Digital Scholarship in the Humanities, 2025, **00**, 1–12
https://doi.org/10.1093/llc/fqaf017
Full Paper

Can professional interpreters truly convey the speaker's sentiment? Exploring the potential of a computational approach

Han Xu[1], Jinghang Gu[1], Kanglong Liu[1], Qinyi Li[2,*]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong, China
[2]School of Information and Communication Technology, Griffith University, Brisbane, 4222, Australia

*Corresponding author. School of Information and Communication Technology, Griffith University, Building N44, 170 Kessels Rd, Nathan QLD 4111, Brisbane 4222, Australia. E-mail: qinyi.li@griffith.edu.au

Abstract

This study investigates whether sentiment analysis, a natural language processing technique, can be used to examine accuracy in interpreting. The data were obtained from a parallel bidirectional corpus of original speeches delivered at the United Nations and their simultaneous renditions provided by professional interpreters. Specifically, this study explores how much sentiment can be conveyed across languages via accurate renditions, how interpreting direction affects the conveyance of sentiment, and how sentiment analysis may help with accuracy assessment. The results show that the sentiment orientation and distribution expressed in the source text can be largely projected into the target text via accurate renditions. This finding confirms the validity of using translational language to create cross-lingual sentiment analysis tools. It also reveals the potential of integrating sentiment analysis into automated interpreting quality assessment frameworks. In addition, this study shows that the amount of sentiment conveyed in each direction seems to vary, suggesting that directionality has an impact on the emotional tone being communicated by the interpreters.

Keywords: accuracy; sentiment analysis; directionality; quality assessment; interpreting.

1. Introduction

In cross-lingual communication, the use of interpreting services is essential to enable effective communication between parties who do not share a common language. Interpreters play a vital role in facilitating mutual understanding by conveying messages accurately. To this end, professional codes of ethics for interpreters worldwide emphasize the importance of accuracy as a critical indicator of interpreting quality (Hale 2007). For example, the International Association of Conference Interpreters (AIIC) states that 'interpreters shall strive to translate the message to be interpreted faithfully and precisely' (AIIC 2022: 3). The Australian Institute of Interpreters and Translators (AUSIT)'s Code of Ethics identifies accuracy as an ethical principle, which requires that 'interpreters and translators use their best professional judgement in remaining faithful at all times to the meaning of texts and messages' (AUSIT 2012: 5). Regarding the achievement of accuracy, AUSIT's Code further explains that it means 'optimal and complete message transfer into the target language preserving the content and intent of the source message or text without omission or distortion' (AUSIT 2012: 5).

While it is critical for interpreters to fulfil their ethical obligation, the codes of professional ethics do not seem to recognize the difficulty involved in producing accurate renditions (Jacobsen 2003; Hale 2007). As a matter of fact, interpreters may not always achieve accuracy in real practice, given that it is a very challenging task (Gile 1995; Hale 2007; Seeber and Zelger 2007; Xu 2022, 2024). The notion of accuracy, as shown in existing studies, features great complexity, and its achievement is subject to the influence of a wide range of factors. These factors may include the interpreter's own attributes, such as level of training, experience, understanding of the professional role, and professional competence (Cheung 2007, 2016; Liu and Hale 2018). Empirical evidence suggests that professionally trained interpreters tend to outperform untrained ad hoc interpreters or learners in producing accurate renditions (Liu and Hale 2018; Stachowiak-Szymczak and Korpal 2019; Xu 2021,

2022; Hale *et al.* 2022a). This is because, due to years of training and experience, professional interpreters tend to develop a better understanding of their ethical role and have more linguistic resources to initiate effective coordination to convey messages across languages successfully. In addition, the achievement of accuracy is also related to various external factors, such as directionality, working conditions, interpreting user's expectations, institutional constraints, and the specific requirement of each interpreting setting (Hale 2007; Xu, Hale, and Stern 2020; Xu 2021). For instance, accuracy may generate varying connotations in different interpreting settings. In court interpreting, accuracy indicates that interpreters should convey not only what is said but also how it is said by the speaker, as both the propositional content and the pragmatic force of the message may reveal the speaker's character and credibility in a legal context (Hale 2004; Liu and Hale 2018). Therefore, it is crucial for court interpreters to strictly follow the accuracy norm without making any unjustified addition, omission, or distortion. Such an interpreting approach includes maintaining discourse markers in the source speech, such as the speaker's tones, hesitations, pauses, hedges, and false starts, in their renditions (Hale 2004; Liu and Hale 2018). In comparison, when working in a conference or business setting, interpreters are often expected to improve on the speaker's speech style by omitting self-corrections or hesitations to make sure the interpreted speech is smooth and fluent (Hale 2007)

Notably, the assessment of accuracy is a crucial step in previous research as its results reflect the interpreter's ability to achieve accuracy. In addition, the way in which accuracy is assessed also has important practical implications for the interpreting profession, informing activities such as training, certification, and recruitment processes (Han 2022). A common approach to assessing accuracy is error-based analysis, which involves identifying and categorizing various types of interpreting errors, including omissions, additions, and distortions. (Gile 1999, 2003, 2011; Napier 2004; Turner, Lai and Huang 2010). This approach is primarily reliant on human assessors' subjective evaluations, which is often a time-consuming and labour-intensive process. Han (2022: 41) once postulated that an ideal assessor should have relevant experience in practising, learning, teaching, and assessing interpreting. Yet, such qualified assessors may not always be available. Multiple assessor-related factors, such as fatigue, time pressure, inter-assessor disparity, inconsistent attention span across different assessment tasks, and order of assessment, may impinge upon the assessment results (Liu 2013; Shlesinger 1994). This makes it necessary to explore methods to assess accuracy in a more systematic and objective manner to increase reliability and rigour, as well as to corroborate the results of existing human-based accuracy assessment approaches.

Against this research background, this study aims to explore the possibility of using sentiment analysis, a natural language processing technique, to examine accuracy in interpreting. Based on a parallel bidirectional corpus of original speech delivered at the United Nations (UN) and their simultaneous renditions provided by highly professional interpreters, this study aims to explore how much sentiment expressed in the source text can be conveyed by the interpreters, how interpreting direction affects the transposition of sentiment across languages, and how the results of sentiment analysis may help to assess accuracy in interpreting. Following this introduction, Section 2 reviews existing literature that investigates accuracy in interpreting. Section 3 introduces the concept of sentiment analysis. Section 4 describes the corpus design, compilation, and data analysis methods. Sections 5 and 6 present the results and discussion, respectively. Section 7 concludes the study by summarising the key findings and pointing out future research directions.

