Chapter 11
Cohesive Relations in Text Comprehension and Production: An Exploratory Study Comparing Translation and Post-Editing

Márcia Schmaltz, Igor A.L. da Silva, Adriana Pagano, Fabio Alves, Ana LuísaV. Leal, Derek F. Wong, Lidia S. Chao, and Paulo Quaresma

Abstract Few studies using Translog-II in conjunction with eye-tracking data in translation studies have focused on languages which use logographic scripts. This chapter reports on an exploratory study of data related to one text included in MS13, contained in CRITT Translation Process Research Database, with a view to investigating the impact of type of cohesive chain on cognitive effort in Portuguese-Chinese translation and post-editing tasks. Eye-tracking and key logging data were assessed by means of a linear mixed-effects regression model. The results point to no impact of task on the dependent variables, but to an impact of the type of cohesive relations on target text reading and production. The chapter also contributes to developing a methodology for processing of Translog-II data involving Chinese.

Keywords Translog-II • Cohesion • Portuguese-Chinese Translation • Post-editing

M. Schmaltz (✉) • A.L.V. Leal
Department of Portuguese, University of Macau (UM), Macau, China
e-mail: schmaltz.marcia@gmail.com

I.A.L. da Silva
Federal University of Uberlandia (UFU), Uberlandia, Brazil

A. Pagano • F. Alves
Laboratory for Experimentation in Translation, Federal University of Minas Gerais (UFMG), Belo Horizonte, Brazil

D.F. Wong • L.S. Chao
Department of Computer and Information Science, University of Macau, Macau, China

P. Quaresma
Department of Informatics, University of Evora (UE), Évora, Portugal

11.1 Introduction

Research using Translog-II (Carl 2012) in conjunction with eye-tracking data in translation studies (Carl and Jakobsen 2009; Jakobsen 2011; Hvelplund 2011; Carl and Dragsted 2012; Sjørup 2013; Balling and Carl 2014; Mesa-Lao 2014; among others) has focused on tasks involving Western European languages and consequently alphabetical scripts. However, studies focusing on languages which use logographic scripts are still incipient. This chapter reports on a study of from-scratch translation and post-editing tasks carried out from Portuguese into Chinese by Chinese translators of Portuguese (L2). Drawing mostly on the methodology used by Sjørup (2013), we carried out a study to examine gaze and key logging data from six participants while translating and six other participants while post-editing a 79-word news report. These data are available in CRITT Translation Process Research Database as MS13 (translation session 16, post-editing session 18).

Building on Halliday and Hasan (1976) and Hasan (1984), referents pertaining to the main cohesive chain of the source text (labelled *chain A*) were defined as our focus of enquiry and contrasted with items in a secondary cohesive chain traceable in the same text (labelled *chain B*). Our assumption was that tracking participants (referents) in chain A would be critical for the 12 translators to build a coherent interpretation of the source text (ST) and would require them to retrieve the identity of what was being talked about by referring to another expression either in the co-text or the context of the situation and culture. A higher number of fixations in eye and keyboard activities were thus expected during reading and production of chain A. A secondary chain, in contrast, would have a lesser contribution to the ST and TT (target text) coherence and would thus demand less attention, as well as fewer keyboard and eye activities.

11.2 Review of the Literature

The role of *cohesion* in the establishment of a coherent interpretation of text is one of the many core questions of reading comprehension in translation tasks (Bell 1991; Hatim and Mason 1990). Since it has to do with translators' active participation in understanding an ST unfolding and in building a TT patterned on it, this is an issue particularly well suited to be approached from a translation process perspective.

Among the different resources playing a part in *texture*, i.e., that which makes a text a text and makes it function "as a unity with respect to its environment" (Halliday and Hasan 1976, 2), cohesive devices are responsible for non-structural relations between items in a text. Such relations are established through the creation of semantic bonds, so that one item is interpreted with reference to the other.

One cohesive relation in particular is especially relevant to discourse coherence in text unfolding: this is *participant tracking*, i.e., the mapping of referents pertaining to the main cohesive chains running along a text. Both grammatical

cohesion (more precisely, reference) and lexical cohesion are recruited in participant tracking. By reference is meant a relationship in meaning construed through the use of a personal reference item (personal pronoun or possessive determiner) that enters into a semantic relationship with an item mentioned either before in the text (anaphora) or afterwards (cataphora). Occasionally, reference is made to entities that cannot be retrieved from the text and need to be established situationally; this is referred to as exophoric reference. When two items share the identity of a referent, this is termed *co-referentiality*. Items sharing identity can also be linked through lexical cohesion, be that repetition, synonymy or hyponymy.

Cohesive ties, i.e. semantically bonded items, are particularly important when they form so-called *cohesive chains*, responsible for strong integration of cohesive ties and a more coherent text. A cohesive chain built on participant tracking may be realized through co-reference or lexical cohesion categories that are valid for language in general but that ultimately need to be interpreted in a particular text. Thus, this type of cohesive chain is crucial to text organization and comprehension. Conversely, secondary chains are not essential to participant tracking and are built upon lexical relations that are not text-specific, but general to the lexicon of the language.

When text is processed in translation tasks, Hatim and Mason (1990) argue, translators rely both on contextual and co-textual cues in order to identify cohesive items deemed relevant to a coherent construction of the TT. These cues can be sought in the immediate co-text or demand integration of items that are more distant in the text.

Cohesion has not been extensively examined in translation process research. Denver (2009) investigated adversative-concessive logical-semantic relations in translations from Spanish into Danish. The author found different right and wrong choices among translators and students, but no trace of mental activity in processing relations realized through conjunctions in Spanish, i.e. no verbalization or keystroke, pause or revision signalling that the relations constituted translation problems for the participants.

Angelone (2010) studied uncertainty management and metacognitive problem solving of a professional translator, two students and a bilingual. He classified the textual level at which the participants' metacognitive activity was employed into lexis, term, collocation, phrasal, syntax, sentential, macro level, and unclassified. The macro level category refers, according to the author, to beyond sentence considerations, such as cohesion, coherence, and gender. Only a small part of the elicited verbalizations fell into the sentential and macro level categories.

Both authors relied on think-aloud protocol data; Denver also used key logging data, and Angelone also used screen recordings. To the best of our knowledge, no other translation process research using eye-tracking has addressed translators' or post-editors' processing of cohesion in STs and TTs.

Staub and Rayner (2007) claim that many eye-tracking studies have focused on syntactic parsing, but few have looked into how discourse processing (including cohesion) affects eye movements in reading. Staub and Rayner (2007, 335) argue that recognizing individual words and analyzing grammatical structures of each

sentence does not suffice to understand a text; the reader "must also maintain a representation of the entities and events that have been mentioned, and relate the information that is currently being processed to this stored representation."

