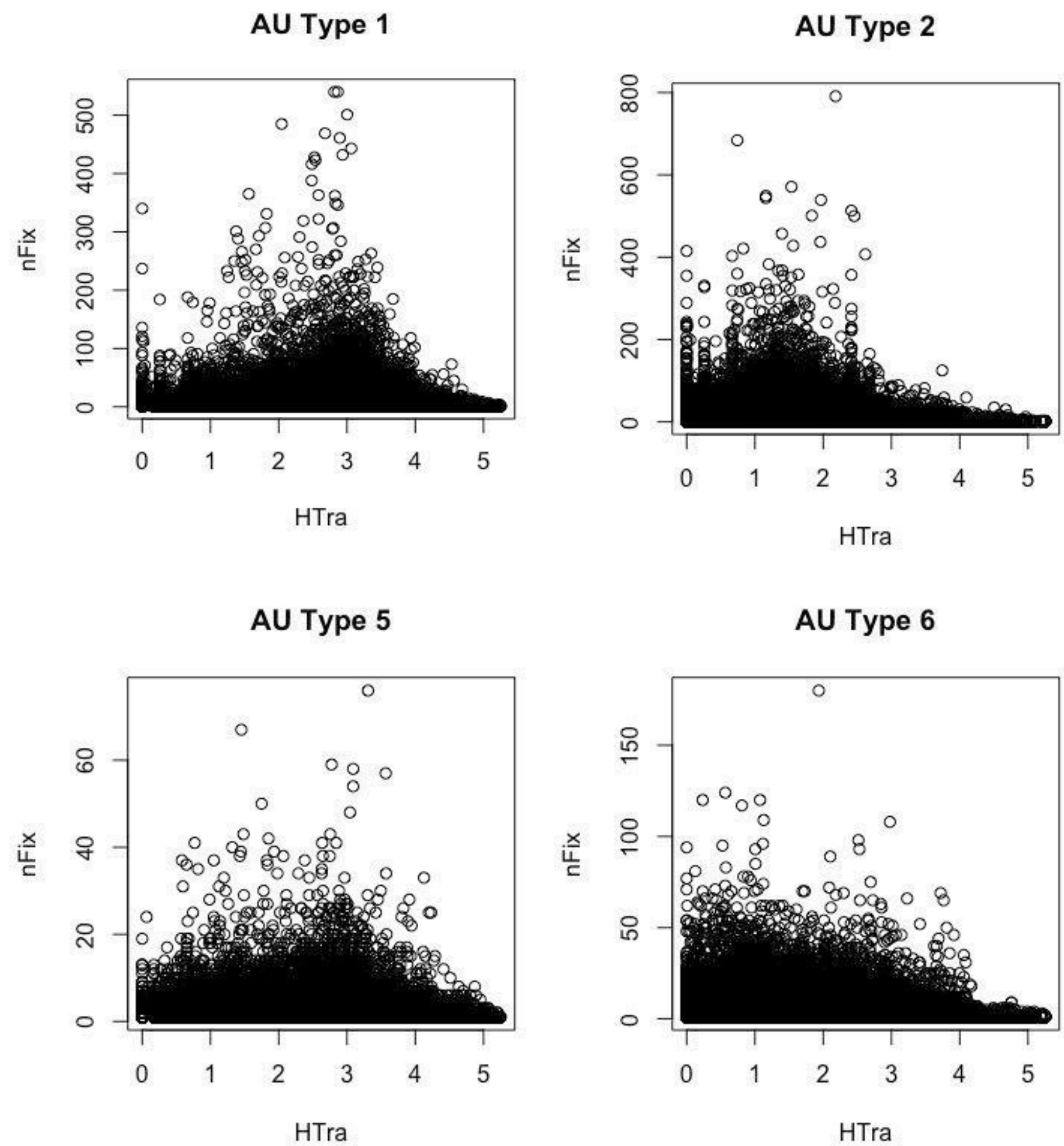


Figure 8.3 Scatter plot for each AU type

8.3.1 Machine Translation Post-editing

As regards the pattern that the AUs with the largest number of fixations correspond to medium-level entropy rather than its maximum, one might be reminded of the findings from Krings' (2001) study on MT post-editing, where the cognitive effort of post-editing is found to reach its highest level not for poor MT output, but for medium-quality translations from the MT (pp. 539-547).

Krings' explanation for this phenomenon in post-editing is that poor MT output causes a strategy shift and a "re-approximation of the post-editing process to a normal translation process" (p. 541). As post-editing MT output of medium-level quality leads to "a disproportionate increase in coordination difficulties among ST-, MT- and TT-related processes", the cognitive effort involved is consequently higher for medium-quality MT output. In contrast, the refocused post-editing process when the output is poor, which approximates to a normal translation process, would require fewer coordination difficulties, and thus less effort, in terms of target text production.

For MT post-editing, if HTra is negatively correlated with MT quality, the above pattern would seem to be another representation of Krings' findings (at a micro level). One might perhaps hypothesize that a higher HTra value for an ST item, which indicates a higher level of translation ambiguity, poses a higher level of difficulty for the MT to produce an appropriate translation, making it more likely to result in errors that need to be post-edited. In this manner, the entropy values would be negatively correlated with MT quality. Then it follows that the instances of poor MT output to which Krings refers would be somewhat equivalent to high entropy values for the ST material, and in this respect, Krings' findings can perhaps be interpreted in terms of entropy as well, in a way which seems consistent with the findings in the above-mentioned study.

This hypothesis regarding HTra and MT output has in fact been supported by empirical evidence in studies of the human translation process and of the errors produced by MT. Translation ambiguity, which reflects the number of possible choices among all TT alternatives, is shown to be correlated with the perplexity of the MT search graph where possible translations are encoded by the engine (Carl & Schaeffer, 2017c), and this perplexity of MT search graph is in turn correlated with post-editing duration (Carl & Schaeffer, 2014). From another perspective, analyses of the types of

MT errors in relation to translation ambiguity (measured by HTra) suggest that "a larger number of translation choices leads to increased (more evident) MT accuracy errors" (Carl & Toledo Báez, 2019, p. 123). In Ogawa et al. (2021), the HTra values calculated from MT output are shown to correlated strongly with the HTra from human translations of the same ST materials. Such findings reveal that translation ambiguity as indicated by HTra, which tends to increase the difficulty for human translation, has a largely similar effect on MT and in turn on the human post-editing process.

Here, as this pattern regarding the relationship between HTra and nFix in AUs is found to be consistent for both the post-editing of MT and translating from scratch (Wei, 2018), it seems that the phenomenon may carry important information that reflects translation or post-editing behaviour in general.

8.3.2 Effect of averaging

It is also important to note that since the HTra of the AU is the mean HTra of the words in the scanpath (see Section 8.1), the pattern as shown in the plots above may also be influenced by the statistical effect of this averaging in the calculation. If this influence is strong enough, then a larger number of fixated words in an AU would mean that this mean HTra value is more likely to approximate the mean HTra for all fixated words in the textual material during the entire task.

Figure 8.4 shows the mean HTra values for all fixations on the ST and on the TT, respectively, which are calculated using the same dataset as in Wei (2018), without excluding outliers (hence inclusion of data points where nFix is excessively large).

It appears that this statistical effect has somehow contributed to the pattern, especially for AU type 2 (TT reading), and especially when data points with excessively large nFix are included (see Figure 8.4). However, since translation behaviours such as eye fixations on the words in the text and the translator's transition between AUs are not randomly occurring phenomena, questions as to what words are fixated in the scanpath, what fixated words fall into which AUs, and what HTra values are associated with these fixated words in corresponding AUs are perhaps far more important than the quantitative pattern of this correlation.

In this regard, the above pattern seems to suggest that when the translator is resolving a problem which arises from a high-HTra item in the text, the high-HTra item is unlikely to be in a long AU where all other items correspond to equally high HTra values: as the AU lengthens, the number of fixations increases together with the AU's duration, and if the fixated items are all at a high level of HTra, the resulting average HTra for the AU would be equally high. This means that the values of both HTra and nFix would be large, which seems rather unlikely based on the plots in Figure 8.2. Instead, the plots show that if the AU is longer, the other items which are fixated in the same AU tend to decrease the average HTra to a (medium) point which may in some way approximate the average HTra for all fixations.

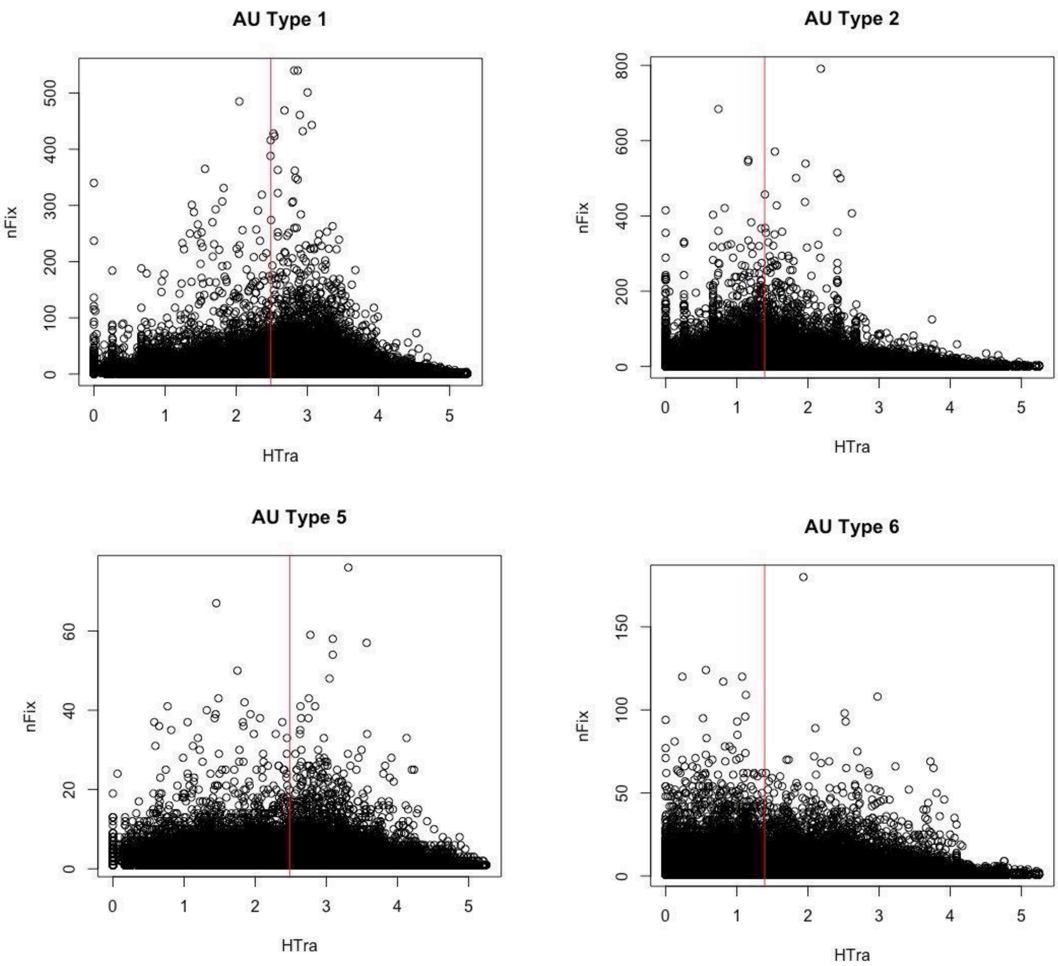


Figure 8.4 Effect of averaging

Mean HTra for all ST fixations: 2.471
Mean HTra for all TT fixations: 1.375

In other words, during the update of the probabilities of the activated (TT) candidates regarding a high-HTra region of the ST (i.e., the decrease of entropy for this region, see Chapter 6), the AU tends to include fixations on relatively low-HTra items and therefore incorporates cognitive processing of the information associated with them. This will be illustrated in a more detailed manner in the following sections, where the scanpath of the translator and the specific entropy values are closely examined for instances where extra processing effort is expended on high-HTra items, using examples of a phrasal verb (*cough up*) which is metaphorical and idiomatic in the ST.

8.3.3 An example of scanpath

The following is an example of ST reading where the scanpath shows expenditure of extra processing effort and where the mean HTra value is at a medium level.

Example 1

As can be seen from Scanpath 1.1 below (see Figure 8.5), when the participant reads along the sentence *British families have to cough up an extra £31,300 a year* and encounters the metaphorical phrase *cough up*, his/her eyes fixate on *cough*, remain on this word for a period of time (indicating extra processing effort), then move back to a previous word *families* and then further back to *in* in the preceding title *Families hit with increase in cost of living*, remain fixated at this point, and then move further back to *increase* (i.e., multiple regressions, which also indicate processing effort). It is at this point that the participant begins to move on to the next AU (see also the progression graph in Figure 8.7).

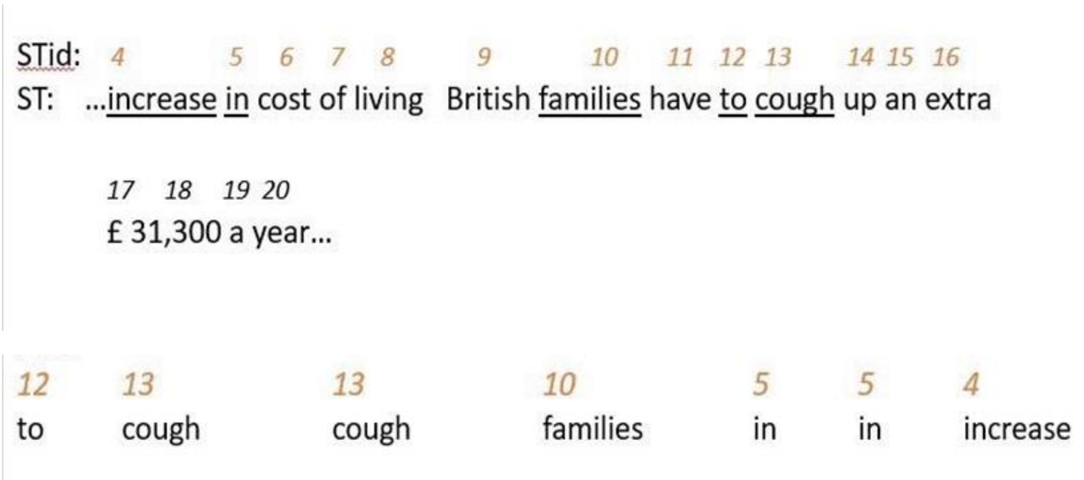


Figure 8.5 Scanpath 1.1 (including STid, ST sentence, and fixated words)

	HTra	HCross
AU 1.1	***2.0357***	1.3734
cough	***3.3755***	2.6082
to	0.7654	1.0144
families	0.2580	0.2580
in	2.0185	0.5132
increase	2.4556	0.4262

Table 8.1 Entropy values for AU 1.1

The HTra and HCross values for this AU (denoted by AU 1.1) are shown at the top of Table 8.1, both of which are at their medium level. The STid in Figure 8.5 is an ID number given to every token in the text, and Scanpath 1.1 (see bottom of Figure 8.5) shows the eye movements of the participant within this unit. HTra and HCross values of each fixated word are also shown in Table 8.1.

Perhaps the initial, lengthened fixation on *cough* in Scanpath 1.1 represents the participant's recognition of a problem regarding the metaphorical use of a word which would otherwise correspond to a different, more commonly used sense and TT equivalent (i.e., a sufficiently unexpected input, see Sections 5.3.1 and 6.2.1), while all the following part of the scanpath contributes to the resolution of the problem where the participant seeks to arrive at a decision regarding the interpretation of the word in face of uncertainty. All the words fixated in this process are somehow related to the metaphor and would help the participant to contextualize and disambiguate the sense of *cough* in *cough up*. At the end of this process, the participant alternates into another type of AU.

Here, what is important is why the entropy of the unit is at a medium level if there is a high level of uncertainty associated with the unexpected input.

Looking at the entropy values of individual words in the scanpath, it is not hard to see that the word which causes the problem (the unexpected input), *cough* (or perhaps more accurately, *cough up*), has relatively high HTra and HCross values (3.3755/2.6082), while the entropy of all the other words on which the participant's eyes fixate are much lower. When the entropies associated with each of these fixations are calculated into a mean, the overall value (2.0357/1.3734) is considerably dragged down by the low-entropy words toward a medium level.

In other words, the scanpath of AU 1.1 is a combination of one word with very high entropy (as the problem) and many low-entropy words (which facilitate problem-solving), resulting in an average at the medium level for the entire unit.[55] It also seems that the process of entropy reduction in the mental states upon encounter of a translation-ambiguous item, as discussed in Section 6.2, might be manifested in behavioural data as an active visual search for low-HTra items which decrease the average HTra of the AU.

8.4 'Cough up': Analysis on activity units

The above analysis should provide a preliminary explanation for the pattern that AUs with larger numbers of fixations tend to correspond to medium-level entropy. However, it is important to examine the translation behaviour pertaining to this example in a detailed and comprehensive manner, to genuinely understand the above-mentioned observation in the scanpath. In the example shown above, what AU does the participant change into after disambiguating the sense in the current scanpath? Does this mark the completion of the activation, disambiguation, and translation selection processes and therefore directly lead to typing (i.e., translation production), or is this only part of the process, so that the participant subsequently switches to the TT for further processing of the disambiguation? How is this behaviour represented in the translation into different languages?

From a broader perspective, it is also important to examine a larger dataset and test if the same pattern in Figure 8.2 exists in the larger dataset regarding more languages and tasks and, more importantly, whether the phenomenon discussed in Section 8.3.3 — that the overall HTra of AUs containing fixations on *cough* is dragged down by lower-HTra items — represents a general pattern in this dataset.

These are important questions, and it is also important to note that the examination of HTra and nFix sheds light on the relationship between uncertainty and

[55] In this regard, it is perhaps interesting, as future work, to analyse the variance (or standard deviation) of these HTra/HCross values, to measure the dispersion of entropy within AUs.

cognitive effort, as well as on the mental processes of disambiguation, lexical selection, and uncertainty resolution.

In order to further analyse the relationship between the number of fixations within AUs and the word translation entropy corresponding to these units, a larger multilingual dataset named "multiLing" in the CRITT TPR-DB is used to investigate the pattern (see Section 8.2).

8.4.1 HTra values for AUs containing "cough"

A look at the HTra of all AUs in which the scanpath contains fixation(s) on this instance of *cough*, and where the total number of fixations in the AU is larger than one, shows that the phenomenon regarding the HTra values of the scanpath in Section 8.3.3 seems to represent a general trend when it comes to the same instance (i.e., *cough*) in the same ST, in view of all participants in each study and in view of different studies on various languages (i.e., the observation is consistent across different tasks, languages, and translation expertise).

Figure 8.6 shows the density plots for such AUs associated with this phrasal verb, in each study and in terms of their HTra values,[56] with the red vertical line in each plot indicating the HTra of *cough* in the corresponding study. As can be seen from the plots, the AUs where *cough* is fixated tend to result in an overall HTra (i.e., average HTra for the fixated words in the scanpath) which is much lower than the HTra value of *cough* itself. This means that the other items which are fixated within the same AUs tend to be lower-HTra words.

[56] Note that these density plots refer to the HTra of the abovementioned AUs, i.e., AUs which contain fixation(s) on *cough* and which contain more than one fixation, rather than the distribution of HTra values of the individual words within specific AUs.

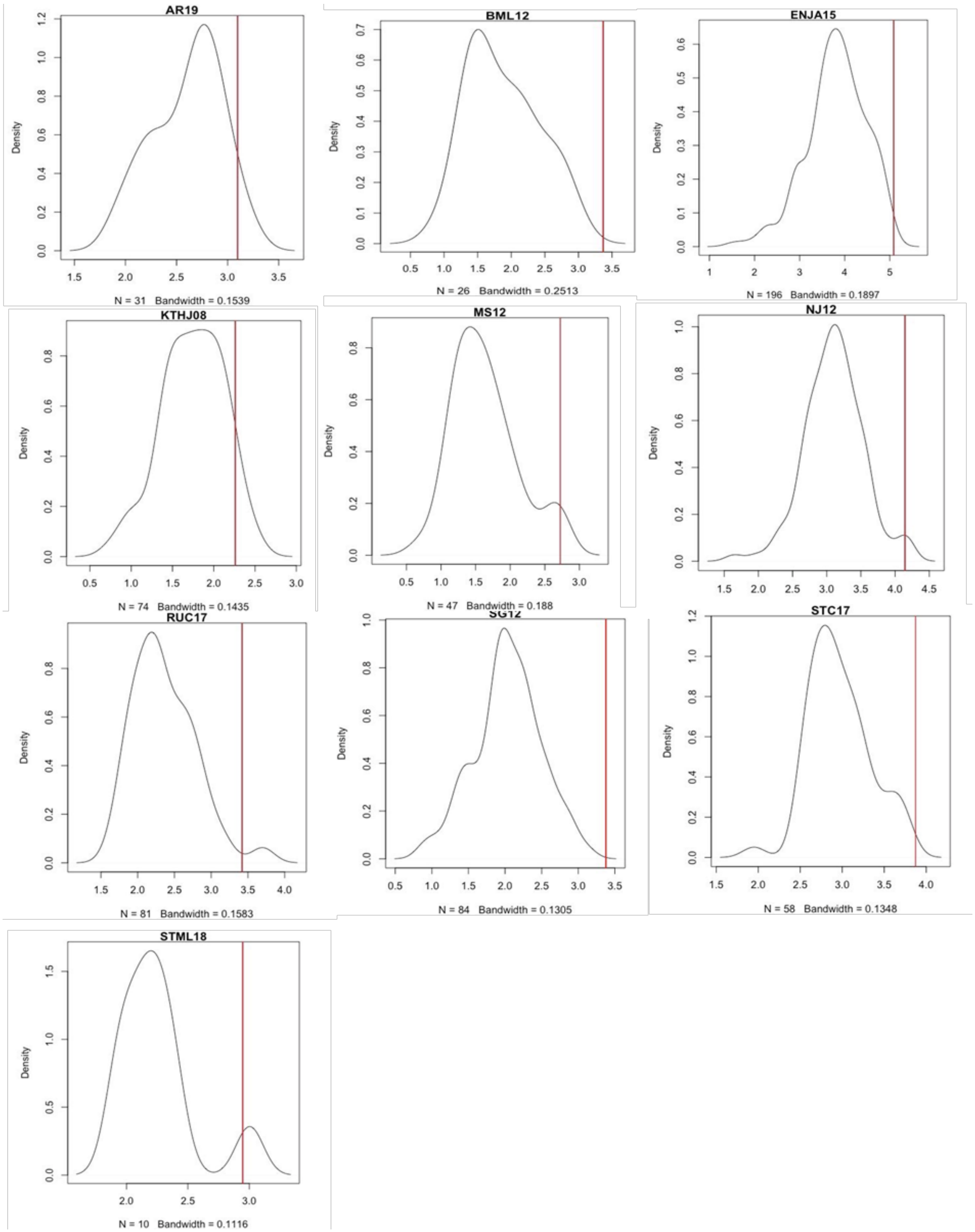


Figure 8.6 HTra values of AUs containing fixation on *cough*. These density plots refer to the HTra of AUs, rather than of words. The red line in each plot indicates the HTra of the word *cough* in the corresponding experimental session. Bandwidth is a real positive number that defines the smoothness of the density plot (in kernel density estimation), with a larger bandwidth indicating a smoother curve.

8.4.2 Progression graph analysis

Regarding the above example of scanpath in Section 8.3.3 (i.e., Scanpath 1.1), a wider scope of the process can be seen in the progression graph below:

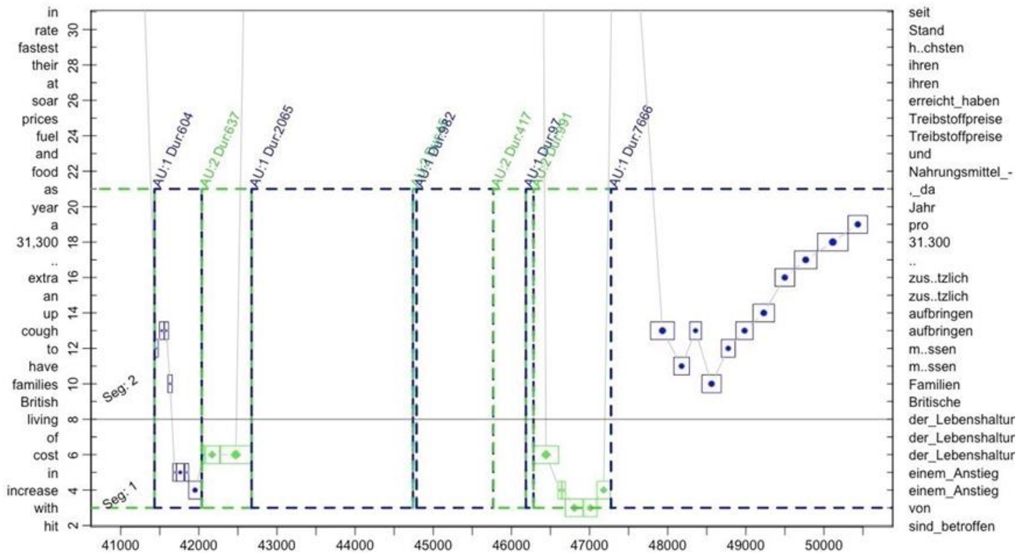


Figure 8.7 Progression graph of Example 1

The task here is MT post-editing, with German as the target language. The scanpath corresponding to the example of ST-reading AU (AU 1.1) in Section 8.3.3 occurs at 41,433 on the horizontal axis (see Figure 8.7).

As can be seen from the graph (starting from 41,433), the participant's eyes fixate on *to*, proceed to *cough*, remain on *cough*, and then move to the preceding words (i.e., the scanpath in AU 1.1 as shown in Section 8.3.3). Those fixations are illustrated with blue dots, and the AU is categorized as Type 1 (i.e., ST reading).

After this encounter of *cough*, as well as the gaze behaviour in AU 1.1, the following AUs seem to be a continuation of the participant's problem-solving process, with fixations on a few other words in the context surrounding the word *cough*, alternating between TT reading and ST reading, before finally coming back to the original word causing the problem (i.e., *cough*) and starting an apparently linear reading process on the ST sentence (i.e., the last AU in Figure 8.7 beginning at 47,275 on the horizontal axis).

In other words, these AUs altogether seem to represent the entire problem-solving process with respect to the issue arising from the metaphorical and idiomatic use of the phrasal verb *cough up* in the ST material.

Specifically, the following AUs are of particular interest:

AU 1.1: ST reading, 41,433–42,038 on horizontal axis (i.e., time)
AU 1.2: TT reading, 42,038–42,675 on horizontal axis
AU 1.3: TT reading, 46,283–47,275 on horizontal axis
AU 1.4: ST reading, 47,275–54,941 on horizontal axis

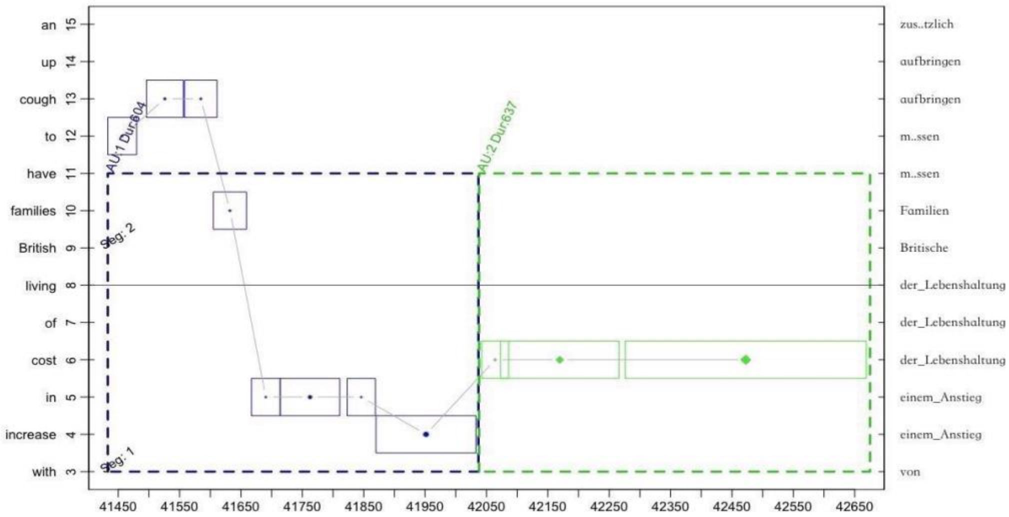


Figure 8.8 Progression graph for AU 1.1 and AU 1.2

Returning to the question raised at the beginning of Section 8.4, Figures 8.7 and 8.8 show that after the scanpath discussed in Section 8.3.3 (i.e., AU 1.1), the participant switches to TT reading on the machine-translated output of a word in the preceding segment, fixating for a relatively long time on *Lebenshaltungskosten*, another word which perhaps helps the participant to contextualize and disambiguate the sense of the metaphor *cough up*.

Then the fixated point is at the very end of the ST for about 4 seconds (see Figure 8.7), which is likely an untargeted gaze while the participant processes the information.

After 46,283 on the horizontal axis (marking the start of AU 1.3, see Figure 8.9), the participant's eyes come back again to the TT token *Lebenshaltungskosten* (‘cost of living’), and then move to a few words in the preceding context, before finally returning to the original position where the problem arises: *cough* in the ST (i.e., the beginning of AU 1.4).

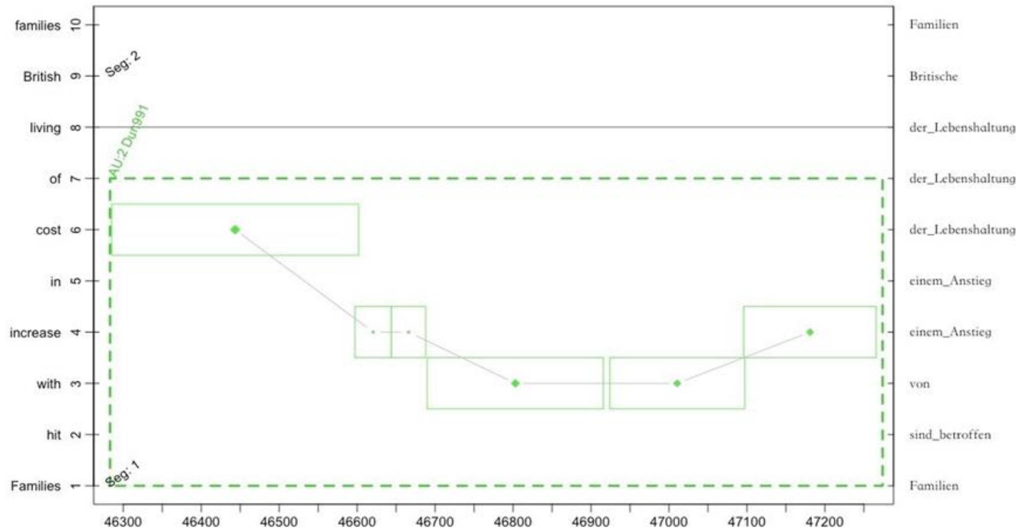


Figure 8.9 Progression graph for AU 1.3

The scanpaths for the two sequences of TT reading (i.e., AU 1.2 and AU 1.3) are shown below, together with the corresponding ST and TT sentences:

(Fixated words are underlined)
ST: Families hit with increase in cost of living
British families have to cough up …

TT: Familien sind von einem Anstieg der Lebenshaltungskosten betroffen
Britische Familien …

Scanpath 1.2: Lebenshaltungskosten -> Lebenshaltungskosten -> Lebenshaltungskosten
Scanpath 1.3: Lebenshaltungskosten -> Anstieg -> Anstieg -> von -> von -> Anstieg

The words fixated in this process, following the ST reading illustrated in Section 8.3.3, seem to have continued to aid the participant to cognitively process the meaning of the word *cough* in the metaphorical phrase *cough up*, and perhaps to disconfirm the alternative interpretations of this word. Therefore, what follows AU 1.3 is continuous reading on the ST, beginning with a gaze pattern similar to the one illustrated in Section 8.3.3, and proceeding with linear reading along the ST sentence (i.e., AU 1.4).

As mentioned, these four AUs constitute the entire process of the participant's recognition of the problem arising from the metaphorical sense of *cough up*, as well as his/her resolution of the problem and verification of the appropriateness of the machine-translated output. It can be seen that the word *cough* has cost considerable effort from the participant to integrate the metaphorical meaning in the context, and under the assumption of non-selective activation of the source and target language (see Sections 6.1 and 6.2), perhaps the encounter of *cough* activates a semantic space which is incompatible with the context and which therefore requires extra processing effort. The long fixation on cough in the initial AU of ST reading (i.e., AU 1.1) signals this, while all the fixations on the other ST words, the alternation between the ST and the TT, the fixation on *cough* again, and the following regression, can perhaps be considered the process in which the probability distribution of the activated items associated with *cough* keeps being updated as the participant attempts to resolve the problem, decreasing the uncertainty to the lowest level and eventually arriving at a choice among the activated items.

Here, what this chapter argues is that this problem, together with the mental processes therein, can be represented and quantified through the lens of entropy (e.g., represented by HTra values which are observed in the text, for word senses and translations), with higher HTra values indicating higher levels of uncertainty pertaining to the words or expressions in question. In the meantime, the resolution of the problem which arises from this uncertainty associated with a particular high-entropy word or expression is achieved through a process which is facilitated by a number of surrounding low-entropy words in the context. This seems to be the behavioural manefistations of the entropy reduction process in the mental states.

In addition to the HTra values of Scanpath 1.1 as shown in Section 8.3.3, the corresponding entropy values for the scanpaths on TT (i.e., AU 1.2 and AU 1.3) are as follows:

	HTra	HCross
Lebenshaltungskosten	0.0435	0.0283
Anstieg	0.6139	0.3239
von	3.0836	2.4349
Scanpath 1.2	0.0435	0.0283
Scanpath 1.3	1.3421	0.9783

Table 8.2 Entropy values for AUs 1.2 and 1.3

For Scanpath 1.2, the entropy for the entire AU is very low, clearly because the fixations are on a very low-entropy word (*Lebenshaltungskosten*). This is not surprising, given that this scanpath, as shown in the analysis above, is part of a larger process of resolving the problem associated with a high-entropy word in the preceding AU.

For Scanpath 1.3, the overall entropy is considerably dragged down by the many fixations on words with very low entropy values, which is consistent with what has been explained in Section 8.3.3.

8.4.3 Translation tasks into different language

As the above example is from a post-editing task, perhaps this discussion merits further analyses on examples of the same problem in translating from scratch. The following are a few examples regarding translation tasks into different languages.

Example 2

Figure 8.10 shows an AU associated with the same phrasal verb, for a different participant translating the same text into German. The AU involves reading on the ST (AU Type 1) during a pause of translation production, where the translator scrutinizes the ST sentence, fixating on *cough* and other items in its context for multiple times, with several regressions toward *cough*, before producing the TT item *müssen* (in *müssen 31,000 £ extra einplanen*).

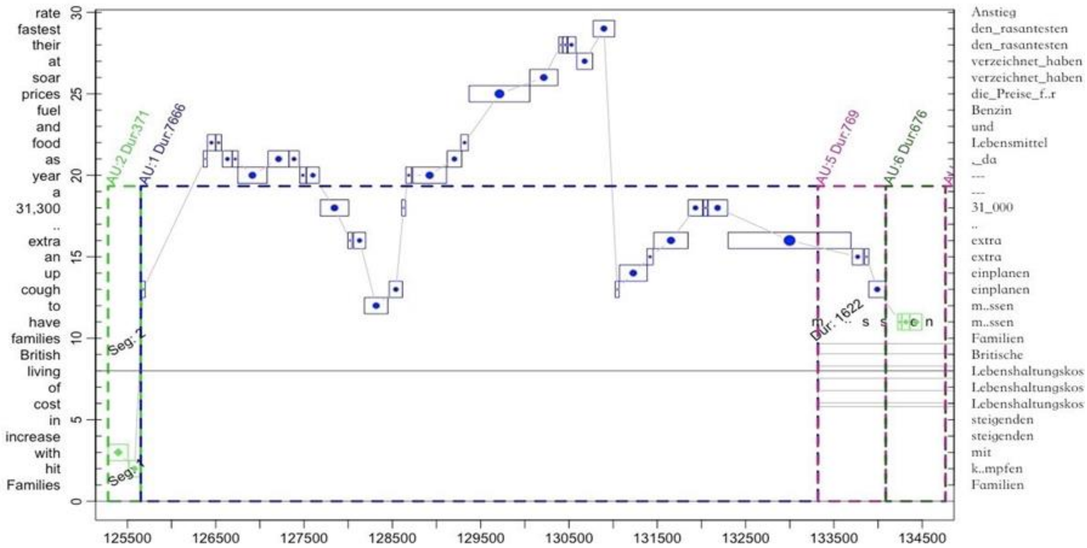


Figure 8.10 Progression graph for Example 2

The ST material, the TT produced by this participant, the translator's scanpath within the AU, and the relevant entropy values in this regard are as follows.

(Fixated words are underlined)

ST: Families hit with increase in cost of living
British families have to cough up an extra £31,300 a year as food and fuel prices soar at their fastest rate in 17 years.

TT: Familien kämpfen mit steigenden Lebenshaltungskosten
Britische Familien müssen 31 000 £ extra einplanen, da die Preise für Lebensmittel und Benzin den rasantesten Anstieg seit 17 Jahren verzeichnet haben.

Scanpath 2: cough -> as -> food -> food -> as -> as -> year -> as -> as -> year -> year -> 31,300 -> extra -> extra -> to -> cough -> 31,300 -> year -> year -> as -> food -> prices -> soar -> their -> their -> their -> at -> fastest -> cough -> up -> an -> extra -> 31,300 -> 31,300 -> 31,300 -> extra

	HTra	HCross
AU 2	***2.5112***	2.2122
cough	***3.3755***	2.6082
as	2.4287	2.0185
food	2.7401	2.599
year	1.0862	1.0862
31,300	1.567	1.3885
extra	2.6914	2.9583
to	0.7654	1.0144
prices	2.4464	1.4708
soar	2.9895	3.2077
their	4.2299	3.2618
at	4.2299	3.5555
fastest	4.2627	3.0994

Table 8.3 Entropy values for AU 2

Similar to the previous example, in this AU, the translator seems to be expending extra processing effort on *cough up*, and in resolving the problem, there is a visual search for a number of items in the context which seem to facilitate disambiguation.

In terms of the entropy values, it is not hard to see that other than five fixations (on three words, namely, *their*, *at*, *fastest*), the vast majority of the fixations in the AU (31 out of 36 fixations) are on words whose HTra values are considerably smaller than *cough*. In other words, the HTra of *cough* is very high, causing the problem and uncertainty, while the rest of the AU tends to be fixations on low-HTra words. This has resulted in the fact that the overall HTra for the unit (2.5112) is lower than that of the word which causes the problem (3.3755). It also shows that the resolution of the uncertainty involved in the disambiguation process is largely facilitated by low-entropy words.

Example 3

Example 3, which illustrates translation into Spanish, is also revealing of the same phenomenon, as can be seen in the following progression graph (Figure 8.11) and the table of the relevant entropy values (Table 8.4).

In this example, it is apparent that the translator is expending extra effort in processing the word *cough* when producing its target translation *asumir*. At 128,265 on the horizontal axis, the translator encounters *cough*, and the typing activity comes to a pause upon completing the production unit of *Las familias británicas*, while the translator scrutinizes the part of the ST containing *cough* before proceeding with a typing burst of *tienen que asumir* (see Figure 8.11).

This pause of typing activity constitutes an AU of ST reading, where the translator fixates on *cough* for a relatively long time (as represented by two consecutive fixations on the same word), searches for other words in the context, re-fixates on *cough*, continues the reading on the following words in the ST, and then engages in a backward eye-movement (i.e., a regression).

Here, the scanpath shows the same behaviour (or behavioural pattern): encountering the problem caused by a high-entropy word and then searching for the low-entropy words surrounding it to facilitate resolution of the problem. In terms of the entropy values, the effect of those low-entropy words results in an average entropy which is at a medium level.

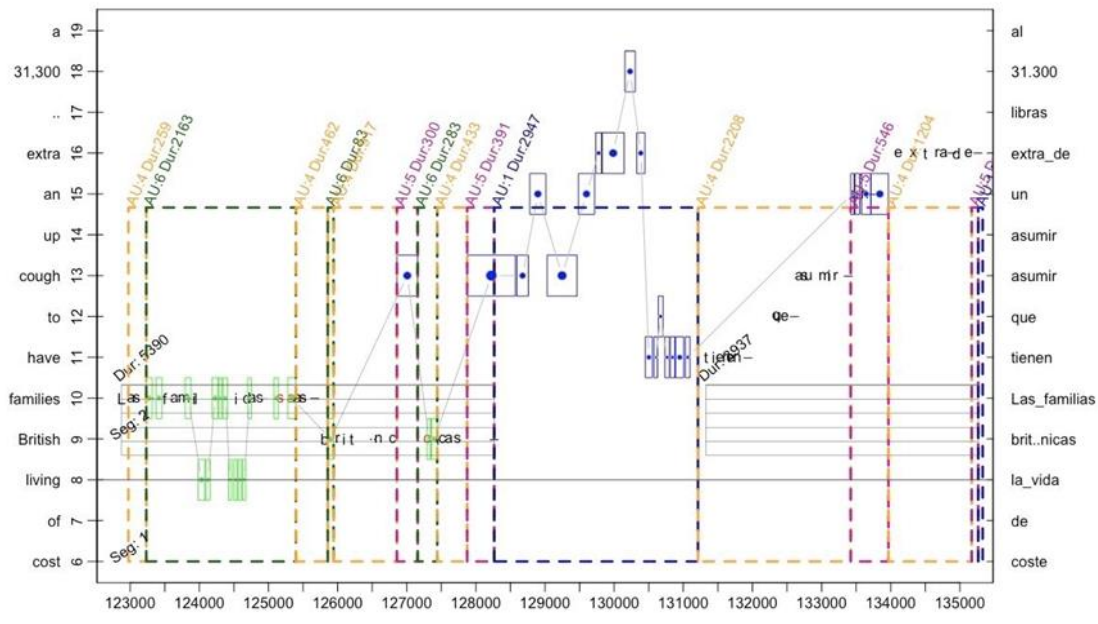


Figure 8.11 Progression graph for Example 3

(Typing bursts are indicated by '///')

ST: British families have to cough up an extra £31,300 a year as …

TT: Las familias británicas /// tienen que asumir /// un extra de 31.300 libras al año …

Scanpath 3: cough -> cough -> an -> cough -> an -> extra -> extra -> 31,300 -> extra -> have -> have -> to-> have-> have -> have -> have

	HTra	HCross
AU 3	***1.8842***	1.1272
cough	***3.373***	1.1055
an	2.0165	1.4071
extra	1.8151	1.8696
31,300	0.8332	1.3039
have	1.3786	0.65
to	1.4453	1.0912

Table 8.4 Entropy values for AU 3

Example 4

Example 4 is from a translation task into Chinese, and the scanpaths of two AUs within pauses of typing show a similar pattern.

Figure 8.12 illustrates the AU (Type 1 — ST reading) between two typing bursts by the translator: "**英国的家庭**"(*British families*) and "**每年**"(*a year*). As can be seen from the progression graph (Figure 8.12), by the time the translator finishes the production of the translation for *British families* (while fixating on the TT), the eyes move back to the ST and encounter the word *cough*. In the meantime, the production of the TT comes to a pause, during which the translator scrutinizes this part of the sentence, his/her eye movements showing a clearly nonlinear reading activity with considerable regression. It is evident that the cognitive effort is elevated at this point, as indicated by the pause of typing and the apparent, multiple regression in the eye movement during this pause (which is also apparent in the previous examples).

At the end of this AU, the translator resumes his/her production of translation while fixating on the TT (see Figure 8.12).

(Typing bursts are indicated by '///')
ST: British families have to cough up an extra £ 31,300 a year as food and fuel prices soar at their fastest rate in 17 years.

TT: 食品和燃料的价格以 17 年中最快的速度增长，导致 /// 英国的家庭 /// 每年 /// 都要额外挤出 31300 英镑的支出。

Scanpath 4.1: cough -> 17 -> £ -> year -> 31,300 -> up -> extra -> extra -> 31,300 -> an

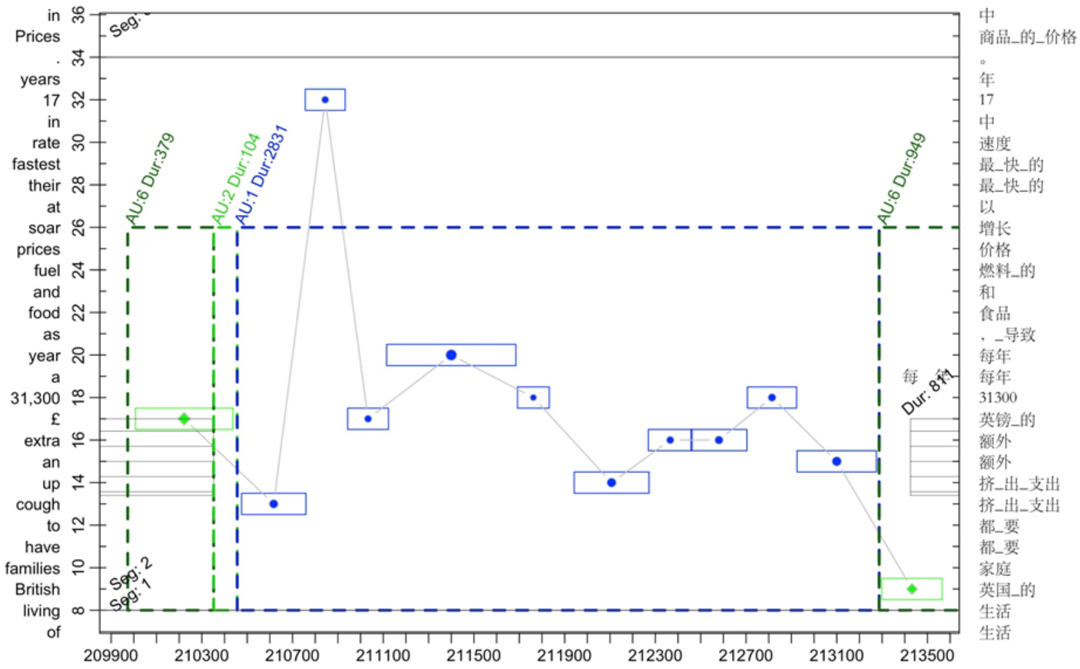


Figure 8.12 Progression graph for AU 4.1

	HTra	HCross
AU 4.1	***2.2962***	2.1219
cough	***3.4226***	2.4255
17	1.8676	2.3935
£	1.5064	1.6017
year	0.549	2.0028
31,300	1.8676	1.5064
up	3.4226	2.4157
extra	2.7539	2.4876
an	2.9511	2.3923

Table 8.5 Entropy values for AU 4.1

Evidently, the HTra value of *cough* (as well as *up*, as part of the phrasal verb) is considerably high compared with all other items which are fixated in this scanpath. The high-HTra item has caused an increase of processing effort, while the scanpath involves other low-HTra items in the context which decrease the overall HTra value of the AU.

This AU is followed by a few typing bursts which are separated by pauses, as can be seen from the progression graph in Figure 8.13. These bursts of typing are ‘每年’ (*a year*), ‘都要’ (*have to*), ‘额外’ (*extra*), and ‘挤出’ (*cough up*).

(Typing bursts are indicated by ‘///’)

ST: British families have to cough up an extra £ 31,300 a year as food and fuel prices soar at their fastest rate in 17 years.

TT: 食品和燃料的价格以 17 年中最快的速度增长，导致英国的家庭 /// 每年 /// 都要 /// 额外 /// 挤出 /// 31300 英镑的支出。

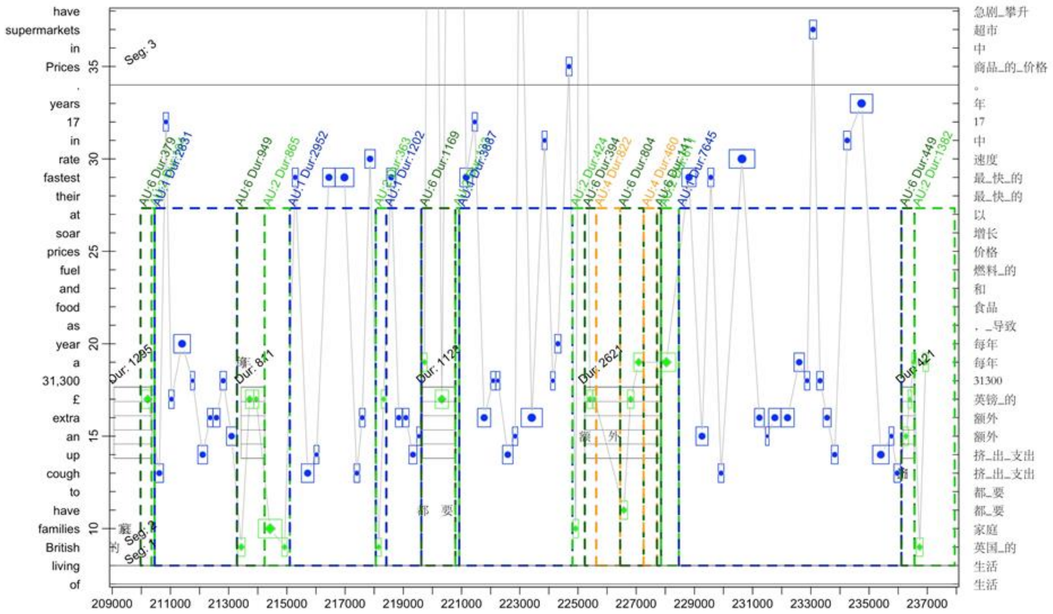


Figure 8.13 Progression graph of the typing bursts after AU 4.1

In this regard, the AU (Type 1 — ST reading) before the typing burst of ‘挤出’ (i.e., the production of the translation for *cough up*) is also particularly relevant to the translator’s processing of the word *cough* in relation to the further steps of word sense disambiguation and — perhaps more importantly — selection of an appropriate TT item among the alternatives which are activated upon encountering this word in AU 4.1 above.