2. The challenge of achieving accuracy in interpreting

The concept of accuracy has been widely examined in interpreting studies, as its conceptualization largely determines the interpreter's practice approach (Pöchhacker 2004/2022). A key concern within this line of inquiry is defining what constitutes accuracy in the context of interpreting. While many theoretical frameworks were proposed, most researchers concur that the lexicosemantic interpreting approach, which concentrates on finding source-target correspondence at the lexical or semantic level, is not sufficient (Gile 1992, 1995; Hale 2004, 2007; Seeber and Zelger 2007). Interpreters need to consider the pragmatic function of the source message and understand the 'text as discourse rather than as words or sentences strung together' (Hale 2007: 23). Gile (1992, 1995), for instance, argues that accurate rendition indicates that both the informational content and the style of the message should be conveyed. Seeber and Zelger (2007: 290) view accuracy as a 'truthful rendition', which means interpreters should convey three principal message components, that is, verbal, semantic, and intentional. However, there are situations where the three levels are not congruent. Interpreters need to assess the weight of each level and prioritize the information that should be conveyed to achieve accuracy. Inadequate evaluation of the weight of each component may lead to renditions that seem accurate at the semantic level but fail to convey the speaker's communicative intention. Similarly, Hale (2004, 2007) made a distinction between semantic and pragmatic interpreting approaches. Semantic rendition only concerns producing a rendition that is accurate at the semantic

level. The rendition may only appear 'correct' on the surface but fail to capture the original intention and illocutionary force of the source text. The pragmatic interpreting approach, on the other hand, not only maintains the propositional content of the source text but also creates the same communicative effect as the source message does.

The many proposed interpreting approaches are undeniably useful in guiding interpreters' practice from a linguistic perspective. However, achieving accuracy in real practice is much more complicated. Simply considering the linguistic components of accuracy can hardly reflect its dynamic nature, given that factors that go beyond the linguistic sphere affect its achievement. Studies have shown that how interpreters perceive their role and how they understand the interpreting user's intention affects their interpreting approach (Hale 2007; Hsieh 2007; Liu and Hale 2018; Xu 2021, 2024). Xu (2021), focusing on interpreted lawyer-client interviews in Australia, found that when interpreters assumed the role of a lawyer's helper, they modified the client's message based on their own understanding of the context. For instance, an interpreter was found to intentionally omit what he believed was 'irrelevant' information from the client in order to 'help' the lawyer 'save time'. However, such a 'mediated' interpreting approach (Hale 2008), that is, interpreters decide on what should or should not be conveyed, was strongly spoken against by the lawyer as it obstructed the lawyer from having direct communication with the client. By contrast, Seeber and Zelger (2007) reported a case of several conference interpreters' intentional omission of offensive remarks from the host towards a head of state. This time, the interpreters' practice was justified because the speaker should not have the intention to insult the head of a state. Therefore, the interpreters' unanimous omission helps to avoid a potential face-threatening act towards a guest of honour, which should be in line with the host's intention in that context.

Existing studies have shown that accurate rendition should involve successful conveyance of the speaker's intention across languages. Interpreters can obtain the intentional component of a message based on contextual manifestations of the speaker's intention, such as their expressed sentiment, attitudes, and emotions. However, unless interpreters are directly informed by the speakers of their intention, interpreters' understanding is always assumptive, which may not be correct all the time (Seeber and Zelger 2007). When there is a mismatch, it can be difficult for interpreters to achieve accuracy. Seen from this perspective, the subjective nature of the speaker's intention adds yet another layer to the complexity of achieving accuracy, which also makes the assessment of accuracy difficult.

3. Sentiment analysis

The present study proposes to use sentiment analysis to examine accuracy in interpreting. Sentiment analysis sits at the intersection of natural language processing, machine learning and computational linguistics (Buscemi and Proverbio 2024). It is the process of using computational methods to examine subjective information, such as opinions, appraisals, feelings, evaluations, and attitudes expressed in texts, sentences or words (Serrano-Guerrero *et al.* 2015; Taboada 2016; Liu and Lei 2018; Mäntylä, Graziotin, and Kuutila 2018). Ever since its emergence in the early 2000s, sentiment analysis has been known by many names, such as opinion mining, opinion extraction, or subjectivity analysis (Nasukawa and Yi 2003; Liu 2022). The aim of sentiment analysis is to identify the semantic orientation of language in use by classifying its sentiment polarity, that is, whether its emotional disposition is positive, negative, or neutral. Largely due to its efficiency in determining subjective information in a systematic and automated way, sentiment analysis has been applied in different domains to address real-world problems. Its applications may include analysing social media commentary, film reviews, or consumer feedback to evaluate the attitudes of the public or to predict the results of certain social events, such as political elections (Medhat, Hassan, and Korashy 2014; Wankhade, Rao, and Kulkarni 2022).

Conventionally, sentiment analysis is conducted via two approaches: lexicon-based and supervised machine learning-based methods. The lexicon-based approach relies on a sentiment lexicon to determine the sentiment polarity of a given text. The lexicon is a list of words that have already been categorized in terms of sentiment polarity and relative strength. A set of linguistic rules are often embedded in the lexicon to increase the accuracy of the analysis. There are both domain-specific lexicons and general-purpose lexicons depending on the target of analysis. Domain-specific lexicon tools are designed specifically for texts in a particular field, while general-purpose tools can be applied across domains but may fail to recognize semantic features unique to certain domains or genres (Lei and Liu 2021; Mukhtar, Khan, and Chiragh 2018; Taboada *et al.* 2011). The supervised machine learning-based approach can be further divided into traditional machine learning methods and deep learning methods. The traditional machine learning methods mainly rely on classification techniques to identify the sentiment polarity of a given text. First, a classifier needs to be built using a set of training data, which contains texts annotated by humans with regard to their sentiment polarity. The classifier can then be used to analyse new unlabelled data, which is called test data, and obtain a sentiment score. Commonly used classification models may include decision trees, random forests, support vector machines (SVMs), and logistic regression.

The machine learning-based approach is highly effective for domain-specific sentiment analysis. However, a classifier that is trained for one domain may not perform well when being used in other domains (Taboada 2016). Deep learning methods, on the other hand, utilize deep neural network architectures that can automatically learn feature representations from raw input data (Prabha and Srikanth 2019). Ain *et al.* (2017) introduced deep learning techniques like Deep Neural Networks (DNN), Convolutional Neural Networks (CNN), and Deep Belief Networks (DBN) for tasks including sentiment classification, cross-lingual problems, and product review analysis. Deep learning models have the ability to learn complex, hierarchical representations of the input data, which can lead to superior performance compared to traditional machine learning methods, especially for large and complex datasets. However, they generally require more training data and computational resources (Zhang, Wang, and Liu 2018).

More recently, pre-trained models, a new paradigm in natural language processing, have gained rapid growth in the field of sentiment analysis (Mathew and Bindu 2020). A pre-trained model is a model that is trained on a massive amount of data containing labelled examples of sentiment polarity. BERT (Devlin *et al.* 2019), RoBERTa (Liu *et al.* 2019), and GPT (Radford and NarasimHan 2018) models are all popular pre-trained models for sentiment analysis. Notably, many pre-trained models are built on Large Language Models (LLMs). LLMs are trained on vast datasets to analyse, understand, and produce human-like language, making them powerful tools in the field of natural language processing (Wei *et al.* 2022). Compared to the conventional machine learning-based approach, the LLM-based sentiment analysis tools show a higher level of contextual understanding and robustness across domains. It is capable of capturing intricate language patterns and semantic nuances for any new data with an increased level of accuracy, reliability, and generalizability. Yet, at the same time, pre-trainered models, including LLM-based sentiment analysis tools, may struggle with specialized terminology or domain-specific contexts that were underrepresented in their training data. Moreover, bias in the training dataset may also impact the results of sentiment predictions.

