The International Journal for Translation & Interpreting Research
trans-int.org

Emotion and interpreting: Does affective language impact interpreting quality?

Paweł Korpal
Adam Mickiewicz University, Poznań, Poland
pkorpal@amu.edu.pl

DOI: 10.12807/ti.118201.2026.a03

Abstract: In recent decades, an increasing number of studies has focused on interpreting emotional language and its potential impact on interpreters' well-being. Interpreters have been shown to respond to emotionally charged content (e.g., Korpal & Jasielska, 2019). However, little is known about the relationship between emotional content and interpreting quality. To fill this empirical gap, interpreting quality for neutral, negative and positive sentences was measured in both L2-to-L1 (English into Polish) and L1-to-L2 (Polish into English) interpreting. Two sets of pre-tested sentences, matched for length and readability index, from psychophysiological studies (Korpal, Jankowiak, & Kaczmarek, 2025; Korpal, Jankowiak, & Kaczmarek, 2026), were used for the present analysis. A propositional accuracy score and an error analysis (in line with Bartłomiejczyk, 2010), were applied to measure interpreting quality. The study results show that interpreting quality may be lower for affect-laden sentences, relative to neutral sentences, which can be mainly observed in a higher number of interpreting errors for affective content. Interpreting into L2 triggers more errors but tends not to impact propositional accuracy scores. Overall, the study contributes to interpreting studies by testing the relationship between affect-laden content and interpreting performance. Study outcomes can be applied in interpreter education by introducing more practice in interpreting potentially challenging affective language.

Keywords: interpreting quality, emotion, valence, directionality, propositional accuracy score, error analysis

1. Introduction

In recent decades there has been a noticeable increase in research interest in the affective aspects of interpreting, with a growing number of studies opting for psychophysiological methods (e.g., Moser-Mercer, Künzli, & Korac, 1998; Roziner & Shlesinger, 2010; Korpal & Jasielska, 2019; Rojo López, Cifuentes Férez, & Espín López, 2021; Spinolo, Olalla-Soler, & Muñoz Martín, 2022). Previous research points to the fact that emotion is indeed implicated in the act of interpreting and interpreters respond to emotionally laden content (e.g., Korpal & Jasielska, 2019; Korpal et al., 2025; Korpal et al., 2026). However, little is known about the potential impact of emotional content on interpreting quality. What is the role of emotion in interpreting performance? Is it more challenging to interpret emotionally laden rather than neutral content? This research aims to answer these questions by analysing interpreting quality for negative, neutral and positive sentences in both L1-to-L2 (Polish into English) and L2-to-L1 (English into Polish) interpreting.

Interpreting quality is one of the central topics in interpreting studies, with various ways to measure it having been developed over the decades. The International Association of Conference Interpreters (AIIC) listed a set of standards of professional interpreting (Déjean Le Féal, 1990, p. 155; see Kurz, 1993). First, an interpretation should produce the same effect on interpretation users as the source language content on its listeners. This reflects a functional approach to translation, where "[t]o translate means to produce a text in a target setting for a target purpose and target addressees in target circumstances" (the Skopos theory; Vermeer, 1987, p. 29). Second, high-quality interpreting "should have the same cognitive content and be presented with equal clarity and precision in the same type of language" (Déjean Le Féal, 1990, p. 155, as quoted in Kurz, 1993), which highlights the importance of accuracy and the intelligibility of an interpreter's rendition. Finally, the "language and oratory quality" of the interpretation should be equal or even greater than that of the source language content (Déjean Le Féal, 1990, p. 155, as quoted in Kurz, 1993). In sum, the AIIC's list of interpreting standards involves a range of factors to be considered, including interpreting accuracy, language correctness, proper delivery, and the same effect on interpretation users. Such components of interpreting quality have also been tested empirically. For example, Kurz (1993) studied expectations for the quality of interpretation users in four groups: medical doctors, engineers, Council of Europe delegates, and interpreters themselves. The three most important factors were: "sense consistency with original message", "logical cohesion of utterance" and "use of correct terminology" (Kurz, 1993, p. 17-18), which points to the importance of the interpreter's accurate and intelligible rendition in the target language.

A range of methods to measure interpreting accuracy have been developed in research with interpreters. One such measurement is a propositional accuracy score, which has been quite extensively applied in interpreting studies (e.g., Kurz & Färber, 2003; Bartłomiejczyk, 2010; Gieshoff & Albl-Mikasa, 2024). In this method, source language content is divided into its smallest meaningful units; it is then verified whether this information has been rendered correctly in the interpretation and points are awarded for accurate renditions. As it is often time-consuming to rate propositional accuracy for an entire interpretation, more salient elements were subsequently identified in some studies and then rated for accuracy, e.g., critical sentences (Liu, Schallert, & Carroll, 2004), complex syntactic structures (Timarová, Čeňková, Meylaerts, Hertog, Szmalec, & Duyck, 2014), and numbers (Korpal, 2017).

