Text characteristics, perceived difficulty and task performance in sight translation: An exploratory study of university-level students

Abstract

This paper reports on an exploratory study examining the relationship between text characteristics, perceived difficulty and task performance in sight translation (ST). Twenty-nine undergraduate interpreters were asked to sight-translate six texts with different properties. Correlation analysis shows that Sophisticated Word Type and Mean Length of a T-unit are, respectively, the lexical and the syntactic variables having the highest correlations with all the three dependent variables (i.e. perceived difficulty, accuracy and fluency in ST performance). Surprisingly, the discoursal variables are weakly or modestly correlated with the dependent variables. Thematic analysis of the students' reflective essays points to two hypothesized causal links among the three Ps in ST: task *properties* may cause decoding difficulties and cognitive overload in the cognitive *process*, which in turn lead to inaccuracy and dysfluency in ST *performance*. The research findings lend empirical support to the "shallow-scan hypothesis" in previous research. Finally, this study proposes a three-tier conceptual framework to inform and guide future research to operationalize variables in ST empirical studies. The pedagogical implications of ST are also discussed.

Key words: text characteristics, perceived difficulty, task performance, sight translation

1. Introduction

Sight translation (ST) is the rendition from a written text in one language to an oral form in another language (Chen 2015). ST is different from sight interpreting (Pöchhacker 2004) or sight interpretation (Lambert 2004) in that interpreters do not have access to aural input when performing ST. Thus, interpreters can orally translate the text at their own speed, without being pressured by the speed of speakers (Lee 2012). In professional settings, ST has been widely used in international conferences, bilateral meetings, exhibitions, hospitals and courts (Chen 2015). In pedagogical settings, ST is an integral part in interpreter training curricula, either as a stand-alone course or as an embedded component of a translation or interpreting course. ST is also an important testing component in aptitude tests (Moser-Mercer 1994) and certification of professional interpreters (Wallace 2013).

As ST requires "double conversion [of] both language and medium" (Setton & Dawrant 2016: 205), efforts have been made to understand the differences between ST, written translation, consecutive interpreting, and simultaneous interpreting (Viezzi 1989; 1990). Researchers have explored the cognitive demands (Agrifoglio 2004; Jiménez Ivars 2008), cognitive process (Shreve et al. 2010, 2011); accuracy rate (Dragsted & Hansen 2007; Lambert 2004), and output speed (Dragsted & Hansen 2009) in ST. These studies are mainly experimental, conducted to examine the ontological features that distinguish ST from other types of translation and interpreting. As regards ST pedagogy, anecdotal suggestions are offered for ST curriculum design (Ersozlu 2005), and empirical studies

have been conducted to show how ST enables students to acquire interpreting skills (Lee 2012), such as sensitizing students to syntactic and stylistic differences between the source and the target languages (Viaggio 1995).

The studies reviewed above primarily focus on the process and/or the performance of ST, without giving due attention to the source text (see Shreve et al. 2010 and Shreve et al. 2011 for two rare exceptions). Campbell's (1999) study indicates that the source text in and of itself is an independent source of translation difficulty. This observation also holds true in ST. After comparing 22 students' ST and written translation performances, Jiménez Ivars (2008) points out that students regard source text comprehension as the top cause of ST problems. However, the existing literature offers very little information about how characteristics of the source text may influence the perception of difficulty and the performance quality in ST. To address this research gap, the current study attempts to explore the tripartite relationship between text characteristics, difficulty perception and task performance. The research findings can inform interpreter trainers to select and adapt materials to match students' developmental levels of ST competency, as unrealistically difficult materials are found to be a demotivating factor among interpreter trainees (Wu 2016). Additionally, based on substantial revision of Robinson's (2011) framework, this paper proposes a three-tier conceptual framework that hopefully will guide future studies on the interrelationship between task complexity, task condition, task difficulty and task performance in ST.

2. Text characteristics, readability and task complexity

2.1 *Text characteristics and readability*

In literacy education, it is important to separate difficult texts from easy ones so that learners can choose appropriate reading materials in accordance with their literacy proficiency. The concept of "readability" was thus proposed to measure "how easily written materials can be read and understood" (Richards et al. 1992: 306). To objectively measure text difficulty, efforts have been made to devise a readability formula by examining text characteristics. By far, the most popular readability formula is the Flesch Reading Ease Readability Formula (Flesch 1948), which uses word length and sentence length to determine text difficulty. While the Flesch score has been widely used to select reading materials in literacy education, it has a weak correlation with perceived difficulty in written translation (Sun & Shreve 2014). Additionally, Liu and Chiu (2009) reported a strange correlation between the Flesch score and consecutive interpreting performance: an easier source text led to poorer interpreting performance. Therefore, it would be interesting to know whether the correlations in written translation and consecutive interpreting tasks can be applied to ST.

2.2 *Text characteristics and task complexity in language tasks*

Task-related variables and task performance have been consistently studied in developmental psychology (Deane et al. 2006), literacy education (Ozuru et al. 2008) and language acquisition (Robinson 2001). To conceptualize task-related variables, Robinson

(2011) proposed a Triadic Componential Framework that includes Task Complexity, Task Condition and Task Difficulty. Task Complexity is "the result of attentional, memory, and other information processing demands imposed by the structure of the task on the language learner" (Robinson 2001: 29); Task Condition is the interactive factor of a task, focusing on how a task is executed; and Task Difficulty is the perceptual dimension of a task, which is influenced by many learner factors, including language proficiency and affective variables.

Text characteristics are an important variable in determining task complexity. In language input and output tasks, it has been found that text characteristics can predict the quality of task performance. For instance, lexical range, density, and diversity and causal content are identified as significant predictors of learners' second language (L2) listening comprehension (Révész & Brunfaut 2013). Additionally, lexical frequency, content word overlap, and syntax similarity are recognized as the three significant predictors of L2 text difficulty (Crossley et al. 2008). As regards language output tasks, Yu (2009) discovered that source text characteristics (such as discourse structure, lexical diversity and percentage of passivization) have great effects on the summary writing performance.

As ST involves both language input and output, it would be interesting to know whether the existing findings can be applied to ST tasks or a different set of text characteristics correlate with ST task performance.