Basically, eye-tracking studies focusing on cohesion have so far shown an increase in fixation times due to: long distance between an anaphor and its antecedent (O'Brien et al. 1997); antecedent being a low-frequency word (van Gompel and Majid 2004); and reading a target word and drawing conclusions that have not been explicitly stated in the text (O'Brien et al. 1988).

11.3 Methodology

The results described in this chapter are part of a larger empirical-experimental project carried out by the AuTema-PostEd Group, which aims at tapping into translation and post-editing processes as a source of insight into the role of translators' understanding in task problem solving. In this chapter, we report the results regarding the Portuguese(L2)-Chinese(L1) translation and post-editing of a text about the China Gold Research Institute.

11.3.1 Equipment and Analysis Tools

Data from gaze and keyboard activity were collected and analyzed using Translog-II (Carl 2012, 2013), version 0.1.0189, connected to a Tobii T120 remote eye tracker. The eye-tracking software application Tobii Studio 3.2.1 was also used as a recorder for the participants' verbalizations and gestures. Calibration was performed in both Translog-II and Tobii Studio, the latter running in the background while the participants worked in Translog-II.

Figure 11.1 shows the screen setting in the post-editing task: the ST appears in the top half of the application window, and the TT in the bottom half, which is empty in the translation task. The ST font was Tahoma, and the TT font was SimSun with font size 17. Both texts were double-spaced.

11.3.2 Participants

Originally, 23 professional translators performed two translation tasks (L1 into L2, and L2 into L1) and two post-editing tasks (one in their L1 and another one in their L2) using machine-translated (MT) input provided by the software
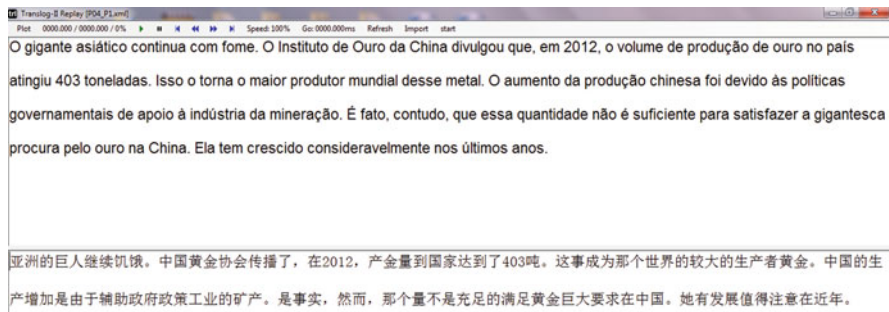
Fig. 11.1 Screenshot of Translog-II user post-editing window

Table 11.1 Results for the three quality assurance criteria

Task	Subject	Mean fixation duration (in ms)	Gaze time on screen (in %)	% of valid win gaze data
Post-editing	P03	317	78.07	70.03
	P09	366	91.98	83.15
	P11	434	93.41	91.80
	P15	439	82.85	77.56
	P19	417	82.28	79.76
	P23	365	86.81	79.37
Translation	P04	421	89.24	76.75
	P06	299	82.98	78.00
	P07	222	63.81	61.95
	P12	309	82.67	88.24
	P18	301	49.59	43.56
	P21	465	83.69	67.16
Post-editing	Mean	389.67	85.90	80.28
Translation		336.17	75.33	69.28

PCT (Portuguese-Chinese Translator[1]) (Wong and Chao 2010). In this chapter, we report the analysis of Text 1 (see Sect. 11.3.4), for which we have data from 17 participants[2] contained in the MS13 study within TPR-DB (translation session 16, post-editing session 18). Only 12 of the participants with more than 1 year of experience and an age range of 23–32 years remained in our sample (see Table 11.1). The others were excluded because of failure to comply with data quality criteria as described in Sect. 11.3.3.

---

[1] Quadrilingual MT. Available at http://nlp2ct.cis.umac.mo/MT/.

[2] Six participants' Text 1 data were excluded because of technical issues related to tokenization or alignment problems (see Sect. 11.3.6).

All participants provided informed consent. They were Chinese nationals and had an undergraduate degree in Portuguese Studies or a Master's degree in Chinese-Portuguese Translation Studies. All had been granted a fellowship to live in a Portuguese speaking country, and all of them used glasses or contact lenses. None of them had experience in post-editing. Each was paid MOP 90.00 to take part in the experiments.

The participants were instructed to sit approximately 55 cm away from the eye-tracker monitor. They were also told that they could move freely, but were told to keep their eyes on the monitor as much as possible.

11.3.3 Data Quality

For data collection, we tried to cope with the numerous factors that may have had an effect on the quality of the gaze data, especially lighting, glasses, and distance from the monitor. However, in order to ensure consistency in the sample, three data quality assurance criteria were observed. Data should comply with at least two of the three criteria to ensure that the results were not skewed by flawed data.

The first criterion was mean fixation duration: following Sjørup (2013) and Hvelplund (2011), our threshold was established at a minimum of 180 ms. The second criterion was gaze time on screen (GTS), that is, the percentage of time spent gazing on the text in relation to the total time of translation production: once again, following Sjørup (2013) and Hvelplund (2011), our threshold was 30 %. The third criterion, called "% of valid win gaze data," was calculated in terms of the percentage of valid gaze data on the ST and TT token considering the attribute "win" in the XML files produced after each Translog session. More specifically, we divided the number of occurrences of win = 1 (gaze on ST) plus win = 2 (gaze on TT) by the total number of "wins", which included both win = 1 and win = 2 and also win = 0 (gaze not ascribed to either the ST or the TT). As, to the best of our knowledge, no study has reported such a measure before, we arbitrarily established our threshold at ≥40 %.

Table 11.1 shows the figures of the remaining data considering the three criteria mentioned above.

11.3.4 The Experimental Text

The ST, Text 1, is a short news report written in Brazilian Portuguese on the increase of gold extraction and consumer market in China (see Appendix 2). Chain A is the main chain, where the participant being tracked is production volume. Chain B is the secondary chain and refers to the country (i.e. China). Table 11.2 shows both chains and their cohesive devices.