Interpreting quality has also been measured by analysing errors (Barik, 1971; Kopczyński, 1980; Altman, 1994), where accuracy is only one criterion, and various classifications of errors have been developed. Barik's (1971) and Altman's (1994) classifications overlap to a large degree as Barik (1975) listed omissions, additions, and substitutions as translation departures, whilst Altman's (1994) classification contained omissions, additions and inaccurate renditions. Chronologically between these two, a slightly more elaborate classification of errors was developed by Kopczyński (1980), who distinguished between errors of competence and errors of performance, as well as included omissions, additions, errors of appropriateness, and errors of translation as potential interpreting infelicities. Kurz and Färber (2003) adopted both of these approaches and combined completeness and grammar scores with the analysis of two kinds of errors: communicative inaccuracies and flaws in delivery (see also Bartłomiejczyk, 2010).

Another way to measure interpreting quality is by recruiting raters, such as interpreter trainers, who provide their assessment based on selected marking

criteria. For example, aspects such as interpreting accuracy, language correctness, and delivery can be considered using such analytical scales. Over the last few decades, various rating scales have been developed. For example, both Zwischenberger (2010) and Lee (2015) used three main categories of interpreting performance: content, form, and delivery, subdivided into more specific criteria for each category. Elsewhere, Han (2015) listed three major categories for assessing interpreting performance, i.e., information completeness (InfoCom), fluency of delivery (FluDel), and target language quality (TLQual). Adopting a rating scale may enable a more objective and holistic assessment of interpreting performance.

In summary, interpreting quality is a broad concept that may be difficult to operationalise. To produce a relatively objective quantification of interpreting performance, a range of methods have been applied in research involving interpreters, including: propositional accuracy scores, error analysis or assessment provided by raters, which are sometimes combined in one study. For example, propositional accuracy scores were used along with error analysis in Kurz and Färber (2003) and Bartłomiejczyk (2010). Another example of triangulating quality assessment methods is a combination of interpreting transcript analysis, interpreters' self-assessment, and audience response, as applied by Hamidi and Pöchhacker (2007).

Outside the context of language, determinants of the quality of performance have also been extensively researched. Psychological literature has long suggested that performance may depend on an individual's emotional states. In line with the widely cited Yerkes-Dodson law (Yerkes & Dodson, 1908), performance depends on the level of an individual's arousal. While performance levels increase to a certain point, too much arousal compromises performance. In other words, a moderate level of arousal tends to lead to optimal performance. Furthermore, emotion has been shown to impact cognitive processes, including memory or reasoning (Tyng, Amin, Saad, & Malik, 2017). These relationships, however, have only been researched to a limited extent in translation and interpreting studies. As such, interpreting quality has to date been rarely studied in the context of affective language in interpreting. Furthermore, it has not been established whether potential challenges in interpreting emotionally laden content (e.g., Ndongo-Keller, 2015; Mehus & Becher, 2016) are reflected in compromised interpreting quality.

However, the relationship between emotion and performance has occasionally been tested in the context of translation. For example, positive feedback provided to student translators was shown to impact their performance (Rojo López & Ramos Caro, 2016). It has also been found that positive feedback may be more encouraging than negative feedback, which is reflected in better translation performance in response to praise (Kimovska & Cvetkoski, 2021). Moreover, it has been shown that the translator's emotional involvement may stimulate creativity in translation (Lehr, 2011; 2013; see also Rojo López, 2017). Elsewhere, positively laden content was found to foster translators' engagement with the task while negatively laden content tended to trigger an emotional response in translators while they were performing the task (Lehr & Hvelplund, 2020). Translation performance has also been shown to be influenced by time pressure. For example, many translation trainees were observed to perform better when given more time to complete a translation task (Hansen, 2006). However, extended time was also found to negatively affect some trainees as they introduced errors in the revision phase when provided with more time to translate a text (Hansen, 2006; see also Rojo López, 2017 for a review of research on emotion and time pressure in translation). In interpreting

studies, some research has focused on a potential relationship between stress or anxiety experienced by interpreters and interpreting performance. For example, a negative correlation was observed between stress and interpreting accuracy (Korpal, 2017). Elsewhere, interpreting learning outcomes were shown to be negatively correlated with student interpreters' foreign language anxiety (Chiang, 2010).