2.3 *Text characteristics and ST task complexity*

While previous studies have pointed out that difficult words and complex sentences cause increased cognitive effort (Shreve et al. 2010) and dysfluencies (Shreve et al. 2011) in ST, to the best of the author's knowledge, there has been no empirical research examining the relationship between text characteristics, difficulty perception and task performance. To compound the problem, the parameters of text characteristics have not been rigorously defined. As Shreve et al. (2010: 70) confessed, "[o]perational definitions of *syntactically complex* and *syntactically non-complex* were informal" (original emphasis). In these circumstances, researchers do not have a reference framework to compare their research findings, because what is considered syntactically complex in one study might not be perceived as such in another. The current study attempts to address this issue by comparing three banks of quantifiable variables: lexical, syntactic and discoursal, to provide a reference framework for future ST studies to compare research results.

3. Research design

3.1 *Research questions*

To explore the relationship between source text properties, perceived difficulty and ST performance, the following research questions are raised:

a) How do source text characteristics correlate with perceived difficulty and ST performance?

b) How do text characteristics, mediated by cognitive process, influence ST performance?

3.2 *Participants*

The current study involved an intact class of 29 undergraduate students at a CIUTI (Conférence Internationale Permanente d'Instituts Universitaires de Traducteurs et Interprètes) member institute in China. At the time of the study, the students had received one semester of training on the basic interpreting skills, such as active listening, use of short-term memory, note-taking, public speaking, and delivery. When the study was conducted, the students were enrolled in the course of *Topic-based Interpreting*, focusing on practical skills to deal with recurring subject domains in interpreting.

3.3 *Directionality*

In this study, the students were asked to sight-translate from English (their L2) into Chinese Mandarin (their L1). The directionality was controlled because it has been reported that interpreting from L1 to L2 is perceived to be more stressful and tiring by spoken language interpreters (Nicodemus & Emmorey 2013). If this study had chosen the L1-to-L2 directionality, the students might have overrated the perceived difficulty. Thus, the L2-to-L1 directionality ensured that the difficulties reported by the students were truly reflective of the source text characteristics, the focus of this study.

3.4 *Source texts*

In this study, the texts for sight translation were all about business and economy. This obviates the possibility that unfamiliarity with a certain subject domain would skew the perceived difficulty and task performance. Topic description and word count of the source texts are provided in Table 1.

Table 1. Basic information about source texts

Text	Topic	Number of words
1	Job description	118
2	Extreme poverty	103
3	Working at home	115
4	Business environment	94
5	Bilateral trade	111
6	Tourism and its economic benefits	102

In addition, the length of the texts was controlled between 90 and 120 words. Although the author shares the concern that the texts are not of the typical length in professional interpreting settings, the decision to control the length was actually justified by a 12-week pilot study, conducted with 10 undergraduates from the same program but from a different class. It was found that when students performed sight translation for more than three minutes (equivalent to 150 words in the source text), half of them reported that their motivation level dropped. As this study does not focus on students' affective variables,

the text length was thus controlled to make sure that all participants had a comparable level of engagement.

Thirdly, a total of six texts were used in this 6-week study (i.e. one text per week). In the pilot study, four out of 10 students exhibited developmental differences in the seventh ST test, and eight of them in the eighth test, evidenced by significant increase of accuracy and/or fluency in ST performance. In other words, by the end of the seventh test, 40% of participants had shown signs of improving their ST competency, and 80% of them in the eighth test. To make sure that the dynamic development of ST competency would not skew the perception of how difficult the source texts were, the study was designed to last for 6 weeks, within which students were relatively stable in terms of their ST competency.

3.5 *Procedure*

Over a period of six weeks, the students were required to sight-translate one text as a warm-up activity at the beginning of each lesson in a language lab. They were given one minute to read the paragraph, and then up to four minutes to sight-translate the text. At the end of each test, the students were prompted to rate their perception of task difficulty by filling out a 5-item 6-point Likert scale questionnaire (see Section 3.6 for details). They were also asked to make a reflective commentary on their performance as a take-home assignment. They were provided the source text and the recording of their ST performance. They were required to compare the source text with their performance and

retrospectively reported all the difficulties they experienced during the task. A total of 174 (29×6) reflective essays were collected, totaling 90 000 words.

3.6 *Variables and analytical procedure*

The selection of independent variables (i.e. text characteristics) was guided by the literature review (see Section 2) and the psycholinguistic theorization of reading as a three-level process: lexical cognition, syntactic parsing, and meaning construction (Crossley et al. 2008; Rayner & Pollatsek 1994). Accordingly, three banks of independent variables were included in this study: lexical, syntactic and discoursal.

Lexical variables. There are many lexical variables that measure various aspects of lexical complexity of a text. Informed by the research results in Lu (2012) and Révész & Brunfaut (2013), this study narrowed these down to three types of measures to assess the lexical complexity of the source texts (see Table 2).

Table 2. Lexical variables examined in this study

Type	Measure	Label
Frequency count	Word type	WT
	Sophisticated word type	SWT
	Word token	WK
	Sophisticated word token	SWK
Lexical density	Lexical density (number of lexical words divided by number of word tokens)	LD

Lexical diversity	Number of word type divided by number of word tokens	TTR

"Token" refers to the number of words, while "type" means the number of distinct words. In this study, sophisticated words are defined as those that are outside the top 2 000 frequent words in the British National Corpus (Leech et al. 2001). Additionally, "lexical words" are defined as nouns, adjectives, verbs (excluding modal verbs and auxiliary verbs) and adverbs with an adjectival base (Engber 1995; Lu 2012).

Syntactic variables. Similarly, there are many syntactic variables that measure syntactic complexity of a text. Informed by the research results in Lu (2011) and Révész and Brunfaut (2013), this study narrowed these down to three types of measures to assess the syntactic complexity of the source texts (see Table 3).

Table 3. Syntactic variables examined in this study

Type	Measure	Label
Length	Mean length of a sentence	MLS
	Mean length of a T-unit	MLT
	Mean length of a clause	MLC
Subordination	Clauses per T-unit	C/T
	T-units per sentence	T/S
Phrasal structure	Complex nominals per T-unit	CN/T

A T-unit is defined as "one main clause plus any subordinate clause or non-clausal structure that is attached to or embedded in it" (Hunt 1970: 4). For example, the

following sentence contains two T-units: “Mary is the general manager of a large company and she is very busy.” Complex nominals refer to: a) nouns modified by adjectives, possessives, relative clauses, prepositional phrases, participles, or appositives; b) noun clauses; and c) gerunds (Cooper 1976).