Figure 8.14 shows this AU (AU 4.2) in detail, and similar to what has been illustrated above, there seems to be an extra processing effort involved in the translator’s production of the translation for *cough up*, as indicated by the pause of production and the nonlinear reading activity on the ST, with the eyes frequently moving back and forth in this region of the sentence.

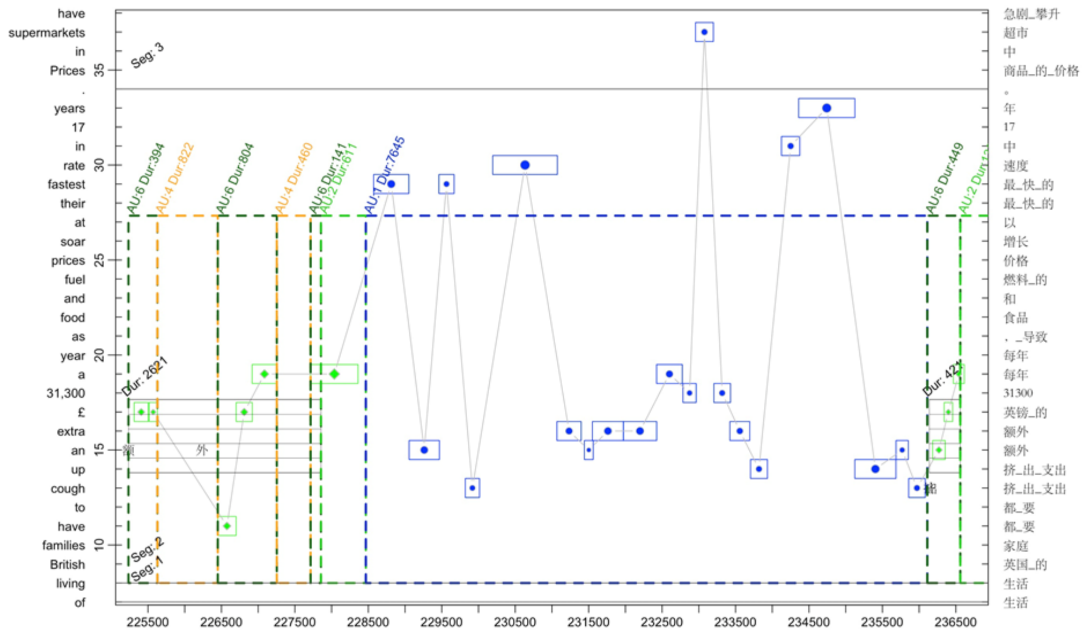


Figure 8.14 Progression graph for AU 4.2

HTra values of the words which are fixated in AU 4.2 display the same pattern as what has been illustrated above (see the scanpath and entropy values below).

ST: British families have to cough up an extra £ 31,300 a year as food and fuel prices soar at their fastest rate in 17 years. Prices in supermarkets have …

TT: 食品和燃料的价格以 17 年中最快的速度增长，导致英国的家庭每年都要额外挤出 31300 英镑的支出。

Scanpath 4.2: fastest -> an -> fastest -> cough -> rate -> extra -> an -> extra -> extra -> a -> 31,300 -> supermarkets -> 31,300 -> extra -> up -> in -> years -> up -> an -> cough

	HTra	HCross
AU 4.2	***2.5603***	2.3389
fastest	2.2485	2.7132
an	2.9511	2.3923
cough	***3.4226***	2.4255
rate	2.485	2.7481
extra	2.7539	2.4876
a	0.549	2.0028

31,300	1.8676	1.5064
supermarkets	0.8181	1.6957
up	3.4226	2.4157
in	3.4585	2.82
years	2.104	2.2624

Table 8.6 Entropy values for AU 4.2

Example 5

In Example 5, the number of fixations (i.e., nFix) in each AU is much smaller than in the examples discussed above, resulting in a less obvious, yet still consistent, pattern of HTra values.

ST: British families have to cough up an extra £ 31,300 a year as...

TT: 由于食物和燃油价格激增，增速为 17 年以来最高，/// 英国家庭 /// 每年需要 /// 额外支付 /// 31, 300 英镑

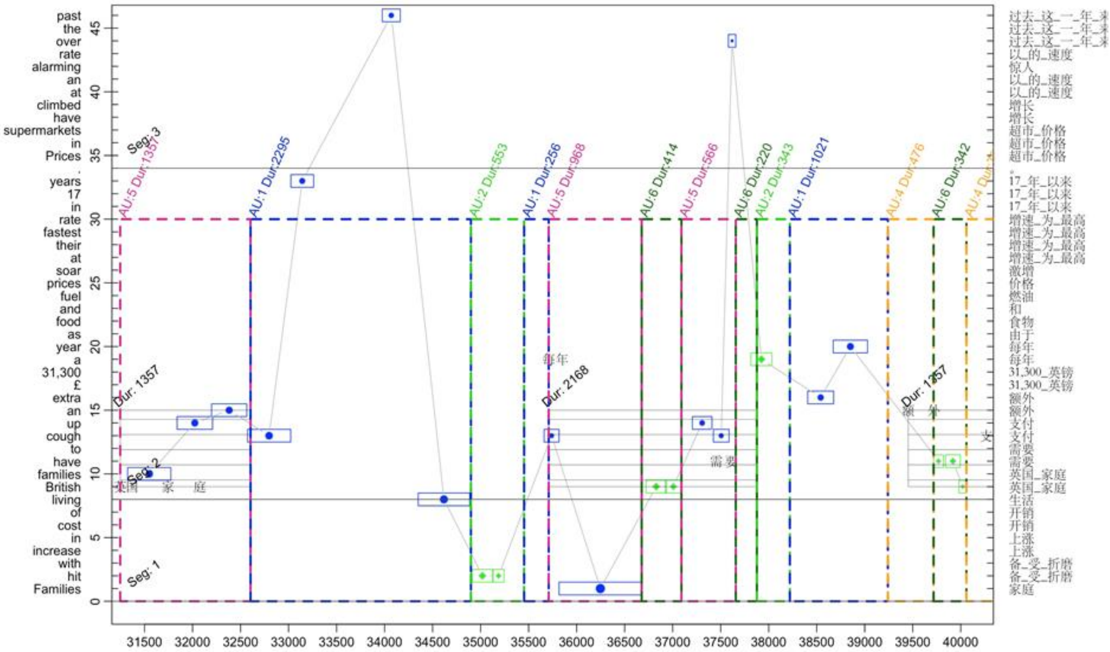


Figure 8.15 Progression graph for Example 5

Starting from the first AU (Type 5, i.e., ST reading and typing) shown in the progression graph in Figure 8.15, the translator produces the TT for *British families* ('英国家庭') while reading the subsequent part of the sentence, and as the eye movement proceeds to the instance of the metaphor (*cough up*), the production of translation is paused. The AU changes into Type 1 (ST reading) where a considerable amount of time is spent in reading the ST at different positions. Then the translator's eyes fixate on the TT for a short time before producing the translation again (see Figure 8.15).

For the first AU at the beginning (AU 5.1 — ST reading and typing), the translator's scanpath and entropy values are as follows:

Scanpath 5.1: families -> up -> an -> cough.

	HTra	HCross
AU 5.1	***3.2817***	2.5464
families	1.7988	2.2169
up	3.875	2.555
an	3.5778	2.8585
cough	***3.875***	2.555

Table 8.7 Entropy values for AU 5.1

Interestingly, a look at the entropy values here shows a phenomenon which is slightly different from the previous pattern. The fixations in this AU are mostly on high-HTra items (*up*, *an*, *cough*), contrary to the examples illustrated above, although the overall HTra for the AU is still smaller than the HTra of *cough*.

This would be understandable if one takes into account the larger context of the unit. As the progression graph and the scanpath both show, this AU is probably a stage where the problem or difficulty arises: the first encounter of the ST item in question, the non-selective activation of a semantic space associated with this item, the translator's subsequent realization that this activation may not be compatible with the context in the ST, and consequently an update in the probability distribution of the activated items where the probabilities seem to have become less concentrated on the (incompatible) ones whose initial probabilities are relatively high. The progression graph indicates that the AU is primarily one which produces the translation of the words prior to *cough* (i.e., *British families*), while the scanpath shows that the fixation on cough occurs at the end of the AU. The translator is already processing the information in the subsequent words (*have to cough up an*) when translating the first phrase *(British families*), and upon encountering the problem *cough*, the typing comes to a pause and the current AU ends, so that the translator focuses on ST reading to resolve the problem associated with *cough*.

For the subsequent AU (AU 5.2 — ST reading) after the typing has paused, the results are as follows:

Scanpath 5.2: cough -> years -> past -> living.

	HTra	HCross
AU 5.2	***2.8870***	2.6442
cough	***3.875***	2.555
years	2.7718	2.6468
past	3.2744	3.5
living	1.6266	1.875

Table 8.8 Entropy values for AU 5.2

Here, the HTra for *cough* is still the highest among all the items fixated in the AU, but the result seems to be somewhat less obvious in terms of the pattern discovered above, i.e., that all the other words in the AU would be associated with very low entropy values.

To explain this, perhaps we can return to the plot in Figure 8.2 (which is discussed in Section 8.3). The plot shows that larger numbers of fixation tend to correspond to medium-level entropy, and as the nFix decreases from the peak, the range of possible HTra values corresponding to the nFix would widen visibly (see Figure 8.2).

In Example 5, the number of fixations in the scanpath is 4, which is rather small especially compared with Example 2 (nFix = 36), Example 3 (nFix = 16), and Example 4 (nFix = 10 for AU 4.1, nFix = 20 for AU 4.2). Therefore, the pattern here is much less obvious, although the pattern itself is still consistent with what has been displayed in the other examples.

In the meantime, a smaller nFix in this case seems to indicate that the translator is expending less effort in this AU than in many AUs of the previous examples, perhaps because the metaphor poses less of a problem for him/her. This is also represented by the fact that the translator quickly switches back to typing activity after fixating very briefly on a limited number of lower-entropy words in the ST and on the most recently typed words ('英国家庭') in the TT (see Figure 8.15).

As mentioned, the translator is already processing the subsequent words when producing the translation of *British families*. In this regard, it would perhaps be meaningful to examine the eye-key span in the data and test if longer eye-key spans tend to influence the number of fixations in AUs. Due to the scale of this study, however, the discussion will focus on entropy and fixations.

Although the nFix value is small in this case, a general tendency similar to the previous examples can still be found regarding the HTra values of the fixated words in the AU.

8.4.4 Dynamic change of HTra within the AU

On the basis of Example 5, the following section provides a close examination of the dynamic change of the average HTra values as the scanpath lengthens in each AU. It can be seen from the progression graph that upon encountering the high-HTra word, the translator pauses translation production (with concurrent reading on the ST) and devotes more attention to reading the ST. AU 5.1 shows an increase of average HTra as the eye movement proceeds, while AU 5.2 shows the opposite.

At the beginning of AU 5.1, the translator processes the information associated with *families* (HTra=1.7988) while producing its translation. The average entropy at this moment would be the entropy of the word itself, i.e., ***1.7988***.

Without pausing the production of translation, his/her eyes move on to the following part of the sentence and fixate on up (HTra=3.875), a high-entropy input which suddenly increases the average entropy of the scanpath: (1.7988 + 3.875)/2 = ***2.8369***. The translator maintains this AU, continuing production of the translation while fixating on more words. The next fixation is on an (HTra, 3.5788), which is again a high-entropy input to further increase the average entropy within the AU: (1.7988 + 3.875 + 3.5788)/3 = ***3.0842***. Then the translator proceeds with the reading and typing activities, this time encountering another high-entropy input — cough (HTra=3.875). The average entropy now becomes ***3.2817***, and at this point, the new entropy value seems to be high enough to trigger a breakdown of the unit, with the translator reallocating cognitive resources to resolve the problem arising from the high-entropy input.

Accordingly, the subsequent AU 5.2 displays a decrease of average entropy as the scanpath lengthens, where the translator attempts to resolve the problem by fixating more words in the context.

While pausing the production of translation, the translator continues fixating *cough* (HTra=3.875), and the average entropy for the current AU is ***3.875***. With the following fixations on *years* (HTra=2.7718), *past* (HTra=3.2744), *living* (HTra=1.6266), the average entropy value for the scanpath is incrementally decreased at each step:

(3.875+2.7718)/2=***3.3234***

(3.875+2.7718+3.2744)/3=***3.3071***

(3.875+2.7718+3.2744+1.6266)/4=***2.8870***.

In short, AU 5.1 is a process of incremental increase of the average entropy, while AU 5.2 is a process of decrease in the average entropy. It seems that once the average entropy increases to a certain point, AU 5.1 breaks down, and the translator starts a new AU where effort is expended to decrease the average entropy toward a lower level.

8.5 Discussion

On the basis of entropy as an indication of uncertainty and cognitive load, Chapters 5 and 6 have discussed from a theoretical perspective how entropy and entropy reduction can describe the cognitive activities, as well as quantify the effort, in the translation and post-editing process. The present chapter have demonstrated with empirical evidence that the expenditure of cognitive effort on ST items of high HTra values constitutes AUs where low-entropy words in the context come into play, thereby decreasing the average entropy of the scanpath in the AU. In the mental processes regarding a high-HTra word, linguistic context is used to gather additional information for clarification or disambiguation, and as a consequence, gaze on other words in the surrounding context is observed. It appears that the ambiguity, uncertainty, or unexpectedness involved in particular lexical items effectively directs the translator to those aspects of context which are relevant to rendering the items unambiguous, to integrating the newly activated information into what has already been activated, to enhancing or suppressing certain activation depending on its relevance to or compatibility with the context, to reducing the uncertainty (and entropy) level in the translator's mind, and to finally arriving at a selection among the co-activated alternative translations for an option which suits the ST and TT context. What seems more meaningful in this observation is that, on the one hand, the high-HTra word has led to an apparent pause of production and nonlinear reading activity (see examples above), and on the other, the words which are fixated in the surrounding context in order to facilitate disambiguation (or resolution of the uncertainty resulting from translation ambiguity) tend to be at a lower level of translation ambiguity. This means that the high-HTra word has resulted in extra processing cost, while this additional effort is expended on searching for the contextual information that is provided by the lower-HTra words surrounding the current high-HTra item.

In other words, the cognitive processes regarding the high-entropy item are facilitated by surrounding low-entropy items, and these cognitive activities can be described as a process of entropy reduction. It is also important to note that this entropy, by definition, captures aspects of TL activation during ST reading, with lower-entropy items activating TL information more rapidly (see also Schaeffer, Dragsted, et al, 2016). In this manner, fixations on the above-mentioned low-entropy items seem to rapidly activate TT co-text in the translator's mind during the visual search, while this activated TT co-text facilitates the selection of a suitable TT option for the item being translated. Therefore, the above observations can perhaps be interpreted as a direct and memory-based search in the mental lexicon for a translation option which is linguistically compatible with the TT co-text, rather than as a mental process that is based on conceptual comprehension. From this perspective, the cognitive processing would perhaps be largely horizontal (as defined in de Groot, 1997; see also Chapter 3).

In the meantime, drawing inferences from the formulation of resource-allocation processing difficulty in psycholinguistics, the Chapter 6 has also provided a theoretical analysis of the cognitive processes in which contextual information is integrated and in which the entropy is decreased in the mental states, while disentangling several important conceptual issues of the translation process through the lens of entropy, such as the decrease of entropy and relative entropy, the entropy in the text and the entropy that describes the mental states, the compatibility of the entropy conceptualization with both serial and parallel processing perspectives of lexical disambiguation, and how HTra and surprisal relate to the cognitive effort expended in these mental processes (see Section 6.2). It has also shown that Carl et al.'s (2016) formulation of word translation entropy (approximated by HTra), if used as a direct quantification of cognitive load, is consistent to the decrease of entropy, rather than the relative entropy, between the initial and final mental states during the process of a translation choice.[57] Importantly, the present chapter has pointed out that when an update of entropy occurs as the translator resolves the problem arising from a highly translation-ambiguous item, this update tends to involve input of information from items at lower levels of translation ambiguity. This results in an active visual search for disambiguating clues from lower-entropy items, thereby decreasing the average HTra value of the AU in question. The entropy reduction in the mental activities seems to be manifested in eye fixations on lower-HTra words and a decrease in the average HTra of the AU.

Upon encounter of a highly translation-ambiguous (i.e., high-HTra) word, it is assumed that multiple TT alternatives for this word are activated non-selectively in early priming processes. As the translator attempts to resolve the uncertainty involved in the high-HTra word and to make a selection among the activated TT items, the pattern of activation, which eventually appears as a probability distribution in the produced text, keeps being updated to concentrate on options which are more probable than others, until a particular translation choice is made regarding this highly translation-ambiguous word. This process of change in the pattern results in a decrease of entropy in the mental state, in terms of both the mathematical expressions for the distribution of probabilities (see Section 6.2) and the notion of entropy as disorder in systems theory (see Section 6.2.7). This decrease of entropy in the mind through cognitive processing appears to be manifested in the behavioural data as an observable visual search for surrounding words at much lower levels of HTra, thereby also decreasing the average HTra of the scanpath in the corresponding AU. From a systems theory perspective, the decrease of entropy (i.e., disorder) of a word translation system

[57] In addition, the HTra in the CRITT TPR-DB is calculated from the text sample in an experimental session.

tends to involve other systems which are much less entropic than the current system in question.

This is supported by a general analysis of the average HTra values for all AUs in the multiLing dataset and more importantly by the detailed discussion in Section 8.4 on the AUs concerning a particular ST item which causes additional processing effort. In addition, Section 8.4 also shows that the pattern regarding HTra values of the scanpaths which are described in Sections 8.3.3, 8.4.2, and 8.4.3 is largely representative of the cognitive processing in general pertaining to the same ST problem.

Specifically, the phrasal verb *cough up* as a metaphor is associated with a relatively large value of word translation entropy, reflecting a high degree of uncertainty in the disambiguation (or selection) process and therefore a higher level of cognitive load in disconfirming the disconfirmable interpretations regarding this ST word. Accordingly, the AU which is primarily associated with the initial encounter of this item in Example 5 has a relatively high entropy value.

When the problem or difficulty regarding the high-entropy word in the ST causes additional processing effort (as indicated by pause of production and nonlinear reading activity), the uncertainty-resolution (disambiguation) process tends to include AUs where the scanpath consists of fixations on the high-entropy word itself and a number of surrounding, low-entropy words. The average entropy in the AU, therefore, is decreased as a result of the impact from these low-entropy words. The more fixations there are in the AU, the more likely it is for this pattern of entropy values to emerge. This seems to indicate that resolution of an instance of uncertainty regarding a translation choice, or the disambiguation of senses, depends on less ambiguous words in the context surrounding the ambiguous word in question.

It can also be seen, from the discussion on scanpaths, that regarding the processing of a translation-ambiguous word within an AU, there would be an input with high enough entropy (i.e., *cough* in the examples above) to trigger the expenditure of additional cognitive effort. This input tends to increase the average entropy of the AU, but as cognitive effort begins to be expended in relation to the high-entropy input, the average entropy for the AU tends to decrease (see Section 8.4.4). In the meantime, when the high-entropy input increases the average entropy of the scanpath to a certain point (which might indicate that it is beyond the cognitive capacity of the translator to maintain the relevant AU, as in AU 5.1 in Example 5), the unit breaks down into smaller ones while the translator switches to a different type of AU, re-allocating cognitive resources to resolve the uncertainty involved in the high-entropy input and, thus, the average entropy in the corresponding AU decreases.

The two effects — the increase of average entropy by the translation-ambiguous input and the decrease of average entropy by the translator's expenditure of cognitive effort — seem to result in a balance in the general average HTra of AUs, so that this value tends to be at a medium level as the AU lengthens.

8.6 Summary

In summary, based on the theoretical discussions of entropy in previous chapters, where inferences are drawn from information theory as well as formulations of resource-allocation processing difficulty in psycholinguistics, this chapter has examined empirically the HTra values of AUs, on the basis of the CRITT TPR-DB, to analyse the behavioural manifestations of the conceptualization. In doing so, it also explains, in terms of HTra values, the manner in which co-textual information is integrated in the cognitive processing of highly translation-ambiguous items.

Chapter 9 Structural analysis

9.1 Introduction

In this chapter, the CRITT TPR-DB is used to investigate the post-editing of ambiguous structures in raw MT output. Through a close examination of eye movement and keyboard behaviour in the CRITT TPR-DB, qualitative and quantitative analyses are conducted on the observable disruptions of cognitive processing when structural ambiguity in the target text is introduced by the MT. Results show that, on the one hand, post-editors seem to display a tendency to parse the ambiguous TT sentence in favour of the interpretation which is semantically consistent with the source, and on the other, disruptions of processing pertaining to the garden-path effect tend to occur not in the later part of the sentence (where the wrong parse is disconfirmed), but in the earlier regions where the most quickly-built analysis is semantically inconsistent with the ST. This indicates that when the MT produces structurally ambiguous translation, the source sentence tends to create a strong bias in the post-editor's mental processes of disambiguation. In the meantime, it also appears that in post-editing, structural disambiguation of the TT is largely suppressed as a result of the biasing effect from the ST, therefore the cognitive resources which are allocated to this aspect of the task seem fairly limited.

Specifically, the chapter analyses in detail a Chinese sentence from the MT output where the main verb is followed by head-final nominal structures with a long "*de*" phrase modifying the head noun, using the post-editing experimental data from a dataset called JIN15. This dataset is selected because, on the one hand, it contains the structural ambiguity (involving a head-final relative clause) discussed in the thesis (see Chapter 2), and on the other, the experiment from which this dataset is collected is well documented in detail (in Huang, 2016).

In the following sections, 7.2 and 7.3 describe the dataset (JIN15) used for this analysis, the processing of the UAD in Translog-II regarding fixation computation and gaze-to-word mapping, especially when such processes are offline. This provides the basis for Section 7.4, where issues in the JIN15 dataset are identified, analysed, and corrected.

Section 7.5 provides a linguistic analysis of a structurally ambiguous sentence from the MT output, and analyses the post-edited result (i.e., translation product) from the 18 participants in the JIN15 study. On this basis, the subsequent sections investigate the eye-movement and key-logging behavioural data of the post-editing process, to shed light on the cognitive processing of the structural ambiguity involved in the text. In Section 7.6, a qualitative analysis of the process is conducted via progression graphs and Translog-II replay sessions, while 7.7 provides detailed statistical discussion, as well as hypothesis testing, on the number of fixations and total reading time of three words that are crucial for structural disambiguation.

9.2 Table generation

Section 8.1 in the previous chapter has provided a brief description of the CRITT TPR-DB and mentioned that other than the processed data (in the format of summary tables), the raw logging files are also available. In this chapter, a dataset (JIN15, see 7.2 below) from the CRITT TPR-DB is used to shed light on the cognitive aspects of post-editing when structural ambiguity is introduced by the MT system (i.e., there are structurally ambiguous sentences in the MT output). This dataset is collected in Translog-II, where gaze sample points from the eye tracker are combined with the keyboard logs for systematic analysis.

As can be seen from the following sections, this chapter involves re-processing of the dataset before the subsequent analyses on structural ambiguity, largely because of issues related to gaze-to-word mapping in this dataset (see Section 7.3 and Appendix A). Therefore, it is worth discussing the way in which raw files are processed before the data tables are generated.

9.2.1 Translog-II and gaze-to-word mapping

Translog-II is a piece of software with which the human reading, writing, and translation processes can be recorded and analysed. It produces log files of UAD in XML format, which can be studied either within Translog-II or via external, more sophisticated tools for rigorous statistical analysis (e.g., R programming on the basis of data tables generated in the form of CRITT TPR-DB).[58] Such UAD contains all relevant information on keystrokes and gaze movements (if connected to an eye-tracker) from the participant, as well as the exact time at which, and the spatial location in which, these keystrokes and gaze points occur.

For keystrokes, the data is classified as insertion, deletion, navigation (cursor movements), copy/cut-and-paste, return key, and mouse operations (Carl, 2012). When connected to an eye-tracker, Translog-II also records gaze-sample points,

[58] Detailed documentation of the Translog-II software can be found on the "Translog-II" page of the CRITT website: https://sites.google.com/site/centretranslationinnovation/translog-ii?authuser=0

computes fixations (i.e., clusters of gaze-samples), and maps these fixations to the closest character on the screen (ibid).

The CRITT website provides a succinct description of the steps by which Translog-II processes the gaze sample points that are produced by the eye-tracker:[59]

1. Check whether the gaze is within the source or the target window
2. Compute fixations based on a variant of a dispersion-based algorithm
3. Map gaze points and fixation on closest character on the screen
4. Find alternative mappings in the line below and above

These steps are usually completed in real time as the experiment proceeds, so that the generated .xml file (i.e., the UAD log file) from Translog-II can be used for analysis immediately after the experimental session, although the fixation-to-word mapping (steps 3 and 4) can be re-computed or manually adjusted offline in view of the noise from the gaze-sample points.

However, for Chinese or Japanese, the Input Method Editor (IME) is not compatible with steps 3 and 4, making online mapping difficult. In this respect, the mapping for such experiments would be computed in a second run after each experiment (i.e., in a "replay" session). When offline mapping is selected before the experiment, Translog-II automatically sets the attribute *Win="-2"* in the log file (this is an attribute that indicates whether the gaze is within the source or target window, represented respectively by *Win="1"* and *Win="2"*), and records the UAD as usual. Figure 9.1 shows four such lines in the log file, containing eye gaze points where Win="-2" (taken from the CRITT website).[60]


```
<Eye Time="31" TT="0"  Win="-2" Xl="678" Yl="277" Xr="678" Yr="277" pl="3.38" pr="3.41" Cursor="0" />
<Eye Time="47" TT="16" Win="-2" Xl="664" Yl="278" Xr="664" Yr="278" pl="3.38" pr="3.43" Cursor="0" />
<Eye Time="62" TT="32" Win="-2" Xl="654" Yl="278" Xr="654" Yr="278" pl="3.38" pr="3.43" Cursor="0" />
<Eye Time="78" TT="49" Win="-2" Xl="658" Yl="279" Xr="658" Yr="279" pl="3.39" pr="3.43" Cursor="0" />
```


Figure 9.1 Eye gaze points in Translog-II file (offline mapping)

This log file can then be opened in Translog-II for replay of the experiment, and as *Win="-2"* is set in the file, the software processes the data offline following steps 1, 2, and 3 above. This re-assigns the window attribute (i.e., whether the gaze

[59] From the "Translog-II" page of the CRITT website (see previous footnote for link).
[60] From the "Translog-II" page of the CRITT website.

points are in the source or target window, as indicated by *Win="1"* or *Win="2"*), computes the fixations, and maps the fixations onto the words. For the above gaze points in Figure 9.1, the re-computed result is as follows (taken from the CRITT website):

```
<Fix Time="30" TT="0" Win="1" Block="1" X="678" Y="272" Dur="352" Cursor="402" />
<Eye Time="31" TT="0" Win="1" Xl="678" Yl="277" Xr="678" Yr="277" pl="3.38" pr="3.41" Cursor="405" />
<Eye Time="47" TT="16" Win="1" Xl="664" Yl="278" Xr="664" Yr="278" pl="3.38" pr="3.43" Cursor="403" />
<Eye Time="62" TT="32" Win="1" Xl="654" Yl="278" Xr="654" Yr="278" pl="3.38" pr="3.43" Cursor="402" />
<Eye Time="78" TT="49" Win="1" Xl="658" Yl="279" Xr="658" Yr="279" pl="3.39" pr="3.43" Cursor="402" />
```

Figure 9.2 Result after re-mapping

As can be seen in Figure 9.2, the window attribute is re-assigned (*Win="1"*) for each gaze points, and the fixation is computed and added to the log file (i.e., the first line in the figure). A replay of this new file will show these fixations and gaze-to-word mappings (with fixations indicated by blue circles and the mapped words represented by violet highlights, see Figure 9.3 from the CRITT website, where the top indicates the gaze points before offline mapping, and the bottom shows the result after fixation computation and gaze-to-word mapping).

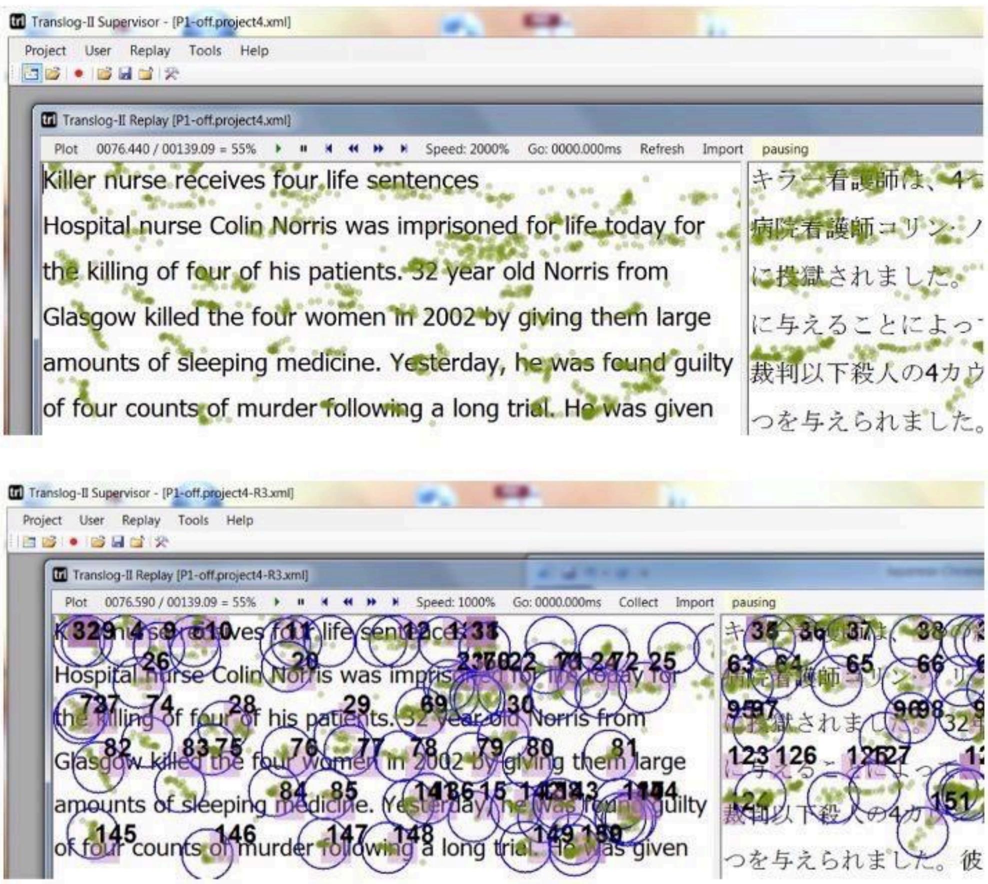


Figure 9.3 Visualization of offline mapping

9.2.2 Generation of summary tables

With the .xml file from Translog-II, the data can be aligned at the word level in terms of the ST and the TT, and further processed to generate TPR-DB tables (also called summary tables) for systematic analysis via statistical means. This generation of data tables can be completed either via the CRITT server (i.e., uploading the raw .xml files within a web browser and then downloading the generated tables), or on a local computer using the perl scripts which are also available from the CRITT website.

For better control of data processing procedures, this thesis uses the perl scripts to generate the tables locally, after re-processing of the raw .xml files in the dataset (see section 9.4).

9.3 Dataset used for analysis

The data used for analysis in this chapter is the "JIN15" dataset in the CRITT TPR-DB (see CRITT@kent, for description of all public studies and their download links), which is collected from Translog-II and Tobii TX300 eye-tracker at a sampling rate of 300 Hz, in a study on the working styles of revision and post-editing (Huang, 2016). The tasks in the study include self-revision (i.e., revision of one's own translation), other-revision (i.e., revision of the translation produced by others), and the post-editing of machine-translated output, from English to Mandarin Chinese (in simplified script). Three comparable texts in English, which are of the same text type (plain text) and length (100 words), are used as ST for the three tasks respectively. For the present chapter, only the post-editing task, as well as the corresponding text for this task, is used for analysis (see Appendix A for the ST and the MT output).

The participants recruited for the study in Huang (2016) are Master of Arts students enrolled in a translation module at Durham University in the United Kingdom, with Chinese as their first language and with "comparable levels of competence" in

English (p. 4).[61] The author describes how 36 participants were initially recruited, but after three rounds of screening "to ensure data quality", the data from 18 participants was selected as "reliable" for analysis (p. 4; pp. 93-118).

For post-editing, a text of 103 words which introduces the University of Warwick is translated using Google Translate. While the author does not mention the details of this translation system (e.g., whether the system uses SMT or NMT as an engine), it can be seen from the label of this dataset in CRITT (i.e., JIN15) that the experiments were conducted in 2015, one year before Google Translate shifted to NMT. Therefore, the text for this data is, presumably, translated from SMT. The post-editing task was conducted in Translog-II, with the ST window on top and the TT window at bottom.

9.4 Data inaccuracy and re-processing

Regarding the JIN15 dataset, there seem to be some issues of concern, primarily in relation to the way in which it is processed. Appendix C provides a detailed description of these issues, the possible reasons behind them, their potential to result in misleading conclusions if not corrected, and the ways in which these issues are dealt with in the present chapter. This dataset is re-processed on the basis of the discussion in Appendix C, and a comparison of the dataset before and after re-processing is also illustrated in Appendix C. With the re-processed data where fixations are re-computed and re-mapped to words, a new set of summary tables are generated using the perl scripts available from CRITT.[62]

The following sections (9.5-9.7) analyse the structural ambiguity involved in the MT output, as well as the behavioural data from the post-editing experiments, on the basis of the tables generated from this re-processed dataset.

---

[61] According to the author (p. 97), all participants who were finally selected for data analysis had IELTS scores of not lower than 7.0, had one year's formal translator training and translation experience, but no post-editing training or experience. Their typing speed was also tested so that the 18 participants had comparable typing skills in Chinese using Sogo (搜狗) IME (ibid).

[62] These scripts, and information about procedures for table generation, are available here: https://sites.google.com/site/centretranslationinnovation/tpr-db/uploading

9.5 Structural ambiguity

In the present section, the linguistic aspects regarding structural ambiguity in the MT output, as well as the post-edited result (i.e., the final translation product after post-editing) from the participants, are discussed. This provides a basis for the following qualitative and statistical analyses on behavioural data (in Sections 9.6 and 9.7 below).

9.5.1 Sentence from MT

(Example 1)

ST: We present our major areas of research strength around key global priorities and challenges currently confronting the world, such as food security and energy.

MT:

我们提出我们主要围绕关键的全球优先事项和当前世界面临的挑战，如粮食安全和能源领域的科研实力。

(Pinyin) *women tichu women zhuyao weirao guanjian de quanqiu youxian shixiang he dangqian shijie mianlin de tiaozhan, ru liangshi anquan he nengyuan lingyu de keyan shili.*

(Back translation of two interpretations)

We propose that we mainly focus on key global priorities and challenges currently confronting the world, such as our research strength in food security and energy.

We present our major areas of research strength around key global priorities and challenges currently confronting the world, such as food security and energy.

In the MT sentence above (see more details in Section 9.5.2), there are two points where a garden path effect is likely to occur regarding two words — A-'提出' (*tichu*, to present / to propose) and B-'围绕' (*weirao*, to surround). The temporary ambiguity here is, on the one hand, directly relevant to the properties of these two words, and on the other hand, also largely the result of the fact that Chinese nominal phrase structures are head-final.

Word A is the main verb of the sentence, and can be followed by either a noun phrase (NP, which would be assigned a theme role from the transitive argument structure), or a sentential complement (an embedded clause which acquires a proposition role from the verb). This is similar to the verb *knew* in *He knew the answer* vs. *He knew the answer was correct*, and to sentences 3, 4, 7, 8, and 9 that are discussed in Chapter 2 (see Sections 2.1 and 2.2). As mentioned in Chapter 2, the structure of the Chinese noun phrase (and perhaps more importantly, relative clause) is head-final, meaning that in this case, any modification of the head noun would be positioned between the main verb (i.e., word A) and the head (see, e.g., sentence 10a in Section 2.1.2). Long and complicated modifications consequently create multiple possibilities regarding the syntactic and semantic relations between word A and the following elements, while it is the head noun, in a very late position of the sentence, that "specifies subcategorizing and adjunctive relations with other elements within the same phrase and gets integrated with linguistic units outside the phrase" (Lin & Bever, 2010, p. 277). In this regard, the modification creates a garden-path effect prior to the head noun if, as believed in incremental processing, arguments are assigned their thematic roles immediately when they are encountered (as argued in, e.g., Section 2.2). The pre-head elements of the noun phrase following the main verb (i.e., following word A) would be analysed preferably as part of the main clause in processing strategies such as minimal attachment, until the head noun disconfirms this interpretation and leads to a re-analysis. If, however, the thematic roles are not assigned until after the head noun is reached (i.e., head-driven approach), multiple possibilities would need to be maintained before the head is arrived at, which means that the temporary ambiguity prior to the head would result in increased processing effort.

Word B here, which occurs in a later region of the sentence than word A, is intrinsically a verb meaning 'to surround', but appears to have also been described as a preposition (meaning 'around') since the beginning of the 20th century, most likely as a result of Western influence on the Chinese language (see, e.g., Peyraube, 1973). The fact that word B can be either a verb or a preposition causes another layer of ambiguity, although some (erroneous) structural interpretations can be disconfirmed by the later part of the sentence. For this word, the temporary ambiguity is also relevant to the head-final structure of Chinese, as prepositional phrases which modify verbs are also in a pre-head position (i.e., before the verb). This means that when word B is encountered, it is unclear whether the word is a main verb of the clause, or a preposition leading the modification of another verb in the later part of the sentence.

In this regard, it is important to note that describing word B as a preposition can be somewhat controversial, and strictly speaking the word seems to be generally considered a verb. However, this does not seem to affect the discussion here, as word B as a verb would be either part of the relative clause for the head noun (i.e., the prepositional use described above), or main verb of the complement clause (i.e., the verb use described above). This creates the same effect as described above.

A compound noun, word C ('科研实力', *keyan shili*, research strength), functions as the head of the NP in one of the interpretations of the sentence (the interpretation which is consistent with the ST). Section 9.5.2 below describes the sentence as well as the different structures possible at each point.

Verb vs. preposition

As mentioned, word B ('围绕', *weirao*) is generally considered a verb in Chinese grammar, but the above description has also somehow considered it a preposition in terms of its grammatical function in the sentence. While this does not affect the analysis below, it is worth making an additional note on this word before proceeding, to accurately describe its linguistic features in this regard.

In linguistics, there seems to be wide disagreement as to whether a particular type of Chinese words, often called co-verbs, should be classified as verbs or prepositions. Although some accounts of the Chinese language consider them explicitly as prepositions (e.g., Li & Thompson, 1974), largely on the basis of the fact that those words frequently have prepositional functions, some other accounts (e.g., Thompson, 1970) argue that such words are simply Chinese transitive verbs which happen to be translated into prepositions in English. Alternatively, more specific accounts in Chinese linguistics tend to classify them depending on individual cases of usage, although not always consistently among different classifications. This controversy is the result of the fact that most Chinese prepositions derive from verbs via grammaticalization (c.f. Jin, 1996), and in the process of this grammaticalization, the verb-derived prepositions are still related to their verbal usage in many ways, making it difficult to distinguish between prepositional and verbal functions (Furukawa, 2008). Not surprisingly, Chinese co-verbs are an example of the words which are, at some point during the evolution of language, "polysemous and functionally ambiguous, simultaneously straddling multiple categories" (Zhu, 2002 p. 1).

In this regard, Fu et al. (1997) point out that under most circumstances, the prepositional functions can actually be differentiated from verbs in a relatively easy manner, and propose that they can be classified on the basis of the following criteria, using X to represent verb/preposition and O to represent the subsequent object (p. 10, see also Furukawa, 2008, for examples on the Chinese co-verb/prepostion '離' li):[63]

1. When there is only one XO in the sentence, X is definitely a verb since a preposition can not be a predicate;

2. When there are two or more XOs in the sentence, the function of the word can be decided by evaluating the sentence after deleting the latter XO: if, after the deletion, the sentence is grammatically valid and semantically unchanged, then the former XO is a verb-object structure; if on the other hand the sentence is not valid, or seemingly valid but semantically changed, then the former XO is a preposition-object structure.

---

[63] Additionally, Furukawa also nicely contrasts the Chinese co-verb 離 (*li*) with the Japanese prepositional particles "から" (*kara*) and "まで" (*made*). It is also worthy of mention that the word 離 here is typed in the form of traditional character, as the paper from Furukawa（古川千春）is published in the traditional Chinese script. The simplified character for this is 离.

While this method is indeed effective under many circumstances, it does not seem very useful for deciding the word class in this case, as word B also involves a '的' (*de*) sequence (see below) making it more complicated. In this chapter, the word '围绕' (*weirao*) which causes additional structural issues other than (but related to) the noun phrase/sentential complement ambiguity, is a transitive verb that might have been, albeit arguably, acquiring prepositional functions in modern Chinese while still maintaining its function as a verb. As mentioned above, most of the prepositions in Chinese are the result of grammaticalization of verbs (e.g., '在' *zai* and '透过' *touguo*), therefore it is possible for such verbs as '围绕' (*weirao*) to acquire some preposition-like usage with its extended meaning. Regarding its word class, however, the present thesis is inclined to considering it intrinsically a verb, even when its grammatical function can be similar to that of a preposition, largely because function words are more or less closed sets, and also because most grammar books and dictionaries tend to mainly consider it a verb.

As a verb, the ambiguity in the sentence can be described in terms of the verb in a pre-head relative clause vs. the main verb in a complement clause, which seems essentially not very different from the above analysis (i.e., preposition vs. verb) for the two interpretations of the sentence.

It is also important to note that in most circumstances, the word can in fact be translated into English as a preposition, and its grammatical function can be very close to that of an English preposition in some aspects. For ease of illustration and discussion, the following description of the sentences would simplify the issue to some extent and adopt the preposition vs. verb distinction for this ambiguity.

9.5.2 Structure of the raw MT output

The Chinese sentence in Example 1 (see Section 9.5.1), if not read together with the ST, is fairly difficult to process because of its structural ambiguity. When it is read alone, perhaps the most straightforward interpretation of the structure would be to regard the second '我们' (*women*, we) — which is immediately after word A, '提出' (*tichu*) — as the beginning of a sentential complement, assigning the proposition role to the following embedded clause:

1a. Sentential complement and preposition (or co-verb)

我们	提出	我们	主要	围绕	…	如…	。	…??
women	*tichu*	*women*	*zhuyao*	*weirao*	...	*ru...*	.	*...??*
We	present/propose	we	mainly	around	…	such as…		…??
[N	V	N	ADV	PREP	NP	PP…].		…VP?

As the processor reaches the later part of the sentence, it would be quick to realize that for this structure, a VP is absent after the PP (with the VP being modified by the PP), which disconfirms the interpretation (see the VP at the end above). Here, this description of the structure classifies word B — '围绕' (*weirao*) — as a preposition. If adopting another (perhaps more widely accepted) perspective regarding Chinese grammar (see previous section), this word would be described as a co-verb[64] rather than a preposition, where, in absence of a following VP to be modified, the structure is then disconfirmed. This is essentially the same scenario.

Other than 1a, a perhaps equally probable alternative interpretation of the sentence is to regard '围绕' (*weirao*) as a verb in the embedded clause, such as the following:

1b. Sentential complement and verb

我们	提出	我们	主要	围绕	…	如…	
women	*tichu*	*women*	*zhuyao*	*weirao*	...	*ru...*	
We	present/propose	we	mainly	surround		…	such as…
[N	V	N	ADV	V		NP	PP]

This seems a plausible interpretation, although not completely consistent with the ST in terms of meaning. (This interpretation can be back-translated as: *We propose that we mainly focus on key global priorities and challenges currently confronting the world, such as our research strength in food security and energy*). Based on the final TT output from the participants, it seems that some of their translations (e.g., from participant P14, see Appendix) are mostly based on this interpretation.

For both 1a and 1b, word B — '围绕' (*werirao*) — is considered as part of the complement clause led by the second '我们' (*women*). This is consistent no matter whether the use of the word as a preposition is recognized (see previous section) — if

[64] That is, a verb that modifies another (following) verb.

it is linguistically classified as purely a verb, then for these two interpretations this verb is part of a complement clause, rather than a relative clause where the head noun is in a later region of the sentence.

In contrast to the sentence complement interpretation, when word A — '提出' (*tichu*) — is analysed as a transitive verb leading a head-final noun phrase (i.e., nominal structure), the interpretation becomes the following, which is consistent with the structure and meaning of the ST (*We present our major areas of research strength around key global priorities and challenges currently confronting the world, such as food security and energy*):

<u>*2a. Noun phrase and preposition (or subject-extracted relative clause)*</u>

我们	提出	我们	主要	围绕	…	的	科研实力
women	*tichu*	*women*	*zhuyao*	*weirao*	...	*de*	*keyanshili*
We	present/propose	our	mainly	around	…	*de*	research strength
[N	V	DET	ADV	PREP	NP	*de*	HN]

Here, considering word B ('围绕', *weirao*) as a preposition, the adjunct to the head noun ('科研实力', *keyan shili*, research strength) is a '*de*' sequence (DeP) which contains a prepositional phrase (PP) and is in a pre-head position (see Figure 9.3).

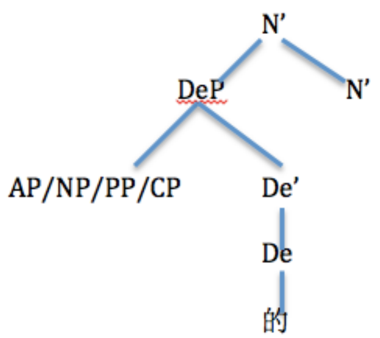


Figure 9.4: DeP as adjunct to the head noun

As the pre-headed DeP in a nominal structure can also contain a relative clause (e.g., '小明提出的建议', 'the suggestion proposed by Xiaoming', involving an object-extracted relative clause which easily creates a superficial NVN sequence if the clause modifies the object of the matrix clause), or an appositive clause (e.g., '人人要自强的论断', 'the statement that everyone needs to be resilient', which also creates a superficial NVN sequence appearing to be argument-complete), perhaps the above sentence from the MT can also be analysed in the following manner with word B being a verb and the DeP containing an appositive clause, although this analysis does not seem semantically appropriate:

2b. Noun phrase and verb

我们	提出	我们	主要	围绕	…	的	科研实力
women	*tichu*	*women*	*zhuyao*	*weirao*	...	*de*	*keyanshili*
We	present/propose	we	mainly	surround	…	*de*	research strength
[N	V	N	ADV	V	NP	*de*	HN].

Again, the linguistic description here has classified word B ('围绕', *weirao*) as a preposition.

If adopting the other perspective (classifying word B as a verb), the nominal structure in 2a would be described as containing a subject-extracted relative clause (with '的', *de*, functioning as a relativizer), and the structure in 2b would still be an appositive clause. From this perspective, there seems to be another possibility — an object-extracted relative clause, which would be easily disconfirmed by the later part of the sentence:

2c. Noun phrase with object-extracted relative clause

我们	提出	我们	主要	围绕	…	的	科研实力
women	*tichu*	*women*	*zhuyao*	*weirao*	...	*de*	*keyanshili*
We	present/propose	we	mainly	surround	…	*de*	research strength
[N	V	N	ADV	V	NP/REL?	*de*	HN].

As the reading proceeds, it would be quick to find that the position where a relativizer (*de*) should occur is now a NP, therefore this interpretation would be disconfirmed at this point.

Similarly, the word '围绕', if analysed as a co-verb, can result in an additional reading:

2d. Noun phrase and co-verb

我们	提出	我们	主要	围绕	…	??	??
women	*tichu*	*women*	*zhuyao*	*weirao*	...	*??*	*??*

We	present/propose	we	mainly	surround	…	??	??
[N	V	N	ADV	Co-verb	NP].	VP?	HN?

This would be disconfirmed only at the end of the sentence, as after the co-verb, it ends without a following VP, a relativizer, and the head noun.