Given the successful utilization of sentiment analysis across fields, there is an emerging research interest in extending its application to the study of languages (Taboada 2016; Wen and Lei 2022). However, most sentiment analysis tools are only available for resource-rich languages, such as English. For low-resource languages, for which there is a scarcity of annotated corpora and sentiment lexicon, machine translation is often used to convert the available sentiment resources into target language to enable cross-lingual sentiment analysis (Vilares, Alonso, and Gómez-Rodríguez 2017: 596; Wan 2008; Xu *et al.* 2022). This approach is based on the presumption that when translation is accurate, sentiment should be conveyed across languages (Lu *et al.* 2011). However, the quality of machine translation varies for different language pairs, which may affect the performance of cross-lingual sentiment analysis tools (Al-Shabi *et al.* 2017).

At the same time, some researchers argued that due to cross-cultural differences, inherent structural disparities between languages, and the context where languages are used, parallel words, sentences or even texts might not share the same sentiment orientation even when they are accurately translated (Demirtas and Pechenizkiy 2013; Ghorbel and Jacot 2011). This is understandable, as sentiment is a culturally and linguistically sensitive conception. A positive expression in one language may be perceived negatively in another linguistic context due to cultural reasons (Lei and Liu 2021). Largely supporting this view, Chen and Zhu (2014) posited that it is more difficult to translate text with more sentiment. Demirtas and Pechenizkiy (2013) even argued that cultural differences generate a greater impact than inaccurate machine translation on the performance of sentiment analysis tools. However, there is a lack of empirical evidence to show how much sentiment can be conveyed via translation and how potential cross-lingual sentiment variation is related to accuracy.

Focusing on simultaneous interpreting, an oral form of translation, this study sets out to fill this research gap by exploring the extent to which sentiment can be conveyed across languages via accurate rendition and discussing its implications for interpreting accuracy assessment. The data were obtained from a corpus consisting of professional interpreters' accurate renditions at UN Security Council meetings. Specifically, this study attempts to address the following research questions:

RQ1: How does the sentiment score of a speech vary when the speech is accurately interpreted into another language?

RQ2: If sentiment score variation is observed, how does interpreting direction affect the variation?

RQ3: How much sentiment can be conveyed across languages via accurate renditions?

4. The study

4.1 Data description

The data used in the present study were obtained from the UN Chinese-English Simultaneous Interpreting

Corpus (UNSI) (Xu and Liu 2024). UNSI collects speeches given by Chinese and English speakers at the UN Security Council meetings and their simultaneous renditions. UN Security Council meetings are high-stakes international meetings where delegates of different countries and regions discuss issues related to global governance, crisis management, and international cooperation. As the delegates speak different languages, simultaneous interpreting is provided to enable real-time communication. Given that miscommunication in such a setting can generate serious consequences, highly competent professional interpreters are engaged to ensure the quality of interpreting services. These interpreters can access the scripts of the speech beforehand to better prepare themselves for the assignment (Cheung 2019). Due to UN interpreters' professional competence and their sufficient pre-interpreting preparation, their performance featured a very high level of accuracy. Since accurate rendition includes a complete transfer of both the propositional content of the message and the speaker's intention (Hale 2007; Seeber and Zelger 2007), it is expected that sentiment should be conveyed across languages.

The UNSI consists of four sub-corpora: original English speeches (NE) and their simultaneous renditions into Chinese (IC), as well as original Chinese speeches (NC) and their simultaneous renditions into English (IE). The English speeches were delivered by delegates from the United Kingdom, while the Chinese speeches were delivered by delegates from China. These speeches were sampled from onsite UN Security Council meetings that featured both British and Chinese delegates, specifically focusing on those that included significant discussions relevant to international and regional affairs. Consequently, the original Chinese and English speeches form pairs, with each pair delivered at the same meeting and addressing the same agenda. This pairing ensures that the four sub-corpora are comparable in terms of genre, topic, and speaker identity. As these speeches reflect the delegates' attitudes, opinions, and perceptions regarding various international and regional affairs, the UNSI serves as a suitable corpus for sentiment analysis. A total of 207 pairs of original speeches and their simultaneous renditions were found from meetings held between 2021 and 2022. The transcripts of all the speeches in their original languages and renditions can be freely downloaded from the UN Digital Library[1]. The accuracy of the transcription was manually verified against the audio files of these meetings. The detailed information about the four sub-corpora is summarized in Table 1.

4.2 Sentiment analysis

The present study employed the leading pre-trained and LLM-based sentiment analysis tools: multilingual BERT (Devlin *et al.* 2019) and Llama2 (Touvron *et al.* 2023). Multilingual BERT is a pre-trained language model built on the transformer encoder architecture developed by Google. Llama2, standing for Large Language Model Meta AI, is a series of LLMs developed by Meta AI. Llama2 has been pre-trained on diverse datasets to enhance its generalization capabilities, making it suitable for tasks such as text classification. These two models were chosen for their strong adaptability to multilingual contexts, reliable performance across domains, and open-access availability (Xu *et al.* 2022; Zhao *et al.* 2024). To provide a fine-grained representation of the semantic polarity expressed in the speeches (Lei and Liu 2021: 14), sentiment analysis was conducted at the sentence level using these two models. The two sets of parallel sub-corpus (NE vs IC and NC vs IE) were aligned at the sentence level. This yields 4400 pairs of parallel sentences for NE and IC, as well as 5891 pairs of parallel sentences for NC and IE. Multilingual BERT and LLaMA2 provide a trisected sentiment polarity for each sentence, which is positive, neutral, and negative. A numerical value is assigned for each sentiment polarity to facilitate the later statistical analysis: positive is 1, neutral is 0, and negative is −1. The overall sentiment of a text was calculated by averaging the sentiment scores of the involved sentences.

To test the accuracy and reliability of the sentiment analysis results, 200 sentences were randomly selected from each sub-corpus for manual sentiment coding. As shown in Table 2, the accuracy performance rate of multilingual BERT was much higher than Llama2 against the human benchmark. It seems despite their wide generalisability across domains, multilingual BERT is more suitable for analysing the bilingual dataset in this study. In addition, multilingual BERT's performance is reliable for both Chinese and English sub-corpora. Interestingly, a slightly higher accuracy performance rate was observed for original speech (NE and NC) compared to interpreted speech (IE and IC). This discrepancy may stem

Table 1. An overview of the UNSI.

Sub-corpus	Texts count	Token no.	Source	Producer
NE	207	100,137	Original English speech	Native English speakers
IE	207	143,498	Interpreted English speech	Interpreters
NC	207	189,866	Original Chinese speech	Native Chinese speakers
IC	207	161,992	Interpreted Chinese	Interpreters

Table 2. Accuracy performance rate of BERT and Llama2.