Discoursal variables. Based on McNamara et al (2014), two discoursal variables were included in this study: referential cohesion (RC) and deep cohesion (DC). Referential cohesion is achieved “when a noun, pronoun, or noun phrase […] refers to another constituent in the text” (McNamara et al. 2014: 49-50). Deep cohesion is measured by the “causal and intentional connectives when there are causal and logical relationships within the text” (McNamara et al. 2014: 85).

Benchmark variable. As explained in Section 2.1, the Flesch score was used as the benchmark variable to see whether the results reported in Sun and Shreve’s (2014) written translation tasks and Liu and Chiu’s (2009) consecutive interpreting tasks could be applied to ST performance. The Flesch scores rated the source texts on a 100-point scale, with easier texts having higher scores.

In this study, three dependent variables were examined: perceived difficulty, accuracy and fluency in task performance.

Perceived difficulty. To obtain quantitative data about perceived difficulty, a questionnaire was adapted from Robinson’s (2001) Task Difficulty Questionnaire with substantial modification. Five questions with a 6-point Likert scale were designed to solicit students’ difficulty perception, with 1 being least difficult and 6 being the most difficult. Five aspects were examined: a) overall difficulty perception of the ST task; b) perceived difficulty during the one-minute preparation; c) perceived readability of the

text; d) perceived difficulty in accurately sight-translating the text; and e) perceived difficulty in fluently sight-translating the text. Cronbach's alpha ($\alpha = 0.893$) showed relatively high internal consistency of items in the questionnaire. The total score of the five items was taken as the perceived difficulty score for each text.

Task performance. Based on Liu (2013) and Cai et al. (2015), students' ST performances were graded on two aspects: accuracy and fluency. To rate accuracy, each text was divided into 20 idea units. When students fully rendered the intended message in each unit, one point was awarded. When they conveyed part of the message (e.g. translating "this stunning progress" as *gai jinbu* ["this progress"]), half a point was awarded. When they misinterpreted or missed the message, no point was given. Fluency was assessed on the speech flow of the ST performance. When the rendition had no incidence of dysfluency features (e.g. false starts, repetition, self-repair), the performance was given one point. When some dysfluency features were present but did not disrupt understanding, half a point was awarded. When dysfluency features disrupted understanding (e.g. an exceptionally long pause), no point was given. Two raters with more than 10 years as interpreter trainers and practising interpreters independently rated students' ST performances. Inter-rater reliability was very high in accuracy ($r = 0.96$, $p < 0.001$), and also high in fluency ($r = 0.87$, $p < 0.001$). The average of the two raters' ratings were taken as the accuracy score and the fluency score.

In summary, the independent variables examined were the text characteristics, which subsumed six lexical variables, six syntactic variables, two discoursal variables, and one benchmark variable. The dependent variables were the perceived difficulty and the ST performance (which subsumed accuracy and fluency). Following Liu and Chiu

(2009) and Liu (2013), the author performed correlation analysis to examine the relationship between all the variables in this study.

As correlation does not mean causation, thematic analysis was performed on students' reflective essays to hypothesize the causal links between text characteristics, cognitive process, and ST performances. An iterative process following the grounded theory approach (Strauss & Corbin 1990) was adopted. In the first stage, all the reflections were read by the author and no coding was performed to avoid biased preconceptions without looking at the bigger picture. In the second stage, all the portions related to text characteristics were highlighted and summarized in tentative codes (e.g. "unknown words" and "long segments"). In the third stage, these tentative codes were examined and grouped thematically. In this process, formal codes were finalized to characterize a certain set of expressed experiences and feelings about text characteristics. In the fourth stage, the author thematically coded students' comments that explained why certain text characteristics caused task difficulty. Four cognitive issues were identified (see Section 4.2). In the final stage, all the codes generated in the third and fourth stages were compared and their causal relationships were established. To check the coding reliability, a second coder (with 15 years of teaching interpreting at the same CUITI institute) independently coded 30% of the comments. Inter-rater agreement for text characteristics and cognitive issues was 97.46% and 94.96% respectively.

Table 4
Correlation Matrix

	WT	SWT	SWK	WK	LD	TTR	MLS	MLT	MLC	C/T	T/S	CN/T	RC	DC	Flesch	PerDiff	Accuracy
SWT	.295**																
SWK	-.282**	.490**															
WK	.160*	-.811**	-.316**														
LD	.331**	.784**	.441**	-.545**													
TTR	.696**	.820**	-.015	-.596**	.644**												
MLS	.263**	.857**	.357**	-.815**	.590**	.812**											
MLT	.251**	.923**	.391**	-.861**	.640**	.832**	.987**										
MLC	.444**	.723**	.420**	-.498**	.843**	.716**	.797**	.767**									
C/T	-.189*	.507**	.042	-.713**	-.070	.375**	.493**	.558**	-.102								
T/S	-.012	-.679**	-.322**	.545**	-.491**	-.385**	-.246**	-.399**	-.078	-.548**							
CN/T	.126	.868**	.446**	-.858**	.510**	.732**	.977**	.981**	.685**	.626**	-.343**						
RC	-.563**	-.496**	.504**	.469**	-.270**	-.803**	-.457**	-.501**	-.212**	-.530**	.409**	-.402**					
DC	.645**	-.521**	-.706**	.709**	-.317**	.020	-.367**	-.447**	-.124	-.542**	.598**	-.514**	-.166*				
Flesh	.040	-.866**	-.577**	.871**	-.838**	-.590**	-.813**	-.850**	-.783**	-.325**	.486**	-.808**	.228**	.672**			
PerDiff	.122	.654**	.255**	-.632**	.424**	.558**	.644**	.673**	.448**	.469**	-.387**	.660**	-.386**	-.374**	-.583**		
Accuracy	-.229**	-.677**	-.278**	.585**	-.527**	-.608**	-.657**	-.681**	-.556**	-.343**	.358**	-.646**	.376**	.302**	.611**	-.541**	
Fluency	-.146	-.556**	-.326**	.480**	-.409**	-.467**	-.587**	-.592**	-.497**	-.272**	.220**	-.585**	.210**	.284**	.522**	-.378**	.578**