Table 11.2 Main chain and cohesive devices of ST (selected tokens in italics)

Type of chain	Tokens in ST	Explicitation of referents as	Co-reference established through
A	*o volume de produção* atingiu 403 toneladas "*the production volume* increased reaching 403 tons"	–	Not applicable (first item)
	Isso "This"	The fact that the production volume increased	Demonstrative pronoun
	O aumento "The increase"	The production increase	Definite article + lexical noun (synonym)
	essa quantidade "such amount"	The amount of production increase	Demonstrative pronoun + lexical noun (superordinate)
B	*O gigante asiático* "The Asian giant"	–	Not applicable (first item)
	n*o país* "in *the country*"	In the country which is the Asian giant	Definite article + lexical noun (synonym)
	o torna "turns *it* into"	Turns the Asian giant into	Personal pronoun
	n*a China* "in China"	In China, which is the Asian giant	Lexical noun (synonym)

The relative position and length of the selected tokens (words and noun groups) in chain A and chain B were accounted for in our statistical analysis (see Sect. 11.4.2).

11.3.5 Task

After a brief warm-up session, which consisted of a copy test before the experiment, each participant was asked to perform four tasks, randomly assigned to participants: two translations (one into their L1 and another one into their L2), and two post-editing tasks (one in their L1 and another one in their L2) using MT output.

Table 11.3 provides the tasks performed by each participant. The analyses in this chapter refer to T1 and P1 highlighted in Table 11.3.

As a brief, the participants were informed that they should render texts aimed at a target audience analogous to that of the ST. They were told to feel free to produce the human TT or post-edit the MT text without any time constraint, but they could not use any kind of translation aids. As they had little to no experience in post-editing, we provided them with guidelines reported in Mesa-Lao (2014, 225), see also chapters 11.3.5, 11.7, 11.8, and 13 in this volume.

After each task, the participants were requested to provide a retrospective protocol, whereby they could explain whatever they felt like concerning their

Table 11.3 Task distribution
across participants

Participant	From scratch		Post-editing	
03	**T1**	T3	P4	P2
04	T2	T4	P3	**P1**
06	T3	T2	**P1**	P4
07	T2	T3	P4	**P1**
09	T3	**T1**	P2	P4
11	**T1**	T3	P2	P4
12	T2	T4	**P1**	P3
14	T3	T2	P4	**P1**
15	**T1**	T4	P3	P2
17	T3	**T1**	P2	P4
18	T4	T2	**P1**	P3
23	**T1**	T3	P2	P4

Note: T = translation; P = post-editing; #1–2 = Portuguese ST; #3–4 = Chinese ST

translation or post-editing, such as difficulties, challenges, strategies, doubts. The retrospective protocols were carried out by means of the Translog-II Supervisor replay function (Jakobsen 2011, 39).

11.3.6   Processing of Chinese Data

In the CRITT TPR-DB are four study folders containing data of translation and post-editing involving both Chinese language and a Latin alphabetic script (see Appendix 1, Chap. 2). As Chinese is a logographic language which does not require blank spaces between the characters (Zang et al. 2011), the processing of the Chinese language data involved additional procedures so that they could be automatically analyzed using the Study Analysis script.[3] These procedures are described in Sects. 11.3.6.1 and 11.3.6.2.

11.3.6.1   Chinese Input System

A logographic language like Chinese requires an input method[4] through a graphic user interface (GUI), which converts sequences of alphabetic letters into Chinese characters. The participants used Sogou[5] as their Chinese input method.

---

[3]Available at https://sites.google.com/site/centretranslationinnovation/translog-ii.

[4]Basically, there are two categories of Chinese input method, i.e. phonetic readings or root shapes. Most of these input methods can be selected directly from the control panel of MS Windows.

[5]Available at http://pinyin.sogou.com/.

Fig. 11.2 Snapshot of a post-editing session showing Sogou's dialog box. Note: *Circle* indicates fixation

Figure 11.2 shows a snapshot of a post-editing session. In the bottom half of the window, Sogou's dialog box pops up below the line where the participant wants to introduce new characters. A zoom-in shows that while the participant types in *pinyin*[6] a series of alphabetic letters, a number of options are shown out of which one may be the desired corresponding character(s). To select the desired characters and insert them in Translog-II, the participant presses the space bar or the corresponding number key.

As shown in Fig. 11.2, the Chinese input system is prone to word gaze error, since the place where pinyin is typed is not the same place where the Chinese character is inserted. Assuming that the Sogou's dialog box pops up right below the space where the character is supposed to be inserted, we manually[7] attribute the fixation

---

[6]*Pinyin* is the official phonetic system for transcribing the *Putonghua* (Mandarin) pronunciations of Chinese characters into the Latin alphabet.

[7]The "StudyAnalysis" scripts include a function to refixate gaze mapping to word. However, several translation drifts remain and need to be corrected manually.

to a specific word through Translog-II Supervisor[8] with support of Tobii Studio replay function.

11.3.6.2 Chinese Tokenization and Alignment

The procedures to analyze Translog-II data can be retrieved from the CRITT website.[9] However, some additional steps were required to tokenize and align the data because of Chinese language specificities. As Chinese texts are written as a stream of characters without blank spaces, there is no explicit delimiter to identify word boundaries and automatically tokenize the data using the "StudyAnalysis.pl tokenize" script. To tackle these problems, we came up with an alternative workflow and developed applications to (semi-)automate the process.

Figure 11.3 shows the conventional and the alternative workflows, which contains additional steps 0, 2.5, and 3 as in Fig. 11.3b. Firstly, we fix the incorrect gaze data (see Sect. 11.3.6.1) and save each log file as a new *.xml file. Secondly, a Chinese tokenization step (Step 2.5) is added after the extraction of text data in Step 2: We use the in-house developed tool, *ChiSegmentor* (Leong et al. 2006; Zeng et al. 2013), to automatically identify the word boundaries, and then we manually revise the output drawing on the *Modern Standard Chinese Dictionary* (Li 2010). This information is recorded in the corresponding log file—*.src or *.tgt.