Sub-corpus	Accuracy performance rate of multilingual BERT	Accuracy performance rate of Llama2
NE	0.80	0.32
IE	0.67	0.25
NC	0.71	0.23
IC	0.66	0.37

from the use of a pre-trained multilingual BERT model, as there were no parallel sentiment-annotated corpora for both original texts and their renditions. The lack of domain-specific fine-tuning likely accounts for the lower accuracy in interpreted texts, as the model may have difficulty capturing the linguistic variations, semantic shifts, and contextual adaptations inherent in translational language. Unlike original speeches, interpreted renditions may involve paraphrasing, condensation, and structural reformulation, which can introduce subtle sentiment modifications that the model is not explicitly trained to recognize. Despite this limitation, the significantly higher accuracy scores for both original and interpreted speeches indicate that multilingual BERT is a suitable model for this analysis. The results from multilingual BERT were exported to a spreadsheet for further statistical analysis to address the relevant research questions.

4.3 Data analysis

The average sentiment score of the texts in the four sub-corpora will be calculated and compared to each other to explore how the sentiment score of a speech varies after it is interpreted and how directionality may affect the variation of sentiment score, addressing the first and the second research questions. Pairwise comparisons will be employed to see whether the observed variation has statistical significance. To explore how much sentiment can be conveyed across languages via accurate renditions, linear regression analysis will be conducted to investigate how the sentiment polarity distribution in the source speech is related to that in the target speech, addressing the third research question. The findings of this study will be discussed in line with previous research to explore the potential of integrating sentiment analysis into the interpreting accuracy assessment process.

5. Results

5.1 Sentiment score variation across the four sub-corpora

The average sentiment score of the texts in the four sub-corpora was calculated. The result is presented in Fig. 1. The descriptive statistical information of the calculation is summarized in Table 3. The result shows that the average sentiment scores at the corpus level are above zero, indicating that speakers maintained a positive tone most of the time in their speeches. This overall positive tone was also maintained in the rendition. In addition, it was found that the average sentiment score of IC ($M = 0.34$, $SD = 0.32$) is close to that of NE ($M = 0.36$, $SD = 0.24$), with a very slight decrease. In the other direction, the average sentiment score of IE ($M = 0.51$, $SD = 0.26$) is notably higher than that of NC ($M = 0.36$, $SD = 0.30$). This result shows that when the same sentiment tool was used to assign sentiment scores for parallel sentences, the values varied between source and target text. This is understandable as the same tool may demonstrate rating differences across languages due to the use of different training routines, data labelling approaches and training data of varying quality (Buscemi and Proverbio 2024). Therefore, the increased or reduced sentiment score found in IC and IE does not necessarily mean the interpreted speech is more or less positive than the original speech.

What is interesting to note is that the amount of sentiment score variation differs in the two interpreting directions. There seems to be a greater sentiment gap between NC and IE compared to an almost negligible one between NE and IC. Pairwise comparisons between the average sentiment scores across the four sub-corpora reveal that the notable difference between NC and IE has statistical significance, while the small difference between NE and ID does not have statistical significance. This is shown in Table 4. This result suggests a sentiment variation disparity between the two directions at the corpus level. To test whether the disparity applies to each pair of source and target text, the difference between the sentiment scores of each pair of source and target texts was calculated and compared via similarity tests. This study first applied the standard Euclidean similarity test, considering the underlying sentiment data are numerical. To avoid potential biases caused by the relatively low dimensions, the Cosine similarity measure was also applied. As shown in Table 5, the two similarity tests confirm a greater sentiment gap between NC-IE, that is when interpreters interpret from Chinese into English. From a statistical point of view, this finding seems to suggest that interpreting direction has an impact on the conveyance of sentiment during interpreting.

5.2 Conveyance of sentiment across languages

To explore how much sentiment can be conveyed across languages via accurate renditions, this study compared the sentiment polarity distribution of the source text and that of the target text. Given accurate rendition includes the successful transfer of the speaker's communicative intention, including their attitudes,

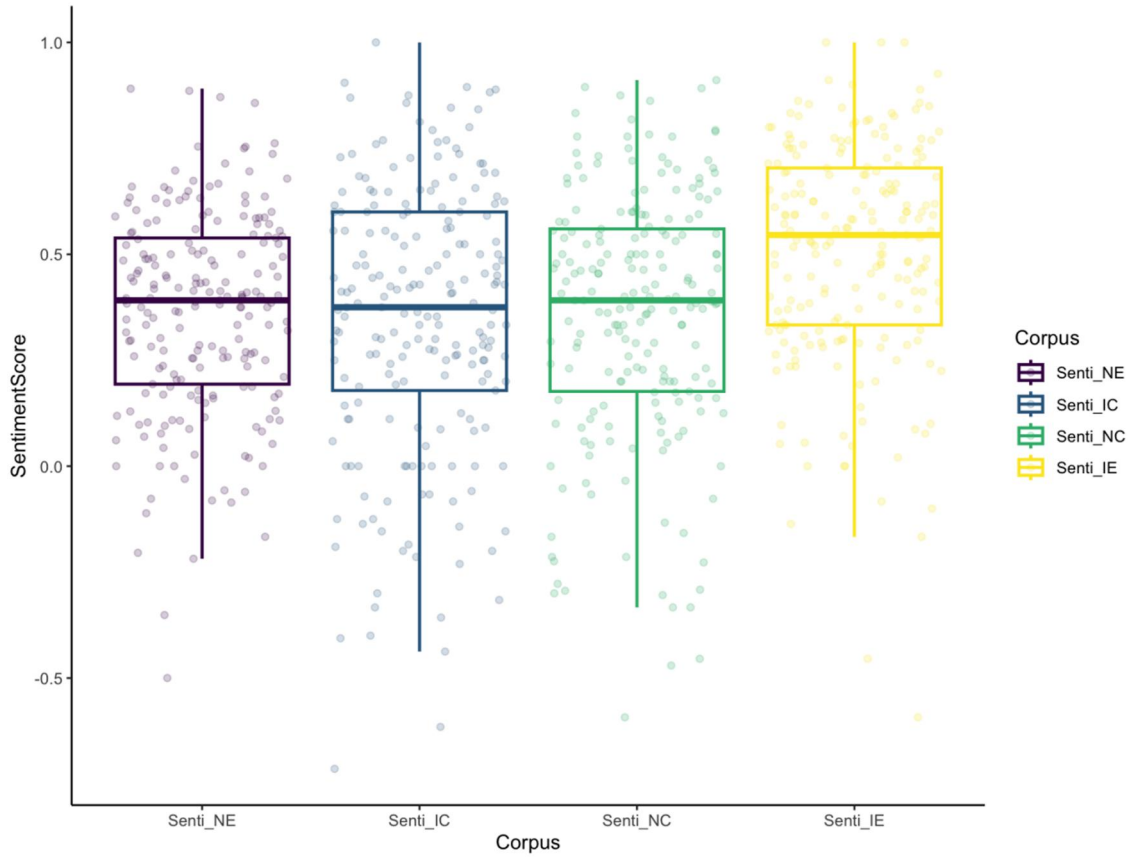

Figure 1. Average sentiment scores across the four sub-corpora.