Although the above research answers the question of whether translators' and interpreters' emotional state may impact their performance, it says comparatively less about the relationship between affect-laden texts or speeches and interpreter performance. To this end, the present study analyses interpreting quality measured by propositional accuracy scores and error analysis depending on the affective characteristics of the source language stimuli. It is hypothesised that interpreting affective language may be more challenging as it involves not only cognitive processing but also deep processing of the affective characteristics of the source language content. It also looks at the potential impact of directionality on the quality of interpreting affective as opposed to neutral language. Most previous research on the role of directionality in interpreting seems to suggest that interpreting into L2 may be more cognitively taxing (e.g., Hyönä, Tommola, & Alaja, 1995; de Bot, 2000; García, Mikulan, & Ibáñez, 2016). On the other hand, interpretations into L2 appear to be more accurate when compared with L2-to-L1 interpreting (Tommola & Helevä, 1998). The potential impact of directionality on translation performance may also depend on text type (Whyatt, 2019), which may also be the case in interpreting. Following this, the present study aims to test the role of directionality in the context of emotional language to see whether interpreting directionality modulates interpreting quality of the source language stimuli differing in emotional valence. Drawing on previous research on the cognitive aspects of directionality in interpreting, it is also hypothesised that interpreting into L2 may be of lesser quality, but not necessarily in terms of accuracy.

2. Methods

2.1. Participants

The analyses below are part of a larger project on interpreting affective language (Korpal et al., 2025; Korpal et al., 2026), consisting of two experimental studies combining psychophysiological, self-report, and behavioural data. In both studies, a sample of thirty professional interpreters was recruited (Study 1: 19 women, 11 men; M_{Age} = 38.5, 95% CI [35.21, 41.79]; Study 2: 19 women, 11 men; M_{Age} = 39.37, 95% CI [36.39, 42.34]). It should be noted that some participants were excluded from final analyses in the psychophysiological experiments, i.e., Study 1 (Korpal et al., 2025) and Study 2 (Korpal et al., 2026), mainly due to ECG abnormalities. However, the analysis of quality is presented here for all the recruited participants. All the participants had Polish as their native language and English as their working language. Participants were highly proficient in English, which was reflected in their LexTALE scores (Lemhöfer & Broersma, 2012); Study 1: M = 92.11, 95% CI [89.56, 94.65]; Study 2: M = 93.55, 95% CI [90.91, 96.19]; as well as in a self-report questionnaire, in which participants rated their proficiency in English on a 7-point scale (1 – very poor, 7 – native-like): Study 1 – reading (M = 6.63, 95% CI [6.44, 6.83]), writing (M = 6.23, 95% CI [6.01, 6.46]), speaking (M = 6.20, 95% CI [5.98, 6.42]), and listening (M = 6.23, 95% CI [5.99, 6.48]); Study 2 – reading (M = 6.55, 95% CI [6.35, 6.75]), writing (M = 6.23, 95% CI [6.06, 6.40]), speaking (M = 6.23, 95%

CI [6.06, 6.40]), and listening (M = 6.23, 95% CI [6.04, 6.43]). To participate in the study, participants were required to have at least 2 years of professional experience in interpreting in the English-Polish language pair (Study 1: M = 13.6, 95% CI [10.29, 16.91]; Study 2: M = 14.55, 95% CI [11.54, 17.56]). All participants had normal or corrected-to-normal vision and hearing, and none reported any neurological, mood, psychiatric, or language disorders. The project was approved by the Ethics Committee for Research Involving Human Participants at Adam Mickiewicz University, Poznań, Poland (Resolution No. 14/2020/2021). Participants provided their written informed consent before commencing participation in the experimental sessions, with participants being remunerated separately for Study 1 and Study 2.

2.2. Materials

45 Polish and 45 English sentences were used in both Study 1 and Study 2, with 15 negatively valenced (e.g., *The terrorists tortured me for a few days.*), 15 positively valenced (e.g., *He hugged her and suddenly she felt safe.),* and 15 neutral sentences (e.g., *He tied his shoes and left the apartment.)* in each language. The sentences in one language were not translation equivalents of the sentences in the other language. Two sets of sentences were created – one for Study 1 and one for Study 2. While in Study 1 the sentences were presented auditorily (Korpal et al., 2025), in Study 2 the speakers were also visible on a screen (Korpal et al., 2026). The speakers in both studies were native speakers of respective languages – one native speaker of English and one native speaker of Polish in both studies. The speakers were instructed not to act out affect-laden sentences; in this way we mitigated the risk of potential variability in the emotional expression between the two speakers in each study. Both sentence length, measured by the number of words per sentence, and readability indexes (English sentences: the Flesch reading ease score – Flesch, 1948; Polish sentences: Jasnopis – Gruszczyński & Ogrodniczuk, 2016) were controlled. Both Polish and English sentences were pre-tested using web-based surveys on sentence valence and arousal. For details on stimuli characteristics and normative studies, please see Korpal et al. (2025) (Study 1) and Korpal et al. (2026) (Study 2).