Another change to the workflow is in the alignment step, for which we use *LexAligner* (Tian et al. 2011) to automatically estimate possible word alignments. To check the alignments, we draw on criteria provided by the *Guidelines for Chinese-English Word Alignment* (Li et al. 2009). Because the translation renditions are a result from the processing of an ST (Mossop 2003), we align all ST and TT tokens.
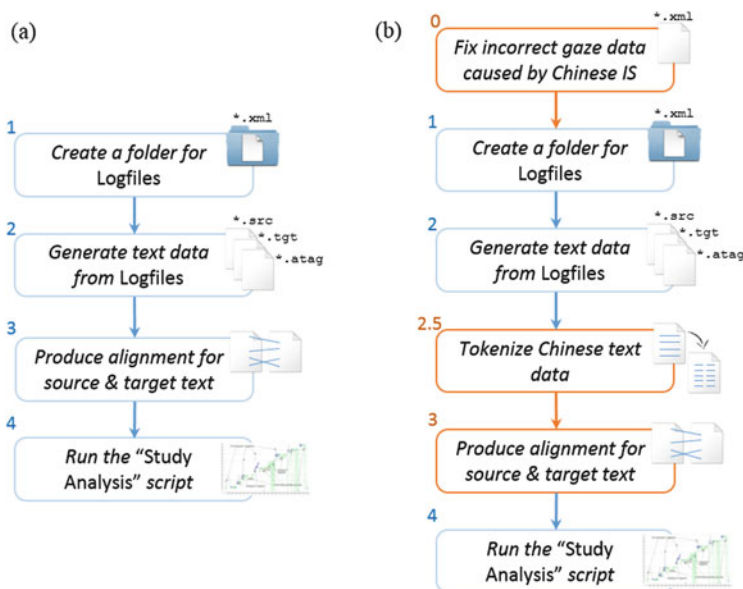
Finally, we run the "StudyAnalysis.pl tables" script to extract several kinds of simple and compound process and product units, which are represented in tables (see Chap. 2 for details). From these units it is possible to generate Translation Progression Graph (TPG) using the R environment for statistical computing.

It is worth noticing that sometimes Chinese and Portuguese tokens have encoding conflict that prevents us from generating TPGs. To overcome this problem, which is identified while running R to generate TPGs, we replace the problematic character with *pinyin* for Chinese and with "a" for Portuguese.[10]

---

[8]Fix Map device, available at https://sites.google.com/site/centretranslationinnovation/translog-ii.

[9]Available at https://sites.google.com/site/centretranslationinnovation/translog-ii.

[10]For further details, please contact the authors.

Fig. 11.3 Procedures to prepare Translog-II data for "Study Analysis": (**a**) Conventional process for Roman-alphabet languages; (**b**) Adapted process for the Chinese language

11.3.6.3 Chinese Production Data Extraction

For technical reasons related to the Chinese input method,[11] which is external to Translog-II, the system logs only the text modifications (delete and space keys), but not the actual keystrokes (see Sect. 11.3.6.1). Table 11.4 shows examples of incorrect number of character insertions in the target tokens (TToken) (see Sect. 2.4): in the column Ins, the number "2" refers to the Chinese characters shown on the screen, whereas the column should provide the number of keystrokes actually typed (five insertions for $TTid_{50}$ and nine insertions for $TTid_{51}$, see Fig. 11.2).

A problem also occurs with the duration of the production time of Chinese characters (as reported in Balling and Carl 2014, 260). As shown in Table 11.5, it is common to find long pauses and short production times (most of them of 1 ms), when in fact the participant had virtually no pause and took longer to produce a given token (TToken).

To solve the aforementioned problems, we use Tobii Studio replay function to identify when exactly the participants started and stopped typing keystrokes corresponding to each word logged in Translog-II. While doing this, we also count

---

[11]There is no immediate connection between the keystrokes and the characters that appear in the text.

Table 11.4 Example of wrong log of keystroke insertions and deletions

TTid	TToken	SToken	Ins	Del	Edit
50	这个	É_fato_,_que_essa	2	0	这个
51	数量	quantidade	2	0	数量

Table 11.5 Example of wrong log of pause and duration (Dur)

TTid	TToken	SToken	Time	Dur	Pause	Edit
50	这个	É_fato_,_que_essa	153911	1	4353	这个
51	数量	quantidade	153913	1	1	数量

Table 11.6 Example of intermediate production alignment problem (Edit1)

TTid	TToken	SToken	Edit1	Time1	Dur1	Pause1	Edit
49	,	,	[那个量]"that amount"	149153	405	10780	那个量
50	这个	É_fato_,_que_essa" It is a fact, that this"		153911	1	4353	这个
51	数量	quantidade "quantitity"		153913		1	数量

Note: Deleted words are in brackets

the actual keystrokes that the participants pressed to produce the characters logged in Translog-II. Then, we manually correct the TT tables.

We also observed that the edited units are coherently aligned to the respective STid and TTid in most of the cases. However, because Translog-II aligns only the initial ST and the final TT, when the characters of the MT text are deleted and/or immediately edited (such as Edit 1 [那个量], which represents the deletion of [that amount] in TTid$_{49}$, Table 11.6), the system does not identify which words they were originally part of (a part of Edit 1 in TTid$_{49}$ should belong to TT$_{50}$ 那个 [that], and another part 量 [amount] should belong to TT$_{51}$). To account for this, we check all Edit1 and Edit2 actual operations to identify the actual ST and TT tokens (STokens and TTokens).

In sum, a substantial part of our production analysis was built on manually processed data. To ensure quality, all manually extracted data were double-checked. These spreadsheets are available in TPR-DB.

11.4 Data Analysis

11.4.1 Research Question

Our main objective in this study is to compare the cognitive effort demanded for translating and post-editing the selected tokens in the main cohesive chain (chain A) in the ST, which is built on participant tracking, with the cognitive effort

demanded for processing selected tokens in a secondary cohesive chain (chain B) (see Sect. 11.3.4).

Our first research question is "Is it cognitively more demanding to understand and produce a cohesive chain that is built on participant tracking than a secondary cohesive chain?" This question is based on the assumption that participant tracking is crucial to construing a coherent representation of a text (Halliday and Hasan 1976).

Our second research question is "In dealing with cohesive chains, is it cognitively more demanding to translate than to post-edit?". Given that translation takes longer than post-editing (Balling and Carl 2014; Mesa-Lao 2014), it is possible that processing cohesive chains during post-editing is also faster or that cohesive chains are processed differently in the two tasks.

11.4.2 Statistical Analysis and Variables

Our investigation is divided into three statistical analyses: (1) eye movements on chain A and chain B in the ST, (2) eye movements on chain A and chain B in the TT, and (3) keyboard movements relating to chain A and chain B.

In the following, we describe the variables of analysis in the order they were included in the statistical model (for details on the model, see Sect. 11.4.3).

The dependent variables[12] for analyses (1) and (2) were:

- Total reading time on ST and TT token (TrtS and TrtT);
- Number of fixations on ST and TT token (FixS and FixT); and
- First pass duration on ST and TT token (FPDurS and FPDurT).

The dependent variable for analysis (3) was TT token total production time. Time was measured considering any pauses preceding a TT token plus duration (Dur) (see Sjørup 2013, 126–127 for further details).

Our analysis investigates how the dependent variables vary as a function of several explanatory variables, as described below. For further details, see Baayen (2008) and Balling (2008).

The first group of explanatory variables consists of random effects. Random factors are not repeatable and are assumed to have been selected randomly from any given population (Baayen 2008, 241). As such, we included the participants and item, i.e., the selected ST and TT tokens of chains A and B (see Sect. 11.3.4).