*$p < 0.05$
**$p < 0.01$
PerDiff = Perceived Difficulty

4. Results

4.1 *Quantitative analysis*

Table 4 shows the correlation matrix for all the variables examined in this study. The majority of variables are significantly correlated with one another to a greater or lesser extent. However, it is important to note that the number of data points is 174 (6×29). With this relatively large sample size, even small correlation coefficients can become significant. This is what Meehl (1990) described as the "Crud Factor"– eventually all variables correlate to some extent with one another. Therefore, it makes better sense to compare the magnitudes of correlation coefficients in this exploratory study. To facilitate comparison, the highest coefficients observed in the lexical and the syntactic variables are underlined and bolded in Table 4. As can be seen, of the six lexical variables, SWT (Sophisticated Word Type) has the highest correlation with all the three dependent variables. Lexical density (measured by LD) and lexical diversity (represented by TTR) are also correlated with the dependent variables, but the correlation coefficients are lower. This seems to suggest that sophisticated words might be the most challenging lexical issue for the students. Of the six syntactic variables, MLT (Mean Length of a T-unit) has the highest correlation with all three dependent variables. Interestingly, the numbers of clauses and T-units (measured by C/T and T/S) have much lower correlations with the dependent variables. This may indicate that the length of a T-unit is more of a problem than the number of a T-unit. This is further corroborated by the high correlation

observed between CN/T (Complex Nominals per T-unit) and the dependent variables, as complex nominals add to the length of a T-unit.

Surprisingly, the two discoursal variables (RC and DC) are weakly or modestly correlated with the dependent variables (with *r* ranging from 0.21 to -0.386). This result is unexpected, because previous studies have shown that cohesion and discoursal structure have a strong impact on reading comprehension (Louwerse & Mitchell 2003; McNamara et al. 2010). One would expect discoursal cohesion to strongly correlate with perceived difficulty and ST performance. A possible reason for the discrepancy is that the students processed the source texts at an intra-sentential not inter-sentential level. This issue will be discussed in Section 5.2 in relation to the cognitive units in ST. Equally unexpectedly, the baseline variable (the Flesch score) correlates with all the dependent variables (with *r* ranging from -0.583 to 0.611), which is different from the findings reported in Liu and Chiu's study (2009) and in Sun and Shreve's (2014) study. These inconsistencies will be discussed in Section 5.3.

As shown in the bottom right of Table 4, perceived difficulty is correlated with ST accuracy ($r = -0.541$, $p < 0.001$) and fluency ($r = -0.378$, $p < 0.001$). This might suggest that the students' perception of ST difficulty is more sensitive to the accuracy issues than the fluency issues. Additionally, ST accuracy is moderately correlated with fluency ($r = 0.578$, $p < 0.001$), which is congruent with the correlation coefficients reported in Yeh and Liu (2006). Liu (2013) noted that there might be a (reverse) halo effect when raters assessed accuracy and delivery and asked whether the two criteria should be judged separately. Although this study does not provide conclusive evidence, it is interesting to note that SWT and MLT are highly correlated ($r = 0.923$, $p < 0.001$), which produces

similar correlation coefficients with accuracy and fluency ($r_{SWT*Accuracy}$ = -0.677, $r_{MLT*Accuracy}$ = -0.681; $r_{SWT*Fluency}$ = -0.556, $r_{MLT*Fluency}$ = -0.592). As such, it is quantitatively reasonable to interpret that SWT and MLT are unidimensional (i.e. measuring the same construct), as are accuracy and fluency (see Liu 2013). However, the qualitative analysis presented in the next subsection will show that SWT and MLT, among other text properties, have different impacts on accuracy and fluency. This may point to the multidimensionality of ST performance, reflective of different cognitive processes evoked by different text properties.

4.2 *Qualitative analysis*

To examine how text characteristics affected the students' perception and performance, their reflective essays were qualitatively analyzed. As reported in Section 3.6, the thematic analysis generated two sets of codes. The first set was about the source text properties perceived by the students as troubling. A total of 1 064 comments were made, which could be grouped into two categories: lexis and syntax. Subcategories were also coded for further analysis. Table 5 shows the frequency and example of each subcategory.

Table 5. Text characteristics causing perceived difficulty

Category	Subcategory	Frequency	Example
Lexis	Unknown words	162	I don't know what "pandemics" means.

Category	Subcategory	Frequency	Example
	Known words with unfamiliar meanings	109	I know what "inclusive" means, but it just doesn't sound right in this context.
	Unfamiliar collocations	72	I know every individual word. But when they are put together, I don't know what they mean.
	Numbers	68	I had trouble translating "7 billion".
	Formal words	45	I tried to search for a more formal word to match the source text.
Syntax	Long segments	274	This segment is very long. I could not read ahead while I did the task.
	Syntactic structures	187	I did not realize there was a prepositional phrase at the end.
	Syntactic order	105	The source text puts the temporal information at the end of the sentence. It sounds strange in Chinese.
	Rhetorical style	42	The speaker used a double negative to emphasize his/her stance on working at home. I hesitated because I wondered whether I should preserve the double negative in sight translation.

The students made more comments about syntactic features (57.14%) than lexical features (42.86%). Scrutiny of Table 5 reveals that unknown words and long segments are the top lexical and syntactic features reported by the students. This corroborates with

the quantitative findings, as SWT and MLT are, respectively, the top lexical and syntactic variables correlating with perceived difficulty. In addition to the factors examined in the quantitative analysis, the qualitative analysis discovered three more factors that might cause difficulties in ST tasks: numbers, formal words, and rhetorical style, although they were less frequently reported by the students.

The second set of coding was performed to uncover why certain text characteristics caused difficulties in ST performance. Totally, the students made 834 comments, which could be grouped into four cognitive issues (see Table 6).