From the above description, it can be seen that the sentence is in itself structurally ambiguous, with interpretations 1b (i.e., sentential complement and verb) and 2a (i.e., head-final noun phrase and preposition) both seeming to be plausible. Between the two, 1b is not only much more fluent than 2a in terms of the overall sentence, but also is an easier, and perhaps immediate parse compared with 2a, given the nature of the head-final structure in the NP. As the pre-head modifications in 2a and 2b are very long and contain multiple '的' (*de*) sequences, it is perhaps safe to assume that when the sentence is processed in isolation, 1a and 1b (i.e., sentential complement readings) would be much more favoured than the other four (i.e., noun phrase readings).

For the sentential complement interpretation (i.e., 1a and 1b), 1a is an immediate parse based on a minimal attachment strategy and is a garden path for 1b. The interpretation would be disconfirmed when the end of the sentence is reached without finding the VP.

For the less-favoured noun phrase interpretation (i.e., 2a, 2b, 2c, and 2d), 2b would be disconfirmed by the relationships between the head noun and the appositive clause, while 2c and 2d would be disconfirmed by the structure of the later part of the sentence as the reading approaches word C. Except for 2a, these are all grammatically invalid structures that are to be disconfirmed in a late region of the sentence.

As mentioned in e.g., Chapter 2, incremental processing and the head-driven approach make different predictions of eye movements in reading, with the former incurring a garden-path effect and predicting disruptions of processing at the point of disconfirmation (see also Chapter 7), while the latter requires extra processing effort in the ambiguous region in order to maintain multiple structures in working memory (see also Chapter 7).

Considering the post-editing process which inevitably receives a contextual input from the ST material, 2a is the interpretation which corresponds to the ST sentence in terms of both structure and meaning. If the ST is provided, it can be assumed that this serves as a bias which leads to the favouring of an interpretation consistent with the ST, i.e., 2a.

The following sections will describe how, during the post-editing process, post-editors seem to still favour the sentential complement interpretation and follow the garden path in 1a. In the meantime, disconfirmation of this garden path seems to be realized not by the later regions of the sentence, but by the ST which is inconsistent with the interpretation. At this point, the post-editor appears to bypass the structural disambiguation or re-analysis of the garden path, but immediately starts to edit the verb and change the TT structure. In other words, it appears that when the input from the ST disconfirms an interpretation, the post-editor tends to disconfirm the translation itself rather than re-analyse the structure for a different interpretation of the MT output.

9.5.3 Structure of the post-edited output

Regarding this sentence, the edited output by the participants falls into the following categories. Again, in the linguistic description below, word B is classified as a preposition (whereas it is perhaps more appropriately classified as a verb or co-verb) when the grammatical function is similar to that of a preposition, for convenience of illustration and discussion.

1. Word B as a preposition which is attached to word A:

a. 我们	围绕…	展现	… 科研实力
women	*weirao…*	*zhanxian*	*… keyan shili*
We	around…	present	… research strength
[N	PP	V	NP]

b. 围绕…	我们	展现	实力
weirao…	*women*	*zhanxian*	*keyan shili*
Around…	we	present	research strength
[PP	N	V	NP]

2. Word B as a verb, with word C as subject (omitting word A, and converting the head-final phrase to a predicate):

我们的　　科研实力　　　围绕…

women de	*keyan shili*	*weirao…*
Our	research strength	surround…
[DET	NP	VP]

3. Merged with the next sentence, but adopting a similar structure to 2:

我们	围绕	…	并…
women	*weirao*	...	*bing…*
We	*surround*	...	*and…*
[N	V	NP	Conjunction]

4. Adopting the original TT syntax for word A (using the sentential complement interpretation), while converting the head-final structure to a clause (i.e. as in 2):

我们	提出	优势领域	是围绕…
women	*tichu*	*youshi lingyu*	*shi weirao…*
We	present/propose	advantageous area	is around…
[N	V	NP	VP]

5. Converting the embedded structure into a paratactic one (where the first chunk contains word A as the verb and word C as the object, and the second chunk is the same as 2):

我们	显示	实力，	这些领域	关乎
women	*xianshi*	*keyan shili,*	*zhexie lingyu*	*guanhu*
We	present	research strength,	these areas	are about
[N	V	NP,	NP	VP]

Regarding the above categories, it is important to mention that they illustrate only the structure of the sentence in the post-editors' final output, and specific lexical choice for A, B and C may vary. For example, the subject in category 2 above may be '领域'" (*lingyu,* area'), '研究' (*yanjiu*, research), or '科研实力' (*keyan shili*, research strength), which is preceded by either '我们' (*women*, our) or other options such as '华威大学的' (*huawei daxue de*, of the University of Warwick). Similarly, the verb here can be '围绕' (*weirao*, to surround) or '涵盖' (*hangai*, to cover/include) in the output from the participants.

It is also worthy of mention that when word B, '围绕' (*weirao*), as a verb, is used together with the particle '着' (*zhe*) to modify another verb, it is considered as Category 1 structure (similar to the prepositional use). If on the other hand '围绕' is a main verb of the clause and followed by the particle (*zhe*) as an aspect marker to indicate an ongoing state, it is considered as Category 2.

A version that adopted excessive re-arrangement of the syntactic elements and that considerably deviates from the ST meaning (i.e., participant P07), as well as another version where there seems to be a lack of verb (participant P22), although appearing to be a category 2 structure, are both marked as 0.

The frequency of these categories are as follows.

Category	0	1	2	3	4	5
Frequency	2	5	8	1	1	1
Proportion	0.111	0.278	0.444	0.056	0.056	0.056

Table 9.1: Frequency of each category of the post-edited result

As can be seen from these frequencies, Category 2 is adopted more frequently than any other structures, and as Categories 3, 4, and 5 are all consistent with Category 2 in terms of the conversion of the head-final nominal structure ('围绕 … 的科研实力', 'around … *de* research strength', research strength around …) into a clause/predicate (e.g., '科研实力围绕 …', 'research strength surrounds …'), in fact the majority of the participants (i.e., 11 out of 18) have adopted such a translation method in handling the head-final noun phrase (involving word B and word C).

It seems that these participants tend to disregard the nominal structure in the TT (i.e., 2a above).

9.6 Qualitative analysis of the process

9.6.1 Keyboard input

To produce these types of translation output, some participants edited the MT output just once, while others edited it two or three times (i.e., had more than one "pass" at the sentence during post-editing), as can be seen from the progression graphs below (Figures 9.5 and 9.6, for participants P15 and P09 respectively; see also Table 9.2 for the number of participants who had each number of "passes" at the sentence). In the graph, the sentence discussed here is marked as Segment 3 ("Seg: 3"). When there are multiple passes on the same segment, the edits made in passes other than the first can be considered as edits on the participant's own version rather than on the raw MT output. As the discussion here is on the structural disambiguation of the raw MT output, it mainly focuses on the eye movement behaviour when the segment is edited for the first time.

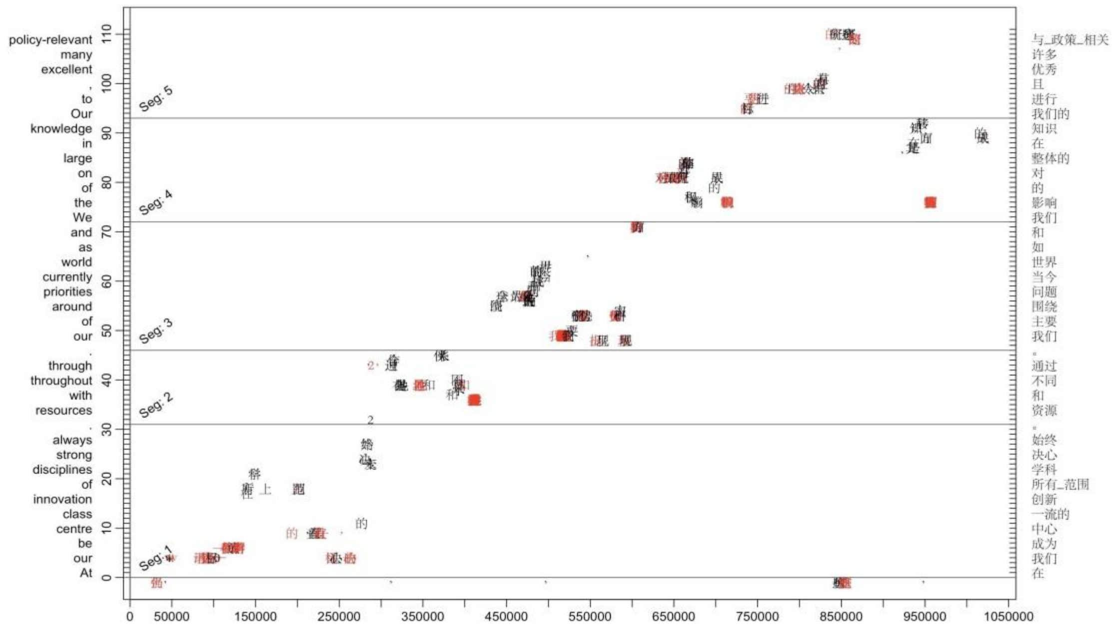


Figure 9.5 Progression graph of a participant who has edited the segment once

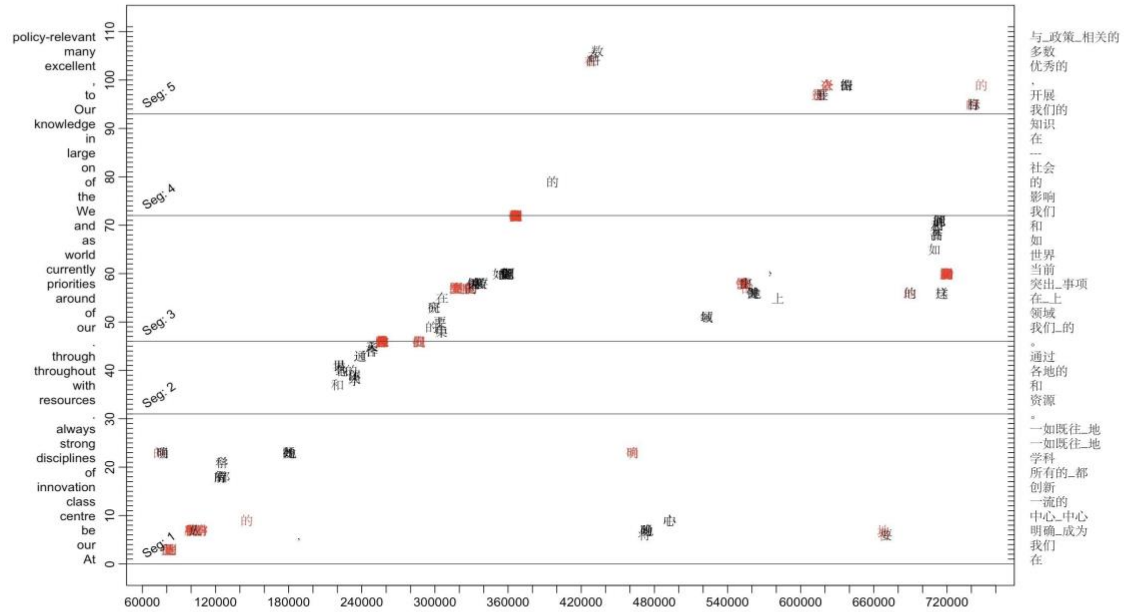


Figure 9.6 Progression graph of a participant who has edited the segment three times

Number of passes	1	2	3	5
Number of participants	4	7	6	1
Proportion	0.222	0.389	0.333	0.056

Table 9.2 Number of participants who had each number of passes for the segment

The following section analyses the eye-movement and key-input behaviour of a participant (P09) who has adopted the structure in Category 2 above. While what is illustrated below in the qualitative analysis is focused on this specific participant, it seems to represent a general pattern in the behavioural data. As mentioned, the participant has edited the sentence three times, but in view of the disambiguation process of the raw MT output, the discussion here is focused on the first time of editing.

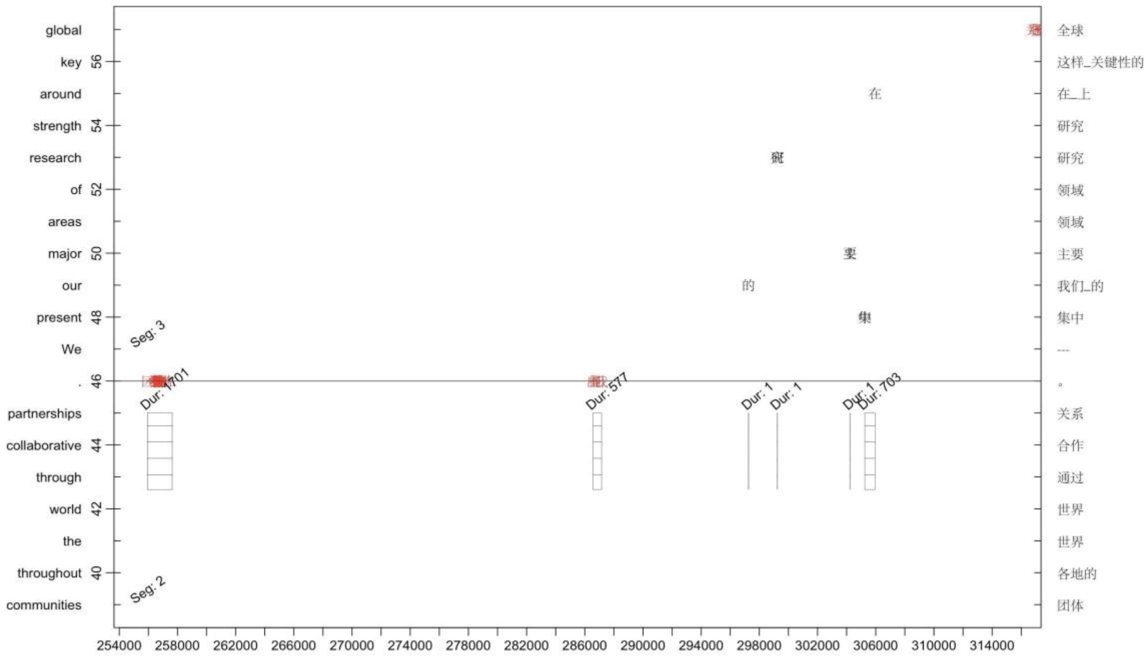


Figure 9.7 Keyboard behaviour of P09 at the start of editing on Sentence 1

The progression graph in Figure 9.7 shows the typing behaviour of the participant when the sentence is edited for the first time. The graph starts with a series of deletions (the deleted sequence is ‘在世界各地建立合作伙伴关系与学术社区’)[65] upon which the editing of the previous sentence (i.e., Segment 2) is completed. Then after a pause, the participant starts editing on the current sentence (i.e., Segment 3) by deleting ‘我们提出’, and proceeds with a few insertions[66] before deleting ‘主要围绕’. These deletions on the current sentence already contain the two critical words in relation to the structural ambiguity (i.e., A ‘提出’ and B ‘围绕’), and combined with the insertions, effectively change the structure of the sentence making other interpretations impossible (see the edited output below).

Edited output regarding Figure 9.7:

我们的	研究	主要	集中在…
women de	*yanjiu*	*zhuyao*	*jizhong zai…*
Our	research	mainly	focuses on…
[DET	NP	ADV	VP]

65 As mentioned previously, deletions are indicated in the progression graph as red characters.

66 These insertions are, respectively, 的, 研究, 主要, 集中, and 在.

As will be described below, the edits here are made by the participant as soon as the earlier regions of the sentence are read, long before fixations occurred on the later parts which facilitate or disconfirm the other interpretations of the structure, and this means that the participant seems to have bypassed any reanalysis or disambiguation regarding the MT output itself, having instead directly disconfirmed, on the basis of the ST, the quickly-built structural analysis (interpretation) of the sentence.

In addition, the key input of the noun '研究' (*yanjiu*, research) here, which corresponds to the original head noun in the MT sentence in a very late region (i.e., word C, '研究实力', *yanjiu shili*, research strength), is inserted into the beginning part of the sentence, and this insertion occurs long before any fixation on word C (see Section 9.6.3).

9.6.2 Eye movements

The eye movements during this process are shown in the following graphs in detail together with the typing bursts of insertion and deletion which are described in Figure 9.7. Figure 9.8.a illustrates the fixations between the two series of deletions — the end of editing on the preceding sentence (sequential deletion of '在世界各地建立合作伙伴关系与学术社区') and the start of editing on the current sentence (deletion of '我们提出'), i.e. 255000-288000 on the horizontal axis. The subsequent fixations together with the input of characters, until '主要围绕' is deleted, are shown in Figure 9.8.b.

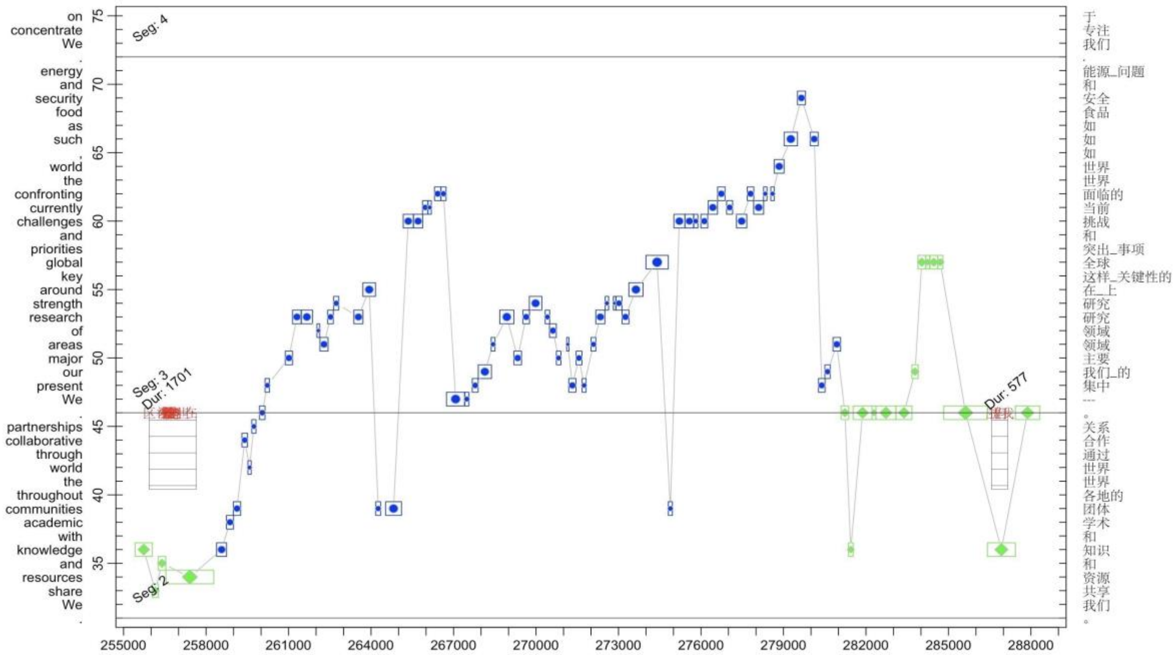


Figure 9.8.a: Progression graph 1 for P09

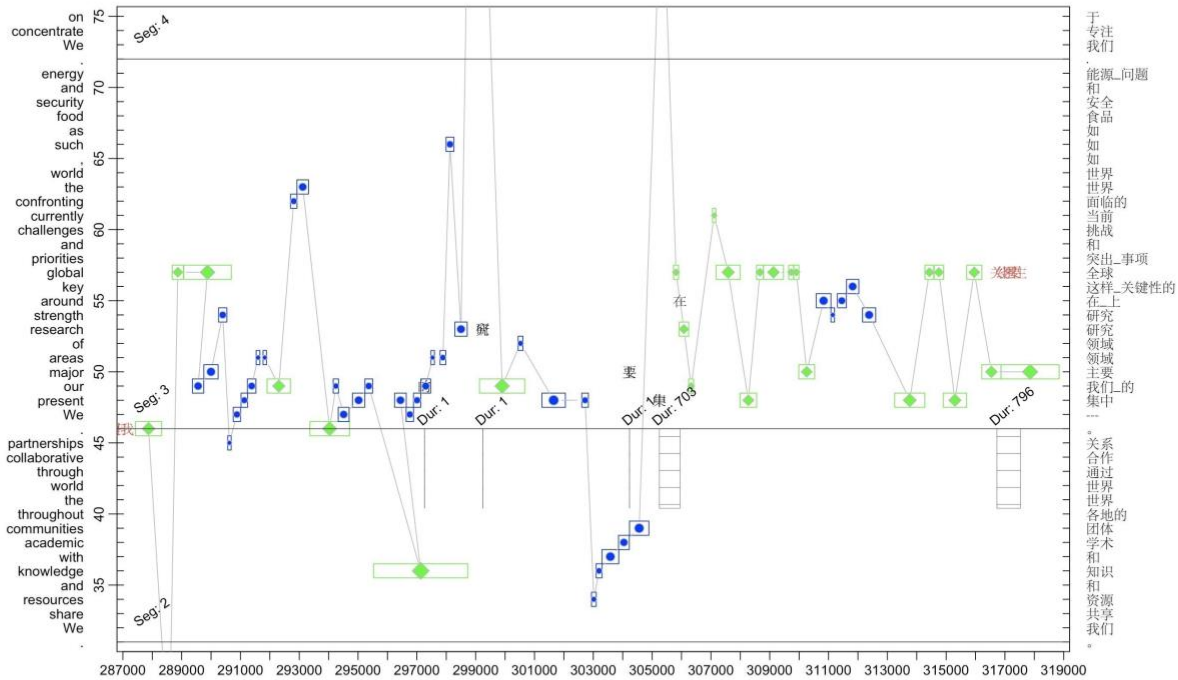


Figure 9.8.b: Progression graph 2 for P09

It can be seen from Figure 9.7.a that after editing the previous segment (deletion, at the beginning of the graph at 25700 on the horizontal axis), the participant reads the ST of the newly-edited segment, proceeds to the reading of the current ST sentence in almost a linear manner, then re-reads the current ST sentence (with a regression from the word *strength* towards *present*, followed by another forward movement between the two, at 27000 on the horizontal axis), and then reads the MT output starting with multiple fixations on the region where four characters are subsequently deleted: '我们提出' (i.e. the typing burst in the graph at approximately 28700 on the horizontal axis). Here, the graph already shows that without reading much of the MT output — especially without fixations on the words associated with disconfirmation of interpretations, the post-editor begins to edit the sentence by deleting the verb (word A), thereby changing its structure completely.

7.6.3 Detailed fixations on TT

As mentioned, in the CRITT TPR-DB, the TT fixations here are mapped to the final edited version of the translation product, therefore the specific details of these fixations on the deleted tokens are not available in the graph above. To illustrate more details of eye movement regarding the TT fixations in Figure 9.8.a, Figure 9.8.c below provides the screenshots from the Translog session before the keystrokes occurred on the current sentence (i.e., deletion of '我们提出'), upon completion of editing on the preceding one. From a replay session in Translog, the specific words associated with the TT fixations (i.e., green diamonds) in Figure 9.8.a before the deletion of '我们提出' (from 28100 to 28700 on the horizontal axis) are shown in the figure below, where each fixation is illustrated as a circle and marked with a number indicating the sequential order of these fixations. These fixations (fixation 1 to 12) are illustrated in three consecutive screenshots in Figure 9.7.c below. The three critical words in relation to the structural ambiguity in the sentence, namely A-'提出' (*tichu*), B-'围绕' (*weirao*), C-'科研实力' (*keyan shili*), are also indicated respectively by A, B, and C in the picture on top.

Figure 9.8.c Translog screenshots for progression graph 1 (Numbers in the figure indicate the sequential order of fixations)

As the graph shows, after reading the ST, the participant fixates on word A, and then the eyes immediately move back to the end of the previous sentence, subsequently going through a process of back-and-forth movement in the region containing word A (see fixations 1-6 in Figure 9.8.c). This seems to be the start of the participant's cognitive processing, with obvious disruption and extra processing effort, regarding the parsing issues arising from word A.[67]

Afterwards, the participant proceeds with the subsequent words in a linear manner, starting with a fixation on '主要' (*zhuyao*, i.e., fixation 7), then on '围绕' (*weirao*, i.e., word B, fixation 8), and then on another two words in this region (fixations 9-10). It is as this point that the participant encounters another disruption of processing, stopping the forward eye-movement and beginning another regression, first towards an earlier word associated with word B — '优先事项' (*youxian shixiang*, priorities, i.e., fixation 11), and then back again to the beginning of the sentence near word A (fixation 12).

9.6.4 Discussion

Following these 12 fixations on the MT output, the participant immediately begins to delete '我们提出' (*women tichu*, we present/propose) (i.e., the typing behaviour at 187000 on the horizontal axis in Figure 9.7 and Figure 9.8.a), where a crucial verb relevant to the structural ambiguity in the sentence is included (i.e. word A). With this deletion, neither the sentential-complement reading nor the noun-phrase reading of the sentence would be possible (i.e., interpretations 1 and 2 in Section 9.5.1 above), while the subsequent change on and around word B (insertions and deletions as shown in Figure 9.7 and Figure 9.8.b) have completely changed the structure of the sentence (see the first subsection of 9.5.2). In the meantime, before this structural change occurs, all the fixations have only been in the early regions of the sentence, with disruptions of processing observed on the words that lead to structural ambiguity rather than in the later regions where some interpretations are disconfirmed. This means that the participant seems to have completely bypassed any disambiguation or reanalysis of the TT structure, in contrast to the cognitive processing of monolingual reading with respect to temporary ambiguity, where eye-movement behaviour would

[67] While the parsing issue seems to be a major cause for the disruption of processing, it can also be aruged that another reason for this could be lexical, as the lexical choice of the verb '提出' (to propose) may not be appropriate for the context here in terms of meaning (to present/demonstrate research strength, rather than to present/propose a design or idea as used in academic writings).

be expected to show disruption of processing much later in the sentence, i.e., on the word which provide disconfirmation of interpretations.

Incremental processing

If one takes an incremental position regarding argument integration, perhaps the above observation can be explained in terms of garden path effect — the post-editor encounters word A, assigns the proposition role to the following sequences as an embedded clause (i.e., sentential-complement reading as in 1a and 1b above), and under the influence of the ST, experiences some hesitation immediately as to its appropriateness since this may not be semantically consistent with the ST (hence the regression and re-fixations around word A as reflected by fixations 1-6 in Figure 9.8.c). This inappropriateness is to be confirmed upon reading the following regions regarding word B. In this aspect, the participant first proceeds with his/her reading of the embedded clause in a linear manner (having assumed the sentential-complement reading), i.e., fixations 7-10, encountering word B as well as the surrounding words in this region. During these fixations, the post-editor experiences the second layer of ambiguity arising from word B, but seems to follow the garden path as in 1a[68] without any observable disruption, until fixations 10 and 11 when the post-editor disconfirms this interpretation based on the realization that the reading was inconsistent with the ST semantically.

This is obviously in contrast to monolingual reading processes (for comprehension), where disruptions of processing would be expected to occur in the region near word C (i.e., the region providing disconfirmation of the garden path), which would be manifested by more fixations, longer fixation duration, and increased regressions of eye movement in this region.

Head-driven approach

If, on the other hand, a head-driven approach is adopted in terms of thematic assignments, perhaps fixations 1-6 indicate the laborious process of maintaining

[68] It is also possible that 1b is adopted, but neither 1a nor 1b is consistent with the ST in terms of meaning, and both would be disconfirmed by the ST. Therefore, the following description is still consistent if 1b is adopted. In the meantime, the fact that the participant does not show observable disruption of processing in this process means that whichever between 1a and 1b is adopted, perhaps only one of them is entertained at this moment.

multiple possibilities with respect to word A, while the relatively less obvious increase of effort regarding word B (fixations 7-12) can be explained as the number of possibilities exceeding the post-editor's processing capacity. However, as these fixations are immediately followed by keyboard input which changes the structure of the sentence effectively — rather than further effort to process the sentence, it does not seem likely that the processing of the sentence at this point is beyond his/her cognitive capacity. In addition, as the ST sentence has already been read twice before the processing of the MT output, it seems safe to assume that the participant has already received clues as to which of the possible structures is more likely to be appropriate for the sentence, which means that the multiple possibilities maintained by the participant should be biased somehow towards at least one of the interpretations. This means that the processing difficulty should be decreased and is more likely to result in linear reading until the end of the sentence. Since this is not the case in the observation, it seems that incremental processing would be a better explanation in this regard.

This is also different from monolingual reading where, adopting a head-driven approach, significantly more cognitive effort would be expected in the region between word B and word C, so as to maintain the multiple analyses of the sentence in the working memory. This sharply increased effort, accordingly, would be manifested by more fixations, longer fixation durations, and regressions of eye movement in this region. On the other hand, in the case of the post-editing processes as described in the previous paragraph, it would be expected that the reading process along the sentence, as the eyes approach word B and word C, becomes smooth and linear, without significant increase of fixations, lengthened fixation duration, or regression of eye movement. As mentioned, this does not seem to be the case given the observation above.

ST-reading and MT-reading

In view of the specific task of post-editing, the above observation shows that when structural ambiguity, in the form of a garden path structure, occurs in the MT output, the reading behaviour shows a sharply different feature from ST reading.

Here, the ST reading is consistent with Krings' (2001, p.361, see also pp. 423-327) description of reading style — reading one sentence at a time to load the information into short-term memory, and reading it more than once where the first reading completes the entire sentence while the subsequent re-readings are of smaller chunks. As shown in Figure 9.8.a, the ST reading for the first time is largely linear, spanning across (almost) the entire sentence, while the second reading can be divided into three activities — reading of the beginning of the sentence (i.e., *We present our major areas of research strength*), another reading of the same chunk (after a regression), and then the reading of the following parts of the sentence. After this second reading, and before the participant begins to edit the MT sentence, the region *We present our* is fixated again in preparation for the post-editing. As Krings describes (p. 325), re-reading of the ST "serves to retain its linguistic form in short-term memory as long as the subject still needs this access to its content to understand the text or for other processes".

However, the reading behaviour on the MT is sharply different from ST reading. It seems that the strategies underlying the processes involved in MT-reading are largely different than in parallel ST-related processes. In Figure 9.8.a and Figure 9.8.b above, the reading behaviour on the MT output shows sharp difference compared with the reading pattern on the ST. Other than their differences at the text level[69], regarding this particular sentence the reading on the TT seems to have bypassed the underlying processes that are typical in ST reading, not completing much of the sentence at all before arriving at an evaluation of the MT, together with the solution to what is considered as a problem in the translation. The MT sentence information that is loaded in the short-term memory seems fairly limited, and it is likely the ST information, combined with a partial interpretation of the MT sentence, that facilitates the processing in this regard.

Syntactic activation and priming

Therefore, in respect of the above explanations for the process of disambiguation, argument integration and thematic assignment, perhaps another, more

[69] The progression graph in view of the entire post-editing process for this participant shows that at the text level, the ST is first read as a whole in an orientation phrase, while the MT text is not; on the other hand, the edited TT is read as a whole in a revision phase at the end, while the ST is not.

reasonable one might be that upon completion of reading on the ST sentence, a TT structure conforming to Category 2 above (see Section 7.4.3) is activated, and during the TT reading process, the structure of the MT output is constantly matched to the activated version rather than parsed in the usual manner as in monolingual sentence processing. From this perspective, the disruption of processing reflected by fixations 1-6 could perhaps be explained by the result of the inconsistency of the current structure (either the garden path or the multiple possibilities) to the structure which is activated in the post-editor's mind. After this disruption, the post-editor starts an attempt to comprehend the MT output adopting incremental processing strategies (in the same manner as described above), and upon the encounter of the region involving word B, realizes the semantic inconsistency to the ST.

This perspective, apparently, is consistent with cross-linguistic priming at the syntactic level, which has received much support from empirical studies in TPR, and is also in line with those models of the translation process that are based on bilingual co-activation, e.g., the shared cognitive representations between SL and TL at the syntactic level in the recursive model mentioned above (see Chapter 4).

9.7 Statistical analysis

The above sections have described the processing issues in relation to the structural ambiguity arising from two words: A ('提出', *tichu*, to present/propose) and B ('围绕', *weirao*, to surround / around), together with the compound noun which plays a crucial role in the disambiguation: C ('科研实力', *keyan shili*, research strength). Statistical analysis of these three words can shed further light on the cognitive aspects of disambiguation, and test hypotheses on the expected eye-movement behaviour mentioned above (see Section 9.6; see also Chapter 7).

However, as these words are from the raw MT output which is tentative and unstable in the post-editing process, i.e., it keeps being changed and updated by the participant, extracting eye-movement information regarding these words is consequently very difficult. As the TT tokens (and Token ID numbers) in the generated tables refer to the tokens in the finalized output from the participants, additional measures are needed to locate the TT tokens which can somehow represent the actual location of the three words in question.

In the following sections, 9.7.1 describes the method of data-labelling in order to locate the TT tokens — in the final post-edited text — on which the fixations can somehow represent the fixations on the three words of interest (i.e., A, B, and C). On this basis, 9.7.2 visualizes these fixations in terms of number of fixations and total reading time, both standardized in view of each experimental session, and analyses the reading process accordingly.

In Section 9.7.3, the fixation data in relation to these three words of interest is analysed in more detail, after further, fine-tuned processing of a combination of summary tables (i.e., TPR-DB tables) — .tt, .au, .kd, and .fd tables, and via more sophisticated statistical means for hypothesis testing.

9.7.1 Method of labelling

If a word in the raw MT output is unedited, it would appear in the final translation product, and the 'TToken' in the table would include the same word. As this word in 'TToken' is directly from the MT output, fixations associated with it would represent the fixations on the original word in the MT output.

On the other hand, if the word has been post-edited, the record of editing on this word would be available in the 'Edit' column of the table, corresponding to a certain token in the final translation which may or may not be the same as the original word, and the edit on this word is considered part of a micro unit in producing the token. Fixations associated with the TT token produced can somehow indicate fixations regarding the word from the MT output.

For example, the following shows an example for the TToken and Edit columns when the participant changes 'ab' (e.g., '全球') into 'AB' (e.g., '国际'):[70]

TToken	Edit
AB	[ba]AB
国际	[球全]国际

[70] The "ab", "AB", and "C" here are symbols for illustration purposes, rather than actual letters that the participant is deleting or inserting. Nor are they related to the words A, B, and C within the MT sentence discussed in previous sections.

AB	AB[C]

Table 9.3: Example of TToken and Edit columns

The Edit column shows the keyboard input in sequential order, with deletions indicated by brackets. In the example here, the participant deletes 'b' ('球') and 'a' ('全'), and then inserts 'A' ('国') and 'B' ('际'), so that 'ab' ('全球') is edited into 'AB' ('国际'). As the original 'ab' is not in the final translation output, it would not be included in the TToken column, therefore the fixations on 'ab' are in turn reflected in the fixations on 'AB'. Note that when a character is deleted without inserting an alternative, such as the deletion of 'C' in the third row above, fixations on 'C' would be represented by 'AB'.

In terms of operationalization, it is important to note the following:

1). When a word in the raw MT output is found in the 'TToken' column (i.e., in the final version by a particular participant), it does not necessarily mean that this word has been left unedited. Nor does it mean that this is a word directly from the MT rather than inserted by the participant. A look at the record of keyboard input is essential for determining if there has been editing on the word, and for identifying what TT Token in the final output is associated with the word in question.

2). If a word (or some characters of the word) in the raw MT output is found in the 'Edit' column and marked as a deletion, it does not necessarily mean that the deletion has occurred on the same word in the original MT output, as the post-editor keeps inserting words during the process. A look at a reasonably large part of the keystroke record is also essential.

For example, the TT sentence from participant P08 includes both word B ('围绕', *weirao*) and word C ('科研实力', *keyan shili*), as shown below, and these two words are therefore included in the TToken column. Further checking on the 'Edit' column shows that word B is directly from the MT output, while word C has been deleted and inserted at different points (see Table 7.4 below). In this regard, word B can be represented by the '围绕' (*weirao*) in the TT output (i.e., Token ID 45), but for word C, it would not be appropriate to use the '科研实力' (*keyan shili*) in the TT output (i.e., Token ID 69) as a representation for the discussion of fixations on C.

TT from P08:

我们围绕着关键的全球优先事项和当前世界面临的挑战，如食品安全和能源领域，体现了我们主要领域的**科研实力**。

ID	TToken	Edit
44	我们	[提出我们主要]
45	**围绕**	**NA**
63	，	，
64	**体现了**	**提出[力实研科的出提]体现了**
65	我们	我们的[的]
66	主要	主要
67	领域	领域
68	的	的
69	**科研**	**科研**
70	实力	**主要[要主]优势的主要领域[势优]实力[域领要主的]**

Table 9.4: TToken and Edit columns for P08

Based on Table 9.4, it is tempting to associate word C with Token ID 64, since '科' (*ke*) and '研' (*yan*) appears as a sequential deletion in one of the micro units for the production of this token. However, further checking on the details of the keyboard input record shows that the insertion of '科研' (*keyan*) occurred before its deletion, with some other keystrokes between the insertion and deletion of the characters (see Table 7.5 below). This means that after it is inserted by the participant, there would be two '科研' (*keyan*) on the screen, one from the raw MT output and the other being newly inserted by the participant. It would be unclear from these tables as to which one is deleted in the later deletion record. For the statistical analysis here, it is important to ensure that the token in discussion corresponds to the token in the raw MT output. In this aspect, a replay of the Translog session would be needed to confirm how the sentence is edited, and to determine which TT Token should be used for word C regarding this participant.

Time	Type	Char	TTid
621446	Mins	要	70
634252	**Mins**	**科**	**69**
634253	**Mins**	**研**	**69**
653504	Mdel	要	70
653675	Mdel	主	70

670507	Mins	优	70
670508	Mins	势	70
671335	Mins	的	70
675390	Mins	主	70
675391	Mins	要	70
679836	Mins	领	70
679837	Mins	域	70
684610	Mdel	的	65
709227	Mdel	粮	58
712082	Mins	品	58
722269	Mdel	势	70
722425	Mdel	优	70
723844	**Mins**	**实**	**70**
723845	**Mins**	**力**	**70**
725576	**Mdel**	**力**	**64**
725748	**Mdel**	**实**	**64**
725935	**Mdel**	**研**	**64**
726232	**Mdel**	**科**	**64**
726544	Mdel	的	64

Table 7.5: Keystroke data for P08

On the basis of the keystroke record and replay of Translog sessions, the TT Tokens associated with the three words in question, namely, ‘提出’ (*tichu*), ‘围绕’ (*weirao*), and ‘科研实力’ (*keyan shili*), are labelled respectively as A, B, and C. They represent, although not strictly, the location of these three words in the raw MT output for each experimental session (i.e., for each participant), and fixations on A, B, and C can therefore be analysed via the fixations on the TT Tokens with corresponding labels.[71]

As words A and B are close to each other in the MT output, in four cases a TT token is associated with the editing on both words (e.g., when they are deleted together in a sequence of keystrokes), and these tokens are labelled as “A_B”. In another special case, a TT token is associated with all three words in question (i.e., A, B, and C), as a result of the participant’s deleting of a long sequence, and this is labelled as “A_B_C”. These cases are relatively rare, as shown in the following table.

Label	Frequency	Proportion

[71] While this is not a strictly accurate representation of the fixations on words of the temporary MT text, it seems to be the best option for systematic statistical analysis.

A (提出)	13	0.029
B (围绕)	13	0.029
C (科研实力)	17	0.038
A_B	4	0.009
A_B_C	1	0.002
No Label	400	0.893

Table 9.6: Frequency and proportion of different labels for TT tokens

With these tokens identified, the corresponding fixation data are used to represent the fixations on words A, B, and C. It is important to note that for each of the tokens in the final TT, the fixations which are mapped to the token itself can refer to any word in any micro unit associated with this token, and the number of micro units, as well as the number of words in the units, can vary from token to token and from participant to participant. Therefore, the fixation duration and counts regarding these tokens are averaged to each character by dividing the original value by the sum of 1). the number of characters in the token, and 2). the number of deleted characters associated with the token (i.e., in all the micro units combined for this token).

In order to statistically analyse the fixation data for all participants (who may display different levels of values from one another), the fixation duration and fixation counts from above are then standardized in view of each experimental session (i.e., each participant), so that the fixation data from these sessions are at the same scale for overall comparison between different tokens.

9.7.2 Reading time and fixation counts

TT tokens along the (ST) sentence

Although each participant produces a different version of the TT, they all work on the same ST, so the TT tokens produced by all participants, which are all aligned to the ST at the token level in the TPR-DB table, can be visualized in the same graph in terms of the aligned ST tokens. This provides a way of analysis for the fixation data corresponding to different regions of the sentence.

In the following figures, the horizontal axis indicates the TT tokens in terms of their aligned STid. In other words, the numbers here refer to the IDs of the ST tokens (or ST token groups) which are aligned to the individual TT tokens in the final, edited translation of the sentence by all participants.

For example, ST ID 53 is the word *research* in the ST sentence, and the versions of TT for this word in the output from different participants include ‘科研’ (*keyan*) and ‘研究’ (*yanjiu*), both aligned to the ST token *research*. Therefore, the ST ID 53 in Figure 9.8.a and Figure 9.8.b refers to all the TT tokens, by all participants (i.e., ‘科研’ and ‘研究’), which are aligned to *research*. Depending on the specific lexical choice and syntactic structure of the TT, this alignment may not be one-to-one between the ST and TT tokens, therefore ST ID 53 in the figures can also appear as combinations of tokens, e.g., 51+52+53+54 (*areas of research strength*), 53+54 (*research strength*).

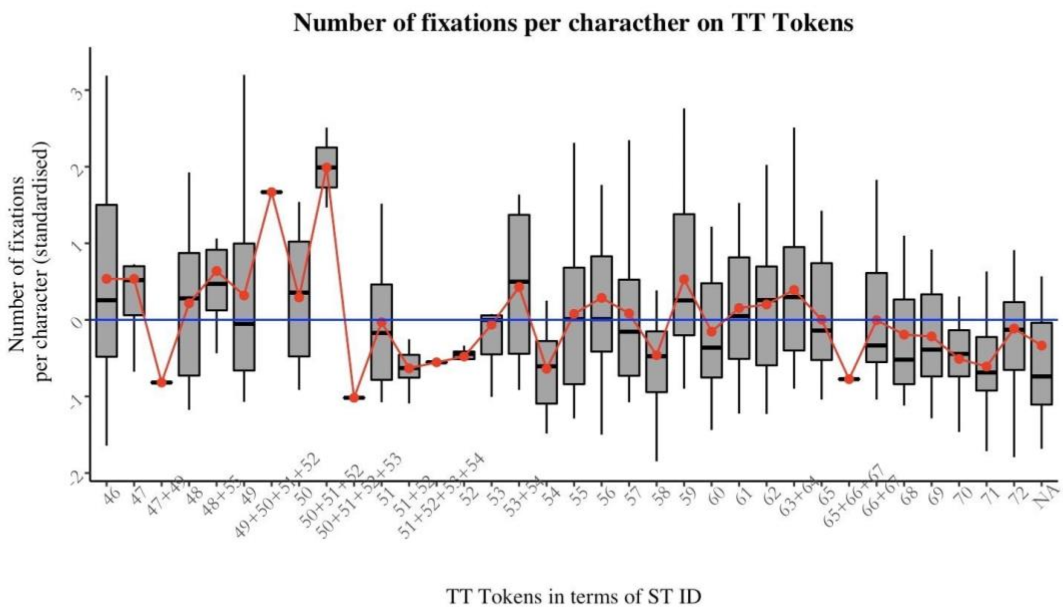


Figure 9.9.a Number of fixations on TT tokens in terms of STid

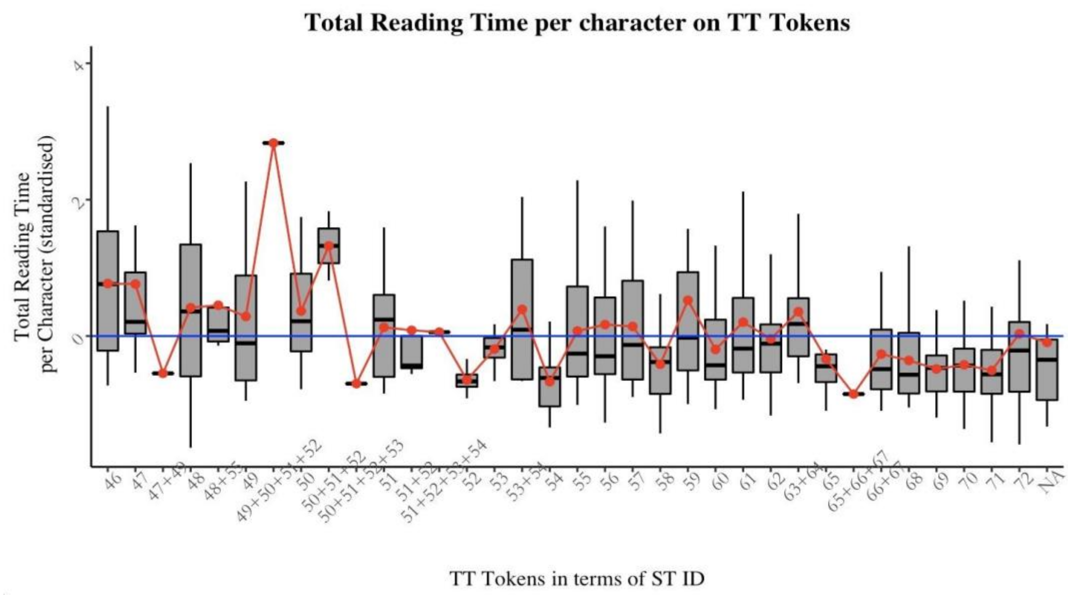


Figure 9.9.b Total reading time on TT tokens in terms of STid

As the order in which these ID numbers are sorted on the horizontal axis is consistent with the order of the ST tokens in the text, the graphs here can show a general trend of nFixT (i.e., number of fixations on the target token) and TrtT (i.e., total reading time on the target token) when each part of the MT sentence is post-edited. In the box plots in these figures, the mean values for the standardized nFixT and TrtT are indictated in red (as connected dots).

It can be seen that the translation for the beginning of the ST sentence, roughly in the region where the STid is between 46 and 50 (i.e., the sequence of ST tokens starting from the punctuation: '*. We present our major*'; in the MT output, this corresponds to '。我们提出我们主要'), receives considerably higher levels of nFixT and TrtT than any other regions of the sentence. In other words, there seems to be a sharp increase of cognitive effort when participants are revising the machine-translated output for this region. As discussed above, this is where the garden path of the TT starts (the region for word A), while the garden-pathed interpretation is inconsistent with the structure and meaning of the ST (this is long before the garden path is disconfirmed by the later part of the TT sentence).

In contrast, the standardized values of nFixT and TrtT tend to be much lower for the other regions of the sentence. Here, of particular interest is the range of 53-64 (i.e., *research strength around key global priorities and challenges currently* *confronting the world*; in the MT output, this region corresponds to two parts — the head noun '科研实力' *keyan shili* and the beginning of the modifier '围绕关键的全球优先事项和当前世界面临的挑战' in a head-final nominal structure), where the values are close to and often slightly above zero.[72] As also mentioned above, the translation of this region is relevant to the other two (temporary) structural interpretations regarding the clause/phrase led by '围绕' (*weirao,* around, i.e., word B), the disambiguation of these two structures provided by the head noun '科研实力' (*keyan shili*, i.e., word C), and the possible disconfirmation of the previous garden path at the end of the sentence.

Other than these two regions, the standardized values of nFixT and TrtT tend to be negative, meaning that the actual number of fixations and total reading time on the TT tokens tend to be below the mean nFixT and TrtT for each participant. This is not surprising, as the MT output for this region is clear in terms of structure and appropriate in terms of meaning.

The above two regions with relatively higher levels of nFixT and TrtT, namely, 1). '*. We present our major*' and 2). '*research strength around key global priorities*', correspond to the following MT:

1). '。我们提出我们主要' (containing A-'提出', *tichu*)

2a). '科研实力' (containing C-'科研实力', *keyan shili*)

2b). '围绕关键的全球优先事项' (containing B-'围绕', *weirao*).

It is not surprising that each of these MT regions contains a word which introduces (i.e., A & B) or resolves (i.e. C) the ambiguity in the sentence structure. The structural issue here appears to result in increased total reading time on the TT.

[72] As these are standardized values (i.e., z-score values), the value being zero does not mean that there is no fixation on the corresponding token, but that the nFixT and TrtT values on the token are equal to the mean value for the entire text. This indicates that there is no *extra* processing effort in the corresponding region.

In monolingual processing (i.e., the processing of TT alone), one can perhaps expect region 2a to be at a much higher level of nFixT and TrtT than 1, 2b, or the rest of the sentence, if adopting incremental processing strategies. For a head-driven approach, the values corresponding to region 2b would be expected to be higher than that of 1, 2a, or the rest of the sentence. This has been mentioned in 9.6.4. In either case, 2a or 2b would be expected to incur more and longer fixations than 1.

Here, however, the graphs apparently show that the values for 2a and 2b both tend to be lower than those for 1. Regarding the difference between 2a and 2b, more specific statistical analysis would be needed, which is described in the sections below.