The second group of explanatory variables consists of fixed effects, which refer to factors with repeatable levels (Baayen 2008, 241). They were used to account for previous studies that have reported their effect on the results or their importance for TT cohesion. Due to space restrictions, we report only on the ones that were significant in our model (see Table 11.10 in Appendix 1).

---

[12]Descriptions for these dependent variables are available in Sect. 2.4.6.

Four fixed effects were used in the analysis of both eye and keyboard movements as proxies for processing of chains A and B in both ST and TT:

- Token Length: The length in characters of the ST and TT tokens of chains A and B was expected to affect the dependent variable, as longer words generally receive longer fixations than shorter words (e.g. Rayner 1998; Hyönä et al. 2003; Staub and Rayner 2007);
- Token Position: The position of the ST and TT tokens of chains A and B in the text was expected to have an effect on the participants' gaze behaviour. It may be due to fatigue (e.g. Rayner 1998; Balling 2008, 2013) and/or to a priming effect (Rayner 1998: 390; Staub and Rayner 2007: 331), which, based on Halliday and Hasan (1976), may imply that the beginning of a text deserves more attention because it will determine the understanding of the remaining of the text and it will have items that will serve as referents for items further in the text (see also Chap. 9);
- Token Unigram Frequency: Readers are expected to fixate longer on low-frequency words than on high-frequency words (e.g. Rayner et al. 2005; Rayner 1998). The Corpus of Portuguese[13] and the Corpus of the Peking University Center for Chinese Linguistics[14] were used to measure frequency;
- Token Trigram Probability: High predictability of word association was expected to have an impact on processing effort (Frisson et al. 2005). The variable was computed following McDonald and Shillcock (2003, 650) and considering the selected token and the two preceding tokens as they occur in the text (Balling 2013).

Two variables were added to the analysis of TT Token production, namely:

- Token Character Count: sum of insertions and deletions of TT Token (see Sect. 11.3.6.3); and
- Correctness of Token in the Chain: TT Tokens were assessed as right or wrong in lieu of the ST Tokens; right TT Tokens were assumed to be as instance in which "patterns of lexical cohesion in texts are maintained, subject to the constraints of particular text norms in particular languages" (Hatim and Mason 1990, 200).

Two other fixed effects were included to directly answer our research questions:

- Task: to investigate the effect of translating or post-editing on ST and TT comprehension and on TT production; and
- Type of Chain: to investigate the effect of chain A and chain B on ST and TT comprehension and on TT production.

All continuous variables were naturally logarithmically transformed.

---

[13]Available at http://www2.lael.pucsp.br/corpora/bp/.

[14]Available at http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp.

11.4.3   Data Analysis: Statistical Models

Following the methods used in Balling and Carl (2014, 250ff.) and Sjørup (2013), we applied a linear mixed-effect regression model (LMER) as implemented in the lme4 package (Bates et al. 2014) in the R environment for statistical computing (version 3.1.2, R development Core Team 2014). The final model, containing only the significant variables, is summarized in the appendix. Table 11.10 in Appendix 1 provides the fixed effects, with variable names in the first column, estimated effect size in the second column, the standard error of this estimate (indicating the amount of variation in the data) in the third column, the t-value in the fourth column, and the associated p-value in the fifth column. We set the significance level at $p \leq 0.05$.

Table 11.11 in Appendix 1 shows the random effects part of the model. Random effects are not associated with p-values, but are included in the model in order to estimate individual effects and dependencies between observations. The standard error of the effects indicates how much variation the different levels capture.

11.5   Results and Discussion

In this session, we report the fixed effects that had a significant impact on the dependent variables.

11.5.1   Comprehension: Eye Movements along Chain A and Chain B in the ST

Table 11.7 summarizes the effects ($\sqrt{}$) that had a significant impact on our dependent variables related to ST comprehension. Type of chain and type of task were non-significant for all dependent variables relating to source text processing.

ST Token length had a significant effect on total reading time on ST Token (Fig. 11.4) and number of fixations on ST Token (Fig. 11.5). This confirms claims in the literature (e.g. Rayner 1998, 387; Sjørup 2013, 140) that the longer the word or words, the longer the gaze time and the higher the number of fixations.

Position of ST Token had a significant effect on total reading time (Fig. 11.6) and number of fixations on ST Token (Fig. 11.7). The ST Tokens in initial positions were gazed longer and more often than tokens towards the final positions. Since the

Table 11.7 Summary of significant results for ST comprehension

| Variable	TrtS	FixS	FPDurS
Token length	$\sqrt{}$	$\sqrt{}$	–
Token position	$\sqrt{}$	$\sqrt{}$	–
Token trigram probability	$\sqrt{}$	$\sqrt{}$	$\sqrt{}$

Fig. 11.4 Total reading time
on ST Token (TrtS, in ms) vs.
length of ST Token in the
chain (character count
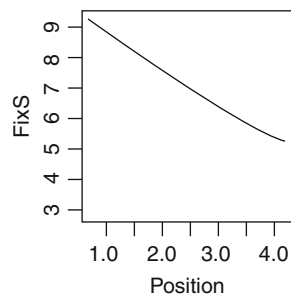naturally logarithmically
transformed)
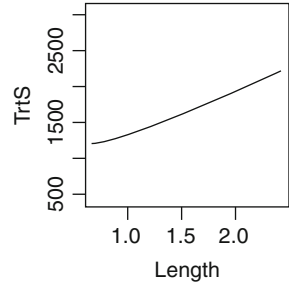
Fig. 11.5 Number of
fixations on ST Token (FixS)
vs. length of ST Token in the
chain (character count
naturally logarithmically
transformed)
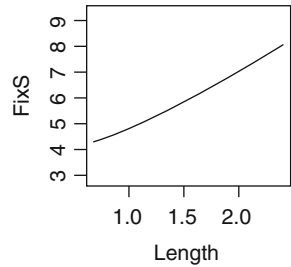
Fig. 11.6 Total reading time
on ST Token (TrtS, in ms.)
vs. position of ST Token
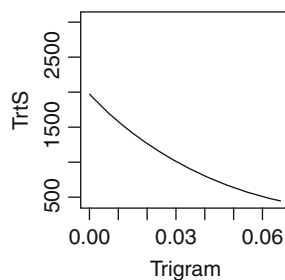(naturally logarithmically
transformed)
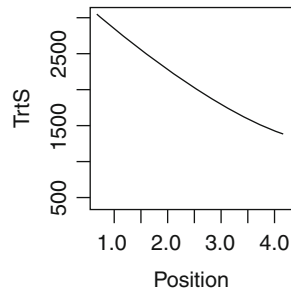
Fig. 11.7 Number of
fixations on ST Token (FixS)
vs. position of ST Token
(naturally logarithmically
transformed)

type of chain had no significant impact on the eye movements on the ST and the text
was relatively short, the results seem to indicate a priming effect: the initial items
in the chains receive more attention from participants because the beginning of the
text is crucial for their orientation in order to understand the entire text (Halliday
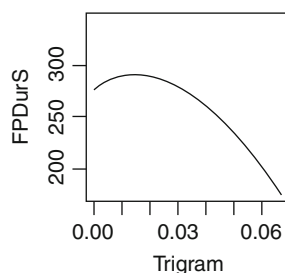and Hasan 1976).