Table 3. Statistical summary of sentiment scores across four sub-corpora.

Sub-corpus	Mean	Standard Dev.	Median	Maximum	Minimum
NE	0.36	0.24	0.39	0.89	−0.5
IE	0.51	0.26	0.55	1.00	−0.59
NC	0.36	0.30	0.39	0.91	−0.59
IC	0.34	0.32	0.38	1.00	−0.72

Table 4. Pairwise comparisons of sentiment scores across four sub-corpora.

Sub-corpus	NE	IE	NC	IC
NE	–			
IE	>**	–		
NC	<	<**	–	
IC	<	<**	<	–

Note.
** $P < .01$

emotions, and perceptions (Hale 2007; Seeber and Zelger 2007), it is expected that the same sentiment polarity distribution in the source text should be maintained in the target text. To this end, two separate linear regression analyses were conducted to examine the relationships between the average sentiment scores of the source and target texts. The results are shown in Table 6 and visualised in Figs. 2 and 3. Statistically significant correlations were found for both analyses. Specifically, the correlation coefficient (r value) in both analyses is above zero, showing that the average sentiment scores of the source texts are positively related to that of the target texts. This result

Table 5. Similarity analysis of the sentiment scores distance.

Direction	Euclidean	Cosine
NE-IC	3.42	0.85
NC-IE	3.58	0.91

Table 6. The results of two linear regression analyses.

Direction	r	p value
NE-IC	0.64	< 0.01
NC-IE	0.78	< 0.01

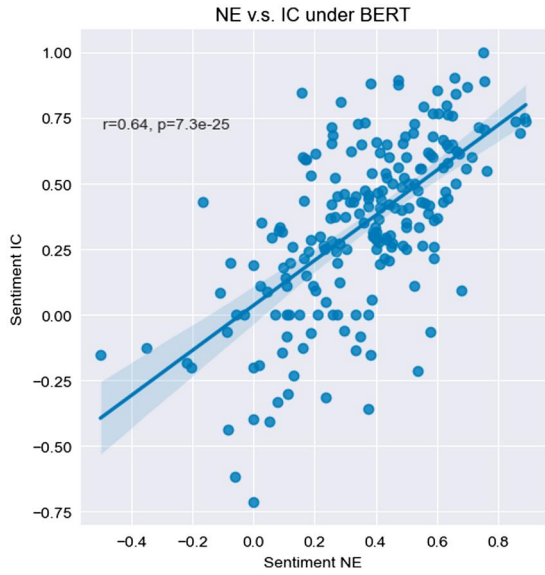

Figure 2. Results of linear regression analysis for NE vs IC.

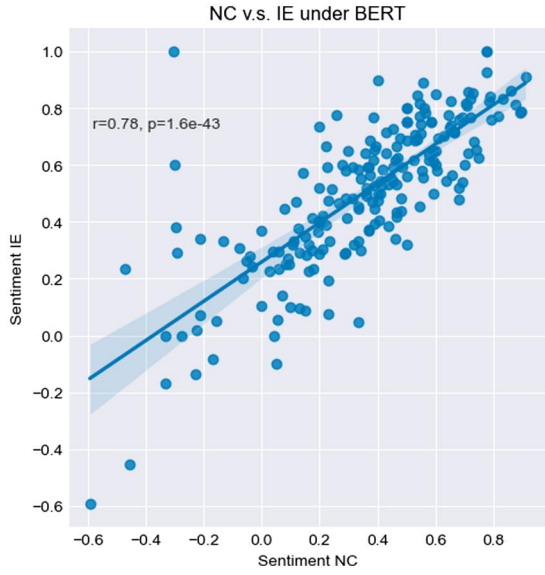

Figure 3. Results of linear regression analysis for NC vs IE.

indicates that the sentiment orientation can be successfully conveyed into the target text. In addition, the r values in both analyses are 0.64 and 0.78, respectively, showing a strong linear relationship between the two variables. This finding suggests that when it comes to accurate rendition, the sentiment orientation and distribution of the source speech can be largely projected into the target texts.

6. Discussion

Generating data from a parallel bidirectional corpus of the original speeches delivered at the UN and their simultaneous renditions provided by highly professional interpreters, this study mainly explored the extent to which sentiment can be conveyed by the interpreters via accurate renditions, how interpreting direction affects the transposition of sentiment across languages, and how the results of sentiment analysis may help to assess accuracy in interpreting.

6.1 Sentiment conveyance across languages

To start with, the present study shows that when it comes to accurate rendition, the sentiment orientation and distribution of the source text can transcend language barriers and be largely projected into the target language. This may indicate that sentiment can be systematically conveyed across languages via accurate translation, providing empirical evidence to the common approach that leverages translation to create cross-lingual sentiment analysis tools (Gopaldas 2014). This forms an interesting comparison to previous research, which shows that certain emotional dispositions may not be easily translated across languages due to cultural or contextual differences (Ghorbel and Jacot 2011; Demirtas and Pechenizkiy 2013). This different result may be because this study examined the speeches delivered at the UN and their simultaneous rendition. These speeches feature explicit expressions of sentiment so that the representatives' attitudes towards various international affairs can be effectively communicated to an international audience. Therefore, this type of text makes it easier for interpreters to convey sentiment across languages than more nuanced and culturally embedded forms of communication. This result has implications for developing and applying cross-lingual sentiment analysis tools. When designing these tools, it is important to use a more varied corpus that covers different genres. For texts that require more contextual and cultural understanding, human annotation may be added to increase the cultural sensitivity of cross-lingual sentiment analysis (Buscemi, A. and Proverbio 2024). When using these tools, it is important to recognize that they may work better for certain text genres, like news articles or business reports, where the sentiment is more overt. However, they may struggle with implicit, contextual, or culturally specific communications, which means the analysis results require careful interpretation and may need to be compared against human benchmarks to ensure the reliability of results.

6.2 Sentiment conveyance and directionality

At the same time, this study shows that while sentiment orientation and distribution can be largely interpreted across languages, the amount of sentiment conveyed in each interpreting direction seems to vary. The impact of directionality on interpreter's performance has been a widely investigated topic. The interpreting profession traditionally holds that interpreters should work into their first language (L1) rather than their second language (L2) (Seleskovitch 1978; Donovan 2004). This preference stems from the recognition that L2 language production demands greater cognitive efforts, leading to challenges such as reduced accuracy and fluency (Ortega 2014). Consequently, interpreters are often seen as having a natural advantage when interpreting into their L1, where they possess greater linguistic and cognitive resources. Over the years, research has largely confirmed the impact of directionality on interpreter's performance. For instance, examining professional interpreters' performance in English-Chinese simultaneous interpreting, Chang and Schallert (2007) found that interpreters adjusted strategies to cope with demands in different directions. When they need to render the message into their L2, for which they may have less linguistic proficiency, interpreters tend to adopt a meaning-based interpreting approach by using generalization, transformation and inferencing. In contrast, when they work into their L1, they rely on existing phrases and idioms to convey meanings rather than relying on generalizations. However, counter-evidence keeps emerging, revealing that the impact of directionality on interpreter's performance may be related to interpreter's qualification. Nicodemus and Emmorey (2015) found that professional interpreters' renditions in both directions are equally good. In the present study, the results show that interpreting direction affects the emotional tone or cultural nuance being communicated by the interpreters. This finding is consistent with previous research, which shows that due to interpreters' asymmetric command of the two working languages, they may present varying performance patterns in different directions (Sandrelli and Bendazzoli 2005; Chang and Schallert 2007; Dayter 2018). These findings underscore the importance of interpreters recognising the impact that direction can have on their performance. Interpreters should be more cognizant of how the intended emotional impact can be preserved in each direction. Given that different cultures may understand sentiment and emotion in distinct ways (Buscemi and Proverbio 2024: 4), this requires interpreters to carefully evaluate how sentiment is perceived by the target audience so that specific strategies can be developed to convey the emotional dispositions across languages.