Categories of TT tokens

Regarding words A, B, and C, the graphs below (Figures 9.10.a and 9.10.b) show the labelled TT tokens in the same manner as Figure 9.9.a and Figure 9.9.b. Here, the horizontal axis is sorted and grouped in terms of the label of the tokens as well as their aligned ST ID. The vertical axis indicates number of fixations (nFixT) and total reading time on the TT (TrtT) which are standardized for each experimental session (i.e., for each participant). It can be seen that the values for ‘A’ (i.e., word A; same below) are visibly high, followed by the ones for ‘B’. The tokens labeled with ‘C’ are much smaller in comparison, somewhat similar but still a bit higher than ‘Other’. Towards the higher end of STid for ‘Other’, the standardized nFixT and TrtT tend to fall down below 0, i.e., the actual nFixT and TrtT values are smaller than the mean of nFixT and TrtT for all tokens.

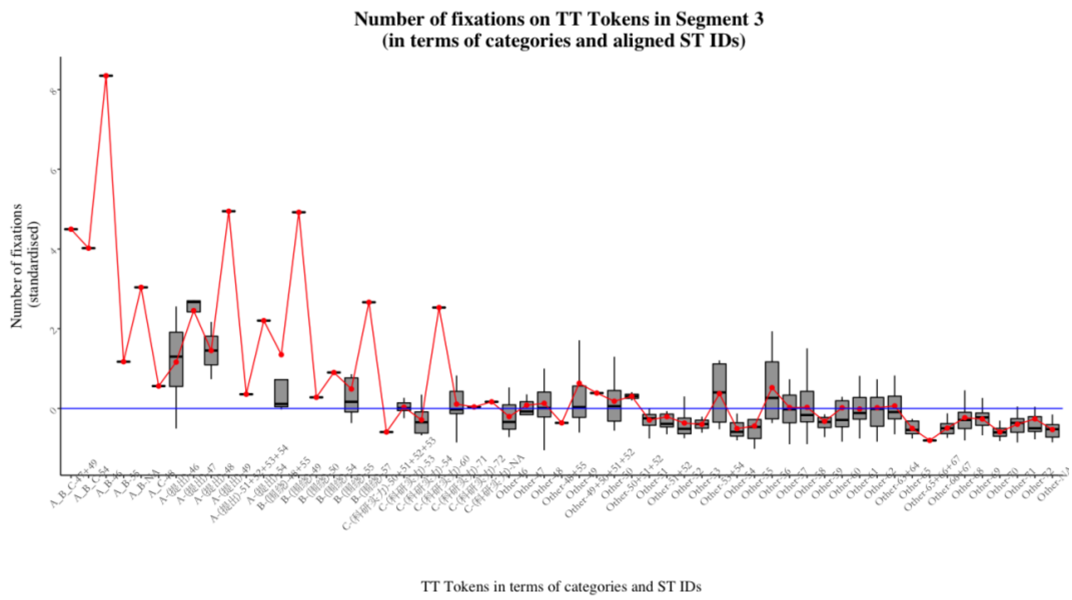


Figure 9.10.a Number of fixations on TT tokens in terms of category and STid

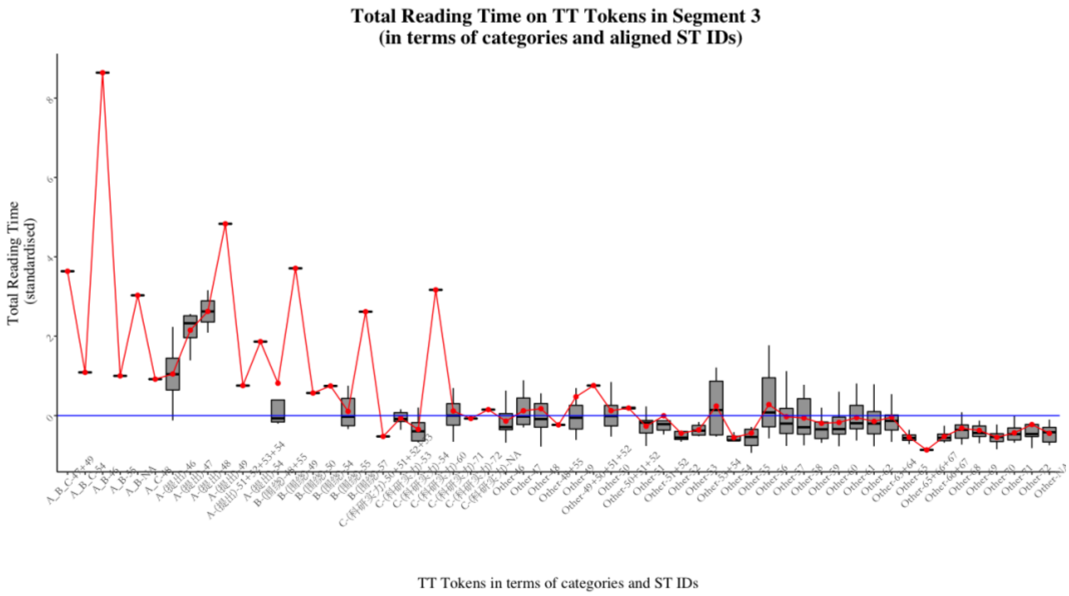


Figure 9.10.b Total reading time on TT tokens in terms of category and STid

The following sections describe statistical analyses of the standardized TrtT and FixT values for categories A, B, and C using t-test and linear mixed effect model.

9.7.3 Statistical analysis

It can be seen from Figure 9.11 that in terms of both standardized TrtT and standardized nFixT, group ‘Other’ corresponds to the lowest level of mean value, and is the only group where the mean is negative (TrtT: *M* = -0.15, *SD* = 0.60; nFixT: *M* = -0.11, *SD* = 0.66). This is followed by group C (‘科研实力’, *keyan shili*) (TrtT: *M* = 0.05, *SD* = 0.77; nFixT: *M* = 0.06, *SD* = 0.68), for which the mean values in terms of both measures are close to and slightly above 0. Other than these two groups, group B (‘围绕’, *weirao*) (TrtT: *M* = 0.78, *SD* = 1.42; nFixT: *M* = 1.18, *SD* = 1.85) and group A (‘提出’, *tichu*) (TrtT: *M* = 1.96, *SD* = 1.23; nFixT: *M* = 1.91, *SD* = 1.36) show a considerable increase of TrtT and nFixT. The order of these groups in terms of the means of standardized TrtT and FixT values, from the smallest to the largest, is ‘Other’ < C < B < A.

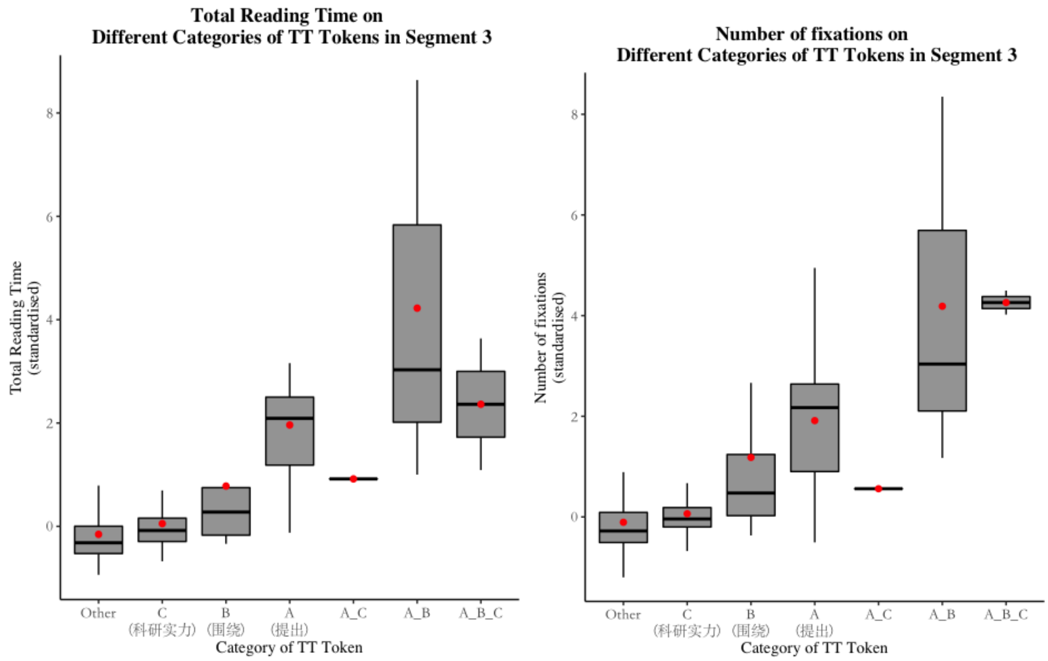


Figure 9.11 TrtT and nFixT on different categories of TT tokens

When the standardized TrtT and nFixT values in each group are compared to the value of 0 using one-sample t-test, results show that the increase or decrease from zero is significant regarding groups A ('提出') (TrtT: $t(12) = 7.73$, $p < 0.001$; nFixT: $t(12) = 5.06$, $p < 0.001$), B ('围绕') (TrtT: $t(15) = 2.20$, $p = 0.04$; nFixT: $t(15) = 2.55$, $p = 0.022$) and 'Other' (TrtT: $t(384) = -4.99$, $p < 0.001$; nFixT: $t(384) = -3.18$, $p = 0.002$). This result means that the tokens associated with '提出' (*tichu*; i.e., word A) and '围绕' (*weirao*; i.e., word B) lead to higher-than-usual levels of total reading time and number of fixations, with '提出' (*tichu*) causing much more disruptions than '围绕' (*weirao*), and that in the same sentence, those tokens which are associated with none of the three words lead to lower-than-usual levels of total reading time and number of fixations.

In contrast, regarding group C ('科研实力', *keyan shili*) the increase from zero is very far from statistically significant, either for TrtT: $t(27) = 0.37$, $p = 0.713$, or for FixT: $t(27) = 0.50$, $p = 0.621$. This means that based on the data analysed here, there is no strong (i.e., statistically significant) evidence to suggest an increased level of total reading time or number of fixations on the TT tokens in relation to ‘科研实力’ (*keyan shili*), and that the positive or negative values for this (i.e., deviation from zero) in the experimental data (and in Figure 9.11) are very likely to be an occurrence by chance (given the very high p-value here).

The above t-test results have already indicated the extent to which TrtT and nFixT are increased (or otherwise) from their usual levels under the influence of each type of token, i.e., ‘Other’ < usual level < B (‘围绕’, *weirao*) < A (‘提出’, *tichu*), whereas no statistically significant effect is observed regarding category C tokens (‘科研实力’, *keyan shili*). To further analyse the relative difference among these categories of TT tokens in terms of their effect (i.e., whether the values for groups A, B and C tend to be higher than those for group ‘Other’), further t-tests between A, B, and C shows that TrtT and nFixT indeed tend to be higher for tokens belonging to categories A (TrtT: $t(12)=6.15$, $p<0.001$; FixT: $t(12)=5.32$, $p<0.001$) and B (TrtT: $t(15)=2.62$, $p=0.019$; FixT: $t(25)=2.77$, $p=0.014$.) than those labelled as ‘Other’. This indicates that the cognitive effort expended on the region of the sentence where either of the two words causing structural ambiguity (i.e., ‘提出’ *tichu* and ‘围绕’ *weirao*) is included tends to be higher in relation to the other part of the same sentence.

Not surprisingly, results from such t-tests also show that this increase in total reading time and number of fixations compared with the rest of the sentence is significant only for categories A and B, with the increase for category A higher and more significant than for category B.[73] For category C, the t-test result in relation to ‘Other’ is $t(29) = 1.40$, $p = 0.172$ for total reading time (TrtT), and $t(30) = 1.29$, $p = 0.208$ for number of fixations (nFixT). Again, no evidence is observed at a statistically significant level to suggest that category C tokens incur more effort of processing in terms of TrtT and FixT. This is consistent with the results mentioned above.

For the groups with multiple words involved (i.e., A_C, A_B, and A_B_C), all values are clearly larger than 0, perhaps because of the presence of ‘提出’ in all of them, and the category of TT tokens which are associated to both of the two words

[73] See the t-test results in the previous paragraph.

causing structural ambiguity, i.e., A_B ('提出'_'围绕'), the TrtT and nFixT values seem to be particularly high. Here, as the data points for them are very few, the meaningfulness of comparison among these categories in terms of their statistical differences would likely to be very limited. Nevertheless, it is important to note that the standardized TrtT and nFixT values of all tokens in this category are considerably larger than 0, and also larger than the mean values for group C and group 'Other'.

In addition, the fixation data on the MT output *before* any editing starts to occur (i.e., the first reading of the raw MT output during post-editing), extracted via R programming on the .tt, .au, .kd, and .fd tables, show consistent results. This provides a more fine-tuned perspective on what has been discussed above. The results of statistical test (using repeated-measures ANOVA) are as follows:

	A *M (SD)*	B *M (SD)*	C *M (SD)*		
TrtT_Before_Edit	15218.11 (13070.63)	8671.83 (12504.26)	3121.61 (5943.27)	$F(2, 34) = 8.46,\ p = .001$	**
nFixT_Before_Edit	23.33 (17.44)	14.94 (15.30)	6.11 (12.50)	$F(2, 34) = 12.54,\ p < .001$	***

Table 9.7: ANOVA results for fixation data before editing

It can be seen that the raw values of the TrtT and nFixT, before the participants' editing on the MT sentence, show a trend of the same relationship among A, B, and C. These results are largely statistically significant.

9.8 Summary

This chapter has investigated the processing of structural ambiguity in the MT output via product and process data in JIN15. After closely examining the dataset, vertical errors in the fixations are identified, analysed, and corrected. With the re-processed dataset and the summary tables (TPR-DB tables) that are generated on the basis of the re-processed raw data files, a sentence that involves structural ambiguity is analysed systematically in view of the sentence itself, the post-edited product, and the behavioural data on eye movement and key-input logs, to shed light on the cognitive processes in post-editing.

The analysis here has pointed out that the participants seem to have disregarded one of the readings and tend to produce translations that dramatically change the structure without much of a disambiguation process on the sentence itself. This is shown via the translation product in 9.5.3, as well as the keyboard input and eye movement in Section 9.6. Regarding the temporary and permanent ambiguity involved in the sentence, disruption of processing does not seem to occur in the disambiguating region (i.e., the word which disconfirms the initial analysis) for garden-path effect. In the meantime, for the permanent ambiguity (i.e., noun phrase vs. sentential complement), the alternative reading other than the simplest analysis seem to be suppressed as the head noun which completes the reading receives very few fixations before a decision is made. On this basis, the discussion in 9.6.4 provides explanation for the observed phenomena in terms of incremental processing vs. head-driven approach, two contrasting perspectives of syntactic processing that make different predictions of eye movement in monolingual reading. This section also explains the observation in terms of the uniqueness of reading for post-editing, where the existence of the ST seems to result in a priming effect and to create a bias in the cognitive processing of the structural ambiguity.

In Section 9.7, rigorous statistical means is used to further analyse the fixations and to test hypotheses regarding the number of fixations on, and the reading time of, different parts of the sentence as qualitatively observed in previous sections. Results here seem to confirm the observation, and the difference in fixations and reading time among three points of ambiguity along the sentence is found to be statistically significant.

Chapter 10 Concluding remarks

10.1 Ambiguity, cognitive models, and statistical approaches

The previous chapters have all explored different aspects of cognitive processing in translation and post-editing in view of lexical and structural ambiguity. Chapters 2, 3, 4, and 5 largely lay the foundation for the subsequent theoretical investigations in Chapters 6, while Chapter 7 provides the basis for the empirical investigations in Chapters 8 and 9.

Chapter 2 has briefly outlined the ambiguity in language, and the psycholinguistic perspectives for the cognitive processing of disambiguation, including the serial vs. parallel processing strategies and the incremental processing vs. head-driven approach. As can be seen in later chapters, discussion on such contrasting views spans across different parts of the thesis, in both theoretical and empirical investigations.

The concept of entropy is introduced in Chapter 3, defined in line with information theory, and various aspects of its implications are explained, including information, surprisal, and uncertainty. As mentioned, the entropy discussed here as a mathematical concept should be differentiated from the operationalized metric that is frequently used in TPR as an approximation (e.g., HTra in the CRITT TPR-DB), which, depending on how it is operationalized, may represent different phenomena, and what is represented by HTra may not strictly reflect the cognitive aspects of a particular individual.

In Chapter 4, models of the translation process, as well as post-editing and MT, are all discussed. Here, the horizontal vs. vertical processing strategies bear resemblance to the parallel and serial processing in psycholinguistics which has been discussed in Chapter 2. Important to mention is that the understanding of serial vs.

parallel processing may be different if one adopts a different perspective (see, e.g., Section 4.6).

Chapter 5 outlines the statistical means in psycholinguistics that have been used for discussion of ambiguity, starting with simple metrics based on frequency of occurrence. As the chapter proceeds, more sophisticated measures are described touching upon entropy and the surprisal framework of (expectation- or experience-based) sentence processing. This starts to move from the statistical features of the text, towards the cognitive processes in the mind. It also sets the stage for a discussion of mental states, entropy and entropy reduction in Chapter 6.

Subsequently, the thesis moves to an empirical side of the discussion, based on eye-movements. Chapter 7 outlines the technique of tracking eye-movements as a method for studying the cognitive processes involved in reading, translation, and post-editing, while also describing the lexical and syntactic issues that influence eye-movements. The empirical findings, as well as the models and theoretical perspectives for explanation of the observed eye-movement behaviour in studies of ambiguity, are also reflections of many issues that have been discussed in preceding chapters, including the parallel vs. serial processing and the incremental processing vs. head-driven approach. As pointed out later in Chapter 9, the investigations on post-editing in this thesis show aspects of considerable difference from these observations.

Chapters 8 and 9 investigate entropy and ambiguity empirically, on the basis of the CRITT TPR-DB, focused respectively on the lexical and structural aspects.

10.2 Entropy

Chapter 6 is a crucial part of the thesis, where the conceptual aspects of entropy in translation (and post-editing, as this perspective can also explain post-editing in the same manner) are discussed systematically. While HTra, in the way it is calculated in the CRITT TPR-DB, is essentially a product-based metric and is by no means a direct measurement of cognitive effort in itself, it has been found in experimental studies to be positively correlated with eye fixations in translation and post-editing. If this metric is somehow related to cognitive effort/load, it would be necessary to discuss how such a statistical measure can reflect a certain aspect of cognitive processing. This chapter looks at the purely conceptual aspect of entropy, not confined to the way in which the produced translations are counted in an experiment (i.e., the way in which HTra is calculated), and in turn considers HTra as an approximation of the (conceptualized) entropy that is directly relevant to the cognitive processes. That entropy, which describes mental activities, would be a direct indication of cognitive effort/load. With HTra being a statistical approximation to this entropy (specifically, a rough approximation of entropy reduction, see Chapter 6), it is not surprising either that the HTra values are sometimes found to have positive, statistically significant correlation to cognitive effort (using eye-fixation data as a proxy, and assuming the eye-mind hypothesis). In this regard, the Chapter draws inferences from the psycholinguistic formulations of resource-allocation processing difficulty and explores the conceptual basis on which entropy and entropy reduction can directly describe the mental activities as well as quantify the cognitive effort involved. This brings the concept of entropy to the assumed mental states from a *process* perspective, rather than considering it a way of counting the items in the product (as most discussions on HTra do), and provides explanations as to how entropy can be a theoretically justifiable means of the quantification of cognitive load.

Several important conceptual issues are also disentangled, e.g., how the HTra values in the CRITT TPR-DB relate to these conceptualizations, how entropy that describes the mental sates differs from the entropy that can be observed in the text via statistical means or the entropy that can be approximated from experimental sessions, what level of entropy it is at the initial state of bilingual lexical activation, how relative entropy, the absolute decrease of entropy, and the surprisal of the lexical selection relate to one another, etc.

All these issues are discussed in Chapter 6, which altogether describe the process of lexical choice (and disambiguation) as a process of entropy reduction, where the effort involved can be quantified from the perspective of this entropy.

Here, it is important to note that the "entropy" under discussion is conceptual and abstract, rather than referring to the result of a metric operationalized to approximate it. As mentioned in Section 6.2.11, this entropy describes mental states and its exact value could be obtained only if one had, somewhat miraculously, access to the items in the mental lexicon of a particular participant, as well as the specific lexical information that is activated in his/her mind, the extent to which this information is active, and the change of this extent through enhancement and suppression, all of which are impossible to obtain.

That being said, the entropy value can be approximated via different means, one of which can be the calculation of HTra in the CRITT TPR-DB. Using corpora of different kinds also results in various approximated values of this entropy, as described in Chapter 6.

Another important issue in this regard is that there are always individual differences from person to person, while this entropy value, approximated via whatever means, would be one single number that seemingly describes every individual. Here, it should be emphasized that, on the one hand, the conceptualization of the mental states, processes, and entropy as abstract concepts would not change from person to person — it would only be the specific values of this entropy that are different. On the other hand, the approximation of the value of this entropy can be different depending on what data is used for calculation. This by analogy would be similar to the relationship between the frequency of occurrence for lexical items and the familiarity of the item for an experimental participant. If the frequency is obtained in a corpus that is maximally similar to the collection of linguistic input that this participant has received all his/her life, it might be reasonable to consider this frequency as a dependable indication of familiarity. Similarly, the approximated value of entropy would be dependent on the specific corpus used, if one considers the genuine level of entropy for an individual. In the meantime, subjective measure of frequency, i.e., rating of familiarity, can also be transformed into entropy values and in turn indicate the actual entropy for a given population of readers.

In terms of the practical use of entropy to predict cognitive effort, Chapter 6 has also provided a detailed analysis on the calculation of relative entropy and decrease of entropy in the lexical selection process, both of which can be theoretically valid means of quantifying the effort needed in the mental processes. In the meantime, the section has also proposed several ways to approximate the theoretical value using corpus-based methods. This is meaningful for the prediction of cognitive effort through statistical means.

While entropy is used to somehow represent and describe mental states, where entropy reduction can be calculated mathematically, the chapter does not assume that the translating/post-editing mind itself is subconsciously conducting these calculations, or that the mental activities are essentially a process of calculating the entropy, the surprisal (i.e., relative entropy), or the reduction of entropy.

In short, the conceptual contribution of this thesis is that it explores, based on the psycholinguistic formulations of resource-allocation processing difficulty, theoretical justifications of entropy as a (probabilistic) description of the mental activities as well as a quantification of cognitive load. As current studies in TPR consider entropy as a feature of the translation product, this thesis brings the concept to the mind, as a feature of the process. It also has the potential of opening new doors to more sophisticated entropy-based frameworks, e.g., the neuroscientific theories such as the free energy principle and active inference (see Sections 1.2 and 3.4.2).

In terms of application, the discussion in this part of the thesis can inform the search for a more dependable means of approximation, to overcome the issues inherent in the metric currently in use (i.e., HTra). Such approximation of entropy would be helpful for the prediction of translation effort.

10.3 Lexical ambiguity and entropy

Other than these conceptual aspects, Chapter 8 has examined the behavioural manifestations of this entropy reduction (i.e., RQ 2) and has observed a tendency of reduction of the average entropy in the activity units of the translation process in the data. This occurs as an active visual search for disambiguating clues in the context, manifested in fixations on lower-entropy items when a high-entropy item introduces much ambiguity (and uncertainty, as indicated by entropy). As stated in the chapter, it appears that the ambiguity, uncertainty, or unexpectedness involved in particular lexical items effectively directs the translator/post-editor to those aspects of context which are relevant to rendering the items unambiguous, to integrating the newly activated information into what has already been activated, to enhancing or suppressing certain activation depending on its relevance to or compatibility with the context, to reducing the uncertainty (and entropy) level in the translator's/post-editor's mind, and to finally arriving at a selection among the co-activated alternative translations for an option which suits the ST and TT context.

In short, focusing on the lexical level, Chapter 8 analyses HTra, a metric operationalized in the CRITT TPR-DB. Examples of eye-movement and keyboard-input are illustrated in detail regarding a high-entropy phrasal verb, showing that it not only leads to extra processing effort, but also involves an active visual search for multiple low-entropy items in the ST to facilitate the resolution of uncertainty. In other words, the cognitive processes regarding the high-entropy ST item are facilitated by surrounding low-entropy items and can be described as a process of entropy reduction. How entropy describes mental states (i.e., RQ 1) has been addressed in Chapter 6, while the behavioural manifestations of the conceptualizations (i.e., RQ 2) are addressed here in Chapter 8.

Additionally, as this entropy, by definition, captures aspects of TL activation during ST reading (see Chapter 6), the above-mentioned low-entropy items seem to represent readily-available TT co-text which is activated in the translator's mind during the visual search. This activated TT co-text facilitates the selection of a suitable TT option for the item being translated. Therefore, if we adopt a conceptual model of the mental lexicon, the above observations can perhaps be interpreted as a direct and memory-based search in the mental lexicon for a translation option which is linguistically compatible with the TT co-text, rather than as a mental process that is based on conceptual comprehension. From this perspective, the cognitive processing would be largely horizontal (as defined in De Groot, 1997). As mentioned in Sections 4.6, 6.2.1 and 6.2.6, however, the entropy perspective seems more relevant to a connectionist model (eg., BIA+), and from that perspective, perhaps the opposition between vertical vs. horizontal in terms of conceptual mediation would no longer be reasonable. In this respect, the above description (of TL activation during the visual search) would mean that TL reformulation processes occur in parallel with the SL comprehension processes, and thus supports parallelism (see also Section 4.6).

This answers RQ 3 at the lexical level.

10.4 Structural ambiguity

It can be seen that the theoretical assumptions in Chapter 6 and empirical analyses in Chapter 8 have endorsed a parallel perspective in terms of the translation process (i.e., between SL and TL), and also a parallel perspective in terms of the activation of multiple candidates for ambiguity at the lexical level (i.e., among the options in the TL).

For structural issues, Chapter 8 conducts detailed and systematic analyses on a sentence from the MT output, in view of the behavioural data in the post-editing process. Here, the parallel perspective for the translation process is also assumed to some extent (in the form of syntactic activation for both languages and cross-language priming effects), but the parallel vs. serial processing strategies for structural disambiguation is in the form of incremental (serial) processing vs. head-driven (parallel) approach, where the incremental (serial) strategy is endorsed. This answers RQ 3 at the syntactic level.

The chapter uses the JIN15 dataset from CRITT TPR-DB, and starts with an examination of the data itself, where errors that can potentially lead to seriously misleading findings are identified, analysed, and corrected through data re-processing. The discussion here is methodologically meaningful, not only for the analysis in the present thesis, but also for TPR research in general, as eye-tracking data is often very noisy and as some studies have already been based on data with errors of considerable scale (see Section 7.3). That being said, the chapter has also pointed out that if the issues causing these errors are successfully resolved, the experiments conducted for the JIN15 dataset are in effect very well-managed and dependable.

Perhaps more importantly, through the present research, this wrongly processed dataset has been re-processed and corrected, so that it can now be used with more confidence in future.

With this re-processed data, the chapter conducts detailed analysis on a sentence involving structural ambiguity, from the perspectives of the linguistic issues in the sentence, the post-edited product from the participants, and the eye-movement and keyboard behaviours. This includes both qualitative and statistical investigations, and results show that the disambiguation processes seem to be bypassed in the post-editing of the sentence. When a garden-path effect is involved, disruptions of processing occur not in the later regions of the sentence where the wrong parse is disconfirmed, but in the earlier regions where the most quickly built analysis is semantically inconsistent with the ST. Results of statistical analysis on fixations and reading time are largely consistent with the observation in the qualitative analysis, and these results are found to be statistically significant.

The Chapter has also attempted to explain the observation in view of different processing strategies, i.e., incremental processing and head-driven approach, endorsing the former, while in the meantime pointing out the difference in the reading behaviour of post-editing in contrast to what would otherwise be expected in monolingual reading.

Here, the methodological aspect of the statistical analysis is meaningful as well (see Section 7.7), as in studies of the post-editing (and translation) process the eye-movement data on the TT, especially at the sub-sentence level, is usually associated with the final post-edited output, rather than the intermediate, temporary, and unstable version of the text which is constantly being edited. The method of locating the fixations on certain tokens in the sentence being edited (see 7.7.1), and calculating the number and duration of fixations on these tokens before editing on the sentence during the participants' drafting (see 7.7.3), using a combination of tables (.tt, .au, .kd, and .fd tables) in the CRITT TPR-DB, can also be a contribution to studies of TT reading in studies of the translation and post-editing processes.

10.5 Summary

Overall, the present thesis has investigated lexical and structural ambiguity in the translation and post-editing process, exploring the conceptual aspects and then using the CRITT TPR-DB for empirical analyses. It draws inferences from psycholinguistics, bilingualism, and entropy-based models of translation cognition. Conceptually, the thesis assumes non-selective activation of both languages (source and target) in the translation and post-editing process, and explores how entropy and entropy reduction can theoretically describe assumed mental states during disambiguation. Empirically, it uses a product-based metric of word translation entropy (HTra), and eye-movement and keystroke data in the CRITT TPR-DB, to shed light on how the conceptual understanding of lexical and structural ambiguity may be manifested by observable behaviour.

At the lexical level, examination of behavioural data pertaining to a high-HTra item from 217 participants translating/post-editing from English into multiple languages shows that the item tends to result in pauses in production and regression of eye movements, and that the translators'/post-editors' corresponding scrutinizing of the ST tends to involve a visual search for lower-HTra words in the co-text and, accordingly, a decrease in the average entropy of the Activity Unit.

Regarding syntax, a Chinese relative clause in the MT output which can involve a garden-path effect is examined in terms of eye movements from 18 participants. Results show that, contrary to monolingual reading, disruptions of processing tend to occur not in the later part of the sentence where the wrong parse is disconfirmed, but in the earlier regions where the most quickly-built analysis is semantically inconsistent with the ST. Structural disambiguation and re-analysis seem to be bypassed. This suggests that, on the one hand, reading for post-editing receives a strong biasing effect from the ST, and on the other, argument integration is more appropriately explained from an incremental processing perspective rather than a head-driven approach, as thematic roles seem to be assigned immediately in reading for post-editing.

While the lexical analysis supports a parallel disambiguation model, the structural analysis seems to support a serial one. In terms of translation models, both analyses in this thesis emphasize the impact of cross-linguistic priming and the presence of considerable horizontality in the translation process.

10.6 Limitations and future work

This thesis has a number of limitations and aspects that would benefit from further exploration in the future. The theoretical discussion on mental states is based on, among others, psycholinguistic models of the bilingual mental lexicon, the formulation of resource-allocation processing difficulty, and TPR models that emphasize both horizontal and vertical processing. While these models are adequate in describing the conceptualization of entropy in the present thesis, there are also other recent theories of cognition that can be explored in the future for more comprehensive conceptual analysis. For example, radical enactivism (Hutto & Myin, 2013), post-cognitivism (Carl, 2021c), and 4EA cognition which situates the mind within its environment (Muñoz, 2016a) can perhaps be touched upon in future work, for analysing the points at which these perspectives and theories are compatible with the conceptualizations in the present thesis.

Even within its own framework, the thesis can also be expanded in many ways. For example, while Chapter 8 has shown that the examples given are representative of a general trend regarding the same phrasal verb — via statistical means on the basis of a large database (see Section 8.4.1) — it is important to note that the analysis in Section 8.4 is indeed confined to one phrasal verb, *cough up*. It remains to be tested whether these findings would be consistent with findings when other high-entropy items, which reflect different linguistic phenomena, are analysed. Therefore, a meaningful avenue for further research would be to systematically analyse different categories of high-entropy items in the ST, in view of the corresponding AUs which are relevant to the resolution of the uncertainty and ambiguity involved in those items. This would contribute to a more comprehensive understanding of the cognitive processing of information regarding translation ambiguity.

In terms of approximating the entropy level that describes the mental state of lexical activation, empirical investigations can be conducted on the basis of parallel corpora and translation behavioural data. The different levels of entropy discussed in Chapter 6 can be approximated using different corpora (e.g., those based on well-balanced general language texts, domain-specific texts, or texts on particular topic, etc.) to indicate the extent to which context restricts lexical activation, while these approximated values can be statistically analysed as an independent variable to examine whether, and how strongly, they predict the fixation- or pause-based measurements of cognitive effort. This would shed light on the possible entropy level that can describe the initial mental state, and on how early sentence context influences bilingual lexical access.

Additionally, entropy and surprisal (relative entropy) can be compared in terms of their correlation with, or prediction of, cognitive effort in the behavioural data, to draw conclusions as to which perspective (relative entropy or absolute decrease of entropy) can better reflect the cognitive effort involved in the process (see 6.2 above).

In Chapter 9, analysis of structural ambiguity is also confined to one example only, and this example is from MT output. While the analysis of this example is detailed and systematic, the research would benefit significantly from expansion to include more examples and more structural issues, and both the ST and the MT output. In this regard, the JIN15 dataset itself can be a starting point, as the MT output contains three sentences where the main verb is followed by head-final nominal structures with a long '*de*' sequence modifying the head noun. Two are ambiguous while the third one is unambiguous, so it would be meaningful to compare these three sentences in terms of both the product and the process, for a more comprehensive investigation. Beyond the dataset, structural issues in the ST can also be analysed in the same manner via experiments specifically designed for this investigation.

References

Abney, S. P. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research, 18*, 51-60.

Aikawa, T., Schwartz, L., King, R., Corston-Oliver, M., & Lozano, C. (2007). Impact of controlled language on translation quality and post-editing in a statistical machine translation environment. *Proceedings of the MT Summit XI*, 1-7.

Alves, F., & Gonçalves, J. L. V. R. (2007). Modelling translator's competence: Relevance and expertise under scrutiny. In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), *Doubts and Directions in Translation Studies: Selected Contributions from the EST Congress* (pp. 41-55). Amsterdam / Philidelphia: John Benjamins Publishing Company.

Alves, F., & Hurtado Albir, A. (2010). Cognitive approaches. In Y. Gambier & L. van Doorslaer (Eds.), *The Handbook of Translation Studies* (Vol. 1, pp. 28-35). Amsterdam: John Benjamins.

Alves, F., & Jakobsen, A. L. (2021). Grounding cognitive translation studies: Goals, commitments and challenges. In F. Alves & A. L. Jakobsen (Eds.), *The Routledge Handbook of Translation and Cognition* (pp. 545-554). New York: Routledge.

Alves, F., & Jakobsen, A. L. (2022). The translation process. In K. Malmkjaer (Ed.), *The Cambridge Handbook of Translation* (pp. 34-54). Cambridge: Cambridge University Press.

Alves, F., & Vale, D. (2009). Probing the unit of translation in time: Aspects of the design and development of a web application for storing, annotating, and querying translation process data. *Across Languages and Cultures, 10*(2), 251-273.

Alves, F., Pagano, A., & da Silva, I. (2011). Towards an investigation of reading modalities in/for translation: An exploratory study using eye-tracking data. In S. O'Brien (Ed.), *Cognitive Explorations of Translation* (pp. 175-196). London and New York: Continuum.

Angelone, E. (2010). Uncertainty, uncertainty management and metacognitive problem solving in the translation task. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 17-40). Amsterdam: John Benjamins.

Arnold, D., Balkan, L., Meijer, S., Humphreys, R., & Sadler, L. (1994). *Machine Translation: An Introductory Guide.* Oxford: NCC Blackwell.

Atkinson, J. (2000). *The Developing Visual Brain*. Oxford: Oxford University Press.

Attneave, F. (1959). *Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods, and Results*. Oxford, England: Henry Holt.

Auerbach, F. (1913). Das Gesetz der Bevölkerungskonzentration. *Petermanns Geographische Mitteilungen, 59*, 74-76.

Ayres, L. P. (1915). *A Measuring Scale for Ability in Spelling*. New York: Division of education, Russell Sage foundation.

Bader, M., & Lasser, I. (1994). *German Verb-Final Clauses And Sentence Processing: Evidence For Immediate Attachment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Baker, M. (1993). Corpus linguistics and translation studies -- Implicantions and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and Technology: In honour of John Sinclair* (pp. 233-250). Amsterdam/Philidelphia: John Benjamins.

Balling, L. W. (2014). Production time across languages and tasks: A large-scale analysis using the CRITT translation process database. In J. W. Schwieter & A. Ferreira (Eds.), *The Development of Translation Competence: Theories and Methodologies from Psycholinguistics and Cognitive Science* (pp. 239-268). Newcastle upon Tyne: Cambridge Scholars Publishing.

Bangalore, S., Behrens, B., Carl, M., Ghankot, M., Heilmann, A., Nitzke, J., . . . Sturm, A. (2016). Syntactic variance and priming effects in translation *New Directions in Empirical Translation Process Research* (pp. 211-238): Springer.

Barkhudarov, L. (1993). The problem of the unit of translation. In P. Zlateva (Ed.), *Translation as Social Action* (pp. 39-46). london and New York: Routledge.

Bazzanella, C. (2011). Redundancy, repetition, and intensity in discourse. *Language Sciences, 33*(2), 243-254. doi:http://dx.doi.org/10.1016/j.langsci.2010.10.002

Bennett, P. (1994). The translation unit in human and machine. *Babel, 40*(1), 12-20.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the Development of Language* (pp. 279-362). New York: Wiley.

Blum-Kulka, S. (1986). Shifts of cohesion and coherence in translation. In J. House & S. Blum-Kulka (Eds.), *Interlingual and Intercultural Communication: Discourse and Cognition in Translation and Second Language Acquisition Studies* (pp. 17-35). Tübingen: Gunter Narr.

Brillouin, L. o. (1962). *Science and Information Theory*. London: Academic Press.

Brown, A. C. (1895). The Relation between the Movements of the Eyes and the Movements of the Head. *Nature, 52*, 184-188.

Brown, P. F., Cocke, J., Pietra, S. A. D., Pietra, V. J. D., Jelinek, F., Lafferty, J. D., . . . Roossin, P. S. (1990). A statistical approach to machine translation. *Computational Linguistics, 16*(2), 79-85.

Brown, P. F., Pietra, V. J. D., Pietra, S. A. D., & Mercer, R. L. (1993). The mathematics of statistical machine translation: Parameter estimation. *Computational Linguistics, 19*(2), 263-311.

Brysbaert, M. (1998). Word recognition in bilinguals: Evidence against the existence of two separate lexicons. *Psychologica Belgica*.

Buswell, G. T. (1935). *How People Look at Pictures: A Study of the Psychology and Perception in Art*. Chicago: University of Chicago Press.

Cappelle, B. (2012). English is less rich in manner-of-motion verbs when translated from French. *Across Languages and Cultures, 13*(2), 173-195. doi:10.1556/Acr.13.2012.2.3

Carl, M. (2012, May). *Translog-II: A program for recording user activity data for empirical reading and writing research.* Paper presented at the Eighth International Conference on Language Resources and Evaluation, Istanbul, Turkey, 23-25 May 2012. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/614_Paper.pdf

Carl, M. (2013). Dynamic programming for re-mapping noisy fixations in translation tasks. *Journal of Eye Movement Research, 6*(2), 1-11. doi:10.16910/jemr.6.2.5

Carl, M. (2021a). Computation and representation in cognitive translation studies. In M. Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 341-355). Switzerland: Springer.

Carl, M. (2021b). Information and entropy measures of rendered literal translation. In M. Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 113-140). Cham, Switzerland: Springer.

Carl, M. (2021c). A radical embodied perspective on the translation process. In M. Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 389-406). Cham, Switzerland: Springer.

Carl, M. (Ed.) (2021d). *Explorations in Empirical Translation Process Research*. Cham, Switzerland: Springer.

Carl, M., & Dragsted, B. (2017). Inside the monitor model: Processes of default and challenged translation production. In O. Czulo & S. Hansen-Schirra (Eds.), *Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation: TC3 II* (pp. 5-30). Berlin: Language Science Press.

Carl, M., & Kay, M. (2011). Gazing and Typing Activities during Translation: A Comparative Study of Translation Units of Professional and Student Translators. *Meta, 56*(4), 952-975.

Carl, M., & Schaeffer, M. (2014). *Word transition entropy as an indicator for expected machine translation quality.* Automatic and Manual Metrics for Operational Translation Evaluation Workshop Programme, 2014.

Carl, M., & Schaeffer, M. (2017a). Models of the translation process. In J. W. Schwieter & A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 50-70). Hoboken: John Wiley & Sons, Incorporated.

Carl, M., & Schaeffer, M. (2017b). Sketch of a Noisy Channel Model for the translation process. *Empirical Modelling of Translation and Interpreting, 7*, 71.

Carl, M., & Schaeffer, M. (2017c). Why translation is difficult: A corpus-based study of non-literality in post-editing and from-scratch translation. *HERMES-Journal of Language and Communication in Business*(56), 43-57.

Carl, M., & Toledo Báez, C. (2019). Machine translation errors and the translation process: A study across different languages. *The Journal of Specialised Translation*(31), 107-132.

Carl, M., Bangalore, S., & Schaeffer, M. (Eds.). (2016). *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB*. Cham, Switzerland: Springer International Publishing.

Carl, M., Schaeffer, M., & Bangalore, S. (2016). The CRITT translation process research database. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research* (pp. 13-54). New York: Springer.

Carl, M., Tonge, A., & Lacruz, I. (2019). A systems theory perspective on the translation process. *Translation, Cognition & Behavior, 2*(2), 211-232. doi:https://doi.org/10.1075/tcb.00026.car

Carpenter, R. H. (1988). *Movements of the Eyes* (2nd ed.). London: Pion Limited.

Chan, S. W. (2004). *A Dictionary of Translation Technology*. Hong Kong: Chinese University Press.

Chang, C. Y. (2009). *Testing Applicability of Eye-tracking and fMRI to Translation and Interpreting Studies: An Investigation into Directionality.* (PhD thesis), Imperial College London, London.

Charniak, E. (1993). *Statistical Language Learning*. Cambridge, MA: The MIT press.

Chesterman, A. (2004a). Beyond the particular. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do They Exist?* (pp. 33-49). Amsterdam: John Benjamins.

Chesterman, A. (2004b). Hypotheses about translation universals. In G. Hansen, K. Malmkjær, & D. Gile (Eds.), *Claims, Changes and Challenges in Translation Studies: Selected Contributions from the EST Congress, Copenhagen 2001* (pp. 1-14). Amsterdam / Philidelphia: John Benjamins Publishing Company.

Chesterman, A. (2007). What is a unique item? In Y. Gambier, M. Shlesinger, & R. Stolze (Eds.), *Doubts and Directions in Translation Studies* (pp. 3-13). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Chiang, D. (2005). *A hierarchical phrase-based model for statistical machine translation.* 43rd Annual Meeting on Association for Computational Linguistics, University of Michigan, Ann Arbor, MI, 2005. Association for Computational Linguistics.

Chmiel, A. (2021). Translation, psycholinguistics and cognition. In F. Alves & A. L. Jakobsen (Eds.), *The Routledge Handbook of Translation and Cognition* (pp. 219-238). New York: Routledge.

Chmiel, A., & Mazur, I. (2013). Eye tracking sight translation performed by trainee interpreters. In C. Way, S. Vandepitte, R. Meylaerts, & M. Bartlomiejczyk (Eds.), *Tracks and Trecks in Translation Studies* (pp. 189-205). Amsterdam / Philidelphia: John Benjamins.

Cho, K., Van Merriënboer, B., Bahdanau, D., & Bengio, Y. (2014). On the properties of neural machine translation: Encoder-decoder approaches. *arXiv preprint arXiv:1409.1259*.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chrisman, L. (1991). Learning recursive distributed representations for holistic computation. *Connection Science, 3*(4), 345-366.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci, 36*(3), 181-204. doi:10.1017/S0140525X12000477

Clifton, C., & Staub, A. (2011). Syntactic influences on eye movements during reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 895-909). Oxford: Oxford University Press.

Collins, M. (2014). Information density and dependency length as complementary cognitive models. *J Psycholinguist Res, 43*(5), 651-681. doi:10.1007/s10936-013-9273-3

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204.

Condon, E. U. (1928). Statistics of vocabulary. *Science, 67*(1733), 300.

Cover, T. M., & Thomas, J. A. (1991). *Elements of Information Theory*. New York: John Wiley & Sons.

Craciunescu, O., Gerding-Salas, C., & Stringer-O′Keeffe, S. (2004). Machine Translation and Computer-Assisted Translation: a new way of translating? *Translation Journal, 8*(3).

CRITT@kent. Public Studies. Retrieved 20 January 2020 from https://sites.google.com/site/centretranslationinnovation/tpr-db/public-studies

Darwin, R. W. (1786). New experiments on the ocular spectra of light and colours. *Philosophical Transactions of the Royal Society, 76*, 313-348.

de Beaugrande, R., & Dressler, W. (1981). *Introduction to Text Linguistics*. New York: Longman.

De Bruijn, E. R., Dijkstra, T., Chwilla, D. J., & Schriefers, H. J. (2001). Language context effects on interlingual homograph recognition: evidence from event-related potentials and response times in semantic priming. *Bilingualism: Language and Cognition, 4*(2), 155-168.

de Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1001-1018.

de Groot, A. M. B. (1997). The cognitive study of translation and interpretation: Three approaches. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive Processes in Translation and Interpreting* (pp. 25-56). Thousand Oaks, CA: Sage Publications Ltd.

de Groot, A. M. B., & Christoffels, I. K. (2006). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition, 9*(2), 189-201. doi:10.1017/s1366728906002537

DePalma, D., Hegde, V., Pielmeier, H., & Stewart, R. (2013). *The Language Services Market: 2013*. Lowell, MA: Common Sense Advisory Report.

Devlin, J., Zbib, R., Huang, Z., Lamar, T., Schwartz, R., & Makhoul, J. (2014). *Fast and robust neural network joint models for statistical machine translation.*

Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014.

Dewey, G. (1923). *Relative Frequency of English Speech Sounds*. Cambridge, MA: Harvard University Press.

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist Connectionist Approaches to Human Cognition* (pp. 189-225). Mahwah, NJ: Lawrence Erlbaum Associates.

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*(3), 175-197. doi:10.1017/s1366728902003012

Dijkstra, T., Timmermans, M., & Schriefers, H. (2000). On being blinded by your other language: Effects of task demands on interlingual homograph recognition. *Journal of Memory and Language, 42*(4), 445-464.

do Carmo, F. E. M. (2017). *Post-editing: A Theoretical and Practical Challenge for Translation Studies and Machine Learning.* (PhD thesis), Universidade do Porto.

Doherty, S., & O′Brien, S. (2012). *A user-based usability assessment of raw machine translated technical instructions*. Paper presented at the The 10th Biennial Conference of the Association for Machine Translations in the Americas (AMTA 2012), San Diego, California, USA.

Doherty, S., Gaspari, F., Groves, D., van Genabith, J., Specia, L., Burchardt, A., . . . Uszkoreit, H. (2013). *Mapping the industry I: Findings on translation technologies and quality assessment*. Retrieved from http://www.qt21.eu/launchpad/sites/default/files/QTLP_Survey2i.pdf

Dong, Y. (2010). 交替传译中的语言转换心理机制:非对称有限并行加工模型 [Language Switching in Consecutive Interpreting: Asymmetrical Limited Parallel Processing]. *中国英语教育 [English Education in China]*(4), 1-11.

Dong, Y., & Lin, J. (2013). Parallel processing of the target language during source language comprehension in interpreting. *Bilingualism: Language and Cognition, 16*(3), 682-692. doi:10.1017/s1366728913000102

Dong, Y., & Wang, B. (2013). 口译过程的两阶段解读 —— 以一般语言理解和产出为参照 [General v. Interpretation-specific Language Comprehension and Production: A Two-Stage Account of the Interpreting Process]. *中国翻译 [Chinese Translators Journal]*(1), 19-24.

Dragsted, B. (2010). Coordination of reading and writing processes in translation: An eye on uncharted terrotory. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 41-62). Amsterdam: John Benjamins.

Dragsted, B., & Hansen, I. (2009). Exploring translation and interpreting hybrids: The case of sight translation. *Meta, 54*(3), 588-604.

Dragsted, B., & Hansen, I. G. (2007). Speaking your translation: Exploiting synergies between translation and interpreting. In F. Pöchhacker, A. L. Jakobsen, & I. M.

Mees (Eds.), *Interpreting Studies and Beyond: A Tribute to Miriam Shlesinger* (pp. 251-274). Copenhagen: Samfundslitteratur.

Dragsted, B., & Hansen, I. G. (2008). Comprehension and production in translation: a pilot study on segmentation and the coordination of reading and writing processes. *Copenhagen Studies in Language*(36), 9-29.

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice* (2nd ed.). London: Springer.

Duffy, S. A., Kambe, G., & Rayner, K. (2001). The effect of prior disambiguating context on the comprehension of ambiguous words: Evidence from eye movements. In G. D. S (Ed.), *On the Consequences of Meaning Selection: Perspectives on Resolving Lexical Ambiguity* (pp. 27-43). Washington: American Pcychological Association.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*(4), 429-446.

Eddington, C. M., & Tokowicz, N. (2012). Examining English–German translation ambiguity using primed translation recognition. *Bilingualism: Language and Cognition, 16*(2), 442-457. doi:10.1017/s1366728912000387

Eldridge, R. C. (1911). *Six Thousand Common English Words: Their Comparative Frequency and what can be Done with Them*. Niagara Falls: Clement Press.

Elman, J. L., Bates, E. A., & Johnson, M. H. (1996). *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT press.

Englund Dimitrova, B. (2005). *Expertise and Explicitation in the Translation Process* (Vol. 64). Amsterdam / Philidelphia: John Benjamins Publishing.

Erdmann, B., & Dodge, R. (1898). *Psychologische Untersuchungen über das Lesen auf experimenteller Grundlage*: Niemeyer.