Fig. 11.8 Total reading time on ST Token (TrtS, in ms.) vs. trigram probability (naturally logarithmically transformed)

Fig. 11.9 Number of fixations on ST Token (FixS) vs. trigram probability (naturally logarithmically transformed)
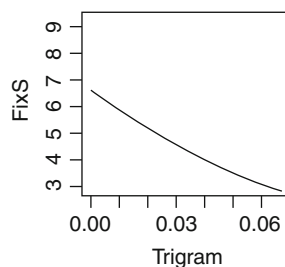
Fig. 11.10 First pass reading time on ST Token (FPDurS) vs. trigram probability (naturally logarithmically transformed)

The probability of word association, measured through trigram probability, had a significant effect on total reading time on ST Token (Fig. 11.8) and number of fixations on ST Token (Fig. 11.9). This confirms findings in the literature, such as Frisson et al.'s (2005) and McDonald and Shillcock's (2003), who found that the more probable or common a word or expression is, the shorter and the less often it is fixated.

The probability of a three-word combination also had a significant effect on first pass duration (Fig. 11.10), which may be indicative of processing of higher level information (see Staub and Rayner 2007, 329). According to our results, the less probable the occurrence of such combination, the more often and the longer it was fixated in subjects' first gaze on it (see the effect of trigram probability on FPDurS).

## 11.5.2 Comprehension: Eye Movements along Chain A and Chain B in the TT

Table 11.8 summarizes the effects ($\checkmark$) that had significant impact on the dependent variables related to TT comprehension. Each variable had an impact on only one of the dependent variables. The type of task had no impact on any of the dependent variables regarding eye movements on the TT. No variable impacted on first pass duration on TT Token.

Token length was significant for number of fixations on TT Token: the longer the token, the greater the number of fixations (Fig. 11.11). Unigram frequency was significant for total reading time on TT Token There seems to be a tendency in fixating more on both the most and the least frequent words; the reason for more fixations on frequent words may be related to their role in the chain, as we are observing instances of participant tracking (Fig. 11.12). Type of chain was significant for total reading time on TT Token: (Figure 11.12): tokens in chain A were fixated longer than those in chain B (Fig. 11.13).
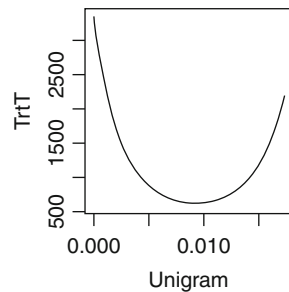
Table 11.8 Summary of significant results for TT comprehension

| Variable	TrtT	FixT	FPDurT
Token Length	–	$\checkmark$	–
Token unigram frequency	$\checkmark$	–	–
Type of chain (A)	$\checkmark$	–	–

Fig. 11.11 Number of fixations on TT Token (FixT) vs. length of TT Token in the chain (naturally logarithmically transformed)
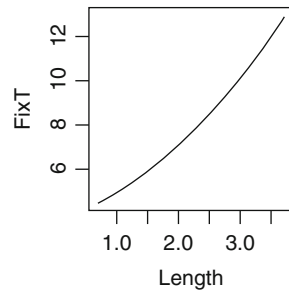
Fig. 11.12 Total reading time on TT Token (TrtT, in ms.) vs. unigram frequency of TT Token (naturally logarithmically transformed)

Fig. 11.13 Total reading time on TT Token (TrtT, in ms.) vs. type of chain (A and B)
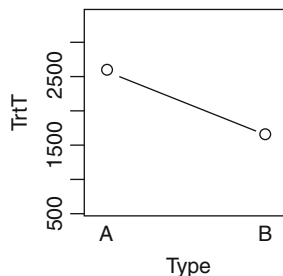
Table 11.9 Summary of significant results for TT production

Variable	Duration of token production time
Character count	✓
Correctness of token (wrong)	✓
Type of chain (A)	✓

11.5.3 *Production: Keyboard Movements for Producing of Chain A and Chain B*

Table 11.9 summarizes the effects (✓) that had significant impact on the dependent variable "duration of token production time."

The more the participants inserted or deleted characters, the longer was the duration of their token production time (Fig. 11.14). The participants took longer to produce a wrong token in the chain than to produce a right item (Fig. 11.15), which may be related to the number of renditions that they provided while being uncertain to what would be an adequate solution. The participants also took longer to produce the items in chain A (Fig. 11.16).

The longer time for producing tokens in chain A may be indicative of hesitation, need for internal support to make decisions, as well as on-line revisions. These results seem to be consistent with our previous findings (Sect. 11.5.2). The participants' verbalizations also showed that they found it difficult to render some items in chain A. The type of task did not have any significant effect on the results. This suggests that processing of cohesive ties is similar in translation from scratch and post-editing.

11.6 Summary and Future Directions

We set out to answer two research questions: (1) Is it cognitively more demanding to understand and produce a cohesive chain that is built on participant tracking than a secondary cohesive chain? And (2), in dealing with cohesive chains, is it cognitively more demanding to translate than to post-edit?
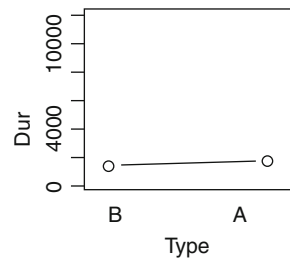
Fig. 11.14 Duration of token production time (Dur, in ms.) vs. character count (naturally logarithmically transformed)
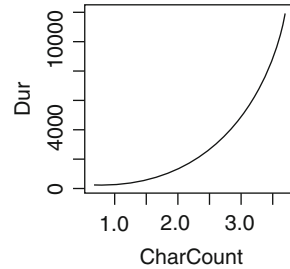
Fig. 11.15 Duration of token production time (Dur, in ms.) vs. correctness of token (*R* right, *W* wrong)
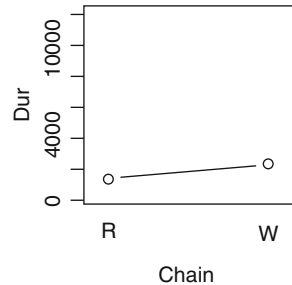
Fig. 11.16 Duration of token production time (Dur, in ms.) vs. type of chain

As for question 1, our results show a significant effect of the type of cohesive chain on eye movements on the TT, which let us infer that it is cognitively more demanding to produce a chain built on participant tracking when it comes to the TT, but no significant effect was observed for the metrics related to the ST.