2.3. Procedure

Study 1 was carried out in a booth in the Psychophysiology of Language and Affect Laboratory and Study 2 – in the Language and Communication Laboratory (both at the Faculty of English, Adam Mickiewicz University, Poznań, Poland). The studies were conducted on a laptop with a screen resolution of 1280×1024 pixels. Throughout the research project, psychophysiological methods were also applied; participants' electrodermal activity and heart rate were monitored while they were interpreting (Korpal et al., 2025; Korpal et al., 2026).

During the experiments, participants were presented with the sentences to be interpreted in the E-Prime 3.0 (Psychology Software Tools, Pittsburgh, USA) software. They interpreted the sentences one by one, consecutively. Each block of sentences was preceded by a 5-minute resting phase which was necessary to collect reliable psychophysiological data. The order of block presentation was randomised, as was the order of 15 sentences presented within a block. Three blocks in one language were presented one after another; the language of the first three blocks vs. the last three blocks was counterbalanced across the participants. Korpal et al. (2025) and Korpal et al. (2026) offer details on psychophysiological, self-report, and behavioural data collection.

2.4. Data analysis

Transcripts of the interpretations were prepared based on audio recordings collected in the experiments. A spreadsheet in Microsoft Excel was prepared, which included sentences in the source language along with their interpretations, so that the raters could evaluate the quality of interpretations and provide their ratings in the same file.

Two methods were applied to measure interpreting quality: a propositional accuracy score and an error analysis. Both methods involved assessment provided by two English-Polish interpreters with experience as interpreting trainers. First, both raters were presented with the methods and assessment criteria. Next, the raters assessed interpreted renditions of a set of neutral sentences from the study's practice block. These sentences had not been used in the experiments proper. This was followed by a discussion of the obtained scores. Finally, raters independently assessed interpreting quality for experiments proper (Study 1 and Study 2). Interrater reliability was tested with the interclass correlation coefficient. The obtained values were: .806 for propositional accuracy scores, .815 for error analysis, and .915 for flaws in delivery, which points to high interrater reliability for all three measurements.

The main aim of adopting propositional rating scores was to measure interpreting accuracy, i.e., the correct rendition of the source language sentences in the target language. Even though the sentences presented to participants did not involve any specialised language, it was assumed that any emotions experienced by interpreters might be correlated with their ability to produce accurate renditions. Transcripts of interpretations were compared with the transcripts of the original sentences, whereby 1 point was awarded for each correct rendition of a proposition, i.e., a meaningful interpreting unit.

For the error analysis, penalty points were given for flaws in delivery (1 pt), lexical mistakes (3 pts), grammatical mistakes (3 pts), omissions, unjustified, additions, and changes in meaning (5 points), as well as incomplete and illogical sentences (10 pts) (slightly adapted from Bartłomiejczyk, 2010). Thus, with this method, not only was interpreting accuracy measured, but also language correctness and target language delivery. A separate analysis was performed on flaws in delivery to study the effect of affective language on the quality of target language delivery in isolation. Following Bartłomiejczyk's (2010, p. 186) definition, flaws in delivery were understood as "repetitions, corrections, false starts, and voiced and silent hesitations". These were identified by the raters based on audio recordings of the participants' renditions.

IBM SPSS Statistics 27 software was used for inferential statistics. For all three measurements, mean results from the two raters were taken for the analyses. Since there were small differences in the total number of propositions (meaningful interpreting units) for all six blocks of sentences (Study 1: negative English sentences – 89, neutral English sentences – 91, positive English sentences – 101, negative Polish sentences – 92, neutral Polish sentences – 96, positive Polish sentences – 98; Study 2: negative English sentences – 91, neutral English sentences – 93, positive English sentences – 95, negative Polish sentences – 88, neutral Polish sentences – 91, positive Polish sentences – 97), propositional accuracy scores were first calculated as percentages (the percentage of correctly rendered propositions within a given block of sentences) and then analysed.

3. Results

3.1. Propositional accuracy scores

ANOVA performed on propositional accuracy scores showed a main effect of sentence valence, $F(2, 236) = 6.311$, $p = .002$, $\eta_p^2 = .051$. Scores for neutral stimuli were the highest. Pairwise comparisons showed that there was a statistically significant difference between neutral ($M = 98.118$, $SE = .154$) and positive sentences ($M = 97.448$, $SE = .190$; $p = .008$) as well as between negative ($M = 97.958$, $SE = .152$) and positive sentences ($p = .025$). There was no statistically significant difference between negative and neutral sentences, $p = 1.000$. There was no effect of interpreting direction (L2-to-L1: $M = 97.802$, $SE = .171$; L1-to-L2*:* $M = 97.880$, $SE = .171$), $p = .748$. The interaction between stimuli valence and interpreting direction was statistically non-significant, $p = .885$. Propositional accuracy scores for all the three valences (positive, neutral, and negative sentences) and both interpreting directions (English into Polish, i.e., L2-to-L1, and Polish into English, i.e., L1-to-L2 interpreting) have been summarised in Figure 1, with higher scores indicating a more accurate interpretation.