Eskola, S. (2004). Untypical frequencies in translated language: A corpus-based study on a literary corpus of translated and non-translated Finnish. In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do they exist?* (pp. 83-99). Amsterdam/Philadelphia: john Benjamins Publishing Company.

Estoup, J.-B. (1902). *Gammes sténographiques. Recueil de textes choisis pour l'acquisition méthodique de la vitesse, précédé d'une introduction par J.-B. Estoup*. Paris: Institut Sténographique.

Fantz, R. L. (1961). The origin of form perception. *Scientific American, 204*(5), 66-73.

Farmer, J. D., & Geanakoplos, J. (2008). *Power Laws in Economics and Elsewhere.* Santa Fe Institute, 2008.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first- and second-language reading of fluent bilinguals. *Mem Cognit, 11*, 565-574.

Ferrer i Cancho, R. (2005a). Decoding least effort and scaling in signal frequency distributions. *Physica A: Statistical Mechanics and its Applications, 345*(1-2), 275-284. doi:10.1016/j.physa.2004.06.158

Ferrer i Cancho, R. (2005b). Hidden communication aspects in the exponent of Zipf's law. *Glottometrics, 11*, 98-119.

Ferrer i Cancho, R., & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences, 100*(3), 788-791.

Findlay, J. M. (1981). Local and global influences on saccadic eye movements. In D. F. Fisher, R. A. Monty, & J. W. Senders (Eds.), *Eye Movements: Cognition and Visual Perception* (pp. 171-179). Hillsdale, NJ: Lawrence Erlbaum Associates.

Findlay, J. M., & Walker, R. (1999). A model of saccade generation based on parallel processing and competitive inhibition. *Behavioral and Brain Sciences, 22*(4), 661-674. doi:undefined

Floridi, L. (2010). *Information: A Very Short Introduction*. New York: Oxford University Press.

Fodor, J. A., Bever, T. G., & Garrett, M. F. (1974). *The Psychology of Language*. New York: McGraw-Hill.

Forcada, M. L. (2017). Making sense of neural machine translation. *Translation Spaces, 6*(2), 291-309.

Forcada, M. L., & Ñeco, R. P. (1997). *Recursive hetero-associative memories for translation.* International Work-Conference on Artificial Neural Networks, 1997. Springer.

Franz, S. I. (1899). After-images. *Psychological Review, Monographs Supplement, 3*, 1.

Frazier, L., & d'Arcais, F. G. (1989). Filler-driven parsing: A study of gap-filling in Dutch. *Journal of Memory and Language, 28*, 331-344.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*(2), 178-210.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat Rev Neurosci, 11*(2), 127-138. doi:10.1038/nrn2787

Friston, K., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017, Oct). Active inference, curiosity and insight. *Neural Comput, 29*(10), 2633-2683. doi:10.1162/neco_a_00999

Friston, K., Schwartenbeck, P., Fitzgerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Front Hum Neurosci, 7*, 598. doi:10.3389/fnhum.2013.00598

Fu, Y., Zhou, X., Li, W., Fan, G., & Jiang, Z. (1997). *現代漢語介詞研究 [Prepositions in modern Chinese]*. 廣州: 中山大學出版社.

Furukawa, C. (2008). 漢語介詞「離」與日語格助詞「から (kara)」、「まで (made)」(Relationship of the Chinese coverb li and the Japanese postpositional particles kara and made). *Language and Linguistics, 9*(2), 271-296.

Galley, M., Graehl, J., Knight, K., Marcu, D., DeNeefe, S., Wang, W., & Thayer, I. (2006). *Scalable inference and training of context-rich syntactic translation models.* 21st International Conference on Computational Linguistics and the

44th annual meeting of the Association for Computational Linguistics, Sydney, 2006. Association for Computational Linguistics.

Galley, M., Hopkins, M., Knight, K., & Marcu, D. (2004). *What's in a translation rule?* Paper presented at the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL, Boston, MA.

Garcia, I. (2011). Translating by post-editing: is it the way forward? *Machine Translation, 25*(3), 217-237.

Garnsey, S. M., Pearlmutter, N. P., Myers, E., & Lotocky, M. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*(1), 58-93.

Gaspari, F. (2007). *The Role of Online MT in Webpage Translation.* (Ph.D), The University of Manchester.

Gaspari, F., Toral, A., Naskar, S. K., Groves, D., & Way, A. (2014). *Perception vs reality: Measuring machine translation post-editing productivity.* Third Workshop on Post-Editing Technology and Practice, 2014.

Geary, D. C. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In J. S. Carlson & J. R. Levin (Eds.), *Psychological Perspectives on Contemporary Educational Issues* (pp. 1-99). Greenwich, CT: Information Age Publishing.

Geary, D. C. (2008). An evolutionarily informed education science. *Educational Psychologist, 43*(4), 179-195. doi:10.1080/00461520802392133

Gerlach, J. (2015). *Improving Statistical Machine Translation of Informal Language: A Rule-based Pre-editing Approach for French Forums*. (PhD thesis), University of Geneva, Geneva. https://archive-ouverte.unige.ch/unige:73226

Gernsbacher, M. A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum.

Gernsbacher, M. A. (1997). Attenuating interference during comprehension: The role of suppression. In M. D. L (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 37, pp. 85-104). San Diego, California, USA.: Academic Press.

Gibson, E. A. F. (1991). *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown.* (PhD thesis), Carnegie Mellon University, Pittsburgh, Pennsylvania.

Gile, D. (2009). *Basic Concepts and Models for Interpreter and Translator Training (Revised edition)*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Goltz, H. C., DeSouza, J. F., Menon, R. S., Tweed, D. B., & Vilis, T. (2003). Interaction of retinal image and eye velocity in motion perception. *Neuron, 39*(3), 569-576.

Gorrel, P. G. (1987). *Studies of Human Syntactic Processing: Ranked-parallel versus Serial Models.* (PhD thesis), The University of Connecticut.

Graham, Y., Baldwin, T., Moffat, A., & Zobel, J. (2014). *Is Machine Translation Getting Better over Time?* EACL, 2014.

Green, S., Heer, J., & Manning, C. D. (2013, April). *The efficacy of human post-editing for language translation.* Paper presented at the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April 2013. ACM.

Grosjean, F. (1997). The bilingual individual. *Interpreting, 2*, 163-187.

Groves, D., & Wicklow, C. (2008). *Bringing humans into the loop: localization with machine translation at Traslan.* Conference of the American Machine Translation Association (AMTA), Waikiki, Hawai'i, 2008. AMTA.

Guerberof, A. (2014). Correlations between productivity and quality when post-editing in a professional context. *Machine Translation, 28*(3-4), 165-186. doi:10.1007/s10590-014-9155-y

Guiraud, P. (1968). The semic matrices of meaning. *Information (International Social Science Council), 7*(2), 131-139.

Gutt, E.-A. (1991). *Translation and Relevance: Cognition and Context*. Oxford: Basil Blackwell.

Hainline, L. (1998). Summary and commentary: eye movements, attention and development. In J. E. Richards (Ed.), *Cognitive Neuroscience of Attention: A Developmental Perspective* (pp. 163-178). Mahwah, NJ: Lawrence Erlbaum Associates.

Hale, J. (2001, June). *A probabilistic Earley parser as a psycholinguistic model.* Paper presented at the Second Meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies, Pittsburgh, Pennsylvania, 1- 7 June 2001. Association for Computational Linguistics.

Halverson, S. L. (2003). The cognitive basis of translation universals. *Target, 15*(2), 197-241.

Halverson, S. L. (2017). Gravitational pull in translation: Testing a revised model. In G. de Sutter, M.-A. Lefer, & I. Delaere (Eds.), *Empirical Translation Studies: New Methodological and Theoretical Traditions* (pp. 9-46). Berlin: De Gruyter.

Halverson, S. L. (2019). 'Default' translation: a construct for cognitive translation and interpreting studies. *Translation, Cognition & Behavior, 2*(2), 187-210. doi:10.1075/tcb.00023.hal

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: cooperative division of labor between visual and phonological processes. *Psychological Review, 111*(3), 662.

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science, 15*(6), 409-414.

He, Y. (2007). A fresh cognitive perspective to horizontal translation. *Journal of Translation Studies, 10*(1), 77-89.

He, Y. (2009). Two encoding mechanisms in translation process: Could there be memory-based encoding in translation? *English Education in China, 2*.

Heilmann, A., & Llorca-Bofí, C. (2021). Analyzing the effects of lexical cognates on translation properties: A multivariate product and process based approach. In

M. Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 203-229). Cham, Switzerland: Springer.

Hernando, A., Puigdomènech, D., Villuendas, D., Vesperinas, C., & Plastino, A. (2009). Zipf's law from a Fisher variational-principle. *Physics Letters A, 374*(1), 18-21.

Hervey, S., Higgins, I., & Loughridge, M. (1995). *Thinking German Translation: A Course in Translation Method: German to English*. London and New York: Routledge.

Hogan, P. C. (2011). *The Cambridge Encyclopedia of the Language Sciences*. Cambridge: Cambridge University Press.

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, *50*(2), 259-285. doi:10.1111/nous.12062

Hornof, A. J., & halverson, T. (2002). Cleaning up systematic error in eye-tracking data by using required fixation locations. *Behavior Research Methods, Instruments, & Computers, 34*(4), 592-604.

Huang, J. (2016). *Working Styles of Student Translators in Revision and Post-editing: an Empirical-Experimental Study with Eye-tracking, Keylogging and Cue-based Retrospection.* (PhD thesis), Durham University.

Huang, J. (2018). Working styles of student translators in self-revision, other-revision and post-editing. In C. Walker & F. Federici (Eds.), *Eye Tracking and Multidisciplinary Studies on Translation* (pp. 145-184). Amsterdam/Philadelphia: John Benjamins.

Huang, L., Knight, K., & Joshi, A. (2006). *Statistical syntax-directed translation with extended domain of locality.* 7th Conference of the Association for Machine Translation of the Americas: Visions for the Future of Machine Translation, Cambridge, MA, 2006. Cambridge, MA.

Huey, E. B. (1908). *The Psychology and Pedagogy of Reading*. New York: The Macmillan Company.

Hutchins, W. J. (2003). Machine translation and computer-based translation tools: What's available and how it's used. *A New Spectrum of Translation Studies*, 13-48.

Hutchins, W. J., & Somers, H. L. (1992). *An Introduction to Machine Translation*. London: Academic Press.

Hutto, D. D., & Myin, E. (2013). *Radicalizing Enactivism: Basic Minds without Content*. Cambridge, MA: The MIT Press.

Hvelplund, K. T. (2016). Cognitive efficiency in translation. In R. Muñoz (Ed.), *Reembedding Translation Process Research* (pp. 149-170). Amsterdam/Philadelphia: John Benjamins.

Hyönä, J., Tommola, J., & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *The Quarterly Journal of Experimental Psychology, 48*(3), 598-612.

Indurkhya, N., & Damerau, F. J. (2010). *Handbook of Natural Language Processing* (Vol. 2): CRC Press.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*(6), 431-439.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Ivir, V. (1981). Formal correspondence vs. translation equivalence revisited. *Poetics Today, 2*(4), 51-59.

Jakobsen, A. L. (2019). Segmentation in translation: a look at expert behaviour. In D. Li, V. L. C. Lei, & Y. He (Eds.), *Researching Cognitive Processes of Translation* (pp. 71-108). Singapore: Springer.

Jakobsen, A. L., & Jensen, K. T. (2008). Eye movement behaviour across four different types of reading task. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Looking at Eyes: Eye-Tracking Studies of Reading and Translation Processing* (Vol. 36, pp. 103-124). Copenhagen: Samfundslitteratur.

James, W. (1890). *The Principles of Psychology* (Vol. 2). New York: Holt.

Jin, C. (1996). 談動詞向介詞的虛化. *漢語學習, 2*, 13-18.

Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., . . . Corrado, G. (2016). Google's multilingual neural machine translation system: enabling zero-shot translation. *arXiv preprint arXiv:1611.04558*.

Juhasz, B. J., & Pollatsek, A. (2011). Lexical influences on eye movements in reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 873-893). Oxford: Oxford University Press.

Juhasz, B. J., & Rayner, K. (2003). Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(6), 1312-1318.

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition, 13*(7-8), 846-863. doi:10.1080/13506280544000075

Juhasz, B. J., Liversedge, S. P., White, S. J., & Rayner, K. (2018). Binocular coordination of the eyes during reading: word frequency and case alternation affect fixation duration but not fixation disparity. *Quarterly Journal of Experimental Psychology, 59*(9), 1614-1625. doi:10.1080/17470210500497722

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*(4), 329-354.

Kaeding, F. W. (Ed.) (1898). *Häufigkeitswörterbuch der deutschen Sprache: Festgestellt durch einen Arbeitsausschuß der deutschen Stenographiesysteme*. Steglitz bei Berlin: Selbstverlag des Herausgebers.

Kalchbrenner, N., & Blunsom, P. (2013). Recurrent continuous translation models. In D. Yarowsky, T. Baldwin, A. Korhonen, K. Livescu, & S. Bethard (Eds.), *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 1700–1709). Seattle, Washington, USA: Association for Computational Linguistics.

Karn, K. S., & Hayhoe, M. M. (2000). Memory representations guide targeting eye movements in a natural task. *Visual Cognition, 7*(6), 673-703.

Kelly, N., DepPalma, D., & Stewart, R. (2012). *The Language Services Market: 2012*. Lowell, MA: Common Sense Advisory.

Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(1), 68-81.

Kenny, D., & Satthachai, M. (2019). Explicitation, unique items and the translation of english passives in Thai legal texts. *Meta, 63*(3), 604-626. doi:10.7202/1060165ar

Kirchhoff, K., Turner, A. M., Axelrod, A., & Saavedra, F. (2011). Application of statistical machine translation to public health information: A feasibility study. *Journal of the American Medical Informatics Association, 18*(4), 473-478.

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*, 1-10.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*(1-2), 262-284.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: the influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135*(1), 12-35.

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge: Cambridge University Press.

Kockelman, P. (2013). Information is the enclosure of meaning: Cybernetics, semiotics, and alternative theories of information. *Language & Communication, 33*(2), 115-127. doi:10.1016/j.langcom.2013.01.002

Koehn, P. (2010). *Statistical Machine Translation*. Cambridge: Cambridge University Press.

Koehn, P. (2020). *Neural Machine Translation.* Cambridge: Cambridge University Press. doi:10.1017/9781108608480

Koehn, P., Och, F. J., & Marcu, D. (2003). *Statistical phrase-based translation.* 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Stroudsburg, PA, 2003. Association for Computational Linguistics.

Kolehmainen, L. (2013). Die Unikat-Hypothese der Translation: Etwas Altes, etwas Neues und etwas Geliehenes [Unique Items Hypothesis: Something Old, Something New, Something Borrowed]. *Trans-kom, 6*(1), 92-114.

Kondo, F. (2007). *Translation units in Japanese-English corpora: The case of frequent nouns.* Proceedings from Corpus Linguistics Conference Series (Birmingham, 27-30 July 2007). Article, 2007.

Krieger, G., Rentschler, I., Hauske, G., Schill, K., & Zetzsche, C. (2000). Object and scene analysis by saccadic eye-movements: an investigation with higher-order statistics. *Spatial Vision, 13*(2), 201-214.

Krings, H. P. (2001). *Repairing Texts: Empirical Investigations of Machine Translation Post-Editing Processes*. Kent, Ohio: Kent State University Press.

Kroll, J., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language, 33*, 149-174.

Kullback, S. (1959). *Information Theory and Statistics*. Hoboken, NJ: John Wiley & Sons.

Kwong, O. Y. (2015). Natural language processing. In S. W. Chan (Ed.), *The Routledge Encyclopedia of Translation Technology* (pp. 563-577). New York: Routledge.

Lacruz, I., Denkowski, M., & Lavie, A. (2014, October). *Cognitive demand and cognitive effort in post-editing.* Paper presented at the 11th Conference of the Association for Machine Translation in the Americas (AMTA), Vancouver, BC, Canada, 22 – 26 October 2014. https://www.amtaweb.org/AMTA2014Proceedings/AMTA2014Proceedings_PEWorkshop_final.pdf

Lacruz, I., Ogawa, H., Yoshida, R., Yamada, M., & Ruiz Martinez, D. (2021). Using a product metric to identify differential cognitive effort in translation from Japanese to English and Spanish. In M. Carl (Ed.), *Explorations in Empirical Translation Process Research*. Cham: Springer.

Land, M. F. (2011). Oculomotor behaviour in vertebrates and invertebrates. In S. Liversedge, I. Gilchrist, & S. everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 3-15). Oxford: Oxford University Press.

Landolt, E. (1879). *A Manual of Examination of the Eyes.* (S. M. Burnett, Trans.). London: Bailliere, Tindall, and Cox.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980a). Expert and novice performance in solving physics problems. *Science, 208*(4450), 1335-1342.

Larkin, J. H., McDermott, J., Simon, D. P., & Simon, H. A. (1980b). Models of competence in solving physics problems. *Cognitive Science, 4*(4), 317-345.

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*(6), 943-951. doi:10.1002/acp.1787

Levelt, W. J. M. (2014). *A History of Psycholinguistics: The pre-Chomskyan era*. Oxford: Oxford University Press.

Levin, P., Dhanuka, N., & Khalilov, M. (2017). Machine translation at Booking. com: Journey and lessons learned. *arXiv preprint arXiv:1707.07911*.

Levý, J. (1967). Translation as a decision process. In J. Levý (Ed.), *To Honor Roman Jakobson: Essays on the Occasion of His Seventieth Birthday* (Vol. II, pp. 1171-1182). The Hague: Mouton.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177.

Levy, R. (2013). Memory and surprisal in human sentence comprehension. In R. P. G. van Gompel (Ed.), *Sentence Processing* (pp. 90-126). London and New York: Psychology Press.

Levy, R., & Gibson, E. (2013). Surprisal, the PDC, and the primary locus of processing difficulty in relative clauses. *Front Psychol, 4*, 229. doi:10.3389/fpsyg.2013.00229

Li, C. N., & Thompson, S. A. (1974). Co-verbs in Mandarin Chinese: Verbs or prepositions? *Journal of Chinese Linguistics, 2*(3), 257-278.

Lima Sanches, C., Augereau, O., & Kise, K. (2016). Vertical error correction of eye trackers in nonrestrictive reading condition. *IPSJ Transactions on Computer Vision and Applications, 8*(1). doi:10.1186/s41074-016-0008-x

Lin, C. J. C., & Bever, T. G. (2010). Garden Path and the Comprehension of Head-Final Relative Clauses. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and Producing Head-final Structures* (pp. 277-297). Springer. doi:10.1007/978-90-481-9213-7_13

Lin, X., & Xiang, X. (2021). 认知翻译加工模式：并行还是串行？[Cognitive processing models of translation: Parallel or serial?]. *语言、翻译与认知 [Studies in Language, Communication & Cognition], 2*, 79-95.

Liu, J., & Xu, Q. (2017). 视译认知加工模式研究 [A proposal for a cognitive model of sight interpretation]. *中国翻译 [Chinese Translators Journal]*(2), 25-30.

Liu, Q., & Zhang, X. (2015). Machine translation: General. In S. W. Chan (Ed.), *The Routledge Encyclopedia of Translation Technology* (pp. 105-119). New York: Routledge.

Liu, Y., Liu, Q., & Lin, S. (2006). *Tree-to-string alignment template for statistical machine translation.* Paper presented at the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, Sydney, Australia, 2006.

Liversedge, S., Gilchrist, I., & Everling, S. (Eds.). (2011). *The Oxford Handbook of Eye Movements*. Oxford: Oxford University Press.

Luong, M.-T., Sutskever, I., Le, Q. V., Vinyals, O., & Zaremba, W. (2014). Addressing the rare word problem in neural machine translation. *arXiv preprint arXiv:1410.8206*.

Macizo, P., & Bajo, M. T. (2004). When translation makes the difference: Sentence processing in reading and translation. *Psicológica, 25*, 181-205.

Macizo, P., & Bajo, M. T. (2006). Reading for repetition and reading for translation: do they involve the same processes? *Cognition, 99*(1), 1-34. doi:10.1016/j.cognition.2004.09.012

MacKay, D. M. (1969). *Information, Mechanism and Meaning*. Cambridge: MIT Press.

Malmkjær, K. (2006). Translation units. In K. Brown (Ed.), *The Encyclopedia of Languages and Linguistics* (pp. 92-93). Amsterdam: Elsevier.

Malmkjær, K. (2007). Norms and nature in translation studies. In G. Anderman & M. Rogers (Eds.), *Incorporating Corpora: The Linguist and the Translator* (pp. 49-59). Clevedon: Multilingual Matters.

Malmkjaer, K. (2011). Translation universals. In K. Malmkjaer & K. Windle (Eds.), *The Oxford Handbook of Translation Studies*.

Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication Theory, 84*, 486-502.

Mandelbrot, B. (1962). On the theory of word frequencies and on related Markovian models of discourse. *Structure of Language and its Mathematical Aspects, 12*, 190-219.

Manin, D. Y. (2008). Zipf's law and avoidance of excessive synonymy. *Cognitive Science, 32*(7), 1075-1098.

Mantegna, R. N., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Peng, C. K., Simons, M., & Stanley, H. E. (1994). Linguistic features of noncoding DNA sequences. *Phys Rev Lett, 73*(23), 3169-3172. doi:10.1103/PhysRevLett.73.3169

Marcu, D., & Wong, W. (2002). *A phrase-based, joint probability model for statistical machine translation.* Paper presented at the ACL-02 conference on Empirical methods in natural language processing, Philadelphia, PA, 2002.

Marcus, M. (1963). An introduction to information theory. *Mathematics Magazine, 36*(4), 207-218. doi:10.2307/2687903

Marcus, M. P. (1980). *A Theory of Syntactic Recognition for Natural Language*. Cambridge, MA: MIT Press.

Marian, V., Spivey, M., & Hirsch, J. (2003). Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging. *Brain and language, 86*(1), 70-82.

Marzouk, S. and Hansen-Schirra, S. (2019). Evaluation of the impact of controlled language on neural machine translation compared to other MT architectures. *Machine Translation* 33, 179–203. doi:10.1007/s10590-019-09233-w

Mauranen, A. (2000). Strange strings in translated language: A study on corpora. In M. Olohan (Ed.), *Intercultural Faultlines: Research Models in Translation Studies v. 1 Textual and Cognitive Aspects* (pp. 119-141). London and New York: Routledge.

McConkie, G. W. (1997). Eye movement contingent display control: Personal reflections and comments. *Scientific Studies of Reading, 1*(4), 303-316.

McConkie, G. W., & Rayner, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics, 17*(6), 578-586.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence coomprehension. *Journal of Memory and Language, 38*, 283-312.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.

Mishra, A., Carl, M., & Bhattacharya, P. (2012). A heuristic-based approach for systematic error correction of gaze data for reading. *Proceedings of the First Workshop on Eye-tracking and Natural Language Processing* (pp. 71-80). Mumbai, India: COLING.

Miyamoto, E. T. (2002). Case markers as clause boundary inducers in Japanese. *Journal of Psycholinguistic Research, 31*(4), 307-347.

Miyamoto, E. T., & Nakamura, M. (2003). Subject/object asymmetries in the processing of relative clauses in Japanese. In G. Garding & M. Tsujimura

(Eds.), *Proceedings of the 22nd West Coast Conference on Formal Linguistics* (pp. 342-355). Somerville, MA: Cascadilla Press.

Miyata, R. and Fujita, A. (2021). *Understanding pre-editing for black-box neural machine translation*. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, April 19 - 23, 2021, 1539–1550.

Moorkens, J. (2018). Eye-tracking as a measure of cognitive effort for post- editing of machine translation. In F. Federici & C. Walker (Eds.), *Eye-tracking Applications in Translation Studies* (pp. 55-69). Amsterdam: John Benjamins.

Moorkens, J., O′Brien, S., da Silva, I. A. L., de Lima Fonseca, N. B., & Alves, F. (2015). Correlations of perceived post-editing effort with measurements of actual effort. *Machine Translation, 29*(3), 1-18. doi:10.1007/s10590-015-9175-2

Moray, N. (1979). *Mental Workload: Its Theory and Mmeasurement*. New York: Plenum.

Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 9-28). Cambridge: Cambridge University Press.

Mossop, B. (2003). An Alternative to "Deverbalization". Retrieved from http://www.yorku.ca/brmossop/Deverbalization.htm

Muñoz, R. (2016a). Of minds and men – computers and translators. *Poznan Studies in Contemporary Linguistics, 52*(2). doi:10.1515/psicl-2016-0013

Muñoz, R. (2016b). Reembedding translation process research. An introduction. In R. Muñoz (Ed.), *Reembedding Translation Process Research* (pp. 1–20). Amsterdam: John Benjamins.

Muñoz, R. (2017). Looking toward the future of cognitive translation studies. In J. W. Schwieter & A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 555-572). Malden, MA: Wiley.

Muñoz, R., & González Fernández, C. A. (2021). Cognitive translatology: A primer, revisited. *语言、翻译与认知 [Studies in Language, Communication & Cognition], 1*(1), 131-165.

Nagao, M. (1989). *Machine Translation: How Far can it Go?* Oxford: Oxford University Press.

Newell, A., & Simon, H. A. (1972). *Human Problem Solving* (Vol. 104): Englewood Cliffs, NJ: Prentice-Hall.

Newhall, S. M. (1928). Instruments for observing ocular movements. *American Journal of Psychology, 40*, 628–629.

Nida, E. A., & Taber, C. R. (1969). *The Theory and Practice of Translation* (Vol. 8). Leiden: Brill.

Nirenburg, S. (1989). Knowledge-based machine translation. *Machine Translation, 4*(1), 5-24.

O'Brien, S. (1993). *Sublanguage, Text Type and Machine Translation.* (M.A. thesis), Dublin City University.

O'Brien, S. (2004). Machine translatability and post-editing effort: How do they relate. *Translating and the Computer*, 26.

O'Brien, S. (2006a). Controlled language and post-editing. *Multilingual, 17*(7), 17-19.

O'Brien, S. (2006b). Eye-tracking and translation memory matches. *Perspectives: Studies in Translatology, 14*(3), 185-205.

O'Brien, S. (2008). Processing fuzzy matches in translation memory tools: An eye-tracking analysis. In S. Göpferich, A. L. Jakobsen, & I. M. Mees (Eds.), *Looking at Eyes — Eye Tracking Studies of Reading and Translation Processing* (Vol. 36, pp. 79-102). Copenhagen: Samfundslitteratur.

O'Brien, S. (2011). Towards predicting post-editing productivity. *Machine Translation, 25*(3), 197-215.

O'Brien, S. (2012). Translation as human–computer interaction. *Translation Spaces, 1*, 101-122. doi:10.1075/ts.1.05obr

O'Brien, S. (2022). How to deal with errors in machine translation: Post-editing. In D. Kenny (ed.) , *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence* (pp. 105–120). Berlin: Language Science Press,

O'Brien, Sharon, & Joss Moorkens. (2014). *Towards intelligent post-editing interfaces.* Paper presented at the FIT XXth World Congress 2014, Berlin, Germany, 4-6 August 2014.

O'Brien, S., & Simard, M. (2014). Introduction to special issue on post- editing. *Machine Translation, 28*(3), 159-164. doi:10.1007/s10590-014-9166-8

Och, F. J., & Ney, H. (2002). *Discriminative training and maximum entropy models for statistical machine translation.* Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, 2002.

Och, F. J., & Ney, H. (2004). The alignment template approach to statistical machine translation. *Computational Linguistics, 30*(4), 417-449.

Och, F. J., & Weber, H. (1998). *Improving statistical natural language translation with categories and rules.* Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, University of Montreal, Quebec, Canada, 1998.

Och, F. J., Tillmann, C., & Ney, H. (1999). *Improved alignment models for statistical machine translation.* 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora, University of Maryland, 1999.

Ogawa, H., Gilbert, D., & Almazroei, S. (2021). redBird: Rendering entropy data and source-text background information into a rich discourse on translation. In M. Carl (Ed.), *Explorations in Empirical Translation Process Research* (pp. 141-163). Cham, Switzerland: Springer. doi:10.1007/978-3-030-69777-8_6

Ogle, K. N., Mussey, F., & De, H. P. A. (1949). Fixation disparity and the fusional processes in binocular single vision. *American Journal of Ophthalmology, 32*(8), 1069-1087.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1-4.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1-8.

Paradis, M. (1994). Toward a neurolinguistic theory of simultaneous translation: The framework. *International Journal of Psycholinguistics, 10*(3), 319-335.

Paradis, M. (2004). *A Neurolinguistic Theory of Bilingualism*. Amsterdam: John Benjamins.

Paradis, M., Goldblum, M.-C., & Abidi, R. (1982). Alternate antagonism with paradoxical translation behavior in two bilingual aphasic patients. *Brain and Language, 15*, 55-69.

Parr, T., & Friston, K. (2017). Uncertainty, epistemics and active inference. *J. R. Soc. Interface, 14*. doi:10.1098/rsif.2017.0376

Parr, T., Pezzulo, G., & Friston, K. (2022). *Active Inference: The Free Energy Principle in Mind, Brain, and Behavior*. MA: The MIT Press.

Patel-Schneider, P. F. (1989). A four-valued semantics for terminological logics. *Artificial Intelligence, 38*(3), 319-351.

Peckham, R. H. (1934). Foveal projection during ductions. *Archives of Ophthalmology, 12*(4), 562-566.

Pérez-Ortiz, J. A., Forcada, M. L., Sánchez-Martínez, F. (2022). How neural machine translation works. In D. Kenny (ed.), *Machine Translation for Everyone: Empowering Users in the Age of Artificial Intelligence* (pp. 141–164). Berlin: Language Science Press.

Peyraube, A. (1973). Westernization of Chinese grammar in the 20th century: Myth or reality? *Journal of Chinese Linguistics, 28*(1), 1-25.

Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: a critical review and future directions. *Psychon Bull Rev, 21*(5), 1112-1130. doi:10.3758/s13423-014-0585-6

Piccinini, G., & Scarantino, A. (2010). Computation vs. information processing: why their difference matters to cognitive science. *Studies in History and Philosophy of Science Part A, 41*(3), 237-246. doi:10.1016/j.shpsa.2010.07.012

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*(3), 427-459. doi:10.1037/0033-2909.134.3.427

Pierce, J. R. (1980). *An Introduction to Information Theory: Symbols, Signals and Noise*. New York: Dover Publications.

Plass, J. L., Moreno, R., & Brünken, R. (Eds.). (2010). *Cognitive Load Theory*. Cambridge: Cambridge University Press.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56.

Porterfield, W. (1737). An essay concerning the motions of our eyes. Part I. Of their external motions. *Edinburgh Medical Essays and Observations, 3*, 160-263.

Potter, M. C., & Lombardi, L. (1998). Syntctic priming in immediate recall of sentences. *Journal of Memory and Language, 38*, 265-282.

Prior, A., Wintner, S., Macwhinney, B., & Lavie, A. (2011). Translation ambiguity in and out of context. *Applied Psycholinguistics, 32*(1), 93-111.

Pritchett, B. L. (1988). Garden path phenomena and the grammatical basis of language processing. *Language, 64*(3), 539-576.

Pritchett, B. L. (1991). Head position and parsing ambiguity. *Journal of Psycholinguistic Research, 20*, 251-270.

Pritchett, B. L. (1992). *Grammatical Competence And Parsing Performance*. Chicago, IL: The University of Chicago Press.

Rabadán, R. (1991). The unit of translation revisited. In M. L. Larson (Ed.), *Translation: Theory and Practice: Tension and Interdependence* (pp. 38-48). New York: State University of New York at Binghamton.

Rao, R. P., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (1996). Modeling saccadic targeting in visual search. *Advances in Neural Information Processing Systems*, *8*, 830-836.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, 85*(3), 618–660. doi:/10.1037/0033-2909.85.3.618

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*(3), 191-201.

Rayner, K., & Fischer, M. H. (1996). Mindless reading revisited: Eye movements during reading and scanning are different. *Perception & Psychophysics, 58*(5), 734-747.

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the EZ Reader model. *Journal of Experimental Psychology: Human Perception and Performance, 30*(4), 720-732.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22*(3), 358-374.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of experimental Psychology: Human Perception and Performance, 22*, 1188-1200.

Reichle, E. D. (2011). Serial-attention models of reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford Handbook of Eye Movements* (pp. 767-786). Oxford: Oxford University Press.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems, 10*(4), 341-350. doi:10.1088/0954-898X_10_4_304

Reiss, K. (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*. Munich: Max Hueber.

Risku, H., & Rogl, R. (2020). Translation and situated, embodied, distributed, embedded and extended cognition. In F. Alves & A. L. Jakobsen (Eds.), *The Routledge Handbook of Translation and Cognition* (pp. 478-499). New York: Routledge.

Ruiz, C., Paredes, N., Macizo, P., & Bajo, M. T. (2008). Activation of lexical and syntactic target language properties in translation. *Acta psychologica, 128*(3), 490-500.

Salapatek, P., & Kessen, W. (1966). Visual scanning of triangles by the human newborn. *Journal of Experimental Child Psychology, 3*(2), 155-167.

Salge, C., Ay, N., Polani, D., & Prokopenko, M. (2015). Zipf's law: balancing signal usage cost and communication efficiency. *PLoS ONE, 10*(10), e0139475.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1*(4), 201-220.

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. *Proceedings of the Eye Tracking Research and Applications Symposium* (pp. 71-78). New York: ACM Press.

Schaeffer, M., & Carl, M. (2013). Shared representations and the translation process: A recursive model. *Translation and Interpreting Studies, 8*(2), 169-190. doi:10.1075/tis.8.2.03sch

Schaeffer, M., Carl, M., Lacruz, I., & Aizawa, A. (2016). *Measuring cognitive translation effort with activity units.* Proceedings of the 19th Annual Conference of the European Association for Machine Translation, 2016.

Schaeffer, M., Dragsted, B., Hvelplund, K. T., Balling, L. W., & Carl, M. (2016). Word translation entropy: Evidence of early target language activation during reading for translation. In M. Carl, S. Bangalore, & M. Schaeffer (Eds.), *New Directions in Empirical Translation Process Research: Exploring the CRITT TPR-DB* (pp. 183-210). Cham: Springer.

Schaeffer, M., Paterson, K. B., McGowan, V. A., White, S. J., & Malmkjaer, K. (2017). Reading for translation. In A. L. Jakobsen & B. Mesa-Lao (Eds.), *Translation in Transition: Between Cognition, Computing and Technology*. Amsterdam/Philadelphia: John Benjamins.

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition, 26*(6), 1270-1281.

Schilperoord, J. (1996). *It's about Time: Temporal Aspects of Cognitive Processes in Text Production* (Vol. 6). Rodopi.

Schultze, F. (1880). *Die Sprache des Kindes: eine Anregung zur Erforschung des Gegenstandes.* Günther.

Schuster, M., Johnson, M., & Thorat, N. (2016). *Zero-shot translation with Google's multilingual neural machine translation system*. Retrieved 17 November 2017 from https://research.googleblog.com/2016/11/zero-shot-translation-with-googles.html

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language, 55*(2), 197-212.

Seeber, K. G. (2015). Cognitive load in simultaneous interpreting: Measures and methods. In M. Ehrensberger-Dow, S. Göpferich, & S. O'Brien (Eds.), *Interdisciplinarity in Translation and Interpreting Process Research* (pp. 18-33). Amsterdam / Philidelphia: John Benjamins.

Seeber, K. G., & Kerzel, D. (2011). Cognitive load in simultaneous interpreting: Model meets data. *International Journal of Bilingualism, 16*(2), 228-242. doi:10.1177/1367006911402982

Shannon, C. E. (1948). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review, 5*(1), 3-55. doi:10.1145/584091.584093

Shannon, C. E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana: University of Illinois Press.

Sharafi, Z., Soh, Z., & Guéhéneuc, Y.-G. (2015). A systematic literature review on the usage of eye-tracking in software engineering. *Information and Software Technology, 67*, 79-107.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237*(4820), 1317-1323.

Shreve, G. M., & Lacruz, I. (2017). Aspects of a cognitive model of translation. In J. W. Schwieter & A. Ferreira (Eds.), *The Handbook of Translation and Cognition* (pp. 127-143). New York: Wiley Blackwell.

Shreve, G. M., Lacruz, I., & Angelone, E. (2010). Cognition effort, syntactic disruption, and visual interference in a sight translation task. In G. M. Shreve & E. Angelone (Eds.), *Translation and Cognition* (pp. 63-84). Amsterdam: John Benjamins.

Shterionov, D., Nagle, P., Casanellas, L., Superbo, R., & O'Dowd, T. (2017). *Empirical evaluation of NMT and PBSMT quality for large-scale translation production.* The 20th Annual Conference of the European Association for Machine Translation (29–31 May 2017, Prague, Czech Republic): Conference Booklet, User Studies and Project/Product Descriptions, 2017.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika, 42*(3/4), 425-440.

Soluch, P., & Tarnowski, A. (2013). Eye-tracking methods and measures. In S. Grucza, M. Płużyczka, & J. Zając (Eds.), *Translation Studies and eye-tracking analysis* (pp. 85-104). Frankfurt am Main: Peter Lang.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and Cognition*. Oxford: Blackwell.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology, 24*(6), 1521-1543.

Stanley, M. H. R., Buldyrev, S. V., Havlin, S., Mantegna, R. N., Salinger, M. A., & Stanley, H. E. (1995). Zipf plots and the size distribution of firms. *Economics Letters, 49*, 453-457.

Staub, A. (2007). The parser doesn′t ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(3), 550-569. doi:10.1037/0278-7393.33.3.550

Stefanics, G., Kremlacek, J., & Czigler, I. (2014). Visual mismatch negativity: a predictive coding view. *Front Hum Neurosci, 8*, 666. doi:10.3389/fnhum.2014.00666

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 27, 2014.

Swadesh, M. (1960). On the unit of translation. *Anthropological Linguistics, 2*(2), 39-42.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257-285.

Sweller, J. (2016). Cognitive Load Theory, Evolutionary Educational Psychology, and Instructional Design. In Geary, D., Berch, D. (eds.), *Evolutionary Perspectives on Child Development and Education* (pp. 291-306). Cham: Springer. doi:10.1007/978-3-319-29986-0_12

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory*. New York, NY: Springer.

Sweller, J., Mawer, R. F., & Ward, M. R. (1983). Development of expertise in mathematical problem solving. *Journal of Experimental Psychology: General, 112*(4), 639.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3).

Tatsumi, M. (2010). *Post-editing Machine Translated Text in a Commercial Setting: Observation and Statistical Analysis.* (PhD thesis), Dublin City University, Dublin, Ireland.

Tatsumi, M., & Roturier, J. (2010). *Source text characteristics and technical and temporal post-editing effort: what is their relationship.* Proceedings of the Second Joint EM+/CNGL Workshop Bringing MT to the User: Research on Integrating MT in the Translation Industry (JEC 10), 2010.

Teixeira, C. d. S. C. (2014). *The Impact of Metadata on Translator Performance: How Translators Work with Translation Memories and Machine Translation.* (PhD thesis), Universitat Rovira i Virgili, Tarragona.

Temnikova, I. P. (2010, May). *Cognitive evaluation approach for a controlled language post-editing experiment.* Paper presented at the Seventh International Conference on Language Resources and Evaluation, Valletta, Malta, 17-23 May 2010. European Language Resources Association (ELRA).

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *Behavioral and Brain Sciences, 24*(4), 629-640.

Theeuwes, J., & Burger, R. (1998). Attentional control during visual search: the effect of irrelevant singletons. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1342.

Thompson, J. C. (1970). The illusory co-verb. *Journal of the Chinese Language Teachers Association, 1*, 15-17.

Thorndike, E. L. (1921). *The Teacher's Word Book*. Teachers College, Columbia University.

Thorndike, E. L. (1932). *A Teacher's Word Book of the Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People*. New York: Bureau of Publications.

Thorndike, E. L. (1937). On the number of words of any given frequency of use. *Psychological Record, 1*, 399-406.

Thorndike, E. L. (1941). *The Thorndike-Century Senior Dictionary*. New York: Appleton-Century.

Thorndike, E. L., & Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. Teachers College, Columbia University.

Tinker, M. A. (1946). The study of eye movements in reading. *Psychological Bulletin, 43*(2), 93–120. doi:10.1037/h0063378

Tirkkonen-Condit, S. (1993). Choice in translation: A challenge to theory and practice. In S. Tirkkonen & J. Laffling (Eds.), *Recent Trends in Empirical Translation Research* (pp. 5-9). Joensuu: Faculty of Arts, University of Joensuu.

Tirkkonen-Condit, S. (2002). Translationese – A myth or an empirical fact? *Target, 14*(2), 207-220.

Tirkkonen-Condit, S. (2004). Unique items — Over- or under-represented in translated language? In A. Mauranen & P. Kujamäki (Eds.), *Translation Universals: Do They Exist?* (pp. 177-184). Amsterdam/Philadelphia: John Benjamins.

Tirkkonen-Condit, S. (2005). The monitor model revisited: Evidence from process research. *Meta, 50*(2), 405-414. doi:10.7202/010990ar

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes, 22*(5), 727-779. doi:10.1080/01690960601057068

Toury, G. (1995). *Descriptive Translation Studies and Beyond*. Amsterdam / Philidelphia: John Benjamins.

Traxler, M. J., & Pickering, M. J. (1998). Adjunct attachment is not a form of lexical ambiguity resolution. *Journal of Memory and Language, 39*, 558-592.

Turner, A. M., Bergman, M., Brownstein, M., Cole, K., & Kirchhoff, K. (2014). A comparison of human and machine translation of health promotion materials for public health practice: time, costs, and quality. *Journal of Public Health Management and Practice: JPHMP, 20*(5), 523-529. doi:10.1097/PHH.0b013e3182a95c87

Underwood, G. & Everatt, J. (1992). The role of eye movements in reading: Some limitations of the eye-mind assumption. In E. Chekaluk & K. R. Llewellyn (Eds.), *The Role of Eye Movements in Perceptual Processes* (pp. 111-169). North-Holland.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing, and Use* (pp. 153-168).

Underwood, N., & Jongejan, B. (2001). *Translatability checker: A tool to help decide whether to use MT.* Proceedings of MT Summit VIII: Machine Translation in the Information Age, 2001.

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*, 458-483.

van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177. doi:10.1007/s10648-005-3951-0

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., Polosukhin, I. (2017). *Attention is all you need*. arXiv. Doi: 10.48550/ARXIV.1706.03762

Vauquois, B. (1968). *A survey of formal grammars and algorithms for recognition and transformation in mechanical translation*. Paper presented at the Internal Federation for Information Processing Congress, Edinburgh, Scotland, UK.

Verheijen, F. (1961). A simple after image method demonstrating the involuntary multidirectional eye movements during fixation. *Optica Acta: International Journal of Optics, 8*(4), 309-312.

Vieira, L. N. (2016). *Cognitive Effort in Post-Editing of Machine Translation: Evidence from Eye Movements, Subjective Ratings, and Think-Aloud Protocols.* (PhD thesis), Newcastle University.

Vilinsky, B. (2012). On the lower frequency of occurrence of Spanish verbal periphrases in translated texts as evidence for the unique items hypothesis. *Across Languages and Cultures, 13*(2), 197-210. doi:10.1556/Acr.13.2012.2.4

Wade, N., & Tatler, B. W. (2005). *The Moving Tablet of the Eye: The Origins of Modern Eye Movement Research*. New York: Oxford University Press.

Wang, F., & Mei, D. (2017). 不同方向的口译过程信息加工与工作记忆的关系———兼议"非对称有限并行模型". *中国翻译 [Chinese Translators Journal], 38*(4), 38-44.

Wang, L., & Liu, S. (2020). 论纵 / 横范式视域下译者双语关联 / 非关联块构表征加工 [On the bilingual (non-) correlated chunk representations processing in interpreting from the perspective of vertical/horizontal processing paradigms]. *解放军外国语学院学报 [Journal of PLA University of Foreign Languages], 43*(3), 145-151.

Wasow, T. (2015). Ambiguity avoidance is overrated. In S. Winkler (Ed.), *Ambiguity: Language and Communication* (pp. 29-47). Berlin: De Gruyter Mouton.

Wei, Y. (2017). *Machine Translation for Scientific Abstracts: A Case Study on Lexical Customization with Applied Optics.* (M.Phil. thesis), The Chinese University of Hong Kong, Hong Kong.

Wei, Y. (2018, November). *Cognitive effort in relation to word translation entropy and syntactic choice entropy.* Paper presented at the 5th International Conference on Cognitive Research on Translation and Interpreting, Renmin University of China, Beijing, China, 3-4 November 2018.

Wells, W. (1794a). Reply to Dr. Darwin on vision. *The Gentleman's Magazine and Historical Chronicle, 64*, 794–797.

Wells, W. (1794b). Reply to Dr. Darwin on vision. *The Gentleman's Magazine and Historical Chronicle, 64*, 905-907.

Wells, W. C. (1792). *An Essay Upon Single Vision with Two Eyes: Together with Experiments and Observations on Several Other Subjects in Optics*. London: Cadell.

Westheimer, G. (1973). Saccadic eye movements. In V. Zikmund (Ed.), *The Oculomotor System and Brain Functions* (pp. 59-77). Sevenoaks, Kent: Butterworth.

Wiener, N. (1948). *Cybernetics, or, Control and Communication in the Animal and the Machine*. New York: MIT Press.

Wiener, N. (1954). *The Human Use of Human Beings: Cybernetics and Society*. New York: Avon Books.

Wigner, E. P. (1990). The unreasonable effectiveness of mathematics in the natural sciences. *Communications on Pure and Applied Mathematics, XIII*, 1-14.

Wilss, W. (1982). *The Science of Translation: Problems and Methods* (Vol. 180). Tübingen: G. Narr.

Wu, D. (1995). *Stochastic inversion transduction grammars, with application to segmentation, bracketing, and alignment of parallel corpora.* 14th International Joint Conference on Artifical Intelligence, San Fransisco, CA, 1995.

Wu, D. (1997). Stochastic inversion transduction grammars and bilingual parsing of parallel corpora. *Computational Linguistics, 23*(3), 377-403.

Wu, Y. J., & Thierry, G. (2010). Chinese-English bilinguals reading English hear Chinese. *J Neurosci, 30*(22), 7646-7651. doi:10.1523/JNEUROSCI.1602-10.2010

Wu, Y. J., & Thierry, G. (2012). Unconscious translation during incidental foreign language processing. *Neuroimage, 59*(4), 3468-3473.

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., . . . Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *arXiv preprint arXiv:1609.08144*.

Xiao, R. (2010). How different is translated Chinese from native Chinese? *International Journal of Corpus Linguistics, 15*(1), 5-35. doi:10.1075/ijcl.15.1.01xia

Xiao, R., & Dai, G. (2014). Lexical and grammatical properties of Translational Chinese: Translation universal hypotheses reevaluated from the Chinese perspective. *Corpus Linguistics and Linguistic Theory, 10*(1), 11-55.

Yamashita, H., Hirose, Y., & Packard, J. L. (Eds.). (2011). *Processing and Producing Head-final Structures*. London and New York: Springer.

Yamaya, A., Topić, G., Martínez-Gómez, P., & Aizawa, A. (2015). Dynamic-programming-based method for fixation-to-word mapping. In R. Neves-Silva, L. C. Jain, & R. J. Howlett (Eds.), *Intelligent Decision Technologies* (pp. 649-659). Cham: Springer.

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology, 97*(2), 259-268.

Young, L. R., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods & Instrumentation, 7*(5), 397-429.

Yu, S., & Bai, X. (2015). Rule-based Machine Translation. In S. W. Chan (Ed.), *The Routledge Encyclopedia of Translation Technology* (pp. 186-200). New York: Routledge.

Yule, C. U. (2014). *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

Zaretskaya, A., Vela, M., Pastor, G. C., & Seghiri, M. (2016). Measuring post-editing time and effort for different types of machine translation errors. *New Voices in Translation Studies* (15), 63-91.

Zhai, F., Zhang, J., Zhou, Y., & Zong, C. (2012). *Machine translation by modeling predicate-argument structure transformation.* Paper presented at the 24th International Conference on Computational Linguistics, Mumbai, India, 2012.

Zhu, Y. (2002). *A Study of the Historical Development of Chinese Coverbs -- with Special Reference to the Grammaticalization of the Coverb Ji.* (PhD thesis), The University of Minnesota.

Zipf, G. K. (1935). *The Psycho-biology of Language: An Introduction to Dynamic Philology*. Cambridge, MA: MIT Press.

Zipf, G. K. (1941). *National Unity and Disunity: The nation as a bio-social organism*. Bloomington: Principia Press.

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, MA: Addison-Wesley Press.

Appendix A Source text and Machine-translated output

(The following are the source text, the machine-translated output, and the tokenized text of the MT output, from the JIN15 dataset used in Chapter 7.)

Source Text

At Warwick, our commitment to be demonstrably a centre of world class research and innovation across all of our academic disciplines remains as strong as it always has been. We share resources and knowledge with academic communities throughout the world through collaborative partnerships. We present our major areas of research strength around key global priorities and challenges currently confronting the world, such as food security and energy. We concentrate on the positive impacts of Warwick research on society at large, particularly in areas of knowledge transfer. Our aim is to undertake exciting, ground-breaking, excellent and in many instances, policy-relevant research.