We expected that the results were also significant for the ST, especially because we assumed that keeping track of participants in the main chain of a text would be challenging both to understand the ST and to produce the TT. However, we should be aware that both translation and post-editing involve transiting gaze from ST to TT, and therefore, the reading of the TT may be closely connected to the understanding of the ST. Future studies should include transiting from ST to TT and vice-versa as also a measure of effort and investigate if ST and TT comprehension should be addressed as one single event. Measures of global processing should also be developed and tried out to account for "relationships between pieces of text information that span relatively long distances in a text" (Hyönä et al. 2003, 314), especially across sentences.

Furthermore, in a larger scale study involving the four tasks for which we collected data, we intend to check if the order in which the task was carried out had an effect on the participants' processing. As we collected all data from each participant on the same day, there might have been an effect of fatigue on the results (e.g., the total time of the last sessions seems to be shorter than that of the first sessions).

As for question 2, the results pointed to no significant impact of type of task on the measures that we assessed in this chapter. We have two potential non-mutually exclusive explanations for this. One reason might be experience—none of the participants had PE experience. Another factor might have been that our design is between subjects, so differences in the results may have to do with differences in the groups. The other potential explanation is that the sample size is small. Considering previous studies that do show significant differences between post-editing and translation (e.g. Balling and Carl 2014; Mesa-Lao 2014), it is possible that either the impact of the type of task is on the transitions across both ST and TT areas or that type of task has an impact on the global processing of the entire text, rather than only on particular cohesive chains. A third possibility concerns to MT quality, as reported in the protocols, participants found the MT text ambiguous at some points and exophoric reference to what MT tokens refer also lacked.

To address some of the aforementioned limitations, in future work we intend to analyse more than one task and use a between subject design, which will allow us to have data for participants that both translated and post-edited. We also intend to compare the tasks considering the entire text and a larger volume of data (including four texts). Following Alves et al. (2014), we also intend to perform a more fine-grained analysis by qualitatively examining the renditions and their processing as shown, for instance, in scan paths.

Besides answering the research questions, the alternatives we had to come up with in order to cope with limitations to process Chinese language data are also a contribution of the present chapter. We hope that the procedures we reported herein contribute to facilitating further studies involving the Chinese language and that our results awake the interest of new scholars to approach language pairs other than those involving only alphabetic scripts. Although the tendency to use the same language pairs and scripts may have methodological advantages (e.g., one language, English, is kept as a standard for comparison's sake, and blank spaces undoubtedly delimit words), not only does it overlook the insights that other language pairs and scripts may add to understanding (non-)language and (non-)script specific cognitive aspects of post-editing and translation, but it also prevents future generalizations based on a comprehensive body of research encompassing multiple languages, language pairs, and scripts.

Appendix 1: Summary of Mixed-Effects Analysis

Table 11.10 Fixed effects in the analysis of ST and TT tokens total reading time, and total production time with estimated effects size, standard error, t- and p-values

Variable	Estimate	Std. error	t	p
Total reading time on ST Token				
Intercept	7.750364	0.310935	24.926	<2e-16
Log token length	0.359864	0.092903	3.874	0.000164
Log token position	−0.223413	0.062151	−3.595	0.000450
Log trigram probability	−22.56563	3.176764	−7.103	5.78e-11
Task (translation)	0.269650	0.301090	0.896	0.387272
Type of chain (A)	−0.008212	0.149548	−0.055	0.956287
Number of fixations on ST Token				
Intercept	1.83671	0.21923	8.378	1.34e-11
Log token length	0.36931	0.07098	5.203	6.90e-07
Log token position	−0.16169	0.04754	−3.401	0.000879
Log trigram probability	−12.67565	2.43980	−5.195	7.18e-07
Task (translation)	0.30240	0.19646	1.539	0.148426
Type of chain (A)	0.01305	0.11462	0.114	0.909544
First pass reading time on ST Token				
Intercept	5.61694	0.09665	58.117	<2e-16
Log trigram probability	−1.39461	0.50621	−2.755	0.00664
Task (translation)	0.18261	0.13136	1.390	0.18839
Total reading time on TT Token				
Intercept	7.2591	0.2175	33.371	<2e-16
Log frequency (poly. 1)	−2.6958	1.1700	−2.304	0.0229
Log frequency (poly. 2)	4.7131	1.1569	4.074	8.18e-05
Task (translation)	0.2772	0.2543	1.090	0.2980
Type of chain (A)	0.4325	0.2063	2.097	0.0381
Number of fixations on TT Token				
Intercept	1.2405	0.2820	4.399	0.000176
Log length	0.3549	0.1034	3.431	0.000806
Type of chain (A)	0.3347	0.2449	1.367	0.219303
TT Token total production time				
Intercept	4.56944	0.19611	23.300	<2e-16
Log character count	1.29942	0.07501	17.323	<2e-16
Correctness of token (wrong)	0.51087	0.13583	3.761	0.000265
Type of chain (A)	0.20504	0.09570	2.142	0.034301

Table 11.11 Random effects in the analysis of ST and TT comprehension and production time

Variable	Random factor	Intercept/level	Standard deviation
Total reading time on ST Token	ST Token	Intercept	0
	Participant	Intercept	0.5097
	Residual		0.6155
Number of fixations on ST Token	ST Token	Intercept	0
	Participant	Intercept	0.3244
	Residual		0.4712
First pass reading time on ST Token	SToken	Intercept	1.565e-08
	Participant	Intercept	1.848e-01
	Residual		5.045e-01
Total reading time on TT Token	TT Token	Intercept	0
	Participant	Intercept	0.2783
	Residual		1.1424
Number of fixations on TT Token	TT Token	Intercept	0.2393
	Participant	Intercept	0
	Residual		0.8528
TT Token total production time	TT Token	Intercept	0
	Participant	Intercept	0.2401
	Residual		0.4782

Appendix 2: Source Text

O gigante asiático (5) continua com fome. O Instituto de Ouro da China divulgou que, em 2012, **o volume de produção (1)** de ouro no país (6) atingiu 403 toneladas. **Isso (2)** o (7) torna o maior produtor mundial desse metal. **O aumento (3)** da produção chinesa foi devido às políticas governamentais de apoio à indústria da mineração. É fato, contudo, que **essa quantidade (4)** não é suficiente para satisfazer a gigantesca procura pelo ouro na China (8). Ela tem crescido consideravelmente nos últimos anos.