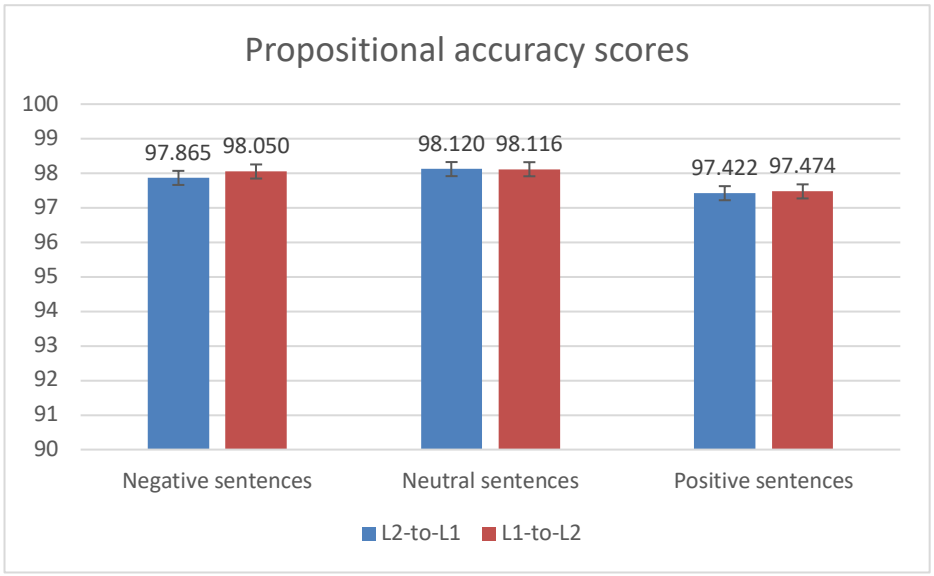

Figure 1. Propositional accuracy scores for negative, neutral, and positive sentences, in both interpreting directions.

Additional analyses were performed for Study 1 (auditory stimuli) and Study 2 (audiovisual stimuli) separately.

ANOVA performed on propositional accuracy scores for Study 1 showed a main effect of sentence valence, $F(2, 116) = 17.616$, $p < .001$, $\eta_p^2 = .233$. Scores for neutral stimuli were the highest. Pairwise comparisons showed that the scores for neutral sentences ($M = 98.363$, $SE = .174$) were statistically different from both negative ($M = 97.421$, $SE = .215$; $p < .001$) and positive sentences ($M = 96.650$, $SE = .303$; $p < .001$). The difference between negative and positive sentences was approaching statistical significance ($p = .063$). There was no effect of interpreting direction (L2-to-L1: $M = 97.504$, $SE = .238$; L1-to-L2*:* $M = 97.453$, $SE = .238$), $p = .881$. The interaction between stimuli valence and interpreting direction was statistically non-significant, $p = .861$.

ANOVA performed on propositional accuracy scores for Study 2 showed a main effect of sentence valence, $F(2, 116) = 3.652$, $p = .029$, $\eta_p^2 = .059$. Scores for neutral stimuli were the lowest. Pairwise comparisons showed that the scores for neutral sentences ($M = 97.873$, $SE = .252$) were statistically different from negative sentences ($M = 98.494$, $SE = .193$; $p = .037$). The difference between neutral and positive sentences ($M = 98.246$, $SE = .181$) was statistically non-significant ($p = .437$), as was the difference between negative and positive sentences ($p = .636$). There was no effect of interpreting direction (L2-to-L1: $M = 98.101$, $SE = .231$; L1-to-L2: $M = 98.308$, $SE = .231$), $p = .530$. The interaction between stimuli valence and interpreting direction was statistically non-significant, $p = .862$.

3.2. Error analysis

ANOVA performed on error analysis showed a main effect of sentence valence, $F(2, 236) = 7.501$, $p < .001$, $\eta_p^2 = .060$. Scores for neutral stimuli were the lowest. Pairwise comparisons showed that the scores for neutral sentences ($M = 11.242$, $SE = .637$) were statistically different from both negative ($M = 13.288$, $SE = .764$; $p = .048$) and positive sentences ($M = 14.275$, $SE = .774$; $p < .001$). The difference between negative and positive sentences was statistically non-significant ($p = .551$). There was also an effect of interpreting direction with more errors in interpreting from L1 to L2 (L2-to-L1: $M = 11.425$, $SE = .796$; L1-to-L2: $M = 14.444$, $SE = .796$), $F(1, 118) = 7.193$, $p = .008$, $\eta_p^2 = .057$. The interaction between stimuli valence and interpreting direction was statistically non-significant, $p = .433$. Figure 2 presents error analysis scores for all the three valences and both interpreting directions, with a higher score for errors indicating lower interpreting quality.