Machine-translated output

在华威仍强，因为它一直是我们的承诺，明确我们在所有学科的世界一流的研究和创新中心的。我们共享资源和知识，在世界各地建立合作伙伴关系与学术社区。我们提出我们主要围绕关键的全球优先事项和当前世界面临的挑战，如粮食安全和能源领域的科研实力。我们专注于华威社会研究，特别是在知识转移方面的积极影响。我们的目的是要进行令人兴奋的，突破性的，优秀的，在许多情况下，与政策相关的研究。

Tokenized (segmented) text

(The following text is the same as the machine-translated output above, and is displayed sentence by sentence.)

在 华威 仍强 ， 因为 它 一直 是 我们 的 承诺 ， 明确 我们 在 所有 学科 的 世界 一流 的 研究 和 创新 中心 的 。

我们 共享资源 和 知识 ， 在 世界各地 建立 合作伙伴 关系 与 学术 社区 。

我们 提出 我们 主要 围绕 关键 的 全球 优先 事项 和 当前 世界 面临 的 挑战 ， 如 粮食安全 和 能源 领域 的 科研 实力 。

我们 专注 于 华威 社会 研究 ， 特别 是 在 知识 转移 方面 的 积极 影响 。

我们 的 目的 是 要 进行 令人兴奋 的 ， 突破性 的 ， 优秀 的 ， 在 许多 情况 下 ， 与 政策 相关 的 研究 。

Appendix B　Post-edited versions from participants

Post-edited versions from all participants of JIN15 (tokenized)

(These texts are shown sentence by sentence below)

P01

在 华威 ， 我们 一如既往地 承诺 并 确立 本 校 在 所有 学科 的 世界 一流 研究 和 创新 中心 地位 。

我们 通过 在 世界 各地 建立 合作 伙伴 与 学术 研究 关系 , 实现 资源 和 知识 的 共享 。

我们 围绕 关键的 全球 重点 问题 和 当前 世界 面临 的 挑战 ， 如 粮食 安全 和 能源 领域 , 展现出 自己的 科研 实力 。

我们 关注 华威 所 取得 研究 在 广大 社会 范围 产生 的 积极 影响 ， 特别 是 在 知识 转移 方面 的 贡献 。

我们的 目标 是 进行 激动人心的 , 突破性的 优秀 研究 ， 以及 在 许多 情况 下 ， 着手 与 政策 相关的 研究 。

P02

华威 大学 一直 致力 于 将 我校 各个 学 科 打造 为 世界 级 研究 与 创新 的 中心 。

我们 通过 建立 合作 关系 , 与 世界 各地 的 学术 社区 共享 资源 和 知识 。

我们 研究 的 重点 领域 是 围绕 国际 上 的 优先 事项 以及 当今 世界 面临的 挑战 ， 比如 粮食 安全 和 能源 问题 。

我们 大体上 关注 华威 研究 对 社会 的 积极 影响 ， 尤其 是 在 知识 转移 这 一 领域 。

我们的 目标 是 要 从事 振奋人心 ， 具有 突破性 而又 出类拔萃 , 在 大多 情 况 下 与 国家 政策 紧密 相连 的 研究 。

P06

在 华威 ， 我们 致力 把 所有 学 科 打造 成为 世界 一流的 研究 和 创新 中心 ， 保持 他 一如既往的 强大 。

我们 与 世界 各地 范围内的 学术 团体 共享 资源 和 知识 ， 建立 合作 伙伴 关系 。

我们 的 科研 领域 主要 围绕着 全球 关键性的 首要 难题 和 当今 世界 面临的 挑战 ， 如 粮食 安全 和 能源 问题 。

我们 专注 于 华威 社会 研究 的 积极 影响 ， 特别 是 在 知识 转移 方面 。

我们 力求 着手 令人 兴奋的 、 突破性的 、 优秀的 、 及 在 多数 情况 下 与 政策 相关的 研究 。

P07

华威 大学 一直 倾力 于 在 各个 学科 成为 世界 一流的 研究 和 创新 中心 。

我们 共享 资源 ， 传播 知识 ， 并 在 世界 各地 建立 合作 伙伴 关系 ， 成立 学术 社区 。

针对 目前 世界 面临的 挑战 ， 如 食品 安全 和 能源 问题 ， 华威 大学 的 主要 优势 在于 占据 着 国际 领先 地位 。

华威 大学 注重 研究 成果 对 社会 的 积极 影响 ， 特别 是 研究 成果 的 实践 价值 。

我们的 目标 是 进行 振奋人心 ， 具突破性 ， 杰出的 并 与 社会 息息相关的 政策性 研究 。

P08

在 华威 ， 我们 致力 在 我们 所有 学科 当中 ， 成为 世界 一流的 研究 和 创新 中心 的 愿望 ， 一直 强烈 如昔 。

我们 通过 协调 的 合作 伙伴 关系 ， 与 世界 各地的 学术 界 共享 资源 和 知识 。

我们 围绕 着 关键的 全球 优先 事项 和 当前 世界 面临的 挑战 ， 如 食品 安全 和 能源 领域 ， 体现了 我们 主要 领域 的 科研 实力 。

我们 大体 专注 于 华威 的 科研 对 社会 ， 特别 是 在 知识 传授 方面 的 积极 影响 。

我们的 目的 是 要 进行 令人兴奋 、 突破性 、 优秀 ， 并 在 许多 情况 下 ， 与 政策 相关 的 研究 。

P09

在 华威 ， 我们 一如既往 地 承诺 , 所有的 学科 都 要 明确 成为 世界 一流的 研究 中心 和 创新 中心 。

我们 和 世界 各地的 学术 团体 通过 合作 关系 共享 资源 和 知识 。

我们 的 研究 领域 主要 集中 在 如 食品 安全 和 能源 问题 这样 关键性的 全球 突出 事项 和 当前 世界 面临的 挑战 上 .。

我们 专注 于 华威 的 社会 研究 ， 特别 是 在 知识 转移 方面 的 积极 影响 。

我们的 目标 是 开展 振奋的 ， 突破性的 ， 优秀的 ， 和 在 多数 情况 下 ， 与 政策 相关的 研究 。

P11

在 华威 ， 我们 旨 在 成为 世界级 的 研究 和 创新 中心 ， 保持 一直 以来 的 优势 地位 。

我们 共享 资源 和 知识 ， 在 世界 各地 建立 合作 伙伴 关系 与 学术 社区 。

我们 主要 围绕 紧要的 全球 事务 和 当今 世界 面临 的 挑战 ， 如 粮食 安全 和 能源 领域 ，。

并且 专注 于 华威 社会 研究 对 知识 迁移 方面 的 积极 影响 。

我们 要 进行 的 是 令人振奋的 ， 具有 突破性的 杰出 研究 ， 这些 研究 在 很多 情况 下 ， 与 政策 相关 。

P12

华威 ， 正如 我们 明确 的 承诺 ， 所有 学科 保持 世界 一流的 研究 和 创新 中心 。

与 世界 各地 建立 合作 ， 我们 在 学术 领域 中 共享 资源 和 知识 。

我们 主要的 研究 领域 是 围绕 重要的 全球 事宜 和 当前 世界 面临 的 挑战 ， 如 粮食 安全 和 能源 领域 的 科研 实力 。

我们 关注 华威 对 整个 社会 的 研究 影响 ， 特别 是 在 知识 传播 方面 的 。

我们的 目标 是 要 进行 活跃的 、 突破性的 、 优异的 和 多条件的 重要 政策 研究 。

P13

辅 以 一 系列 学科 的 创新 研究 ， 和 世界 一流的 研究 中心 ， 华威 大学 将 兑现 曾经 提出 的 口号 。

我们 同 世界 各地 的 学术 组织 建立 合作 关系 ， 并 共享 资源 和 知识 。

我们的 科研 实力 主要 涵盖 食品 安全 ， 能源 等 一些 全球 重要 领域 和 当前 世界 面临 的 挑战 。

我们 专注 于 华威 大学 的 研究 在 社会 上 的 积极 影响 ， 特别 是 在 知识 转移 方面 的 影响 。

我们的 目的 是 要 进行 震撼人心 ， 具有 突破性的 ， 卓越的 ， 并 在 很多 时候 与 政策 相关的 研究 。

P14

在 华威 ， 我们 致力 于 成为 一个 世界 一流的 、 囊括 所有 学科 的 研究 和 创新 中心 ， 并 不断 保持 这 一 优势 。

我们 与 学术 团体 共享 资源 和 知识 ， 在 世界 各地 建立 合作 伙伴 关系 。

我们 提出 我们 主要的 研究 优势 领域 都 是 围绕 主要的 全球 事务 和 当前 世界 面临 的 挑战 ， 如 粮食 安全 和 能源 。

我们 专注 于 华威 在 社会 上 的 积极 影响 ， 特别 是 在 知识 传播 领域 。

我们的 目标 是 要 进行 激动人心的 ， 突破性的 ， 优秀的 ， 而且 大多 与 政策 相关的 研究 。

P15

在 华威 , 我们 在 所有 学科 范围 成为 世界 一流的 研究 和 创新 中心 的 决心 始终 未 变 。

通过 合作 伙伴 关系 ， 我们 和 不同 学术 团体 共享 资源 和 知识 。

我们 围绕 全 社会 最为 关注的 问题 和 当今 世界 面临 的 挑战 ， 展现 我们 主要 科研 实力 ， 如 粮食 安全 和 能源 领域 方面 。

我们 专注 于华威 的 研究 成果 ， 尤其 是 在 知识 转移 方面 的 成果 ， 对 社会 整体的 积极 影响 。

我们的 目标 是 进行 令人兴奋 且 具有 突破性的 优秀 研究 ， 在 许多 情况 下 ， 这些 研究 与 政策 相关 。

P16

我们 承诺 将 华威 大学 展示 成 一个 在 所有 学科 上 世界 一流的 研究 和 创新 中心 ， 这 一 承诺 至今 仍然 强烈 。

通过 在 世界 各地 建立 合作 伙伴 关系 ， 我们 与 其他 学术 团体 共享 资源 和 知识 。

我们 围绕 关键的 全球 问题 和 当前 世界 面临的 挑战 展示 研究 优势 ， 如 在 粮食 安全 和 能源 领域 。

我们 专注 于 华威 的 研究 对 整个 社会 产生 的 积极 影响 ， 特别 是 在 知识 转移 方面 。

我们的 目的 是 要 进行 令人兴奋的 ， 具有 突破性的 ， 优秀的 ， 并且 在 许多 情况 下 ， 与 政策 相关的 研究 。

P17

华威 大学 承诺 始终 保持 我们 一直 以来 在 所有 学术 领域 中 世界 一流的 研究 创新 中心 的 地位 。

通过 协同 合作 ， 我们 与 世界 各地的 学术 社区 共享 资源 与 知识 。

我们 的 优势 研究 领域 主要 围绕 全球 重点 问题 和 当前 世界 面临的 挑战 ， 如 食品 安全 和 能源 问题 。

华威 的 研究 专注 于 对 社会 产生 广泛 正面的 影响 ， 尤其 是 在 知识 迁移 领域 。

我们的 目标 是 开展 令人振奋的 、 突破性的 杰出 研究 ， 大多 情况 下 ， 这些 研究 与 政策 相关 。

P18

华威 大学 承诺 ， 本校 在 所有 学科 的 研究 和 创新 将 一如既往地 保持 世界 一流 水平 。

我们 通过 与 世界 各地的 学术 机构 建立 合作 伙伴 关系 ， 共享 资源 和 知识 。

本校 研究 主要 围绕 关键的 国际 事务 和 当前 世界 面临的 挑战 ， 如 粮食 安全 和 能源 问题 。

我们 专注 于 华威 社 科 研究 的 广泛 积极 作用 ， 尤其 在 知识 交流 领域 的 影响 。

本校 旨 在 进行 令人兴奋的 、 具有突破性的 、 优秀的 、 以及 在 许多 情况 下 与 政策 相关的 研究 。

P19

将 华威 发展 为 覆盖 所有 学科 的 世界 一流的 研究 和 创新的 中心 ， 是 我们 自始至终的 追求 。

我们 通过 和 世界 各地的 学术 组织 建立 合作 关系 ， 从而 实现 资源 和 知识 的 共享 。

我们 显示 了 在 主要 研究 领域 的 科研 实力 ， 这些 领域 关乎 当今 世界 面临的 主要的 全球 问题 和 挑战 。

我们 专注 于 华威的 研究 对 社会 ，特别 是 在 知识 传递 方面 所 产生的 积极 影响 。

我们 意 在 致力 于 进行 振奋人心的 ，突破性的 ，优秀的 ，并 在 多数 情况 下 与 政策 相关的 研究 。

P20

在 华威 ，要 使 我们的 学术 领域 成为 世界 一流的 研究 和 创新 中心 的 承诺 依然 坚定 如 初 。

通过 建立 广泛的 合作 关系，我们 与 世界 各地的 学术 团体 共享 资源 和 知识 。

我们 的 主要 研究 领域 围绕 关键的 全球 热点 问题 和 当前 世界 面临的 挑战 展开 ，如 食品 安全 和 能源 问题 等 。

我们 专注 于 华威的 研究 对 整个 社会 ，特别 是 对 知识 传播 方面 产生的 积极 影响 。

我们的 目的 是 要 进行 激动人心的 ，突破性的 ，优秀的 ，以及 在 许多 情况 下 ，与 政策 相关的 研究 。

P21

在 华威 ，我们 依旧 努力 在 各 学 科 创建 世界 一流 研究 与 创新 中心 。

通过 建立 合作 伙伴 关系 ，我们 与 世界 各地的 学术 团体 共享 资源 与 知识 。

围绕 关键的 全球 议题 和 当前 世界 面临的 挑战 ，如 粮食 安全 以及 能源 问题 ，我们 展示出 自己 主要的 研究 强项 。

我们 专注 于 对 社会 研究 带来的 积极 影响 ，特别 是 在 知识 传递 方面 。

我们 致力 于 开展 激动人心 、突破性 、精彩的 以及 在 多数 情况 下 与 政策 相关的 研究 。

P22

在 华威 大学 ， 我们 始终 承诺 ， 在 所有 学科 领域 都 成为 世界 一流的 研究 和 创新 的 中心 。

我们 与 全 世界 的 学术 机构 建立 合作 伙伴 关系 , 共享 资源 和 知识 。

我们的 科研 实力 全球 优先 事项 和 当前 世界 面临的 挑战 等 主要 领域 ， 如 粮食 安全 和 能源 。

我们 专注 于 广泛的 社会 研究 ， 特别 是 知识 转移 领域 研究 所 带来的 积极 影响 。

我们的 目的 是 要 进行 令人兴奋 的 ， 突破性的 ， 优秀的 并且 在 许多 情况 下 与 政策 相关的 研究 。

Appendix C Data inaccuracy and re-processing of JIN15

(The following describes the issues in the JIN15 dataset, the possible reasons behind them, their potential to result in misleading conclusions if not corrected, and the ways in which these issues are dealt with in Chapter 7.)

As the data is collected from the experiments in Huang (2016), a look at the progression graphs in this work (mostly in Chapter 7 of Huang, 2016 — on "Working styles of student translators", pp. 189-231, which constitutes a major part of its analysis, findings, and conclusions) is perhaps the most reasonable starting point for this discussion. These graphs reveal a systematic drift of the gaze points in the data. The same is also reflected in the progression graphs in Huang (2018), which is based on the same dataset, and in the graphs which can be generated using the dataset downloadable from the TPR-DB website.

For example, Figure C.1 below — taken from Huang (2016, p. 227) — shows the process of an experimental session in terms of the fixations on the ST (blue dots), fixations on the TT (green diamonds), and the keyboard behaviour (insertions indicated by black characters and deletions by red characters). As described in Section 8.1, the horizontal axis indicates time during the experiment (in milliseconds) while the vertical axis (on the left side) describes the token ID along the ST sentence from the beginning to the end (i.e., the1st, 2nd, 3rd, …, 10th, …, 20th, …, 30th, etc. token of the ST) as well as the TT token ID numbers (on the right side) which are sorted in the order of the ST on the basis of word-level alignment. The position of each of the fixations (blue dots and green diamonds) and keyboard inputs (read/black characters) in the graph indicates both the time at which it occurs (horizontal axis) and the part of the ST/TT sentence in which it occurs (vertical axis). In this figure, it is obvious that the TT fixations (green diamonds) have systematically drifted downwards to a significant extent, with all these fixations vertically lower than the corresponding editing behaviour (red and black characters) as well as the corresponding ST fixations (blue dots). See the annotation in Figure C.2 for a better illustration of the vertical drift.

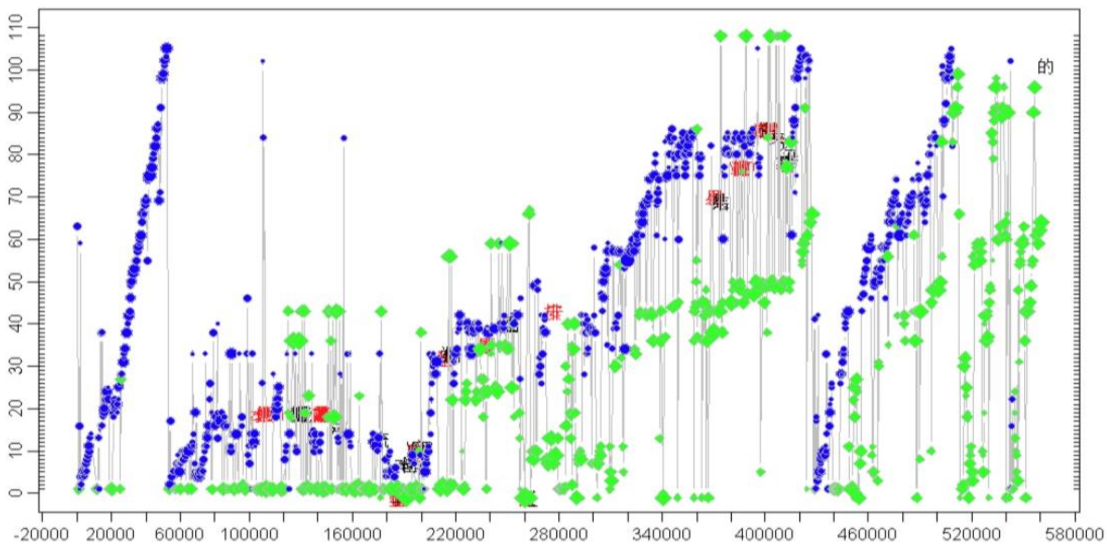


Figure C.1 Progression graph 1 for JIN15 dataset (from Huang, 2016, p. 227)

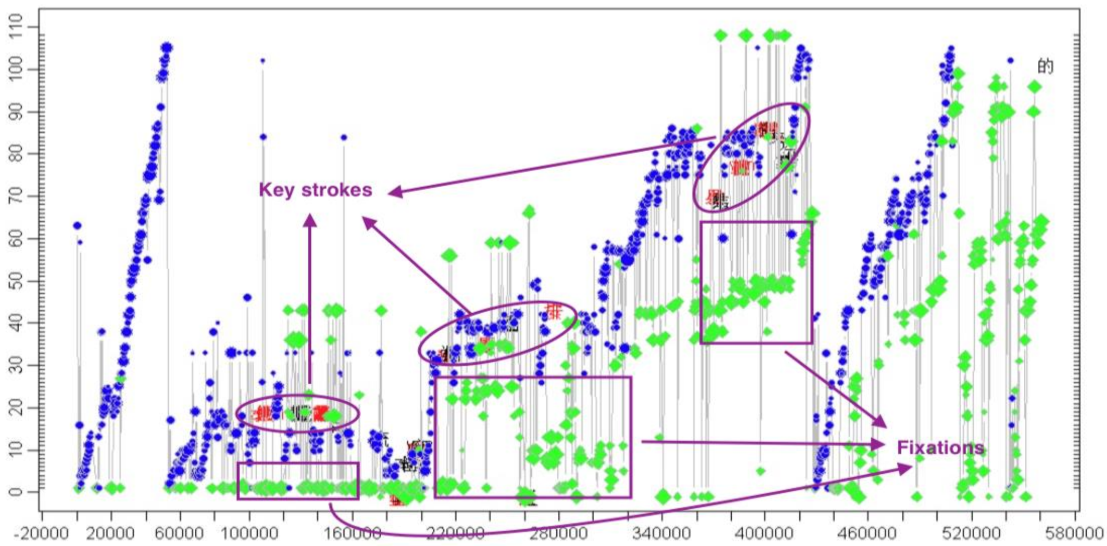


Figure C.2 Vertical drift in fixations in JIN15 dataset

The same can be seen from most of the other progression graphs in Huang (2016), as shown in Figures A.3 to A.6, with the TT (and sometimes ST) fixations systematically and consistently displaced from the sentence being edited (and mostly likely, the sentence being cognitively processed). See the green diamonds in Figures A.3, A.4, A.5, and A.6, as well as the blue dots in Figures A.5 and A.6.

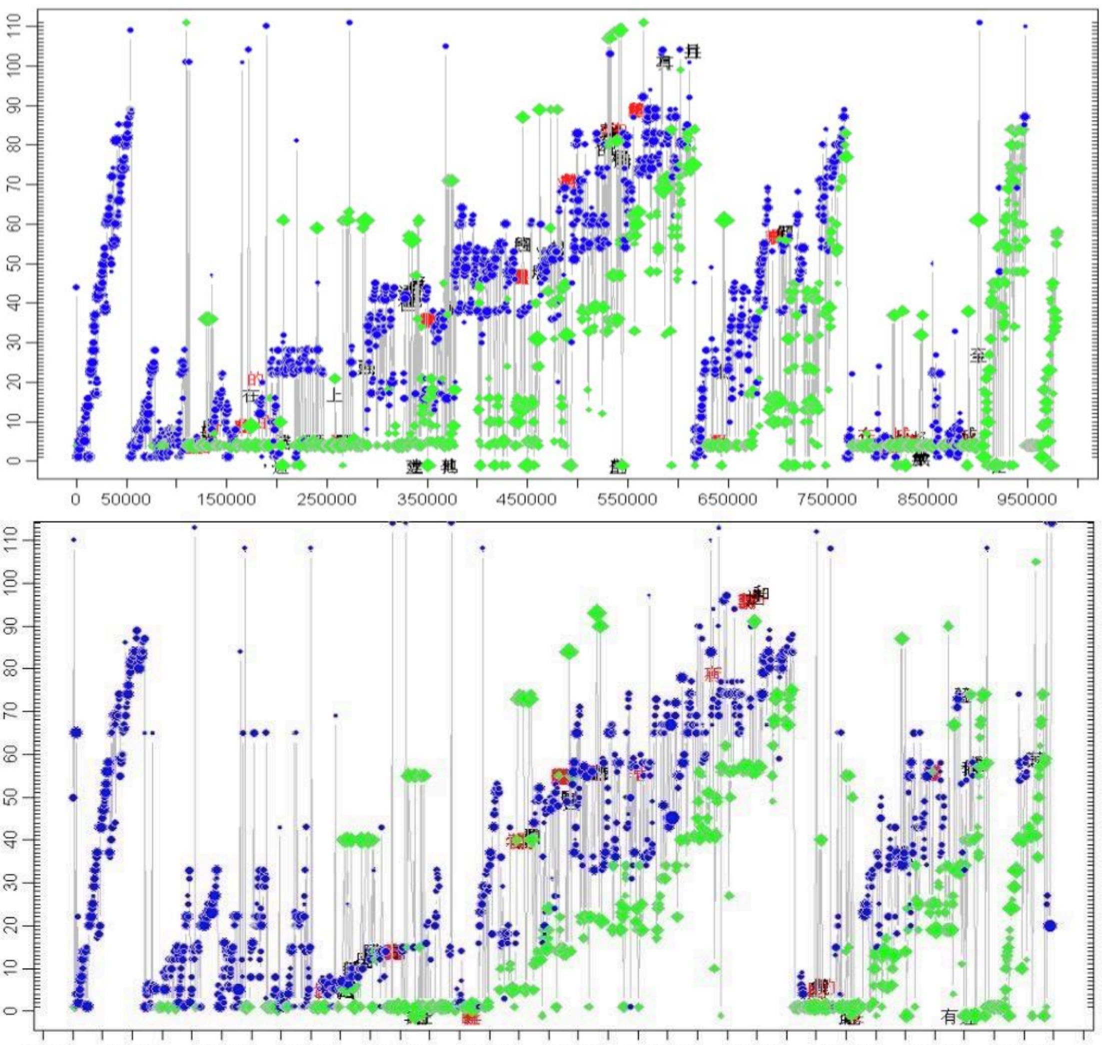


Figure C.3 Progression graphs 2 & 3 for JIN15 dataset (from Huang, 2016, also p. 227)

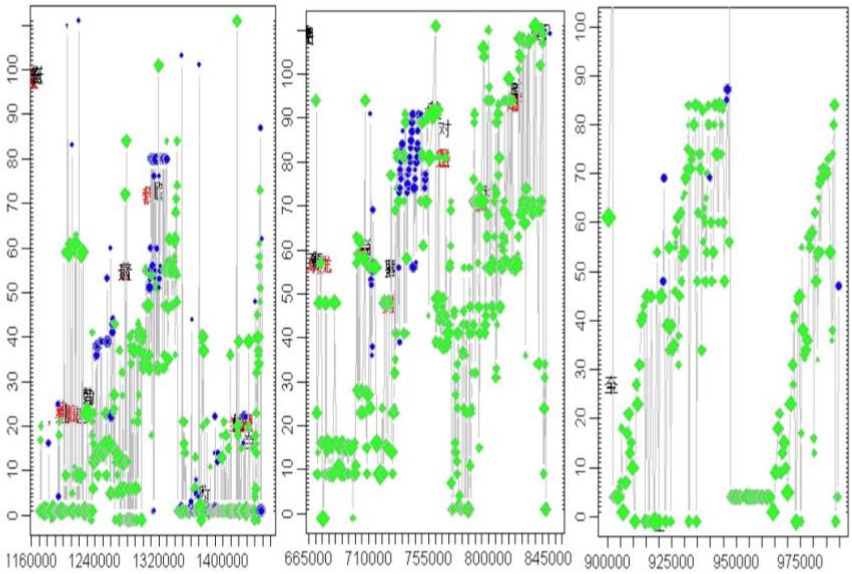


Figure C.4 Progression graph 4 for JIN15 dataset (from Huang, 2016, p. 213)

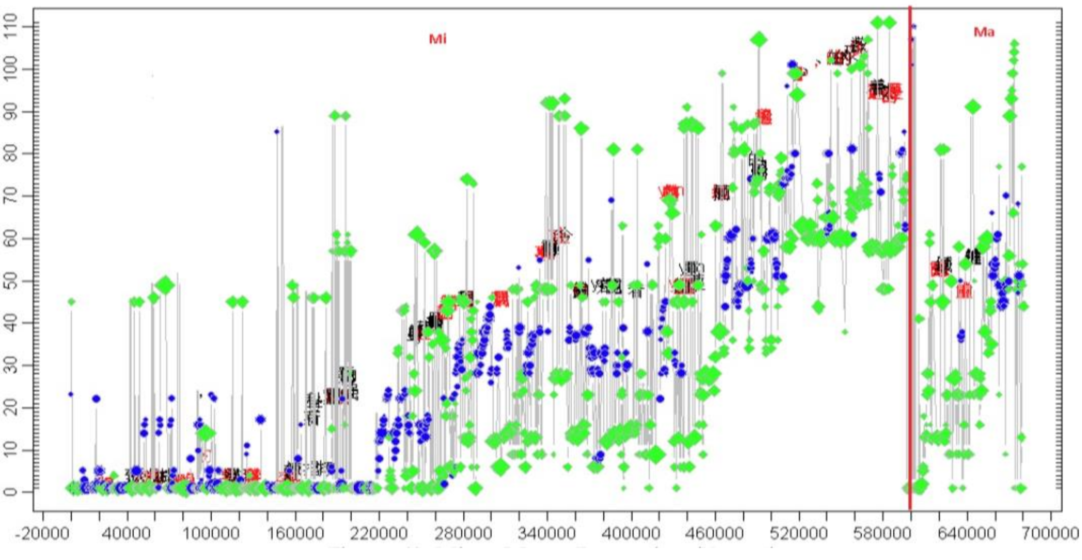


Figure C.5 Progression graphs 5 for JIN15 dataset (from Huang, 2016, p. 221)

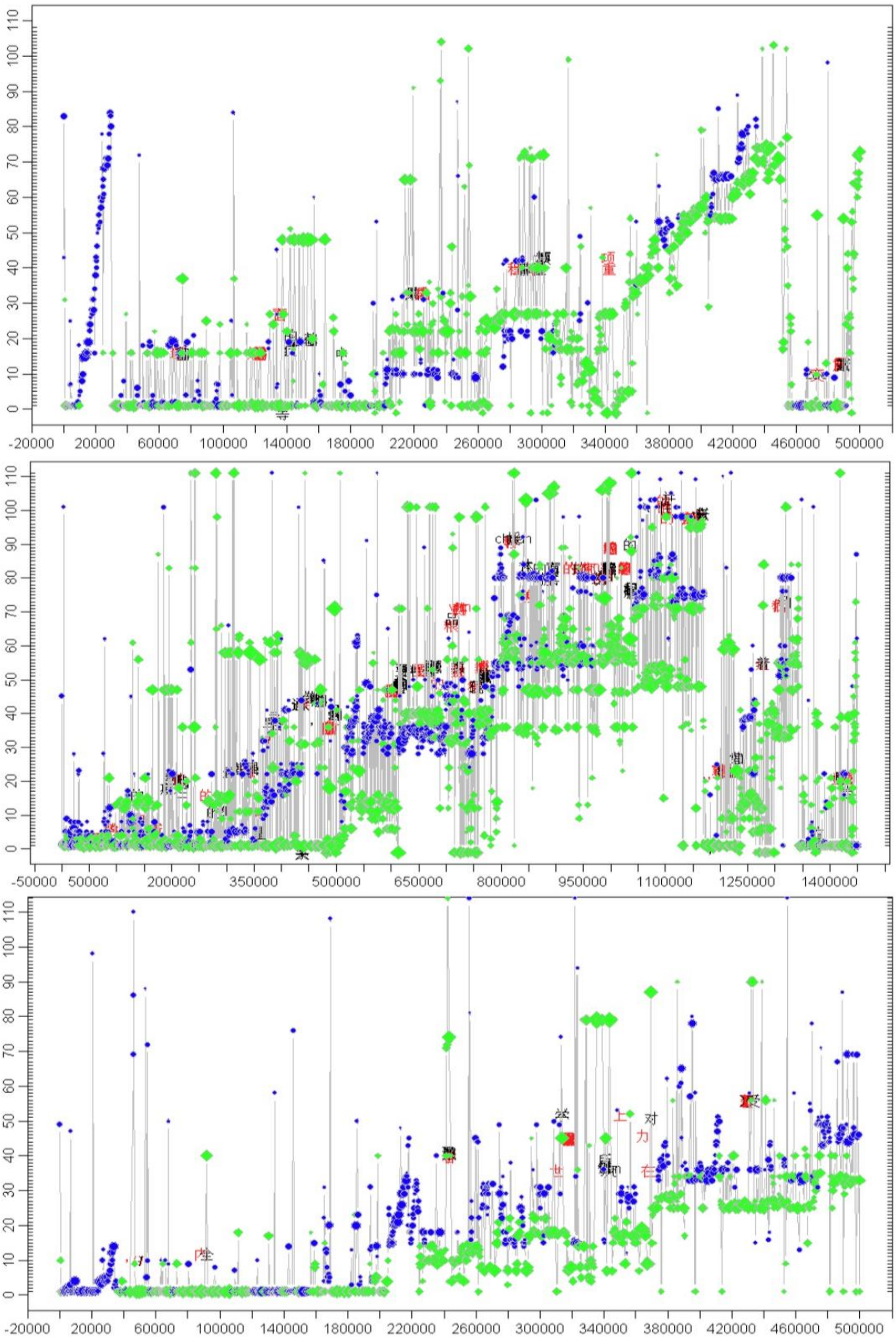


Figure C.6 Progression graphs 6-8 for JIN15 dataset (from Huang, 2016, p. 228)

Other than these, most of the progression graphs in Huang (2018), on the basis of which major conclusions on the working styles of student translators are drawn, show the same issue in relation to the displacement of the fixation points. On the other hand, there are a few graphs in Huang (2016) where these fixations are in the appropriate position, as can be seen in Figures A.7 and A.8.

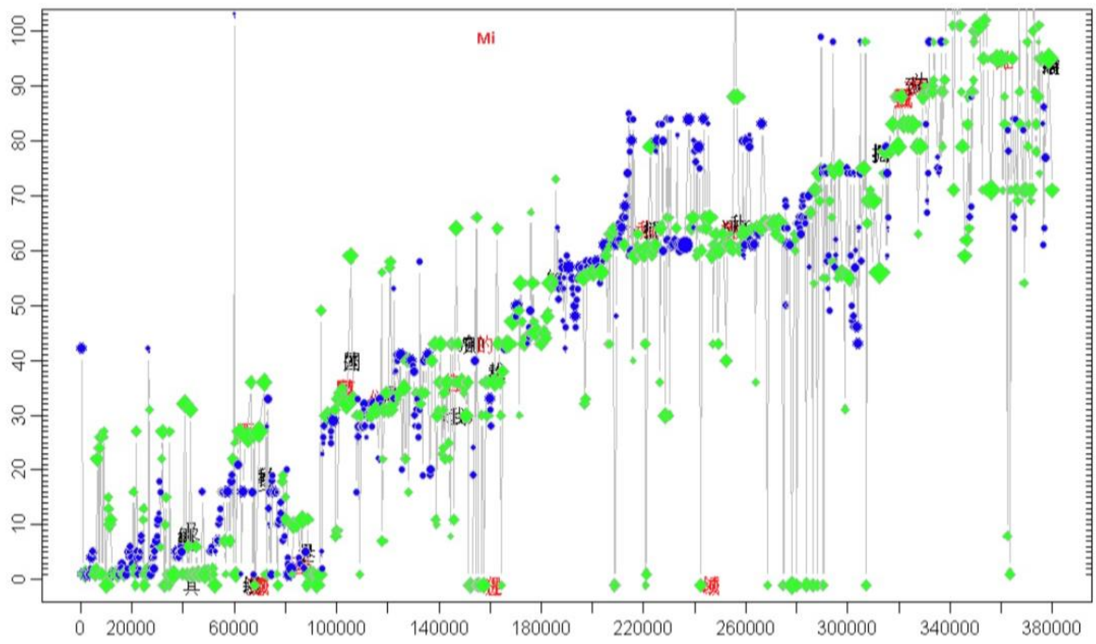


Figure C.7 Progression graph 9 for JIN15 dataset (from Huang, 2016, p. 223)

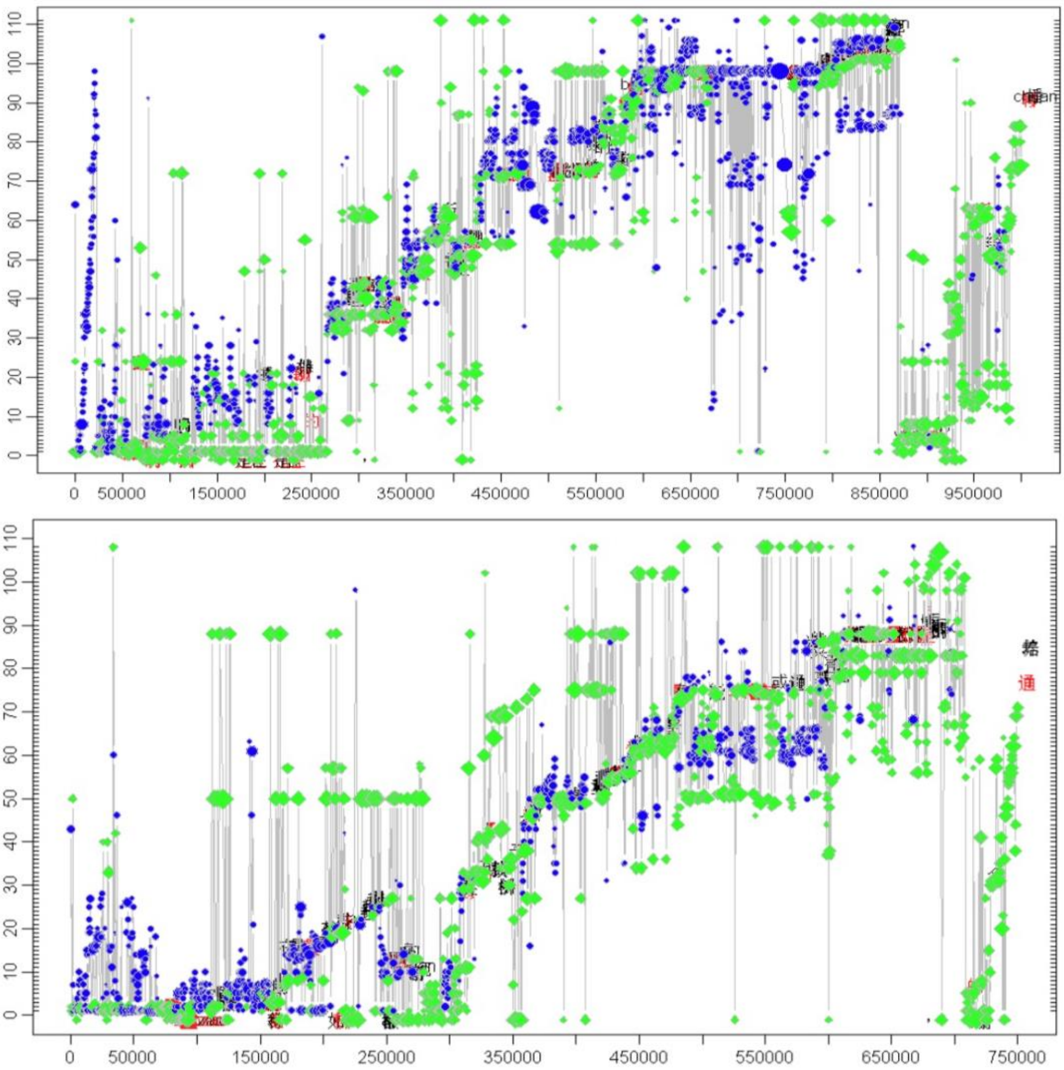


Figure C.8 Progression graphs 10 & 11 for JIN15 dataset (from Huang, 2016, p. 229)

In Figures A.7 and A.8, it can be seen that the fixations on the ST (blue dots), the fixations on the TT (green diamonds), and the keyboard records (red and black characters) are mostly aligned to one another. This means that when the participant is working on a particular sentence, the eye fixations are on the same sentence, not too far from the current word/phrase being edited. In Figures A.1, A.3, A.4, A.5, and A.6 shown above, however, the eye fixations are consistently on the preceding sentence when each sentence is being edited.

While this difference between the aligned and mis-aligned fixations is not addressed explicitly in Huang (2016) or Huang (2018), it seems that the vertical displacement of the gaze points (which is prominent in the progression graphs above) is somehow recognized as “inaccurate data” (Huang, 2016, p.117) in Huang’s description of a pilot study for the project (using a screenshot from Translog-II, see Section C.2 below) but was nevertheless ignored in the analyses on the progression graphs and largely used as a basis for drawing major conclusions on the translation behaviour, not considering the above progression graphs (and the issues therein) as problematic. In turn, these drifts of gaze points (i.e., editing a particular TT sentence while reading on the sentence prior to the one being edited, as represented in the progression graph by the green diamonds being considerably lower than the red/black characters) is somehow (perhaps wrongly) interpreted as “narrow-context planning” and “backtracking” (p. 203), as shown in Figure C.9 below, a figure taken from Huang (2016, p. 203) with her own annotations on the progression graph. The following sections will show that such interpretation can be seriously misleading, and that the drift in gaze data seems to be largely the result of inappropriate data processing.

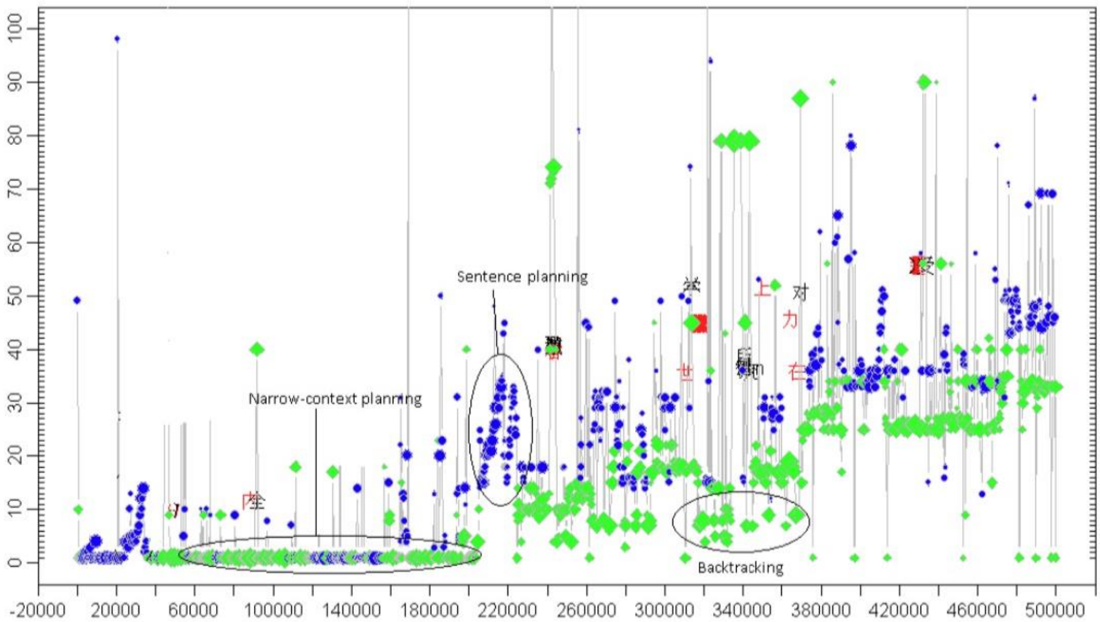


Figure C.9 Huang’s interpretation of the fixation points (from Huang, 2016, p. 203)

C.1 Translog replay analysis

While it is tempting to consider the observed pattern — of fixations being in an earlier position of the text than the part being edited — as a behavioural phenomenon worthy of scientific discussion (e.g., in terms of context planning and backtracking, as Huang 2016 seems to do), it is also important to closely and critically examine the dataset to see if this is, after all, a misleading presentation of the data, an artefact resulting from the way it is processed, or simply an error in the data itself.

This is important for the analysis in the present chapter, because if there are vertical errors in the dataset, such errors might lead to incorrect findings and conclusions.

In this regard, as the .xml file from Translog contains all the fixation information, the keyboard records, and the time stamps for each eye-movement and key-input activity, the experimental session can be conveniently replayed to show the details of the editing process from the participant in real time. This would be helpful for scrutinising the issue mentioned above and, more importantly, for clarifying whether the issue is a genuine reflection of the participants' behaviour or simply an effect from the data processing procedure, and for analysing the possible causes of the issue, on the basis of which solutions are arrived at in this chapter.

The following figures show the screenshots of a replay session in Translog-II (regarding the post-editing process of participant P16 as an example), using the downloadable JIN15 dataset from CRITT. Figure C.10 is the Translog-II user interface when the post-editing data of the participant (P16) is loaded to the software. This is the interface that the participant would see when the experimental session begins, and its layout is the same as the screenshot from Huang (2016, p. 87) as shown in Figure C.11. The upper window contains the ST, while the lower window shows the raw translation to be edited.

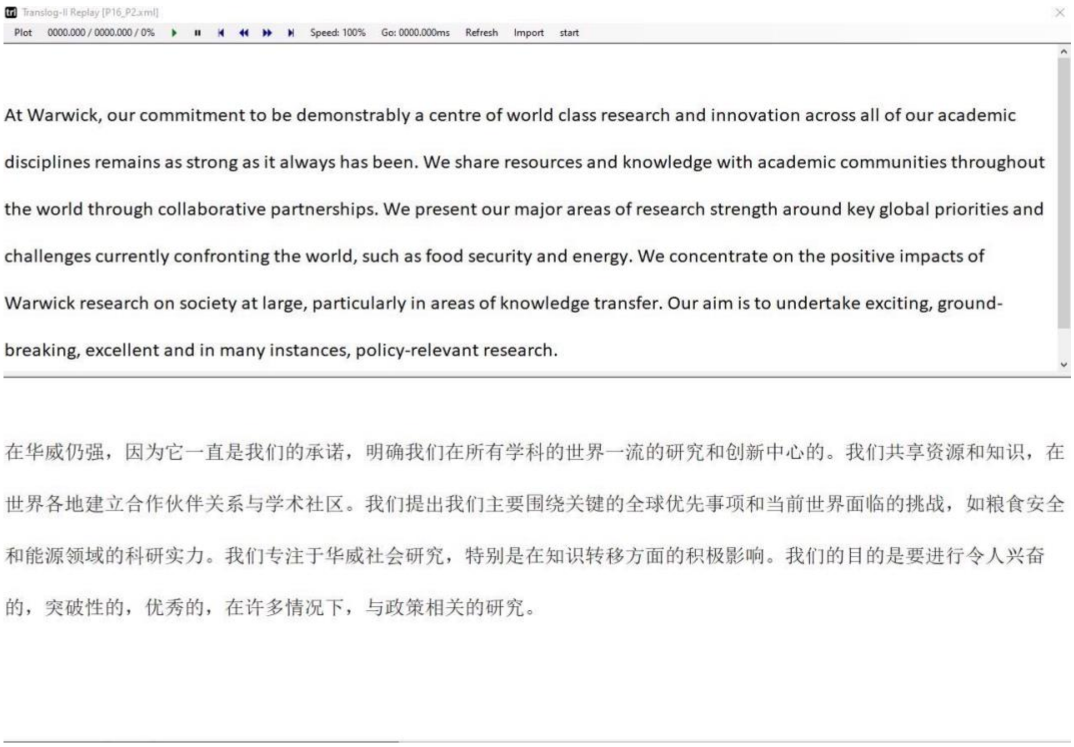


Figure C.10 Screenshot for the replay in Translog-II

Figure C.11 Huang’s Screenshot of Translog-II user interface (p. 87)

Replay of editing on sentence 1

When the participant is editing the first sentence in the TT window (as annotated in Figure C.12 for both the ST in the upper window and the TT in the lower window), the fixations from the beginning of his/her editing until moving on to the second sentence are shown in the following Figure C.13. Here, the red and green dots indicate the gaze points from each of the participant's eyes, while the blue circles are the fixations which are calculated from these gaze points. The violet squares (i.e., the highlights of the words in ST and TT) represent the characters onto which the fixations are mapped — essentially the closest characters to the fixation centres but adjustable via an algorithm which considers the translation behaviour (Carl, 2013), or alternatively via a manual process of fixing the word-to-fixation mappings in Translog-II (called FixMap).