References

Alves, F., Pagano, A. S., & da Silva, I. A. L. (2014). Effortful text production in translation. *Translation and Interpreting Studies, 9*(1), 25–51.

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.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to Statistical Using R*. Cambridge: Cambridge University Press.

Balling, L. W. (2008). A brief introduction to regression designs and mixed-effects modelling by a recent convert. In S. Göpferich, A. L. Jakobsen, & I. Mees (Eds.), *Looking at eyes: Eye-*

tracking studies of reading and translation processing (Copenhagen studies in language, Vol. 36, pp. 175–192). Frederiksberg: Samfundslitteratur.

Balling, L. W. (2013). Reading authentic texts: What counts as cognate? *Bilingualism: Language and Cognition, 16*(3), 637–653.

Balling, L., & Carl, M. (2014). Production time across language and tasks: A large-scale analysis using the CRITT translation process database. In J. Schwieter & A. Ferreira (Eds.), *The development of translation competence: Theories and methodologies from psycholinguistics and cognitive science* (pp. 239–268). Cambridge: Cambridge Scholar Publishing.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4. R package version 3.1.2*. Available at http://CRAN.R-project.org/package=lme4

Bell, R. T. (1991). *Translation and translating: Theory and practice*. London: Longman.

Carl, M. (2012). Translog-II: A program for recording user activity data for empirical reading and writing research. In *Proceedings of the eighth international conference on language resource and evaluation* (pp. 4108–4112). Istanbul: European Language Resources Association.

Carl, M. (2013). Feature representation in the translation process research DB. In R. Bonk, V. Alabau, M. Carl, & P. Koehn (Eds.), *D5.3: Beta release of Casmacat workbench*. Available at http://www.casmacat.eu/uploads/Deliverables/d5.3.pdf

Carl, M., & Dragsted, B. (2012). Inside the monitor model: Process of default and challenged translation production. *Translation: Corpora, Computation, Cognition, 2*(1), 127–145. Special issue on the Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation.

Carl, M., & Jakobsen, A. L. (2009). Towards statistical modelling of translator's activity data. *International Journal of Speech Technology, 12*(4), 125–138.

Denver, L. (2009). Unique items in translation. In S. Göpferich, A. Jakobsen, & I. Mees (Eds.), *Behind the mind: Methods, models and results in translation process research* (pp. 125–148). Copenhagen: Samfundslitteratur.

Frisson, S., Rayner, K., & Pickering, M. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(5), 862–877.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. New York: Longman.

Hasan, R. (1984). Coherence and cohesive harmony. In J. Flood (Ed.), *Understanding reading comprehension: Cognition, language and the structure of prose* (pp. 181–219). Newark, DE: International Reading Association.

Hatim, B., & Mason, I. (1990). *Discourse and the translator*. New York: Longman.

Hvelplund, K. T. (2011). *Allocation of cognitive resources in translation: An eye-tracking and key-logging study*. Published PhD thesis, Copenhagen Business School, Copenhagen.

Hyönä, J., Lorch, R. F., Jr., & Rinck, M. (2003). Eye movements measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Amsterdam: North-Holland.

Jakobsen, A. L. (2011). Tracking translators' keystrokes and eye movements with Translog. In C. Alvstad, A. Hild, & E. Tiselius (Eds.), *Methods and strategies of process research: Integrative approaches in translation studies* (pp. 37–55). Amsterdam: John Benjamins.

Leong, K. S., Wong, F. D., Tang, C. W., & Dong, M. (2006). CSAT: A Chinese segmentation and tagging module based on the interpolated probabilistic model. In Z. H. Yuan & M. W. Yao (Eds.), *Computational methods in engineering and science* (pp. 1092–1098). Sanya: Tsinghua University Press/Springer.

Li, X. J. (Ed.). (2010). 现代汉语规范词典 *[Modern Standard Chinese Dictionary]*. Beijing: Foreign Language Teaching and Research Press.

Li, X. S., Grimes, S., & Strassel, S. (2009). *Linguistic data consortium. Guidelines for Chinese-English word alignment, version 4.0*. Philadelphia, PA: Linguistic Data Consortium. Available via https://catalog.ldc.upenn.edu/docs/LDC2012T16/GALE_Chinese_alignment_guidelines_v4.0.pdf

McDonald, S. A., & Shillcock, R. (2003). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14*(6), 648–652.

Mesa-Lao, B. (2014). Gaze behavior on source texts: An exploratory study comparing translation and post-editing. In S. O'Brien, L. W. Balling, M. Carl, M. Simard, & L. Specia (Eds.), *Post-editing of machine translation* (pp. 219–245). Newcastle upon Tyne: Cambridge Scholar Publishing.

Mossop, B. (2003). *An Alternative to "Deverbalization"*. http://www.yorku.ca/brmossop/Deverbalization.htm

O'Brien, E. J., Raney, G. E., Albrecht, J. E., & Rayner, K. (1997). Processes involved in the resolution of explicit anaphors. *Discourse Processes, 23*, 1–24.

O'Brien, E. J., Shank, D. M., Myers, J. L., & Rayner, K. (1988). Elaborative inferences during reading: Do they occur on-line? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 410–420.

Rayner, K. (1998). Eye movements in reading and information processing. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K., Li, X., Juhasz, B. Z., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review, 12*(6), 1089–1093.

Sjørup, A. C. (2013). *Cognitive effort in metaphor translation: An eye-tracking and key-logging study.* Published thesis, Copenhagen Business School, Copenhagen.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 327–342). Oxford: Oxford University Press.

Tian, L., Wong, F., & Chao, S. (2011). Word alignment using GIZA++ on Windows. *Machine Translation Summit, 13*, 369–372.

van Gompel, R. P. G., & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition, 90*, 255–264.

Wong, D. F., & Chao, L. S. (2010). PCT: Portuguese-Chinese machine translation systems. *Journal of Translation Studies, 13*(1–2), 181–196.

Zang, C. L., Liversedge, S. P., Bai, X. J., & Yan, G. (2011). Eye movements during Chinese reading. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 961–978). Oxford: Oxford University Press.

Zeng, X. D., Wong, D. F., Chao, S., & Trancoso, I. (2013). Graph-based semi-supervised model for joint Chinese word segmentation and part-of-speech tagging. In *Proceedings of the 51st annual meeting of the association for computational linguistics (ACL 2013)* (pp. 770–779). Sofia, Bulgaria: Association for Computational Linguistics.