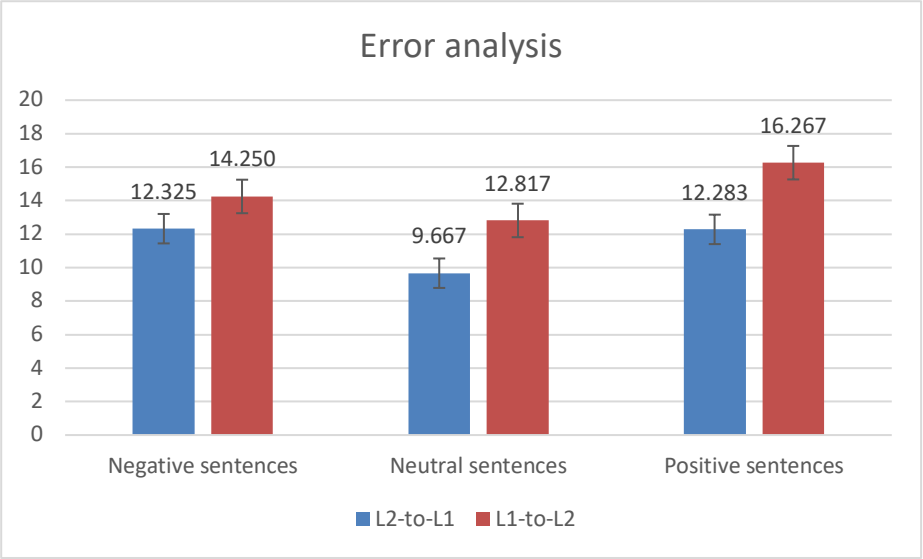

Figure 2. Error analysis scores for negative, neutral, and positive sentences, in both interpreting directions.

Additional analyses were performed for Study 1 (auditory stimuli) and Study 2 (audiovisual stimuli) separately.

ANOVA performed for error analysis for Study 1 showed a main effect of sentence valence, $F(2, 116) = 19.909$, $p < .001$, $\eta_p^2 = .256$, while scores for neutral stimuli were the lowest. Pairwise comparisons showed that the scores

for neutral sentences (M = 10.358, SE = .749) were statistically different from both negative (M = 15.950, SE = 1.127; p < .001) and positive sentences (M = 16.808, SE = 1.111; p < .001). The difference between negative and positive sentences was statistically non-significant (p = 1.000). There was also an effect of interpreting direction with more errors in interpreting from L1 to L2 (L2-to-L1: M = 12.278, SE = 1.105; L1-to-L2: M = 16.467, SE = 1.105), F(1, 58) = 7.181, p = .010, η_p^2 = .110. The interaction between stimuli valence and interpreting direction was statistically non-significant, p = .098.

ANOVA performed for error analysis for Study 2 showed no main effect of sentence valence: negative sentences (M = 10.625, SE = .926), neutral sentences (M = 12.125, SE = 1.005), positive sentences (M = 11.742, SE = .634), p = .981. There was no effect of interpreting direction (L2-to-L1: M = 10.572, SE = 1.091; L1-to-L2: M = 12.422, SE = 1.091), p = .236. The interaction between stimuli valence and interpreting direction was statistically non-significant, p = .598.

3.3. Flaws in delivery

A separate analysis was performed on flaws in delivery – one of the components of error analysis. ANOVA performed on flaws in delivery showed a main effect of sentence valence, F(2, 236) = 6.463, p = .002, η_p^2 = .052. Scores for neutral stimuli were the lowest. Pairwise comparisons showed that the scores for neutral sentences (M = 1.475, SE = .110) were statistically different from both negative (M = 2.058, SE = .178; p = .002) and positive sentences (M = 1.900, SE = .168; p = .026). The difference between negative and positive sentences was statistically non-significant (p = 1.000). There was also an effect of interpreting direction with more flaws in delivery in interpreting from L1 to L2 (L2-to-L1: M = 1.322, SE = .172; L1-to-L2: M = 2.300, SE = .172), F(1, 118) = 16.238, p < .001, η_p^2 = .121. The interaction between stimuli valence and interpreting direction was statistically non-significant, p = .804. The flaws in delivery scores for all the three valences and both interpreting directions are summarised in Figure 3; a higher score indicates a greater frequency of the delivery flaws and, consequently, lower interpreting quality.