Table 6. Cognitive issues caused by text characteristics

Cognitive Issues	Text characteristics	
	Lexis	Syntax
Decoding	211	117
Failure to read ahead	52	148
Searching for information	34	93
Shallow processing	85	93

It is noted that decoding issues were primarily caused by difficult lexis. Decoding issues include problems of comprehending an unknown word or syntactic structure and problems of misunderstanding a known word or syntactic structure. For instance, Jacky (all pseudonyms) noted how a sophisticated word had caused a decoding issue in ST: "*I simply do not recognize the word 'pandemic', so I have to skip this part"* (Reflection #2-27). Similarly, Cindy reported a decoding issue involving syntax: *"I thought the relative*

clause modified the preceding phrase 'roads and hotels.' Actually, it modified 'Tourism' located half a sentence away" (Refection #5-18).

Additionally, the cognitive overload issues (represented by failure to read ahead and searching for information) were largely caused by difficult syntax. Sam described his frustration when he dealt with the complex syntactic structure: "*The gerund in this sentence was a real pain in the neck. I had to pause and force myself to search for the logical subject [of the gerund] and put it first in the sight translation"* (Reflection #4-19). Similarly, Cathy explained why she repaired her rendition: *"[…] 'time and money spent on travelling', two nouns modified by a phrase […] I was not able to read ahead and place it before the nouns in Chinese, so I had to repeat the nouns in the sight translation"* (Reflection #3-07). These extracts lend further support to the quantitative finding that CN/T is highly correlated with all the dependent variables. When the student interpreters were forced to unpack the complex structures, they struggled to manage their cognitive resources and could not make any further effort to read ahead or search for information.

Finally, it is noted that both lexis and syntax could cause the issue of shallow processing, whereby the students opted for literal translation as a cognitively less challenging rendition. For instance, Kate wrote, *"I literally translated all the pronouns (*he *and* she*). I should have explicated what* he *and* she *referred to, but I had little energy to figure them out"* (Reflection #1-12). Benjamin also noted, *"The sentence begins with the parenthesis 'when done carefully'; I was not sure what it meant, so I translated it word for word"* (Reflection #5-21). Susan summed up the issue of shallow processing thus*: "word-for-word translation seems to be my automatic response […] I have to pay extra attention and effort if I want to refrain from it"* (Reflection #3-08). These

comments are congruent with Lee's (2012) observation that students have troubles distancing themselves from the source texts and processing the texts at a deeper level.

5. Discussion

5.1 *Relationship between text properties, cognitive process and ST performance*

Although the results are obtained from a group of university-level students and cannot be generalized to include a wider range of ST activities, the tentative findings can be used to hypothesize the causal relationships between the three Ps in ST: task *properties*, cognitive *process* and ST *performance*. The first hypothesized causal link is represented in the middle panel of Figure 1. In the correlation analysis, SWT was found to highly correlate with ST accuracy. Close examination of the students' reflective essays further showed that decoding difficult words was a major cause of mistranslation. This result is in line with Lee's study (2012), in which student interpreters usually got stuck on difficult or unfamiliar words. Previous studies also show that lexis has significant impacts on L2 reading comprehension (Jeon & Yamashita 2014; Rydland et al. 2012). This is why decoding issues may be caused by infrequent words and/or meanings, which in turn may lead to inaccuracy in ST performance.

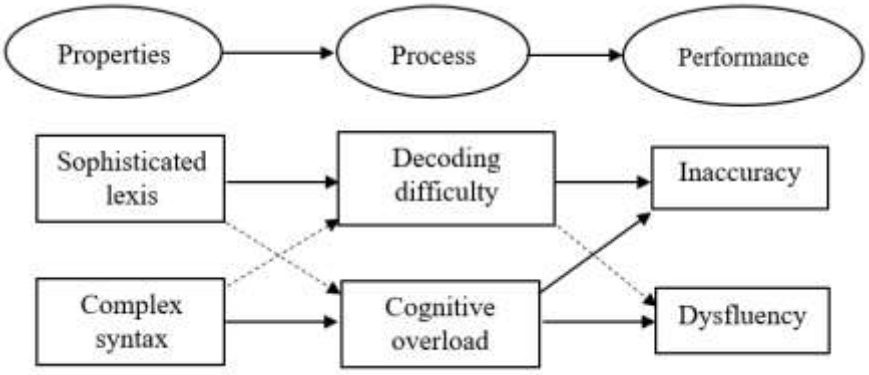

Figure 1. Hypothesized relationships between task properties, cognitive process and performance product

The second hypothesized causal link is represented in the bottom panel of Figure 1. As discovered in the correlation analysis, MLT and CN/C were highly correlated with perceived difficulty and ST performance. This relationship may be caused by the cognitive overload issue on the part of the students. For instance, Mary noted, "*I was too preoccupied translating the noun. I did not read ahead and mistakenly assumed the prepositional phrase was to modify the next clause"* (Reflection #6-09). Helen also explained, *"I stuttered, because I felt my brain was drained, trying to figure out the logical subject of the gerund [...]"* (Reflection #6-28). These reflections lend empirical support to Agrifoglio's (2004) effort model and Gile's revised model (2009) for ST. Originally, in Gile's model (1997: 198), only efforts in reading and producing are required (ST = R + P), while in Agrifoglio's model (2004: 47) and Gile's revised model (2009: 179), efforts in reading, producing, short-term memory and coordination are competing for a finite amount of cognitive capacity. The retrospective analysis offered by the students unequivocally showed that they struggled to manage their cognitive

resources in ST tasks. When they came across a long T-unit or a complex nominal, they should have retained certain information (e.g. the noun being modified) in their short-term memory and simultaneously read ahead (for the modifiers). The complex syntactic features led to ST inaccuracy and disfluency, because they demanded more efforts than simple syntax (Shreve et al. 2010) and the students could not devote any effort to short-term memory or coordination.

Finally, in the reflective essays, it was found that sophisticated words could also give rise to cognitive overload and that complex syntax could lead to decoding problems, although the reported frequencies were not very high (see Table 6). Thus, in Figure 1, the solid arrows represent relatively stronger causal links, while the dotted arrows represent relatively weaker causal links. In terms of magnitude, while both sophisticated lexis and complex syntax pose challenges to students in ST, the latter seems to have a more pronounced effect on the cognitive overload, subsequently leading to inaccuracy and disfluency.