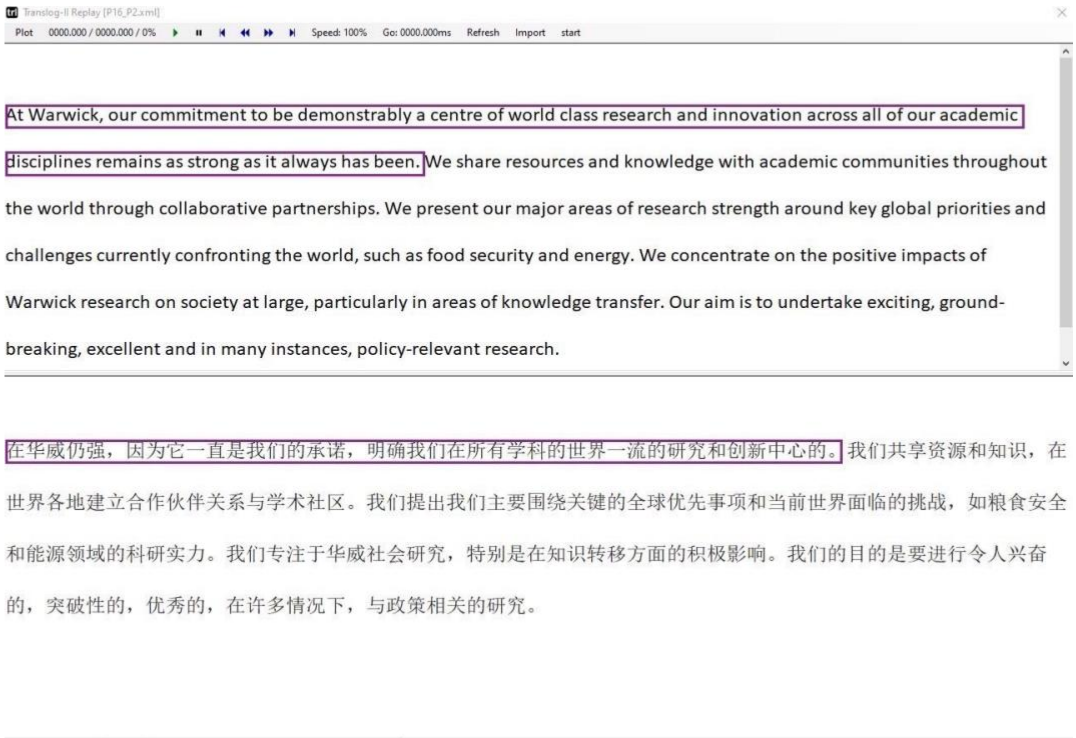


Figure C.12 Editing on the first sentence

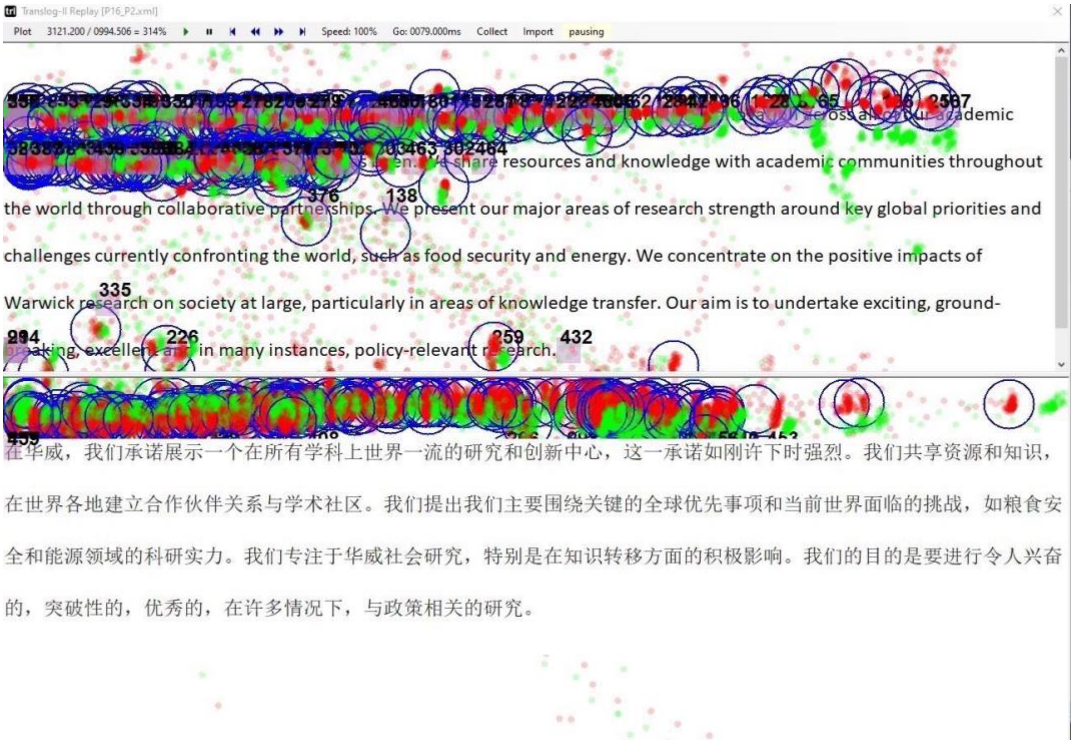


Figure C.13 Fixations during editing of the first sentence

It can be seen that the participant's reading on the ST follows exactly the position of the characters in the first ST sentence. This is stable, consistent, and reliable, following a straight line as the gaze path continues, and with this line of gaze path overlapping with the characters on the screen. Accordingly, the fixation-to-word mapping is accurately onto the characters in this sentence. When it comes to TT window, however, the position of the participant's eyes is apparently one line above the position of the TT sentence, located in the blank space above the text in the TT window. This gaze path on the TT, similarly, follows a straight line without any gradual drifting as the reading continues. A detailed look at the process of editing here also shows that the participant is indeed reading the first sentence of the TT, with the edited words consistently aligned with the TT fixations horizontally and consistently below these fixations vertically. When the fixation is above a particular TT token (often with editing behaviour regarding that token), the preceding and subsequent ST reading behaviour also involves the ST regions corresponding to the TT token in question. This means that the participant is in effect fixating the words in the first sentence of the TT, and that these TT fixations are systematically displaced to a line above the text. As this is the first sentence in the TT window, all these fixations are mapped to the beginning of the TT, i.e., the character "在", as can be seen from the violet highlight of the TT in Figure C.13.

As a result, this reading process would not be captured after such raw data is processed, and when shown in the progression graph, it is manifested by a number of TT fixations (green diamonds) on the first character of the TT sentence (i.e., a series of data points mostly at the bottom of the graph, as this is mostly aligned to the beginning of the ST as well). For this participant, specifically, Figure C.14 shows the displacement of fixations in the experimental session, while the progression graph resulting from this data (for the first sentence), is shown in Figure C.15, generated in RStudio using the raw JIN15 dataset (i.e., the same raw data used for the Translog replay above). See the green diamonds at the bottom in Figure C.15.

This issue is also apparent in Figures A.1 to A.8, with TT fixations located at the bottom of the progression graph when the editing is on the first sentence).

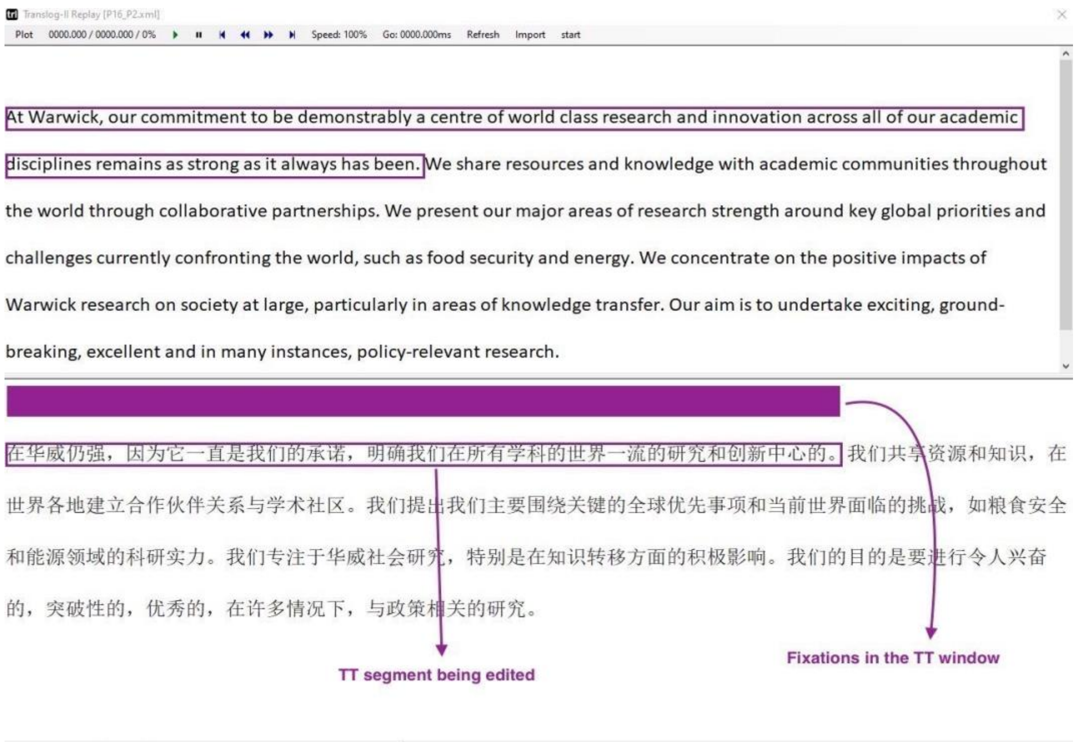


Figure C.14 The displacement of fixations for the first sentence

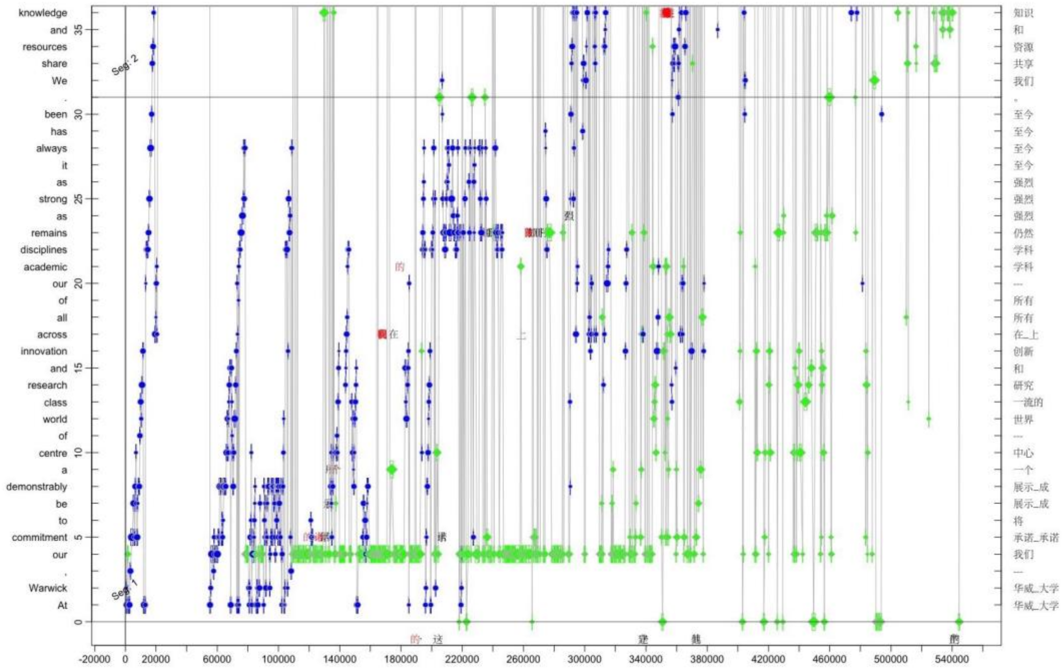


Figure C.15 Progression graph of P16 for the first sentence

Looking back at Figure C.9 with respect to the interpretation of these fixation points (see above), it seems obvious that considering this process as context-planning is flawed. Other than the fact that the TT reading behaviour here is directly related to editing (i.e., typing) behaviour when examined from the Translog replay session (and that in the graph, these TT fixations should overlap with the key strokes and the ST fixations), interpreting the green diamonds at the bottom of the graph as "narrow-context planning", when it is clearly and incorrectly just the first character of the text that is shown to be fixated without any indication — from the graph itself — of reading anywhere else until after the completion of editing on the sentence, would appear to be problematic.

More importantly, crucial information regarding the TT reading processes is lost, when all the fixations on different positions of the first sentence are mapped to just one single character (the beginning of the sentence). In addition, a perhaps far more severe result is that with this systematic drift, the fixations on the subsequent sentence would be incorrectly taken as fixations on the current one, which is shown below in more detail. If the problem with the first sentence is loss of information, then the issue regarding subsequent sentences would be that the data being analysed is entirely wrong and seriously misleading.

Replay of editing on sentence 2 and subsequent sentences

For the second sentence, Figure C.16 shows the position of the ST and TT while Figure C.17 shows the gaze points, fixations, and the characters onto which the fixations are mapped, while this sentence is being edited. Here, the ST fixations show slight and gradual drifting as the reading continues, with the latter part of the gaze path slightly moving upwards as it moves from the left to the right side of the window. Accordingly, the mapped words on the right-hand side start to become inaccurate as the fixations approach this area, if the simplest, "naïve" mapping algorithm (on the basis of the closest characters alone) is used. This is certainly a source of noise and another misleading aspect of the conclusions that can be drawn from the dataset used in Huang (2016). However, the drift here is very mild and can be easily corrected via the re-mapping algorithm from Carl (2013), which is shown to be effective and is incorporated in the Translog-II software.[74]

Regarding the TT fixations, on the other hand, the problem is far more severe and misleading. As can be seen from Figure C.17, during the reading on the second TT sentence, the fixations are systematically and consistently displaced one line above this sentence, starting from the blank space above the text (see fixations on the right side of the TT window in the figure) and proceeding to the first line of the TT, with the fixations first mapped onto the first TT character and then to the subsequent characters in the first line (see also Figure C.18 for illustration). The same is observed in all subsequent sentences in the experimental session.

[74] This algorithm is based on the following criteria (Carl, 2013):
1). successive fixations are more likely on neighbouring words than in the lines above or below;
2). translators are likely to read passages of source text words which they are currently translating;
3). the distance between the fixation centre and the fixated characters should be minimal.

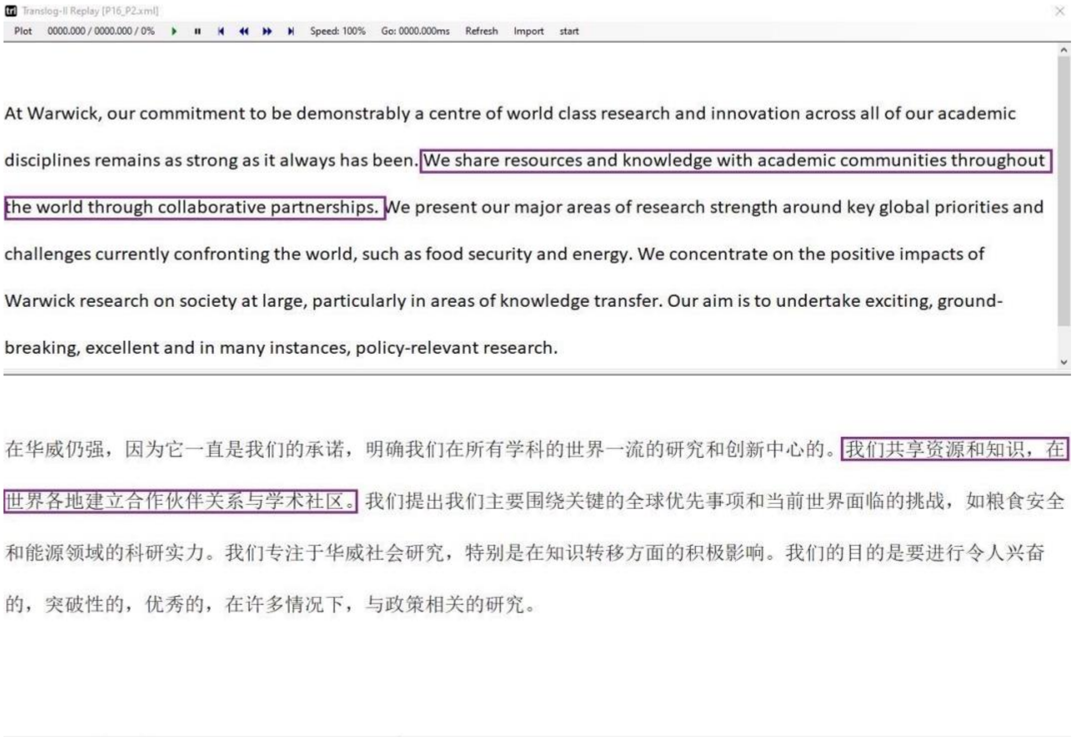


Figure C.16 Editing on the second sentence

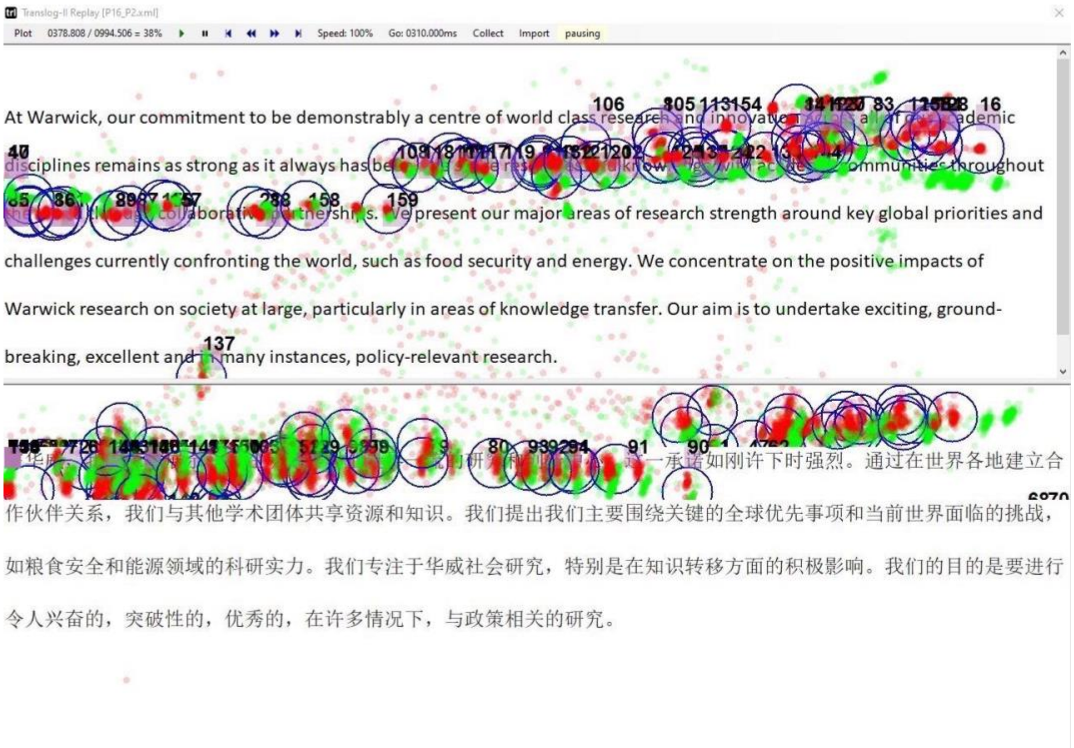


Figure C.17 Fixations during editing of the second sentence

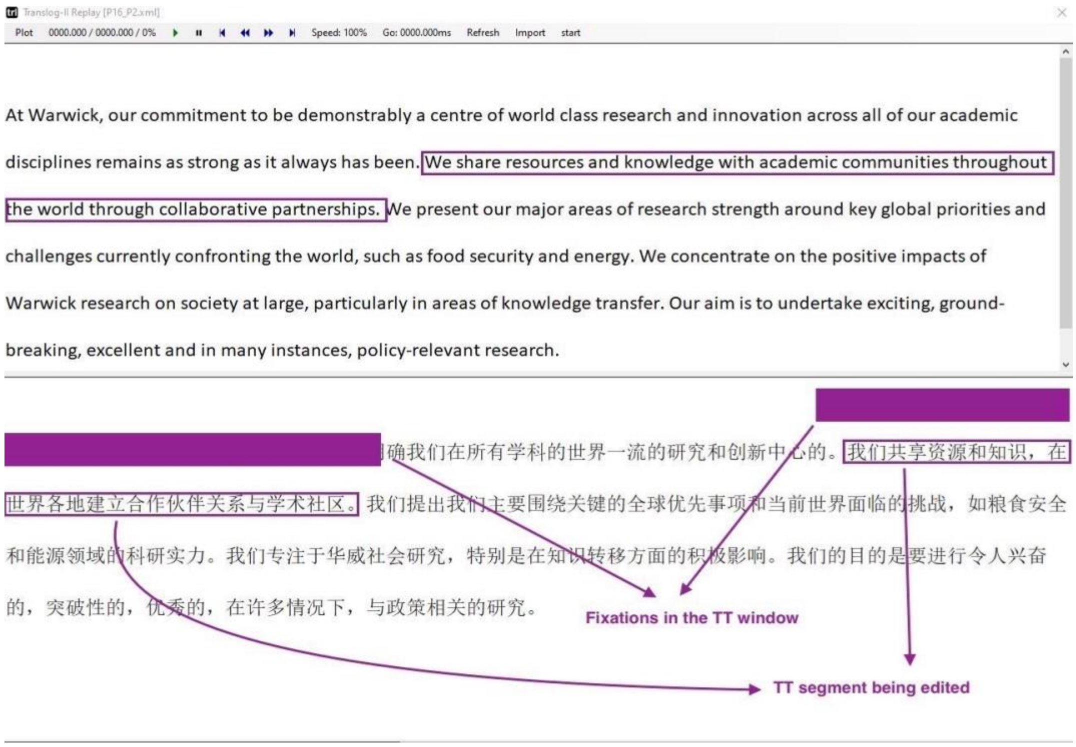


Figure C.18 The displacement of fixations for the second sentence

This means that in the processed data, the fixations that are in fact on a token in the second line would be taken as on a token in the first line. This can be some 30 tokens apart, as can be seen from the progression graphs above (see the numbers on the vertical axis). In other words, for the processed data, when fixation metrics (number or duration of such fixations) on a particular token (or sentence) are extracted, the corresponding numbers do not reflect reading behaviour on that token (or sentence), but are in fact associated with a token which is about 30 tokens before it (and effectively a sentence before the current one). With such a scale of error introduced in every sentence of the entire experimental session, the progression graph analysis, and more importantly, statistical investigations of the fixation data, even at the sentence level, would be seriously flawed and, as Carl (2013) rightly points out, such issues "may falsify major parts of the findings" (p. 1).

In terms of the interpretation of such fixations in Figure C.9, again, this process is considered in Huang (2016) as "backtracking", as can be seen in Figure C.9 (reproduced below). Based on the discussion above, it seems that this interpretation is also flawed.

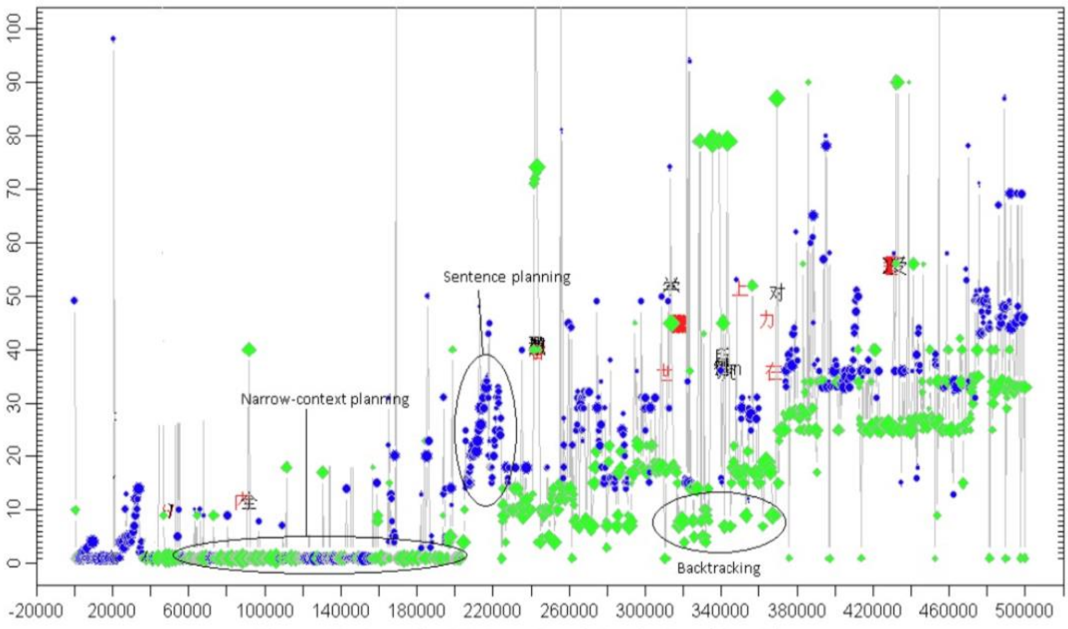


Figure C.9 Huang's interpretation of the fixation points (from Huang, 2016, p. 203)

As mentioned, this vertical displacement of the fixations occurs not only in the two sentences above, but also all through the experimental session. In other words, after such raw data is processed the participant's actual fixations on the first TT sentence would be mapped to the beginning of the TT (the first token), the fixations on the second TT sentence would be mapped to the first TT sentence, those on the third sentence would be mapped to the second, etc., in the entire experiment. This can also be seen from the progression graph in Figure C.19.

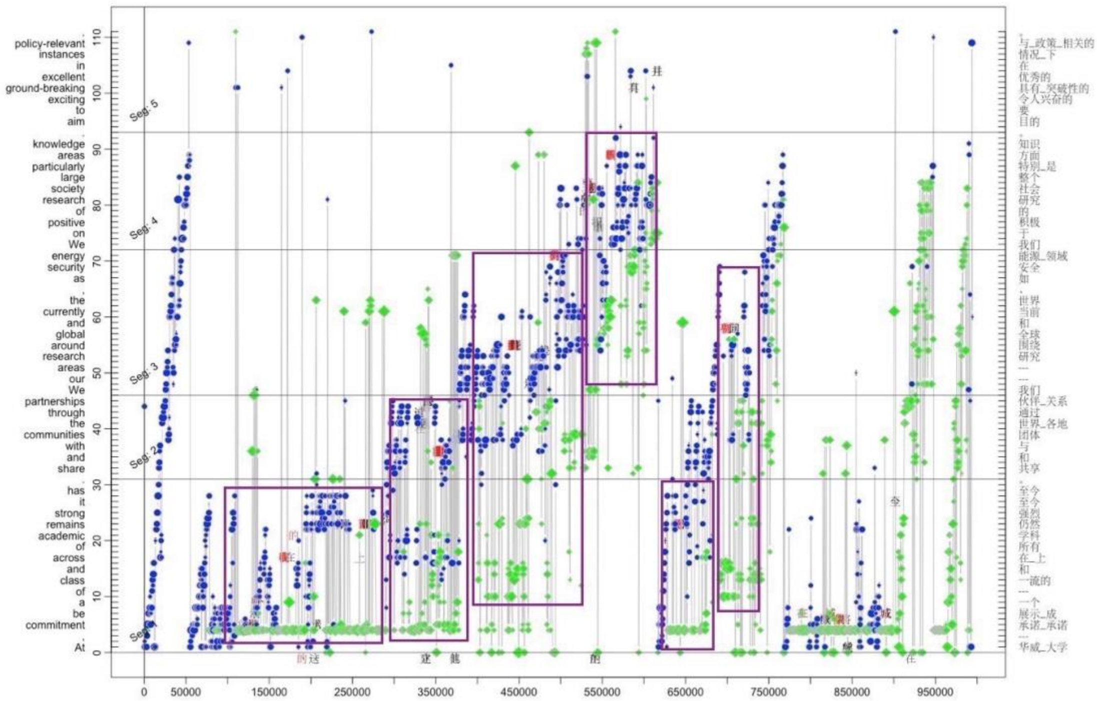


Figure C.19 Progression graph of P16 for the entire experimental session

Here, the horizontal lines in the graph indicate the boundaries of the sentences, so the area at the bottom (from 0 to 31 along the vertical axis) represents the first sentence, the second area upwards indicates the second sentence, and so forth. In the graph, when editing and ST reading is on the first sentence, the TT fixations are at the beginning; when the editing and ST reading is on the second sentence, the TT fixations are on the first sentence and begins to move to the latter part of this sentence. This is what has been illustrated above for the two sentences. For the subsequent sentences, this sentence-level displacement consistently occurs. This displacement, if adopting the interpretation mentioned above, would all be mistakenly considered "backtracking".

Figure C.2 above, together with most other progression graphs, also shows this sentence-level mis-alignment (Figure C.2 is re-produced below, with more annotations, as Figure C.20). For example, in the figure below, F3 is roughly in the same vertical position as K2, F2 is mostly in the same vertical region as K1, and F1 is at the beginning of the vertical axis. This already indicates that the displacement is serious enough to result in sentence-level errors in the fixation data.

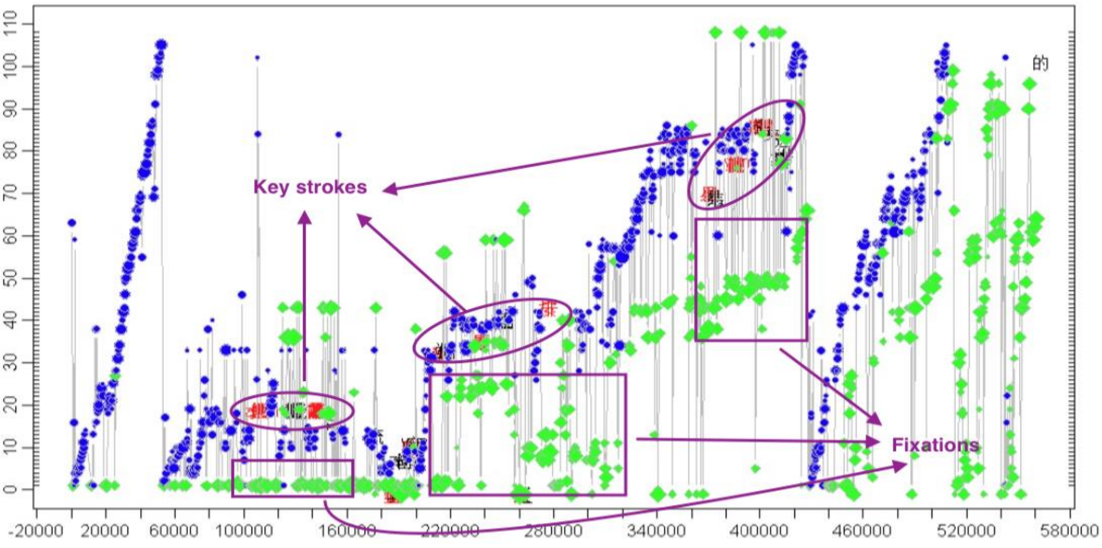


Figure C.20 Vertical drift in fixations in JIN15 dataset

Regarding the present chapter, the issue mentioned here would be crucial, which in the absence of proper resolution would result in equally misleading conclusions and falsify major parts of the findings. Therefore, this dataset (i.e., the generated tables containing relevant numbers in relation to fixation and keyboard data) would need to be re-processed (i.e., corrected), from the raw .xml files, before any analysis is conducted in view of structural ambiguity. The following sections address the possible causes of the issue, which provides the basis for the way in which the dataset is re-processed.

C.2 Possible causes of the issue

Huang's experimental design

Here, the issue of the gaze points being systematically higher than the actual line of words in the text seems to be, to some extent, recognized by Huang (2016) as inaccurate data in her description of a pilot study, as mentioned above. On page 117, Huang provides the following screenshot for this issue regarding ST reading (see Figure C.21 below):

Figure C.21 Huang's illustration of 'inaccurate data sample' (taken from p.117)

It can be seen that the ST gaze points recorded are one line above the actual loci of the eyes. In her own description, it seems, this was considered the result of inaccurate detection of eye movement from the eye tracker:

> … as the participants gradually began to concentrate more on the tasks, they would unconsciously lean their body and head forward towards the computer screen and physically attain their normal working position. This might cause two problems: data loss (gaze data on the corners of the screen/stimulus cannot be tracked) and data inaccuracy. (ibid, p. 116)

The "data inaccuracy" here is illustrated in Huang (2016) via the screenshot in Figure 48 above. Huang explains that the such data "had to be filtered out" (p. 116), that any distance of less than 65cm from the eye tracker to the participant would "increase the eye-camera angulation and lead to data inaccuracy" (p. 116), and that this problem was tackled in the following manner:

> … the participant under study was advised to remain at a constant distance of 65cm from the eye tracker, as the optimal eye-camera angulation is 35° maximum, to ensure data accuracy. (p. 116) …before calibration, participants were asked to move back and forth to find their most comfortable and relaxed posture, and to make sure they were within the optimal tracking zone, even if they moved slightly backwards or forwards. They were then informed of the range of space within which they could lean slightly backwards or forwards, although the ideal status would have been to sit still. However, the participants were not forced to stay completely still in front of the computer because this may have distracted them from performing the tasks. (p. 117)

It appears that in the experiment, the calibration, the lighting control, and the eye-camera angulation are carefully managed, together with careful management of the other phases of the study, so that the eye-tracking accuracy is maintained to a reasonable extent. In addition, as Huang describes, such inaccurate data resulting from eye-tracking issues are filtered out already. It is perhaps for this reason, among others, that after three rounds of screening where different aspects of the data are considered (see pp. 93-123; 153-155), only 18 participants out of the initial 36 are used (the JIN15 dataset are from these 18 participants), since only these participants "produced satisfactory data that were considered reliable for analysis" (p. 96).

For these 18 participants, the data produced also went through an "eye-tracking data quality control in the post-experiment phase" (ibid, pp. 153-155), where the mean fixation duration and the gaze percentage on screen are examined and are shown to be satisfactory.

It seems reasonable to trust the experiment conducted and to consider the issue mentioned in the previous section somehow not related to the experimental design or the management of the eye-tracker and the experimental settings. After all, if what Huang describes as the source of inaccuracy is indeed involved in this dataset, then the resulting issue would likely be that the collected gaze points gradually or suddenly drift towards the wrong line as the eye-movement proceeds from left to right towards then end of the line (see, e.g., Mishra et al., 2012), rather than that these points form straight lines which are stably and consistently one line above or below the actual loci of the eyes. The cause of this displacement as shown in the previous section is likely to be something else.

As mentioned, Huang (2016) recognizes the displacement of fixations as inaccurate data, and multiple sections in her work emphasize the importance of data quality as well as the careful effort that was spent in the research design to ensure the accuracy of the data collected. Since this inaccuracy is considered in Huang (2016) to be primarily due to the eye-tracking experimental setting rather than the data processing techniques, it is no wonder that most of such effort, although very effective in avoiding the kind of inaccuracy and noise that can be introduced from the experiment itself, did not eliminate the issues shown above, issues that are likely to be the result of the *processing*, rather than the collection, of this dataset, as will be shown below. On the other hand, such effort in carefully designing and conducting the eye-tracking and key-logging experiments means that the raw data itself, if processed properly, would be very reliable and useful for the analysis to be conducted in this chapter.

In this regard, the progression graphs and Translog screenshots shown above already indicate that the gaze points collected do not seem to unstably drift towards the wrong line as the eye-movement proceeds — they start from one line, and stick to that line, even if the line itself is wrong. If these lines are systematically moved to the correct position, the dataset would be truly dependable.

Fixation computation and fixation-to-word mapping

In terms of data processing, the fixation-to-word mapping can be distorted by a number of issues, especially given the noise that is often involved in eye-tracking data. These issues are well-known and have been discussed from different aspects with regards to methods of (vertical) error correction (e.g., Carl, 2013; Hornof & halverson, 2002; Lima Sanches et al., 2016; Mishra et al., 2012; Yamaya et al., 2015). Some of these issues are described in Carl (2013, p. 3) as the following:

> 1). due to calibration difficulties, free head movement or changes in light or other conditions, the gaze sample points which are recorded may not exactly correspond to the place which was gaze at;

2). the choices that are made when computing the fixation, e.g., based on the left or the right eye gaze sample, their average, how proximity or saccades between successive gaze samples are defined, etc.;

3). the computation of the closest character for a given x/y position depends on which part of the character is taken as a reference, e.g., the upper left corner, or the centre of the character, etc.

The source of inaccuracy described in Huang (2016) and mentioned above is clearly the same as issue 1) here. In this regard, the experiments conducted for the project seem to be largely adequate in avoiding this issue, and the progression graphs and the Translog-II screenshots shown above already demonstrate that this issue is less likely to be the cause for the problem. The other issues, on the other hand, have to do with the processing of the raw data — the computation of fixations based on gaze points, and the mapping of the computed fixations onto the characters.

The computation of fixations (i.e., the second issue above) does not seem to be the likely cause of problems identified above either. As can be seen from the screenshots of replay in Translog-II (see Figure C.13 and Figure C.17), the computed fixations (blue circles in the pictures) do not drift away from the gaze sample points (red and green dots). It seems that the computation from the dispersion-based algorithm (Salvucci & Goldberg, 2000) in Translog-II is dependable for the dataset here.

Regarding the mapping of the computed fixations onto the characters, the problem in the JIN15 dataset does not seem to be the same as the third issue above either. The part of the character which is taken as a reference for computation, it seems, would be a prominent issue only when the fixations are between two lines, rather than when they fall onto a completely different line.

To illustrate this, consider Figure C.22 and Figure C.23, both taken from Carl (2013).

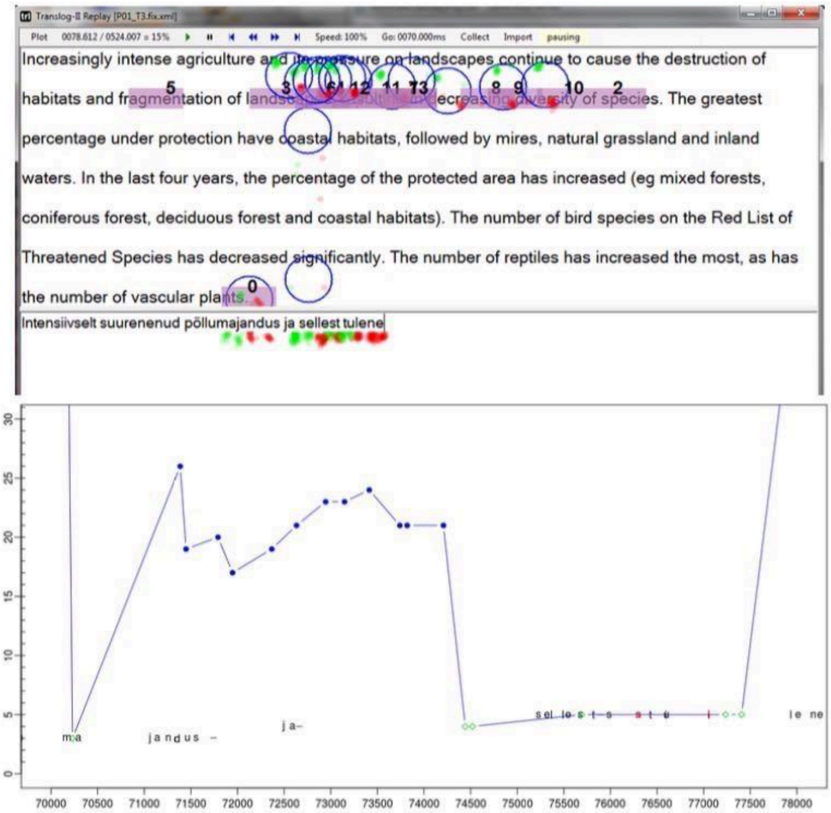


Figure C.22 Carl's illustration of the error in naïve mapping (2013, p. 3)

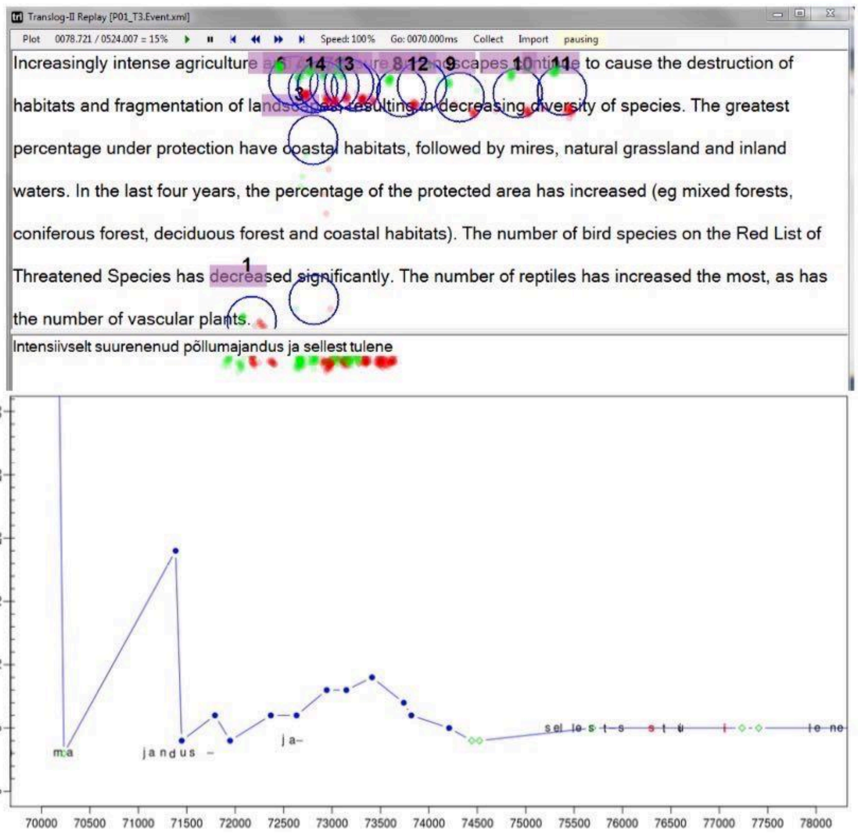


Figure C.23 Carl's illustration of the re-computed mapping (2013, p. 4)

These figures show the replay of the same editing and eye-movement behaviour, but with different mapping results. In each figure, the screenshot from Translog-II is shown together with the corresponding progression graph. Figure C.22 is the result using naïve mapping, while the mapping result of the same fixations from a dynamic programming method is shown in Figure C.23. Apparently, the errors in the naïve mapping are corrected in the re-computed result, which is also visible in the progression graphs in Figure C.24 (top — naïve mapping; bottom — re-mapped version of the same segment).

Here, the difference between Figure C.22 and Figure C.23 is mainly about whether the fixations (blue circles) are mapped onto the first line or the second line of the text. This noise occurs primarily because of the fact that these fixations are located between the first and the second line, so that which line they are mapped onto can be largely dependent on which part of the character is used as a reference for computing the closest character to each fixation (i.e., the third issue mentioned above).

For the JIN15 dataset, the situation seems to be somewhat different. The Translog-II replay screenshots above (Figures A.13, A.17, and A.21) show fixations clearly on a certain line (only with very few exceptions), rather than between two lines creating confusion for the computation of the closest characters. The character which is computed to be closest to the fixations would be likely to remain the same even if a different part of the character (e.g., centre or upper left corner) is used as a reference, and this would still be a character in the wrong line. The minimal-distance character from each fixation would be unambiguously the wrong one in the computation of the mapping, which the dynamic programming in Translog-II does not seem to correct adequately.[75] Therefore, the issue 3) above is unlikely to be the cause of the problem here.

On the other hand, even when some fixations involve this issue (as some do in Figure C.17, for example), the algorithm in Translog-II (Carl, 2013) should have effectively corrected them in the mapping process (see also Figures A.22 and A.23 above, and Figure C.24 below for this correction).

[75] See also footnote above, for the criteria of mapping in this algorithm.

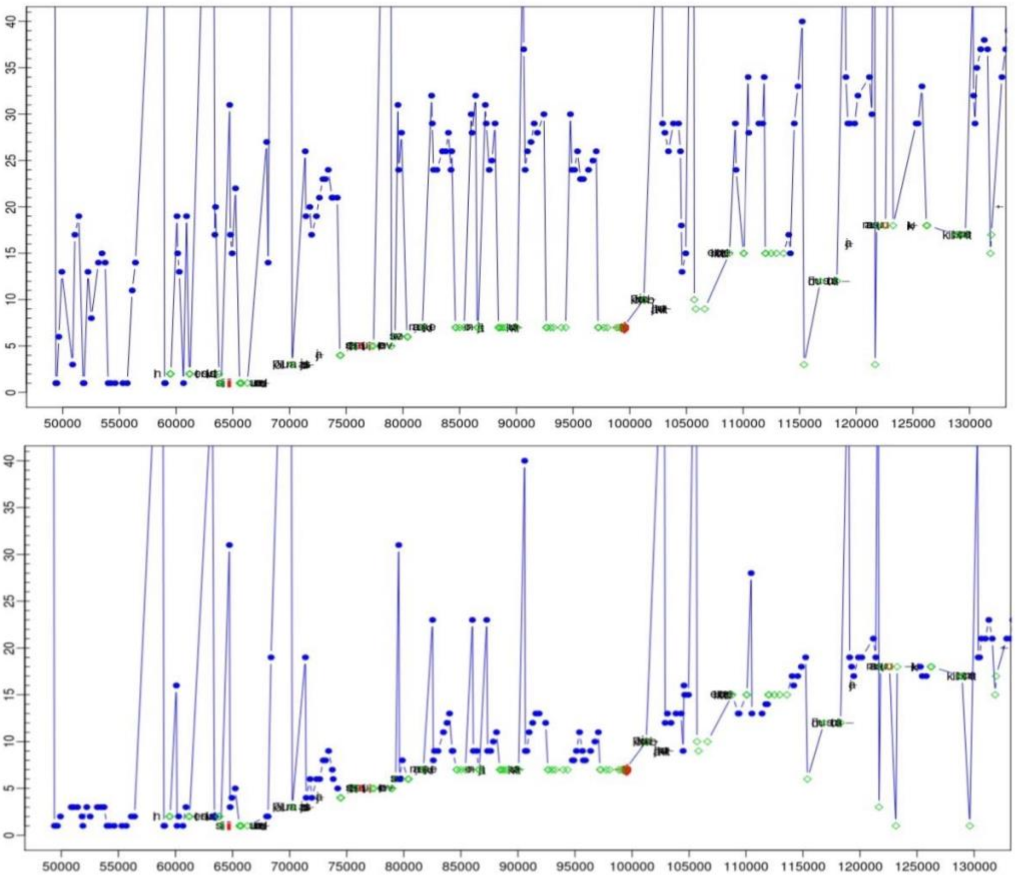


Figure C.24 Carl's illustration of naïve mapping vs. re-mapped version

Offline mapping

The above sections have described several possible causes for the systematic error in the JIN15 dataset, and it seems that none of them is likely to be the genuine cause. In the following section, explanation is provided from a different perspective, and solutions on this basis show considerable improvement in the data.

It is important to note that the fixation-to-word mapping in the JIN15 dataset was not conducted online (in real time as the experiment is being conducted), but offline (in a replay session after the experiment),[76] and that the displacement of the fixations occurs mostly in the TT window (see above discussion). Assuming that the experiment is properly conducted and that the calibration is adequate, the displacement of the fixations may not be intrinsically an issue of the noise in mapping, but is likely

[76] This mapping occurs only in the replay session (for Chinese and Japanese) due to the necessary use of IME when typing.

to be a problem of possible inconsistency in the visualization (i.e., location displayed) of the words and of the gaze points in the replay session.

As the gaze points are recorded and displayed in terms of X-Y coordinates in Translog-II, while the TT words are displayed in the replay session with reference to the TT window, a change in the position or size of the TT window (or the position of the scrollbar on the side) in the replay session would lead to a corresponding change in the position of words, but not to any change in the position of the gaze points (and the fixations computed). This means that when there is, for some reason, a change in the replay session in terms of the position of the text (i.e., inconsistency with the experimental session), then the gaze points (based on X-Y coordinates) in this replay session would no longer be positioned in the same place as the corresponding characters. In a properly conducted experiment, the gaze points that are collected from the eye-tracker overlap with the words that the participant is actually fixating, but this overlap would stop in the replay session if there is such a change.

For example, the following figures show that moving the bar separating the ST and TT window would result in a corresponding change of position of the TT, but the gaze points and fixations remain unchanged (since they are recorded and displayed on the basis of the X-Y coordinates). Figure C.13 (reproduced below for convenience) is the replay session of the participant's (P16) editing on the first sentence as discussed above, and for the same replay, moving the bar upwards for a distance of one line of text results in Figure C.25, where the fixations overlap with the text in the TT window.

The same can also be seen if the scrollbar on the side is moved.