6.3 Sentiment conveyance and accuracy assessment

In addition, this study reveals that sentiment analysis is effective in detecting the systematic conveyance of sentiment in accurate renditions. This aligns with the theoretical conception of accuracy, which delineates that interpreters should convey the intentional content of the message in addition to its semantic content (Hale 2007; Seeber and Zelger 2007). This finding has practical implications for research that explores automated approaches to assessing interpreting quality (Yu and van Heuven 2017; Ouyang, Lv, and Liang 2021; Lu and Han 2023). An important line in this research direction is to use linguistic or paralinguistic features that can be automatically extracted from interpreted speech to predict certain aspects of quality (Yu and van Heuven 2017; Ouyang, Lv, and Liang 2021). Considering that sentiment analysis is effective in measuring how much sentiment can be conveyed across languages, the sentiment score of a given speech may be used as an indicator to reflect its level of accuracy, which is a major measure of interpreting quality. Yet, it is worth pointing out that the sentiment score can hardly serve as a standalone indicator for accuracy. This is because sentiment can only reflect whether the semantic polarity is conveyed rather than the transfer of the remaining information contained in a message. Messages with comparable sentiment levels may still differ tremendously in their semantic meaning and substance. Therefore, it is essential for automated quality assessment models to include multiple indicators to accommodate the various dimensions of accuracy.

7. Conclusion

Sentiment analysis has been widely adopted across a variety of domains to address real-world problems, yet its application to study the use of language in multilingual contexts is only an emerging area of research. Adopting a corpus-based computational approach, this study explored the potential of using sentiment analysis to objectively evaluate the transfer of semantic polarity across language barriers. Based on a parallel bidirectional corpus consisting of speeches delivered at the UN and their simultaneous renditions, the study shows that despite interpreters' varying performance in different directions, the sentiment orientation and distribution expressed in the source text can be largely projected into the target language via accurate renditions. This finding shows the effectiveness of sentiment analysis in measuring the transfer of the speaker's communicative intention, an important component of accuracy. It highlights the promise of integrating sentiment analysis into interpreting accuracy assessment frameworks and advances the use of computational linguistic methods to assess quality automatically. In addition, the findings of this study

hold significance for the field of digital humanities as it bridges the gap between natural language processing and the nuanced understanding of human sentiment in cross-linguistic communication. By employing sentiment analysis to examine the accuracy of professional interpreters' renditions, this research contributes to the growing body of knowledge that intersects technology with humanistic inquiry. The findings underscore the importance of integrating computational methods into the analysis of translational language, thus facilitating a deeper understanding of how sentiment is conveyed across languages.

The present study has its limitations. It focuses on a single domain where speakers' sentiments and attitudes are stated clearly. It would be interesting to test the effectiveness of sentiment analysis in assessing interpreted speech that contains irony, sarcasm, or other contextually ambiguous information. In addition, this study only examined one language pair, namely English and Chinese, using two sentiment analysis tools. Both Chinese and English are resource-rich languages for sentiment analysis (Xu *et al.* 2022). Further research is needed to explore the generalizability of these findings across diverse interpreting contexts, language combinations, and sentiment analysis tools. Moreover, due to the inherent limitations of sentiment analysis, such as bias introduced by the training data and potentially insufficient understanding of deep semantic and pragmatic meaning beyond superficial sentiment (Liu 2022), the results of sentiment analysis should always be interpreted with caution. Therefore, ongoing research efforts are required to explore the predictive power of accuracy as compared to human-assigned accuracy scores.

Author contributions

Han Xu (Conceptualization, Data curation, Funding acquisition, Methodology, Supervision), Jinghang Gu (Formal analysis, Methodology, Software, Validation), Kanglong Liu (Conceptualization, Investigation, Methodology), Qinyi Li (Conceptualization, Formal analysis, Investigation, Methodology, Visualization)

Note

1. UN Digital Library can be accessed via https://digitallibrary.un.org/?ln=en.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Funding

This study was funded by The Hong Kong Polytechnic University (Projects No. P0043847, P0051009).

References

AIIC. (2022) 'AIIC Code of Professional Ethics', https://aiic.org/document/10277/CODE_2022_EandF_final.pdf

Ain, Q. T. *et al.* (2017) 'Sentiment Analysis Using Deep Learning Techniques: a Review', *International Journal of Advanced Computer Science and Applications*, **8**.

Al-Shabi, A. *et al.* (2017) 'Cross-Lingual Sentiment Classification from English to Arabic using Machine Translation', *International Journal of Advanced Computer Science and Applications*, **8**: 434–40. https://doi.org/10.14569/ijacsa.2017.081257

AUSIT. (2012) 'AUSIT Code of Ethics and Code of Conduct', https://ausit.org/wp-content/uploads/2020/02/Code_Of_Ethics_Full.pdf

Buscemi, A., and Proverbio, D. (2024) ChatGPT vs Gemini vs LLaMA on Multilingual Sentiment Analysis. *arXiv preprint arXiv: 2402.01715*. https://doi.org/10.48550/arxiv.2402.01715

Chang, C. C., and Schallert, D. L. (2007) 'The Impact of Directionality on Chinese/English Simultaneous Interpreting', *Interpreting : International Journal of Research and Practice in Interpreting*, **9**: 137–76. https://doi.org/10.1075/intp.9.2.02cha

Chen, B., and Zhu, X. (2014) 'Bilingual sentiment consistency for statistical machine translation', in *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, pp. 607–15. Gothenburg, Sweden: Association for Computational Linguistics. https://doi.org/10.3115/v1/e14-1064

Cheung, A. K. F. (2007) 'The Effectiveness of Summary Training in Consecutive Interpreting (CI) Delivery', *FORUM/Revue internationale d'interprétation et de traduction/International Journal of Interpretation and Translation*, **5**: 1–23. https://doi.org/10.1075/forum.5.2.01che

Cheung, A. K. F. (2016) 'Paraphrasing Exercises and Training for Chinese-to-English Consecutive Interpreting', *FORUM. Revue internationale d'interprétation et de traduction/International Journal of Interpretation and Translation*, **14**: 1–18. https://doi.org/10.1075/forum.14.1.01che

Cheung, A. K. F. (2019) 'The Hidden Curriculum Revealed in Study Trip Reflective Essays', in F. Austermühl, V. Enríquez Raído, and D. B. Sawyer (eds) *The Evolving Curriculum in Interpreter and Translator Education*, pp. 393–408. John Benjamins: Amsterdam/Philadelphia.