5.2 *Discoursal variables*

Surprisingly, discoursal features (i.e. referential cohesion and deep cohesion) weakly or modestly correlate with the three dependent variables. In the students' reflective essays, no incidences were reported about the discoursal variables causing ST difficulties (see Table 5). Discoursal variables were consistently *not* recognized as salient factors in either the quantitative or the qualitative analysis for at least two reasons. First, although cohesive texts are easier to comprehend (Louwerse & Mitchell 2003; McNamara et al.

2010), studies also show that cohesive texts with difficult words are still challenging to read (McNamara et al. 2012). This could explain why no discoursal variables, compared with the lexical variables, were recognized as challenging factors. Second, as textual cohesion indexes a deep and global level of cognition (Graesser et al 2004), it can be surmised that the students processed the source texts at a shallow level, without attending to the textual cohesion in ST. For instance, the quantitative analysis found that the mean length of a T-unit was the syntactic variable having the highest correlations with all the dependent variables. This was corroborated by the students' reflective essays. For instance, as Calvin confessed, *"Honestly, I did not try to connect the dots in the text […] because I was still able to translate the text segment by segment, with the script at hand […]"* (Reflection #6-16). This extract resonates with Dragsted and Hansen's (2009) study, where eye-tracking data indicate that the eye movement and gaze pattern of ST interpreters are linear, compared with the scattered gaze fixation and backtracking behaviors of translators. Taken together, this study lends empirical support to the "shallow-scan hypothesis" in ST (Shreve et al. 2010: 65) – students tend to process the ST texts shallowly and progress linearly in a small cognitive unit (e.g. a T-unit), overlooking the idea progression or cohesion in the text.

5.3 *Flesch score*

Different from the previous research (Sun & Shreve 2014), this study found that the Flesch score was strongly correlated with perceived difficulty. This inconsistency can be explained in at least two ways. First, in Sun and Shreve's (2014) study, the participants

were allowed to consult dictionaries in the written translation tests. As such, when they came across words with unknown or unfamiliar meanings, they could look them up, thereby solving the decoding issues. In the current study, the students had no access to reference materials, so difficult lexis had a large impact on perceived difficulty. Second, as Shreve et al. (2010: 83) point out, compared with ST tasks, written translation "places less pressure on working memory". Therefore, students can allocate more attentional efforts to comprehend and analyze sentences in translation tasks, so sentence length is less disruptive in written translation tasks. This further explains why the Flesch score was weakly correlated with perceived difficulty in written translation (Sun & Shreve 2014), but strongly correlated in the current study.

Previous research also reported a peculiar relationship between the Flesch score and consecutive interpreting performance: "the more difficult a source material is, the better the performance in interpreting" (Liu & Chiu 2009: 256). However, this study found that the Flesch score was negatively correlated with perceived difficulty and positively correlated with ST performance. One possible explanation for the divergent findings is that students in ST are usually sensitive to the surface structure of a text, as they are engaged in a local and shallow level of cognition. In consecutive interpreting, interpreters focus more on "the global comprehension of the input" (Liu & Chiu 2009: 259), so they are less attentive to the surface elements of a text (i.e. word length and sentence length), as judged by the Flesch score.

Although strong correlations were observed between the Flesch score and the dependent variables in this study, the author shares Liu and Chiu's (2009) caution that trainers and test developers should not uncritically use the readability formula to

determine the difficulty level of a source text. As reported in Table 4 and corroborated by the reflective essays, the students were perhaps more sensitive to SWT and MLT in ST. Therefore, long words and long sentences (two elements measured by the Flesch score) may not mean difficult words and sentences, but unknown words/meanings and long T-units spell real trouble for ST interpreters.

5.4 *Pedagogical implications*

5.4.1 Difficulty manipulation

Authentic materials have been suggested in interpreter training (Gillies 2013; Setton & Dawrant 2016). However, there are times when authentic materials do not match the competency level of interpreter trainees. The findings in this study may inform trainers to manipulate the complexity of source texts to fit the training needs. For example, if trainers want to make the ST task more difficult, they can increase the mean length of a T-unit, and/or substitute simple words with sophisticated ones. Conversely, trainers can reduce the complexity of source texts by decreasing the mean length of a T-unit and/or replace infrequent words with frequent ones. As an example, in this study, the students were asked to sight-translate the following sentence: “In fact, most people are not able to work just at home without meeting their colleagues face to face sometimes.” This is one long sentence, and also one long T-unit. Trainers may reduce the syntactic complexity by reducing the mean length of a T-unit: “In fact, most people are not able to work just at home because they need to meet their colleagues face to face sometimes.” The sentence

itself is still long, but now it comprises two T-units, so the mean length of a T-unit and the syntactic complexity are significantly reduced.

5.4.2 Limitations of ST in interpreter training

While ST has been suggested as a valid training method to acquire the necessary skills in interpreting, such as chunking, condensing and paraphrasing information (González et al. 2012), ST may not be a satisfactory method to train students to analyze the source text or process information at a deep level. As argued in the "shallow-scan hypothesis" (Shreve et al. 2010) and confirmed by this study, student interpreters process the source texts at the lexical and intra-sentential levels, instead of the inter-sentential or discoursal level. The visual presence of the source texts allows students to process the texts linearly and shallowly and in small cognitive units. Therefore, interpreter trainers should be alerted to the possibility that student interpreters may not develop the desired source text analysis skills when exposed to ST training. Trainers need to create a pedagogical space to train students to detach from the source text (Lee 2012; Setton & Dawrant 2016) so that students can see both the whole and the parts.

5.5 ***Research implications***

The existing ST scholarship predominantly focuses on the process and the performance of ST, without giving due attention to the task properties. To address this research gap, a three-tier conceptual framework is proposed, with substantial revisions of Robinson's

Triadic Componential Framework (2011). The three-tier conceptual framework consists of three first-tier categories (properties, process and performance), and four second-tier categories (Task Complexity, Task Condition, Task Difficulty and Task Performance). Figure 2 summarizes some of the possible variables that can be included in each category (with asterisks indicating variables examined in this study).