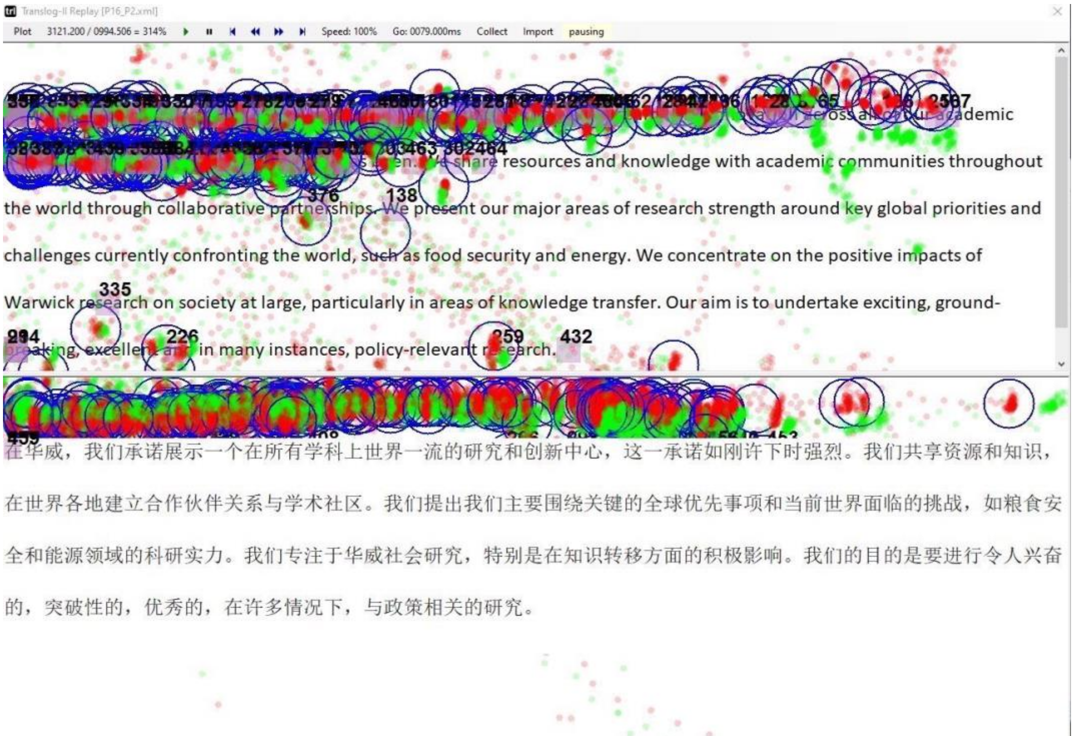


Figure C.13 Fixations during editing of the first sentence

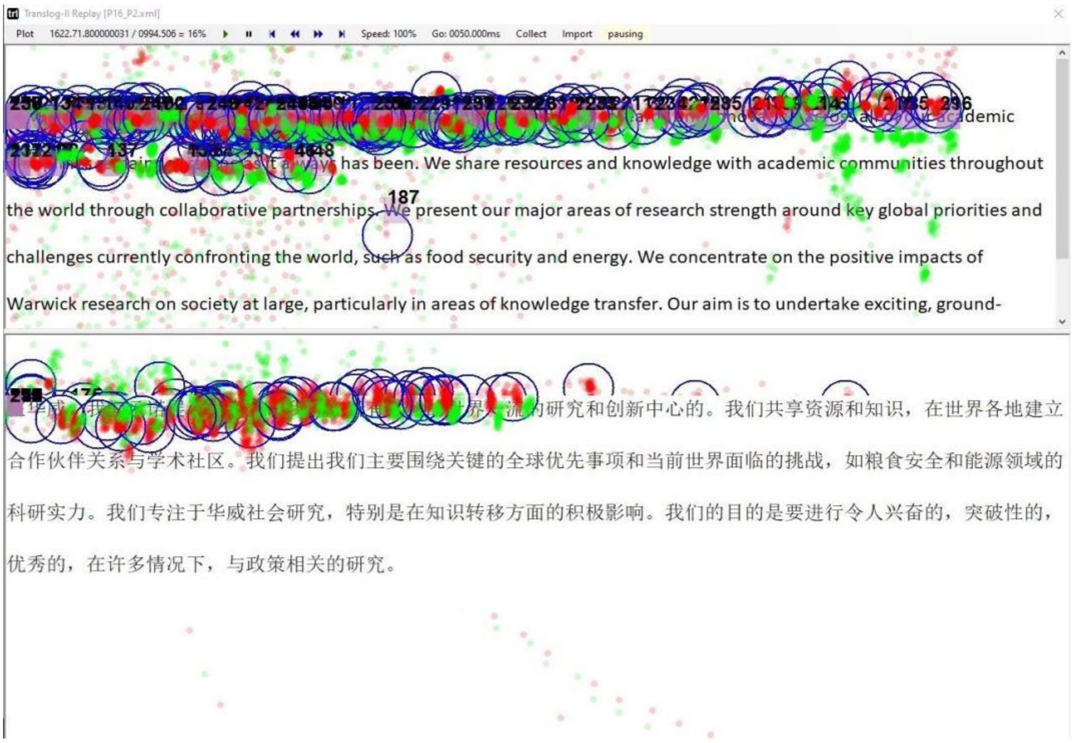


Figure C.25 Fixations when the bar is moved upward

In essence, this is merely a visualization issue, and would not be a problem for statistical analysis when gaze-to-word mapping is real-time during the experiment, since the mapping process would have already been completed before this change occurs. However, for experiments where the target language is Chinese or Japanese (which require an IME for typing), gaze-to-word mapping has to be offline in the replay session, and with unmapped data, this discrepancy in visualization (i.e., the position of the words) might potentially cause errors in the mapping result, such as the one illustrated in the JIN15 dataset above.

It seems here that the displacement of fixations in JIN15 dataset is largely due to this issue when the fixations are mapped onto the TT words offline.

To better explain this, Figure C.11 is reproduced below, and as comparison, Figure C.26 shows the replay session (both taken from Huang 2016).

Figure C.11 Huang's screenshot of Translog-II user interface (p. 87)

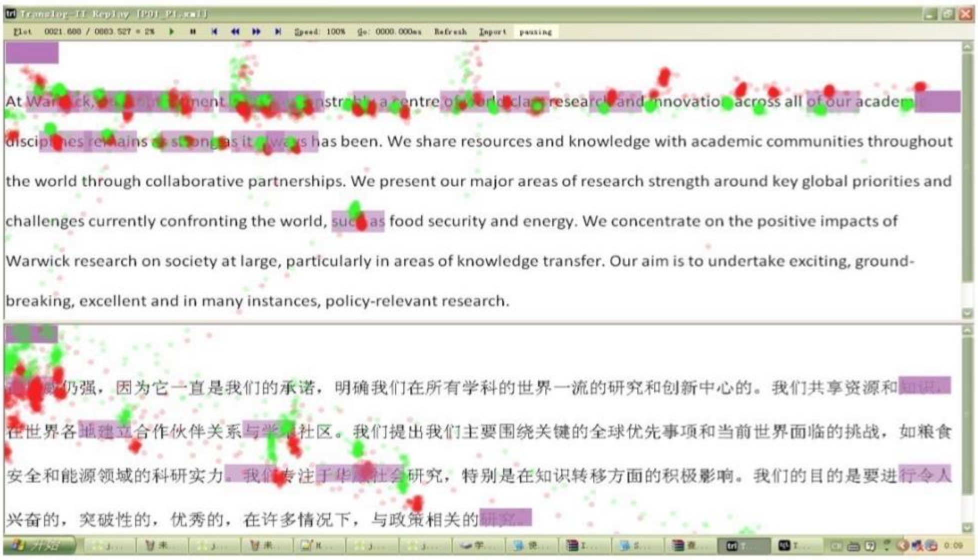


Figure C.26 Huang's screenshot of the replay session (p. 91)

It can be seen that in Huang's illustration, the experimental session (Figure C.11) and the replay session (Figure C.26) apparently show some inconsistency in terms of the size of the TT window — in the latter the window is visibly smaller than the former. The difference in window size (height) appears to be roughly the height of the (Windows system) task bar at the bottom. This can also be shown by the scrollbar in Figure C.26 in comparison with Figure C.11. Such inconsistency might be the result of the researcher doing the experiment on one computer (with a larger screen), and replaying the session on another (with a much smaller screen), with the (TT) window somehow adjusted (either manually or automatically) and being different in size.

For the JIN15 dataset, it seems that while the gaze points were collected in an interface in Figure C.11, the fixation-to-word mapping were completed in the replay session in the interface in Figure C.26. A closer look at the scrollbar in the latter shows that the space for scrolling in the TT window is roughly equivalent to the width of one line of the text. Therefore, it is not surprising that in the JIN15 dataset (see discussion above), the displacement of TT fixations is consistently one line away from where they are most likely to be. This also seems to explain the fact that most of the problems with the dataset have to do with TT fixations while the ST fixations show far better (mostly satisfactory) quality (see the progression graphs at the beginning of Appendix A, and the Translog-II replay screenshots in Section C.1).

Here, any change in the position of the bar separating ST and TT windows, or in the position of the scrollbar in either window, would lead to a corresponding change in the position of the ST/TT words, but the gaze points (and the calculated fixations from these gaze points) would remain in the same position. In the replay session with offline fixation-to-word mapping, this change means that the fixations would likely be mapped to words in the wrong line, introducing vertical error in the processed eye-movement data.

It is also important to note that when the raw data is loaded to Translog-II, the layout of the ST and TT windows is roughly the same as Huang's screenshot, as mentioned (see the beginning of the "Translog replay" section). This can be clearly seen by a comparison between Figure C.10 and Figure C.11.

C.3 Data re-processing and error correction

On the basis of the analysis above, the raw files of the JIN15 dataset were re-processed in order to correct the error before the data was used for analysing structural ambiguity. While Translog-II provides a function for manually correcting the fixation-to-word mapping (called “FixMap”), this would be excessively laborious to carry out on the entire dataset. At the same time, the above analysis on the vertical error in the dataset has already shown that the displacement is largely systematic, consistent, and stable, almost always showing one-line discrepancy, which means that the correction can be automated and that the manual correction process is unnecessary in this regard.

The way in which this dataset is re-processed is described below.

Each experimental session is replayed in Translog-II and the gaze points (fixations) are checked to see if they match the position of the text, if the mapped words in the text are in the same line as the words being edited (i.e., correspond to the actual words likely being fixated), and how much adjustment is needed to shift the text upwards. For the sessions that show displacement of TT fixations, each of the relevant raw .xml files is processed in the following manner:

1. For all lines where win=“2”, change this value to “-2” (i.e., replace win=“2” to win=“-2”).
2. Delete the lines containing the computed fixations for the TT window.
3. Load the file to Translog-II as a replay session, and adjust the position of the TT window (shifting the bar towards the appropriate position) so that the words and the corresponding gaze points would be in the same line.
4. Replay the session without interruption, observing the replay in the meantime to ensure that the gaze points are on the correct line of the text; if there is still considerable displacement, close the session without saving, re-open the raw file in Translog-II, re-adjust the TT window, and then repeat this step.
5. When the replay session is satisfactory, save the .xml file.

In this process, the fixations in the TT window are re-computed from the gaze sample points, and these fixations are mapped to the appropriate characters in the corresponding line. The fixations and mapping results for the ST window, however, remain unchanged.

As this modification made in the .xml file is only on the processed information from Translog-II (i.e., whether the gaze points are in the ST or TT window, the computed fixations from the gaze points, and fixation-to-word mapping), rather than the information associated to the eye positions detected by the eye-tracker (in the form of X-Y coordinates) or to the keyboard logging, it does not change the recorded eye-movement and key-input behaviour itself. In other words, after the above procedures, the raw data is only re-processed in terms of offline fixation computation and fixation-to-word mapping, without introducing any change in the user activity.

After the above procedures, the ST gaze points are also checked for inaccurate positions, and similar procedures are also followed (regarding lines where win="1"). Here, as the present chapter is mainly focused on reading behaviour on the TT, the correction on the ST fixations is not strictly necessary.

This modification has effectively corrected the error in the JIN15 dataset, as can be seen from the progression graphs below. In each of these figures, the graph before correction is shown on top, and the one from the corrected data is at the bottom. A comparison between the two easily reveals the effect of this correction.

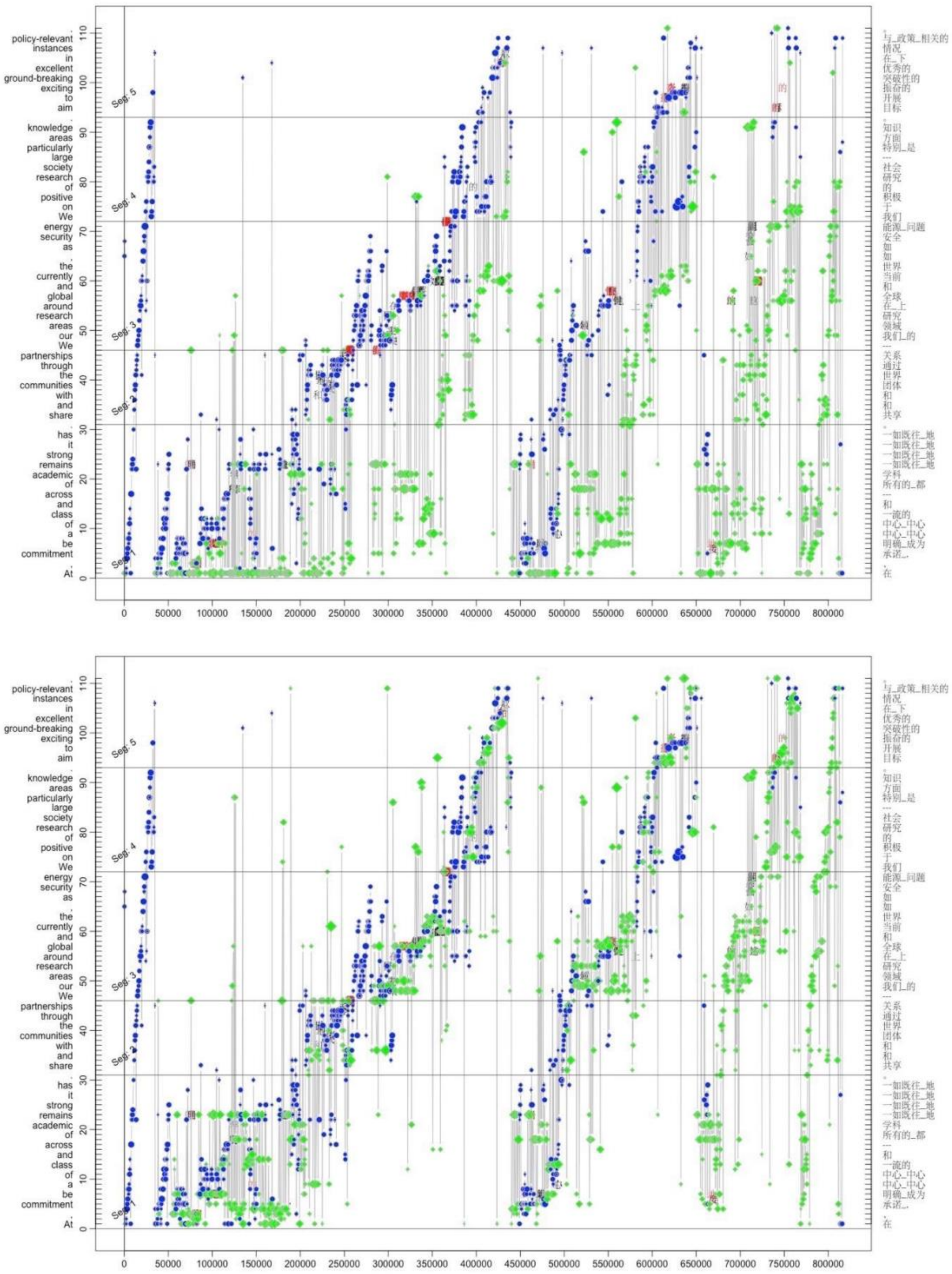


Figure C.27 Data correction for P09 (top: before correction; bottom: after correction)

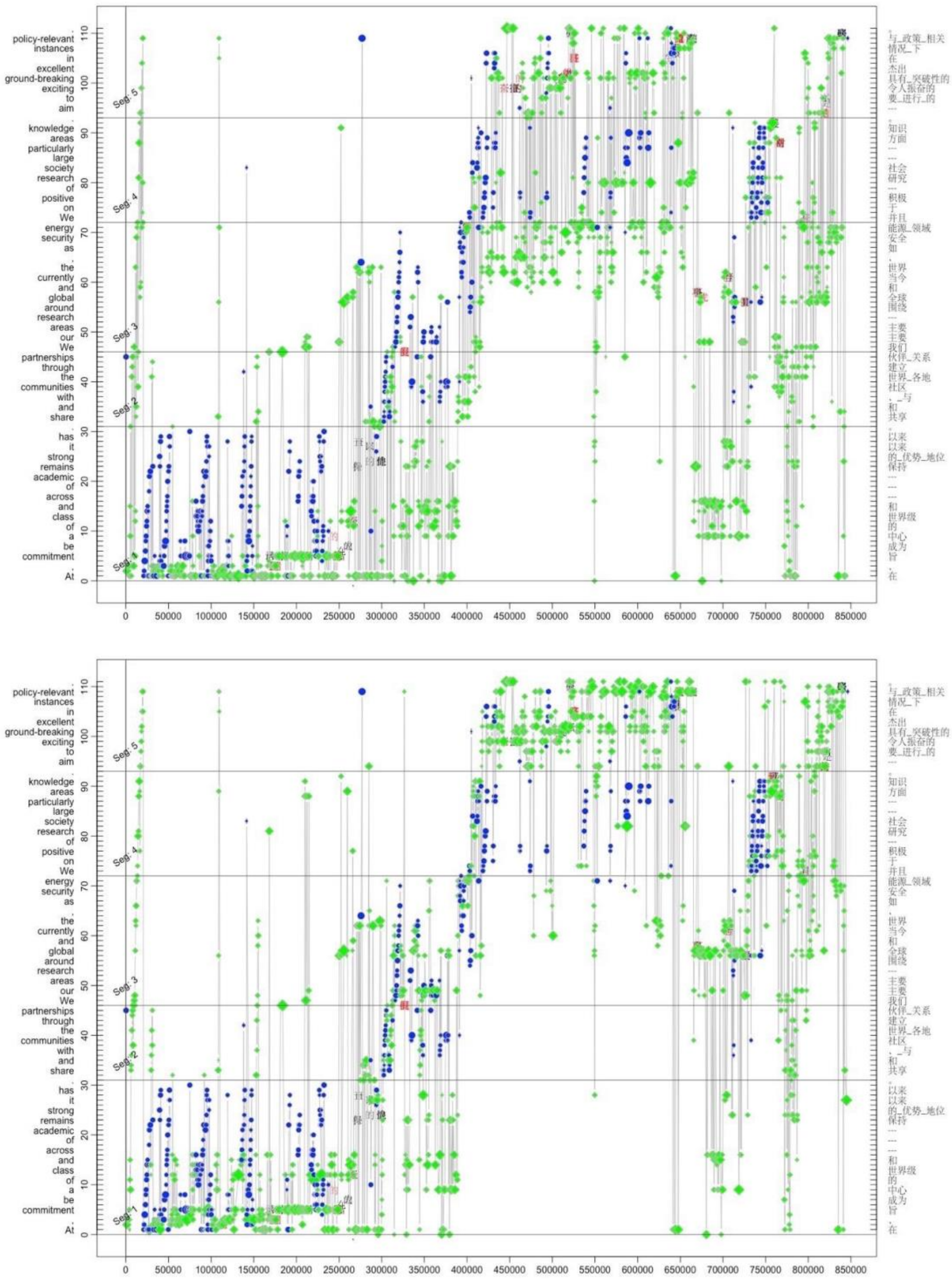


Figure C.28 Data correction for P11 (top: before correction; bottom: after correction)

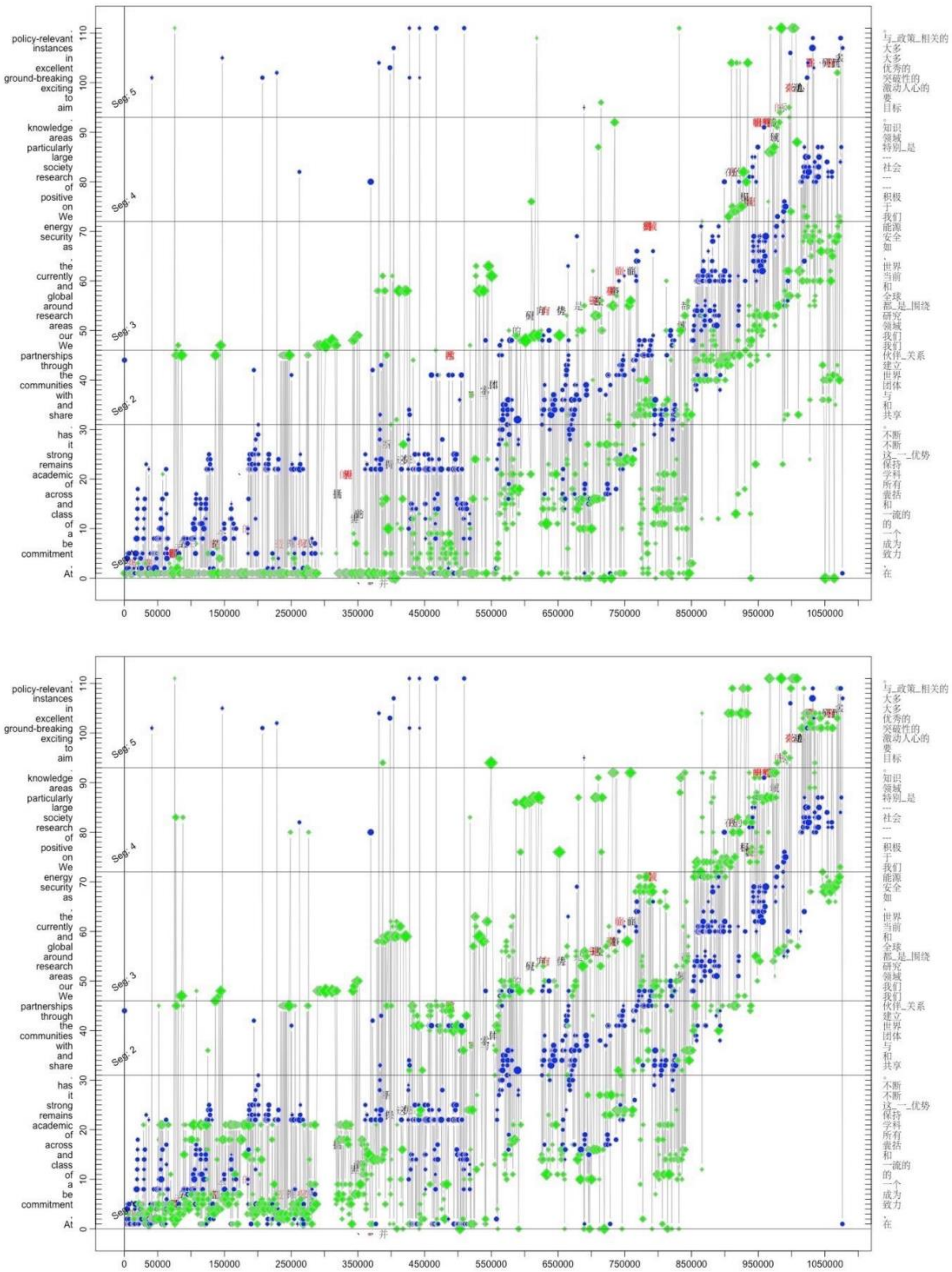


Figure C.29 Data correction for P14 (top: before correction; bottom: after correction)

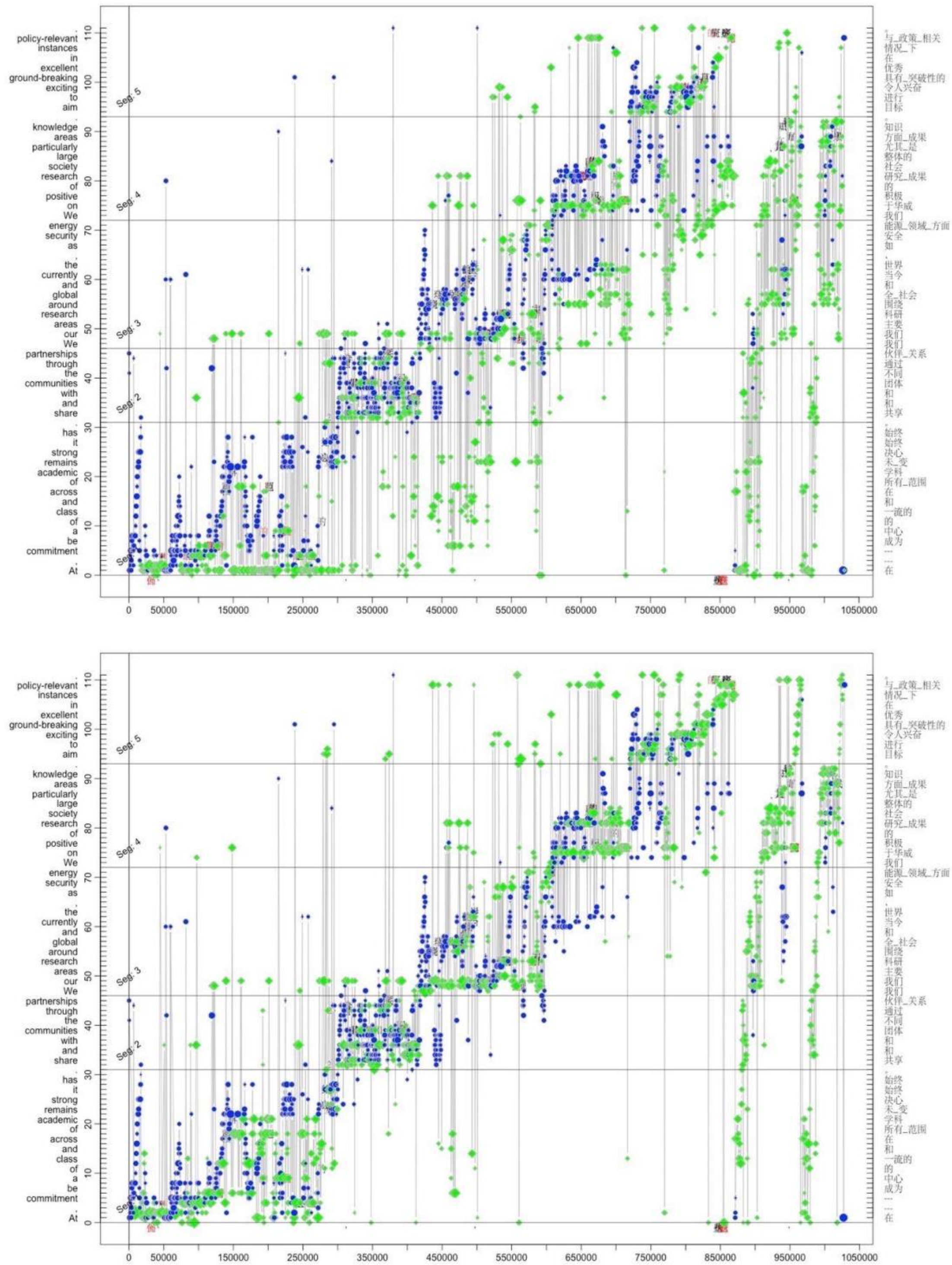


Figure C.30 Data correction for P15 (top: before correction; bottom: after correction)

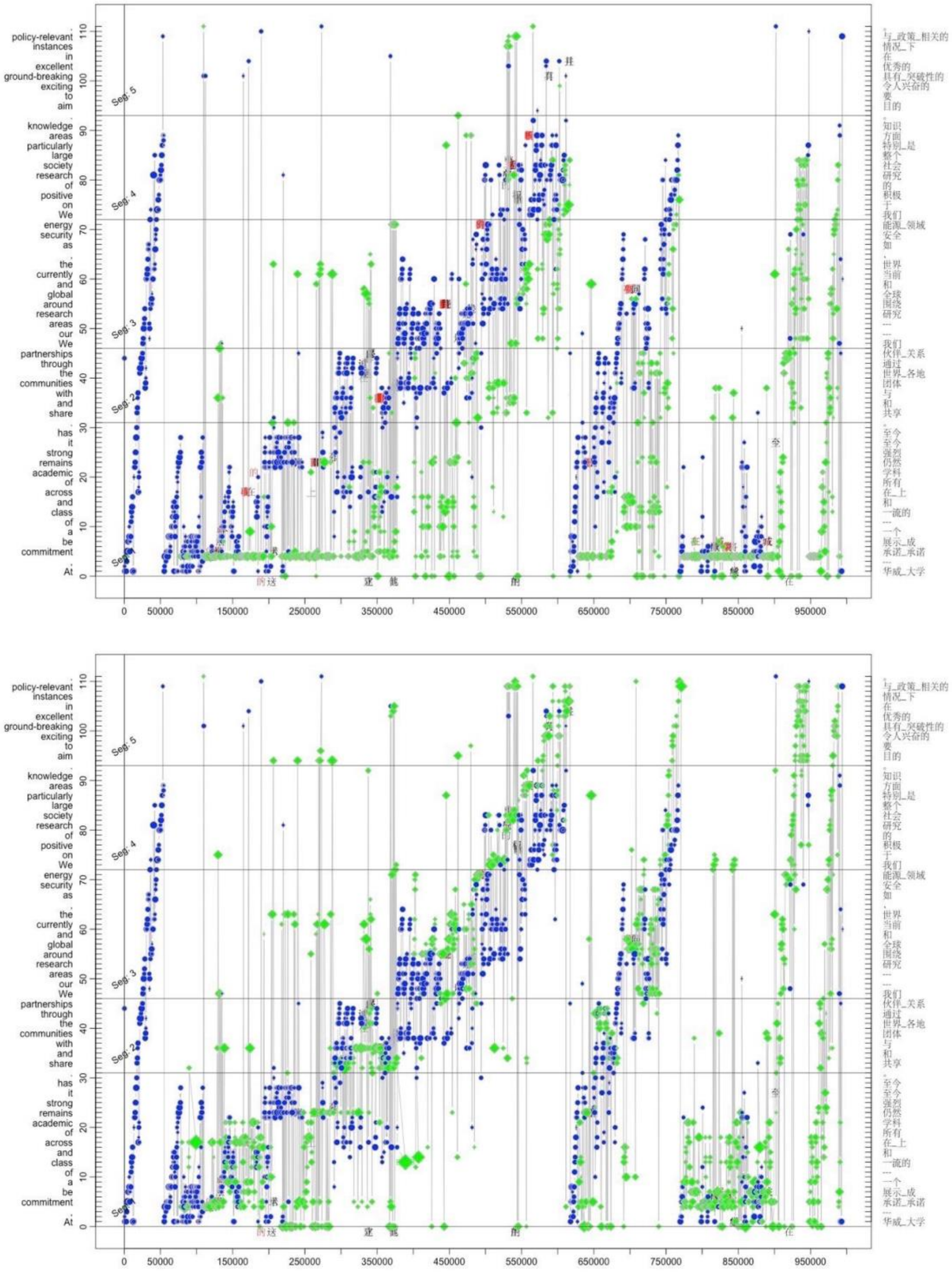
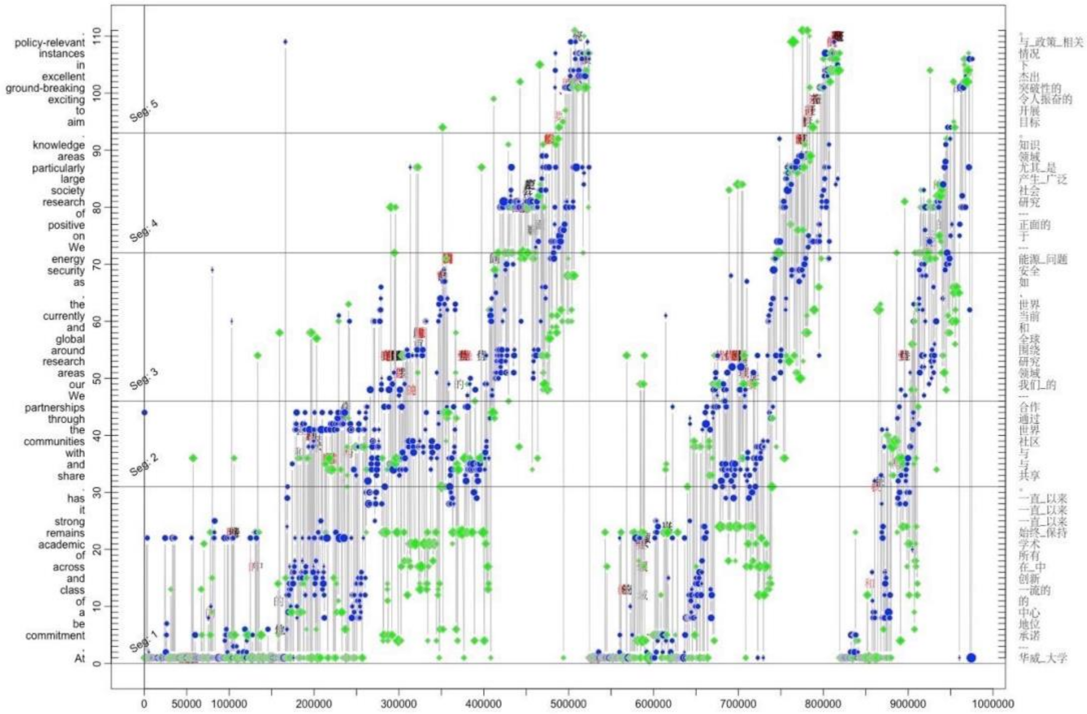


Figure C.31 Data correction for P16 (top: before correction; bottom: after correction)



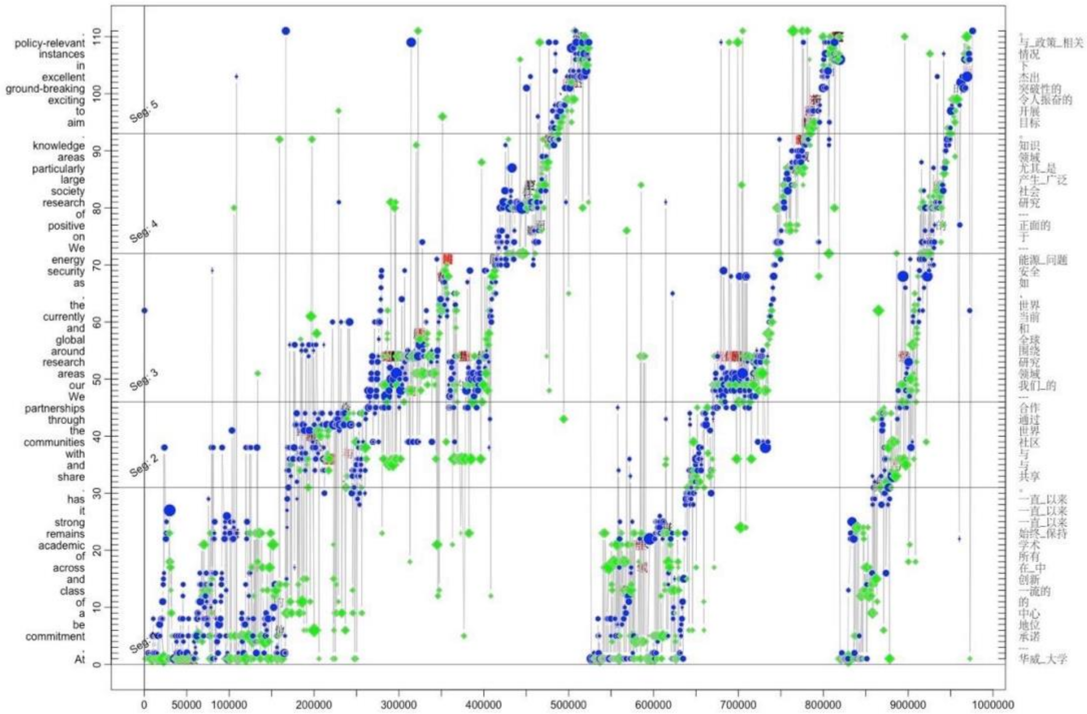


Figure C.32 Data correction for P17 (top: before correction; bottom: after correction)

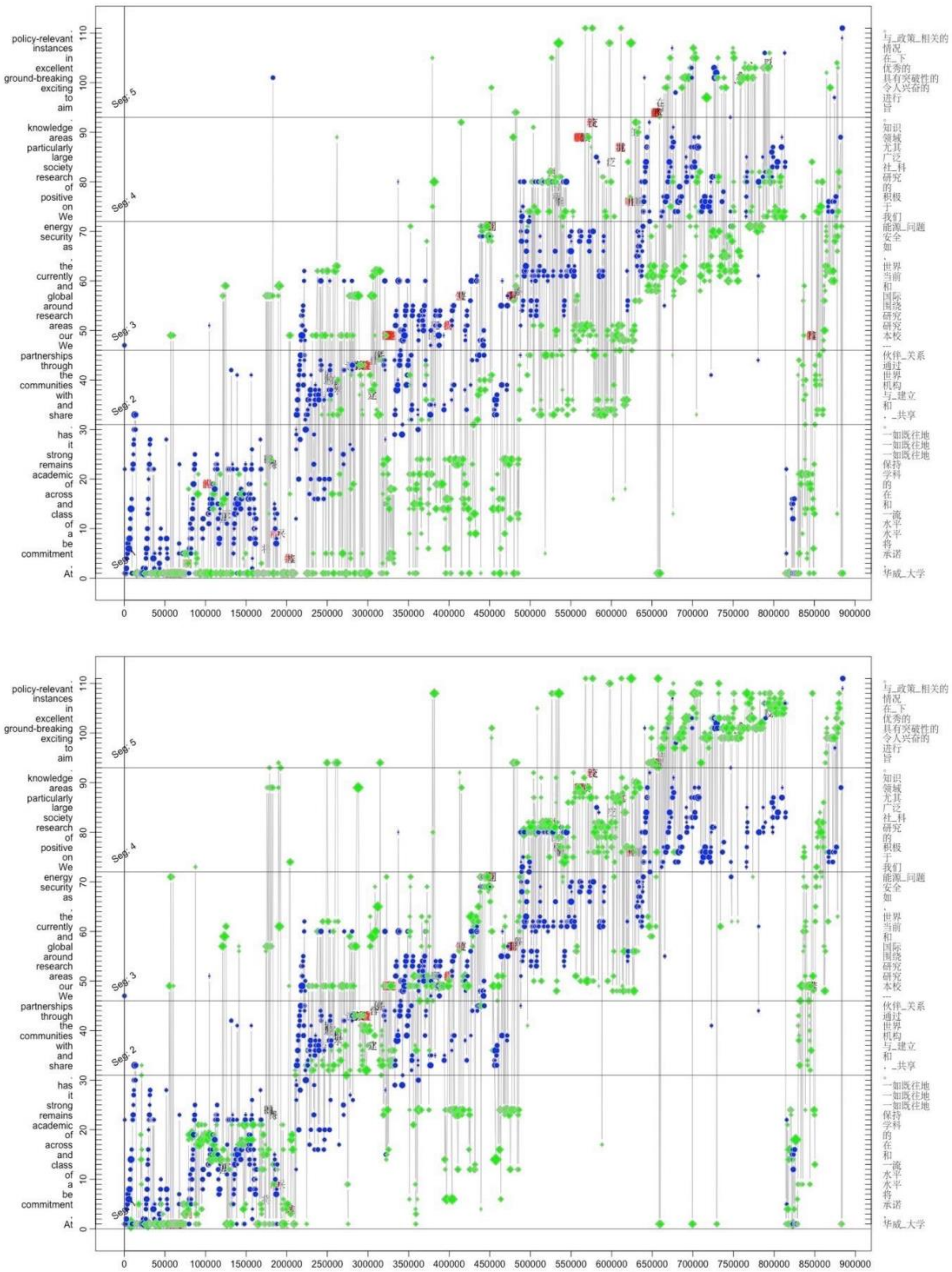


Figure C.33 Data correction for P18 (top: before correction; bottom: after correction)

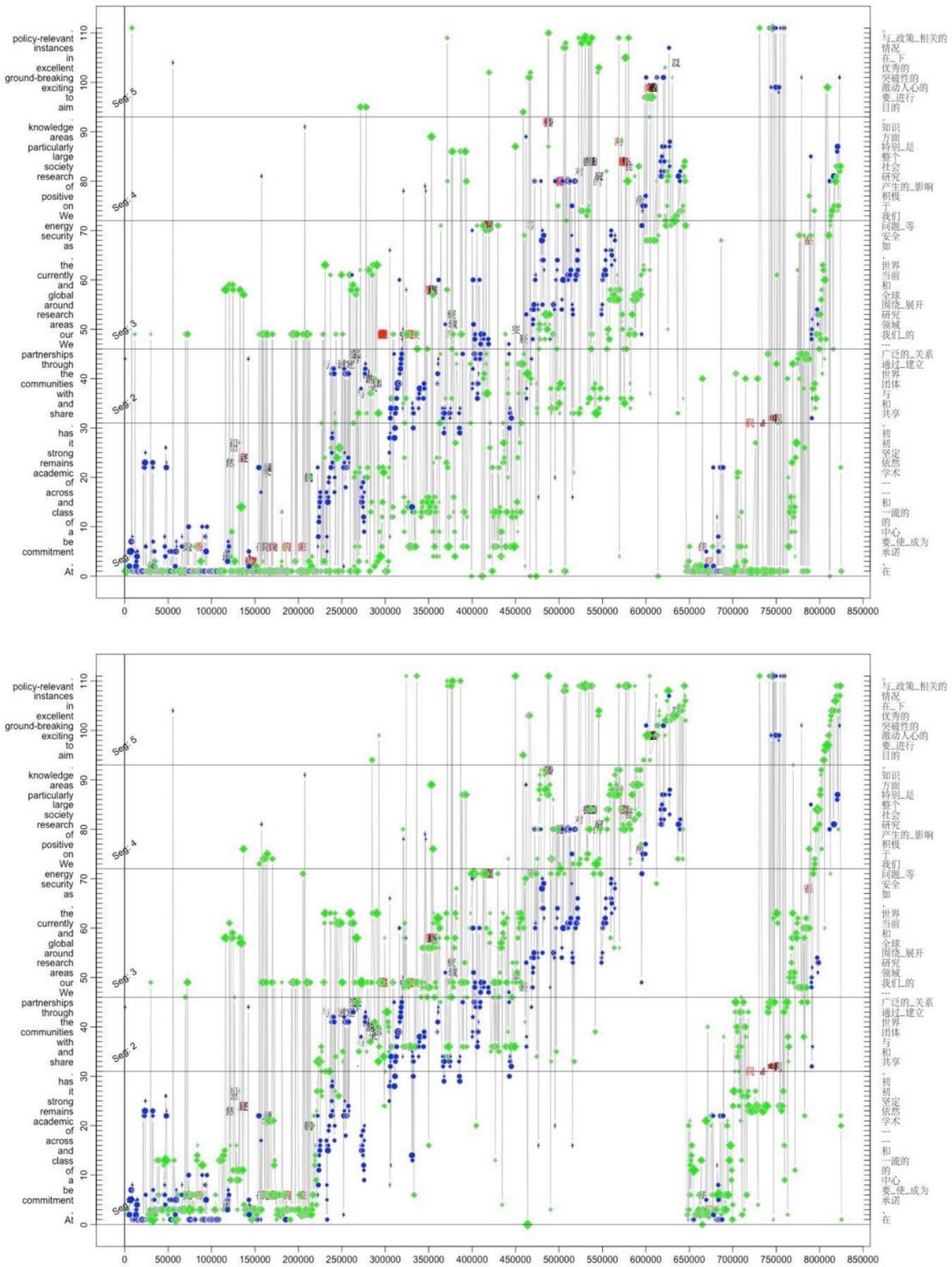


Figure C.34 Data correction for P20 (top: before correction; bottom: after correction)

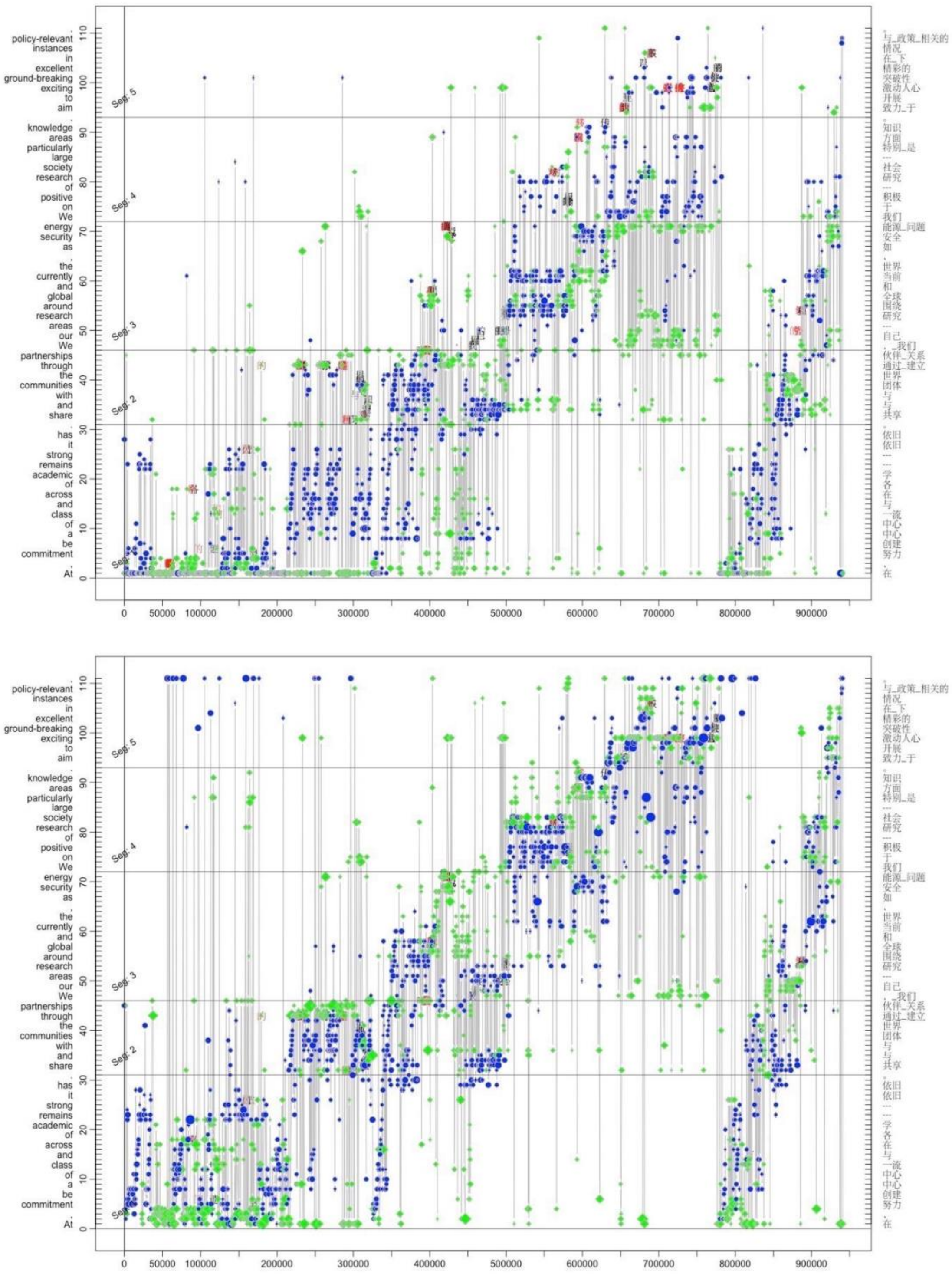


Figure C.35 Data correction for P21 (top: before correction; bottom: after correction)

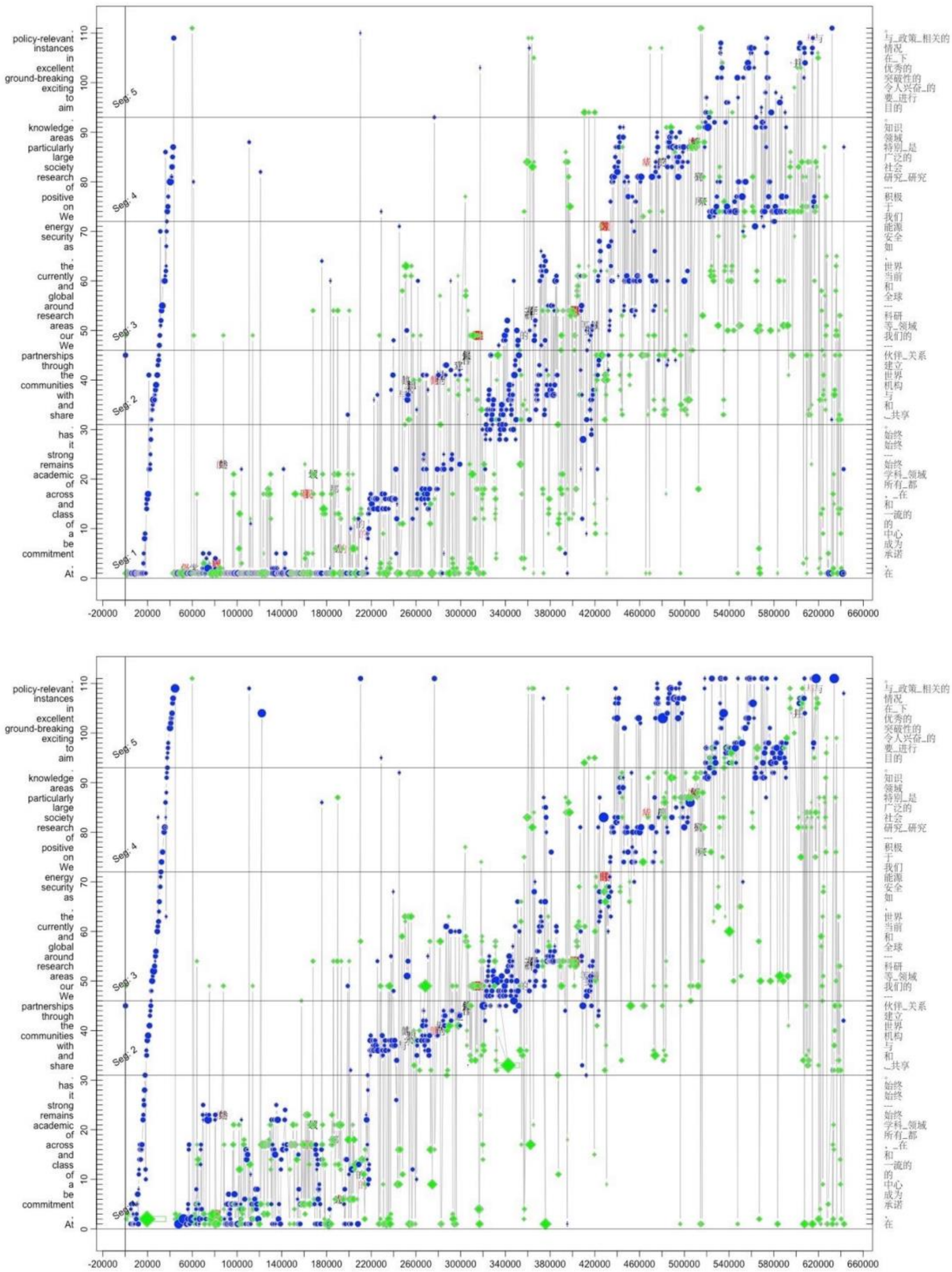


Figure C.36 Data correction for P22 (top: before correction; bottom: after correction)