Dayter, D. (2018) 'Describing Lexical Patterns in Simultaneously Interpreted Discourse in a Parallel-Aligned Corpus of Russian-English Interpreting (SIREN)', *FORUM. Revue Internationale d'interprétation et de Traduction/International Journal of Interpretation and Translation*, **16**: 241–64. https://doi.org/10.1075/forum.17004.day

Demirtas, E., and Pechenizkiy, M. (2013) 'Cross-lingual polarity detection with machine translation', In *Proceedings of the Second International Workshop on Issues of Sentiment Discovery and Opinion Mining*, pp. 1–8. https://doi.org/10.1145/2502069.2502078

Devlin, J. *et al.* (2019) 'Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding', in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, volume 1, pp. 4171–86.

Donovan, C. (2004) 'European Masters Project Group: Teaching Simultaneous Interpretation into a B Language: Preliminary Findings', *Interpreting*, **6**: 205–16.

Ghorbel, H., and Jacot, D. (2011) 'Sentiment Analysis of French Movie Reviews', in V. Pallotta, A. Soro, and E. Vargiu (eds) *Advances in Distributed Agent-Based Retrieval Tools*, pp. 97–108. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21384-7_7

Gile, D. (1992) 'Predictable sentence endings in Japanese and conference interpretation', The Interpreter's Newsletter.

Gile, D. (1995) 'Fidelity Assessment in Consecutive Interpretation: An Experiment', *Target. International Journal of Translation Studies*, **7**: 151–64.

Gile, D. (1999) 'Variability in the Perception of Fidelity in Simultaneous Interpretation', *HERMES-Journal of Language and Communication in Business*, **12**: 51–79. https://doi.org/10.7146/hjlcb.v12i22.25493

Gile, D. (2003) 'Justifying the Deverbalization Approach in the Interpreting and Translation Classroom', *FORUM. Revue internationale d'interprétation et de traduction/International Journal of Interpretation and Translation*, **1**: 47–63. https://doi.org/10.1075/forum.1.2.03gil

Gile, D. (2011) 'Errors, Omissions and Infelicities in Broadcast Interpreting', in C. Alvstad, A. Hild, and E. Tiselius (eds) *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies*, pp. 201–18. https://doi.org/10.1075/btl.94.15gil

Gopaldas, A. (2014) 'Marketplace Sentiments', *Journal of Consumer Research*, **41**: 995–1014. https://doi.org/10.1086/678034

Hale, Sandra Beatriz (2004) *The Discourse of Court Interpreting : Discourse Practices of the Law, the Witness, and the Interpreter*. Amsterdam: J. Benjamins Publishing.

Hale, Sandra Beatriz (2007) *Community Interpreting*. London: Palgrave Macmillan.

Hale, S. B. *et al.* (2022) 'Does Interpreter Location Make A Difference? A Study of Remote Vs Face-to-Face Interpreting in Simulated Police Interviews', *Interpreting*, **24**: 221–53. https://doi.org/10.1075/intp.00077.hal

Hale, S. (2008) 'Crossing Borders in Community Interpreting: Definitions and Dilemmas', in C. Valero Garcés and A. Martin (eds) *Controversies Over the Role of the Court Interpreter*, pp. 99, 121, 1. Amsterdam: John Benjamins.

Han, C. (2022) 'Interpreting Testing and Assessment: A State-of-the-Art Review', *Language Testing*, **39**: 30–55. https://doi.org/10.1177/02655322211036100

Hsieh, E. (2007) 'Interpreters as Co-Diagnosticians: Overlapping Roles and Services Between Providers and Interpreters', *Social science and medicine*, **64**: 924–37. https://doi.org/10.1016/j.socscimed.2006.10.015

Jacobsen, B. (2003) 'Pragmatic Meaning in Court Interpreting: An Empirical Study of Additions in Consecutively Interpreted Question-Answer Dialogues', *International Journal of Speech, Language and the Law*, **11**: 165–9. https://doi.org/10.1558/ijsll.v11i1.165

Lei, L., and Liu, D. (2021) *Conducting Sentiment Analysis*. Cambridge, United Kingdom: Cambridge University Press.

Liu, B. (2022) *Sentiment Analysis and Opinion Mining*. Switzerland: Morgan and Claypool Publisher.

Liu, D., and Lei L. (2018) 'The Appeal to Political Sentiment: An Analysis of Donald Trump's and Hillary Clinton's Speech Themes and Discourse Strategies in the 2016 US Presidential Election', *Discourse, Context and Media*, **25**: 143–52. https://doi.org/10.1016/j.dcm.2018.05.001

Liu, M-H. (2013) 'Design and Analysis of Taiwan's Interpretation Certification Examination', in D. Tsagari and R. van Deemter (eds) *Assessment Issues in Language Translation and Interpreting*, pp. 163–78 Frankfurt: Peter Lang.

Liu, X., and Hale, S. (2018) 'Achieving Accuracy in A Bilingual Courtroom: The Effectiveness of Specialised Legal Interpreter Training', *The interpreter and translator trainer*, **12**: 299–321. https://doi.org/10.1080/1750399X.2018.1501649

Liu, Y. *et al.* (2019) RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv.Org*. https://doi.org/10.48550/arxiv.1907.11692

Lu, B. *et al.* (2011) 'Joint bilingual sentiment classification with unlabelled parallel corpora', in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pp. 320–30. Portland, Oregon: ACL.

Lu, X., and Han, C. (2023) 'Automatic Assessment of Spoken-Language Interpreting Based on Machine-Translation Evaluation Metrics: A Multi-Scenario Exploratory Study', *Interpreting*, **25**: 109–43. https://doi.org/10.1075/intp.00076.lu

Mäntylä, M. V., Graziotin, D., and Kuutila, M. (2018) 'The Evolution of Sentiment Analysis: A Review of Research Topics, Venues, and Top Cited Papers', *Computer Science Review*, **27**: 16–32. https://doi.org/10.1016/j.cosrev.2017.10.002

Mathew, L., and Bindu, V. R. (2020) 'A review of natural language processing techniques for sentiment analysis using pre-trained models', In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)*, pp. 340–5. Erode, India: IEEE.

Medhat, W., Hassan, A., and Korashy, H. (2014) 'Sentiment Analysis Algorithms and Applications: A Survey', *Ain Shams engineering journal*, **5**: 1093–113. https://doi.org/10.1016/j.asej.2014.04.011

Mukhtar, N., Khan, M. A., and Chiragh, N. (2018) 'Lexicon-Based Approach Outperforms Supervised Machine Learning Approach for Urdu Sentiment Analysis in Multiple Domains'. *Telematics and Informatics*, **35**: 2173–83. https://doi.org/10.1016/j.tele.2018.08.003

Napier, J. (2004) 'Interpreting Omissions: A New Perspective'. *Interpreting*, **6**: 117–42.