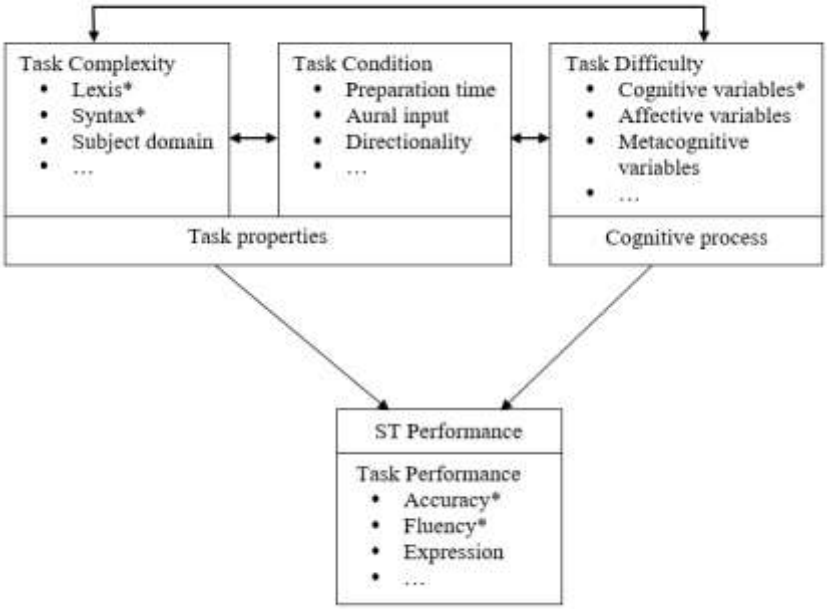

Figure 2. A three-tier conceptual framework for researching ST

It is noted that the definitions of the second-tier categories are different from what has been proposed by Robinson (2011; see also Section 2.2). Specifically, ST Task Complexity is determined by the source text characteristics (including the variables of lexis, syntax, subject domain, etc.). Task Condition refers to the demands of the task, such as preparation time, aural input, and directionality. In Figure 2, Task Complexity and Task Condition are grouped under Task Properties because the variables in these two

subcategories all reflect the nature of the task itself. Task Difficulty is decided by the nature of interpreters: their cognitive factors (e.g. decoding ability, working memory, language proficiency, ST competency, etc.), affective factors (e.g. motivation, engagement, attitude, emotions, etc.) and metacognitive factors (e.g. coping strategies). Finally, as shown at the bottom box of Figure 2, Task Performance can be measured by the accuracy, fluency, expression and other quality-oriented variables. The bidirectional arrows between Task Complexity, Task Condition and Task Difficulty indicate that different combination of variables in these three areas may have a confounding effect on ST performance. The unidirectional arrows indicate how Task Properties and Cognitive Process impact ST Performance.

Hopefully, this conceptual framework can inform and guide researchers in their research design. For instance, this study hypothesizes two causal links between text properties and ST performance, mediated by cognitive process, as represented by the arrows from Task Complexity to Task Difficulty and then to ST Performance (see Figure 2). This is a mediation model between text properties and ST performance. Future research could study a direct effect model in which no mediators are specified, as represented by the arrow from Task Complexity to ST Performance (see Figure 2). Another line of future research might be to examine how the variables in Task Difficulty and/or Task Condition influence ST performance. As the perception of task difficulty is relative, researchers can control the task complexity variables (e.g. given an 800-word text) and explore how ST interpreters at different competency levels experience the "fatigue effect" (Shreve et al. 2010: 80), subsequently leading to varying degrees of quality decline as the task progresses. Alternatively, researchers can manipulate the task

condition variables, such as making aural input available, to explore how the presence of speed pressure accelerates the onset of the "fatigue effect" in ST performance.

6. Conclusion

Although limited by the number of source texts, this exploratory study is the first-ever attempt to address the interplay of text characteristics, difficulty perception and task performance. Based on the correlation analysis and the students' reflective reports, two causal links are hypothesized to describe the relationship between the three Ps (task properties, cognitive process and ST performance). The tentative findings show that lexical complexity and syntactic complexity may cause decoding difficulties and cognitive overload, which subsequently lead to inaccuracy and dysfluency in ST performance. The study also reveals that university-level ST interpreters linearly process the source texts in small cognitive units. These findings lend empirical support to the shallow processing hypothesis in ST (Agrifoglio 2004; Shreve et al. 2010) and bring to light the potential limitations of using ST to train students' information processing skills. Finally, this study proposes a conceptual framework and points to the need for further research that investigates the relationships between variables in task complexity, task condition, task difficulty and task performance. Sight translation continues to occupy an important place in interpreting and translation curricular. For the benefits of this widely used training and testing method to be truly realized, further evidence about the interplay among the three Ps is needed to inform practice.

References

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting* 6 (1), 43–67.

Cai, R., Dong, Y., Zhao, N. & Lin, J. (2015). Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters. *The Interpreter and Translator Trainer* 9 (1), 104–120.

Campbell, S. (1999). A cognitive approach to source text difficulty in translation. *Target* 11 (1), 33–63.

Chen, C. (2015). Sight translation. In H. Mikkelson & R. Jourdenais (Eds.), *The Routledge handbook of interpreting*. London/New York: Routledge, 144–153.

Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research* 69, 176–183

Crossley, S. A., Greenfield, J. & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly* 42 (3), 475–493.

Deane, P., Sheehan, K. M., Sabatini, J., Futagi, Y. & Kostin, I. (2006). Differences in text structure and its implications for assessment of struggling readers. *Scientific Studies of Reading* 10, 257–275.

Dragsted, B. & Hansen, I. (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*. Copenhagen: Samfundslitteratur, 251–274.

Dragsted, B. & Hansen, I. (2009). Exploring translation and interpreting hybrids: The case of sight translation. *Meta* 54 (3), 588–604.

Engber, C. A. (1995). The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing* 4 (2), 139–155.

Ersozlu, E. (2005). Training of interpreters: Some suggestions on sight translation teaching. *Translation Journal* 9 (4). http://www.translationjournal.net/journal/34sighttrans.htm (accessed 17 May 2019).

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology* 32, 221–233.

Gile, D. (1997). Conference interpreting as a cognitive management problem. In J. Danks, G. Shreve, S. Fountain & M. McBeath (Eds.), *Cognitive processes in translation and interpreting*. Thousand Oaks, CA: Sage, 196–214.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training* (Rev. ed.). Amsterdam: John Benjamins.