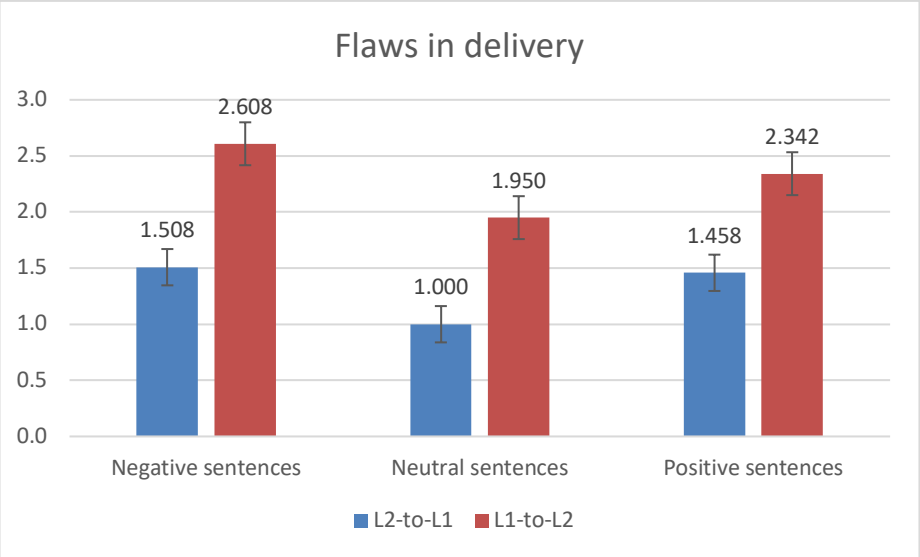

Figure 3. Flaws in delivery: Scores for negative, neutral, and positive sentences, in both interpreting directions.

Additional analyses were performed for Study 1 (auditory stimuli) and Study 2 (audiovisual stimuli) separately.

ANOVA performed on flaws in delivery for Study 1 showed a main effect of sentence valence, $F(2, 116) = 7.213$, $p < .001$, $\eta_p^2 = .239$, while scores for neutral stimuli were the lowest. Pairwise comparisons showed that the scores for neutral sentences ($M = 1.150$, $SE = .116$) were statistically different from both negative ($M = 1.875$, $SE = .242$; $p = .006$) and positive sentences ($M = 1.950$, $SE = .235$; $p = .001$). The difference between negative and positive sentences was statistically non-significant ($p = 1.000$). There was also an effect of interpreting direction with more flaws in delivery in interpreting from L1 to L2 (L2-to-L1: $M = 1.228$, $SE = .221$; L1-to-L2*:* $M = 2.089$, $SE = .221$), $F(1, 58) = 7.584$, $p = .008$, $\eta_p^2 = .116$. The interaction between stimuli valence and interpreting direction was statistically non-significant, $p = .414$.

ANOVA performed on flaws in delivery for Study 2 showed no main effect of sentence valence: negative sentences ($M = 2.242$, $SE = .263$), neutral sentences ($M = 1.800$, $SE = .176$), positive sentences ($M = 1.850$, $SE = .245$), $p = .132$. There was an effect of interpreting direction with more flaws in delivery in interpreting from L1 to L2 (L2-to-L1: $M = 1.417$, $SE = .263$; L1-to-L2*:* $M = 2.511$, $SE = .263$), $F(1, 58) = 8.673$, $p = .005$, $\eta_p^2 = .130$. The interaction between stimuli valence and interpreting direction was statistically non-significant, $p = .563$.

4. Discussion

Previous research has suggested that interpreters may respond to affect-laden content (e.g., Korpal & Jasielska, 2019; Korpal et al., 2025; Korpal et al., 2026). The main aim of the present analysis was to test whether such affective language can also impact quality in both L1-to-L2 and L2-to-L1 interpreting. It was hypothesised that interpreting emotionally laden content may be more challenging as it involves processing emotion on top of the other cognitive mechanisms involved in the task. To offer more detailed insights into the relationship between affect-laden content and interpreting performance, propositional accuracy scores were complemented with the analysis of errors globally and flaws in delivery more specifically.

Overall, the study findings suggest that emotional content modulates interpreting quality. For most measurements, lower quality ratings were observed for interpreting affect-laden, relative to neutral sentences. This is visible in interpreting accuracy (propositional accuracy scores) in Study 1, holistic error analysis including language correctness, and an analysis of flaws in delivery. These findings appear to apply to both positive and negative sentences. For most measurements no statistically significant differences were observed between negative and positive stimuli, while slightly lower propositional accuracy scores were observed for positive relative to negative sentences. The obtained results are in line with existing research suggesting that emotional arousal may compromise interpreting accuracy (e.g., Korpal, 2017) and extend this research by including an analysis of the affectivity of source language material. Interpreting emotionally laden content may be more taxing as it not only involves language and cognitive processing, but also an analysis of affective cues to be able to render them appropriately in the target language.