Nasukawa, T., and Yi, J. (2003) 'Sentiment analysis: Capturing Favorability using Natural Language Processing'. International Conference On Knowledge Capture: Proceedings of the 2nd International Conference on Knowledge Capture; 23–25 Oct. **2003**, 70–7. https://doi.org/10.1145/945645.945658

Nicodemus, B. and Emmorey, K. (2015) 'Directionality in ASL-English Interpreting: Accuracy and Articulation Quality in L1 and L2', *Interpreting*, **17**: 145–66.

Ouyang, L., Lv, Q. and Liang, J. (2021) 'Coh-Metrix Model-Based Automatic Assessment of Interpreting Quality', in: *Testing and Assessment of Interpreting*, pp. 179–200. Singapore: Springer. https://doi.org/10.1007/978-981-15-8554-8_9

Prabha, M. I., and Srikanth, G. U. (2019) 'Survey of sentiment analysis using deep learning techniques', In *2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT)*, pp. 1–9. Chennai, India: IEEE. https://doi.org/10.1109/ICIICT1.2019.8741438.

Pöchhacker, F (2004/2022). *Introducing Interpreting Studies*. London: Routledge.

Radford, A., and Narasimhan, K. (2018) Improving Language Understanding by Generative Pre-Training.

Sandrelli, A., and Bendazzoli, C. (2005) 'Lexical patterns in simultaneous interpreting: A preliminary investigation of EPIC (European Parliament Interpreting Corpus)', in *Proceedings from the Corpus Linguistics Conference Series*, 1/1, Birmingham, United Kingdom: University of Birmingham.

Seeber, K., and Zelger, C. (2007) 'Betrayal–Vice or Virtue? An Ethical Perspective on Accuracy in Simultaneous Interpreting', *Meta*, **52**: 290–8. https://doi.org/10.7202/016071ar

Seleskovitch, D. (1978). 'Language and cognition', in D. Gerver and H. W. Sinaiko (eds) *Language Interpretation and Communication*, pp. 333–41. Boston, MA: Springer.

Serrano-Guerrero, J. *et al.* (2015) 'Sentiment Analysis: A Review and Comparative Analysis of Web Services', *Information Sciences*, **311**: 18–38. https://doi.org/10.1016/j.ins.2015.03.040

Shlesinger, M. (1994). 'Intonation in the Production and Perception of Simultaneous Interpretation', in S. Lambert and B. Moser-Mercer (eds) *Bridging the Gap: Empirical Research in Simultaneous Interpretation*, pp. 225–36. Amsterdam, The Netherlands: John Benjamins Publishing Co.

Stachowiak-Szymczak, K., and Korpal P. (2019) 'Interpreting Accuracy and Visual Processing of Numbers in Professional and Student Interpreters: An Eye-tracking Study', *Across Languages and Cultures*, **20**: 235–51. https://doi.org/10.1556/084.2019.20.2.5

Taboada, M. (2016) 'Sentiment Analysis: An Overview from Linguistics', *Annual Review of Linguistics*, **2**: 325–47. https://doi.org/10.1146/annurev-linguistics-011415-040518

Taboada, M. *et al.* (2011) Lexicon-Based Methods for Sentiment Analysis. *Computational linguistics*, **37**: 267–307. https://doi.org/10.1162/COLI_a_00049

Touvron, H. *et al.* (2023) 'Llama 2: Open foundation and fine-tuned chat models', arXiv preprint, arXiv:2307.09288.

Turner, B., Lai, M., and Huang, N. (2010) 'Error Deduction and Descriptors: A Comparison of Two Methods of Translation Test Assessment', *Translation and Interpreting: The International Journal of Translation and Interpreting Research*, **2**: 11–23.

Vilares, D., Alonso, M. A., and Gómez-Rodríguez, C. (2017) 'Supervised Sentiment Analysis in Multilingual Environments', *Information Processing and Management*, **53**: 595–607. https://doi.org/10.1016/j.ipm.2017.01.004

Wan, X. (2008) 'Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis', in *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pp. 553–61. Honolulu: Association for Computational Linguistics.

Wankhade, M., Rao, A. C. S., and Kulkarni, C. (2022) 'A Survey on Sentiment Analysis Methods, Applications, and Challenges', *Artificial Intelligence Review*, **55**: 5731–80. https://doi.org/10.1007/s10462-022-10144-1

Wei, J. *et al.* (2022) 'Emergent abilities of large language models', arXiv preprint, arXiv:2206.07682.

Wen, J. U., and Lei, L. (2022) 'Linguistic Positivity Bias in Academic Writing: A Large-Scale Diachronic Study in Life Sciences Across 50 Years', *Applied Linguistics*, **43**: 340–64. https://doi.org/10.1093/applin/amab037

Xu, H. (2022) 'A Survey Study of Lawyers' and Interpreters' Approaches to Interactional Management in Interpreted Lawyer-Client Interviews in Australia', *Across Languages and Cultures*, **23**: 226–44. https://doi.org/10.1556/084.2022.00190

Xu, H. (2021) 'Roles, Ethics, and Lawyers' Reactions: An Ethnographic Study of Interpreters' Role Performance in Interpreted Lawyer-Client Interviews', *Multilingua*, **40**: 617–46. https://doi.org/10.1515/multi-2020-0108

Xu, H. (2024) '"Please Make Sure We Don't Get This Interpreter Again" Australian Legal Aid Lawyers' Experience of Working with Interpreters', *Translation and Interpreting Studies,* **19**: 257–76. https://doi.org/10.1075/tis.21070.xu

Xu, H., Hale, S., and Stern, L. (2020) 'Telephone Interpreting in Lawyer-Client Interviews: An Observational Study'. *Translation and Interpreting: The International Journal of Translation and Interpreting Research*, **12**: 18–36. https://doi.org/10.12807/TI.112201.2020.A02

Xu, Y. *et al.* (2022) 'A Survey of Cross-Lingual Sentiment Analysis: Methodologies, Models, and Evaluations', *Data Science and Engineering*, **7**: 279–99. https://doi.org/10.1007/s41019-022-00187-3

Xu, H., and Liu, K. (2024) 'The Impact of Directionality on Interpreters' Syntactic Processing: Insights from Syntactic Dependency Relation Measures', *Lingua*, **308**: 103778.

Yu, W., and van Heuven, V. J. (2017) 'Predicting Judged Fluency of Consecutive Interpreting from Acoustic Measures', *Interpreting*, **19**: 47–68. https://doi.org/10.1075/intp.19.1.03yu

Zhang, L., Wang, S., and Liu, B. (2018) 'Deep Learning for Sentiment Analysis: A Survey', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **8**: e1253. https://doi.org/10.1002/widm.1253

Zhao, J. *et al.* (2024) 'Llama beyond English: An empirical study on language capability transfer', arXiv preprint, arXiv:2401.01055.

Digital Scholarship in the Humanities, 2025, 00, 1–12
https://doi.org/10.1093/llc/fqaf017
Full Paper