Gillies, A. (2013). *Conference interpreting: A student's practice book*. London: Routledge.

González, R. D., Vásquez, V. F. & Mikkelson, H. (2012). *Fundamentals of court interpretation*. Durham, NC: Carolina Academic Press.

Graesser, A. C., McNamara, D. S., Louwerse, M. M. & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers* 36, 193–202.

Hunt, K. W. (1970). Do sentences in the second language grow like those in the first? *TESOL Quarterly* 4, 195–202.

Jeon, E. H. & Yamashita, J. (2014). L2 reading comprehension and its correlates: A Meta–analysis. *Language Learning* 64 (1), 160–212.

Jiménez Ivars, A. (2008). Sight translation and written translation. A comparative analysis of causes of problems, strategies and translation errors within the PACTE translation competence model. *Forum* 6 (2), 79–103.

Lambert, S. (2004). Shared attention during sight translation, sight interpretation and simultaneous interpretation. *Meta* 49 (2), 294–306.

Lee, J. (2012). What skills do student interpreters need to learn in sight translation training? *Meta* 57 (3), 694–714.

Leech, G., Rayson, P. & Wilson, A. (2001). *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.

Liu, M. (2013). Design and analysis of Taiwan's interpretation certification examination. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting*. Frankfurt: Peter Lang, 163–178.

Liu, M., & Chiu, Y. H. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting* 11 (2), 244–266.

Louwerse, M. M. & Mitchell, H. H. (2003). Toward a taxonomy of a set of discourse markers in dialog: A theoretical and computational linguistic account. *Discourse processes* 35 (3), 199–239.

Lu, X. (2011). A corpus–based evaluation of syntactic complexity measures as indices of college–level ESL writers' language development. *TESOL Quarterly* 45 (1), 36–62.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Journal* 96 (2), 190–208.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M. & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes* 47 (4), 292–330.

McNamara, D. S., Graesser, A. C. & Louwerse, M. M. (2012). Sources of text difficulty: Across genres and grades. In J. P. Sabatini, E. Albro & T. O'Reilly (Eds.), *Measuring up: Advances in how we access reading ability*. Plymouth, UK: Rowman & Littlefield Education, 89–116.

McNamara, D. S., Graesser, A. C., McCarthy, P. M. & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge: Cambridge University Press.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports* 66 (Monograph Suppl. 1–V66), 195–244.

Moser-Mercer, B. (1994). Aptitude testing for conference interpreting: Why, when and how. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation*. Amsterdam: John Benjamins, 57–68.

Nicodemus, B. & Emmorey, K. (2013). Direction asymmetries in spoken and signed language interpreting. *Bilingualism: Language and Cognition* 16 (3), 624–636.

Ozuru, Y., Rowe, M., O'Reilly, T. & McNamara, D. S. (2008). Where's the difficulty in standardized reading tests: The passage or the question? *Behavior Research Methods* 40 (4), 1001–1015.

Pöchhacker, F. 2004. *Introducing interpreting studies*. London/New York: Routledge.

Rayner, K. & Pollatsek, A. (1994). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Révész, A. & Brunfaut, T. (2013). Text characteristics of task input and difficulty in second language listening comprehension. *Studies in Second Language Acquisition* 35 (1), 31–65.

Richards, J. C., Platt, J. & Platt, H. (1992). *Longman dictionary of language teaching and applied linguistics*. London: Longman.

Robinson, P. (2001). Task complexity, task difficulty, and task production: Exploring interactions in a componential framework. *Applied Linguistics* 22, 27–57.

Robinson, P. (Ed.). (2011). *Second language task complexity: Researching the cognition hypothesis of language learning and performance*. Amsterdam: John Benjamins.

Rydland, V., Aukrust, V. G. & Fulland, H. (2012). How word decoding, vocabulary and prior topic knowledge predict reading comprehension. A study of language-minority students in Norwegian fifth grade classrooms. *Reading and writing* 25 (2), 465–482.

Setton, R. & Dawrant, A. (2016). *Conference interpreting: A complete course*. Amsterdam: John Benjamins.

Shreve, G., Lacruz, I. & Angelone, E. (2010). Cognitive effort, syntactic disruption, and visual interface in a sight translation task. In E. Angelone & G. Shreve. (Eds.),.

Translation and cognition: Recent developments. Amsterdam: John Benjamins, 63–84.

Shreve, G., Lacruz, I. & Angelone, E. 2011. Sight translation and speech disfluency: Performance analysis as a window to cognitive translation processes. In C. Alvstad, A. Hild & E. Tiselius. (Eds.), *Methods and strategies of process research*. Amsterdam: John Benjamins, 93–120.

Strauss, A. & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Thousand Oaks, CA: Sage.

Sun, S. & Shreve, G. M. (2014). Measuring translation difficulty: An empirical study. *Target* 26 (1), 98–127.

Viaggio, S. (1995). The praise of sight translation (and squeezing the last drop there out of). *The Interpreters' Newsletter* 6, 33–42.

Viezzi, M. (1989). Information retention as a parameter for the comparison of sight translation and simultaneous interpretation: An experimental study. *The Interpreters' Newsletter* 2, 65–69.

Viezzi, M. (1990). Sight translation, simultaneous interpretation and information retention. In L. Gran & C. Taylor (Eds.), *Aspects of applied and experimental research on conference interpretation*. Udine: Campanotto, 54–60.

Wallace, M. (2013). Rethinking bifurcated testing models in the court interpreter certification process. In D. Tsagari & R. van Deemter (Eds.), *Assessment issues in language translation and interpreting*. Frankfurt: Peter Lang, 67–83.

Wu, Z. (2016). Towards understanding interpreter trainees' (de)motivation: An exploratory study. *Translation & Interpreting* 8 (2), 13–25.

Yeh, S. & Liu, M. (2006). A more objective approach to interpretation evaluation: Exploring the use of scoring rubrics. *Journal of the National Institute for Compilation and Translation* 34 (4), 57–78.

Yu, G. (2009). The shifting sands in the effects of source text summarizability on summary writing. *Assessing Writing* 14 (2), 116–137.