The study also aimed to test the effect of directionality in the context of interpreting affective language. There were no statistically significant differences between L2-to-L1 and L1-to-L2 interpreting for propositional

accuracy scores. However, more errors – including for delivery – were identified when participants were interpreting into the foreign language, except for error analysis for Study 2. Since interpreting accuracy was also included in the holistic error analysis, the discrepancy in the results between propositional accuracy scores and error analysis appears to result mainly from a higher number of language errors and delivery flaws in L1-to-L2 interpreting. The study results are largely consistent with research on the role of directionality in general, albeit not in the context of emotionally laden content (e.g., Hyönä et al., 1995; de Bot, 2000; García et al., 2016), suggesting that interpreting into L2 may be cognitively more demanding than interpreting into L1. Study outcomes also partly confirm the results of a study by Tommola and Helevä (1998) by showing that L1-to-L2 interpreting does not have to be less accurate than L2-to-L1 interpreting. In their study, a reverse trend was even observed with retour interpreting being more accurate. In general, it appears that the complex nature of retour interpreting stems from the fact that it involves speech production in a foreign language, which is more likely to trigger grammatical and pronunciation mistakes, as well as flaws in delivery. However, since in retour interpreting the interpreter receives the input in their native language, comprehension and – in turn – accurate rendition in the target language is not compromised.

The project reported here involved two studies; in Study 1 sentences were presented in the auditory modality, while in Study 2 participants could see the speakers on the screen. Additional analyses were performed for the two studies separately to see whether stimulus modality might be implicated in the quality of interpreting affective language. Differences in the effect of valence were observed for error analysis and flaws in delivery, in that a main effect of valence was only observed in Study 1. This may potentially be because emotion was easier to process when there was visual access to the speaker. In other words, processing emotion was more challenging when participants only received auditory stimuli (Study 1), as they were lacking visual cues that could help them process the valence of the sentences to be interpreted. Interestingly, when participants received audiovisual stimuli (Study 2), propositional accuracy scores were slightly higher for negative sentences, relative to neutral stimuli. This may again suggest that visual access to the speaker may modulate emotion processing in interpreting, make the interpreting task easier, and thus smooth out the differences in interpreting quality depending on the valence of the source language content. As for directionality, differences were observed for error analysis – an effect of directionality was observed only for auditory stimuli, which further suggests that visual access to the speaker may facilitate retour interpreting, which results in similar scores for L1-to-L2 and L2-to-L1 interpreting. Further research is needed to gain a better understanding of the potential role of modality in interpreting affective language.

The main limitation of the study is that sentences were used as stimuli, as opposed to longer narratives, which may not be most natural for professional interpreters. The rationale for applying single sentences as stimuli was the experimental control, where stimuli were pre-tested for both valence and arousal as well as controlled for their length and readability index. Adopting a longer narrative might have introduced more confounding variables in the study. Moreover, there was a slight overlap between the two methods of assessment, namely propositional accuracy scores and error analysis, as the latter included an accuracy rating. Finally, results for the propositional accuracy scores were in general very high which might have flattened the effect. However, the adoption of well-matched single sentences was important for the psychophysiological part of the project. Even so, further research could use

more complex stimuli to test the effect of emotional content on interpreting performance in more challenging interpreting contexts (see also Korpal et al., 2025).

The present analyses provide insights into the interrelationship between affect-laden content and interpreting quality by adopting methods used to measure interpreting accuracy, target language correctness, and delivery in an experimental psycholinguistic study. To avoid confounding variables, stimuli were well-matched and pre-tested in norming studies (see Korpal et al., 2025; Korpal et al., 2026). Future research could extend this research to in-situ interpreting, e.g., in community settings where the content to be interpreted would be role-played. Although experimental control could be more of a challenge in such a scenario, a study of that nature would test the outcomes obtained here in a more natural interpreting setting. Further research could also compare professional interpreters with interpreting trainees to study whether interpreting experience modulates the quality of affective language interpreting. This could serve as a further incentive to introduce more practice of interpreting emotionally charged conversations in interpreter training. The outcomes of the study may also be relevant for community interpreting settings, which have been shown to be emotionally challenging (e.g., Doherty, MacIntyre, & Wyne, 2010; Mehus & Becher, 2016; Leanza, Pointurier, & Duchesne, 2025), and where compromised interpreting accuracy can potentially have serious consequences. Since affective language transpires to be more challenging to interpret, more attention should be devoted to affect-laden content in interpreter training, as well as to how strategies to deal with interpreting affective language could be introduced in interpreter education.

5. Conclusion

There has been an increasing interest in the role of emotion in interpreting in the last few decades and emotion has been identified as potentially impacting the interpreter's well-being. However, the impact of affective language on interpreting quality has been scarcely studied in research involving professional interpreters. To this end, the analysis for this paper was performed to study the role of emotional valence and directionality in interpreters' performance. Overall, the findings suggest that affective language is more challenging to interpret than neutral content. Although not fully conclusive, the present analysis suggests that it is more cognitively taxing to interpret into one's L2, which is mainly reflected in interpreting errors, including flaws in delivery, but not in propositional accuracy scores. In sum, the results presented here suggest that emotion is implicated in interpreting performance. From a practical perspective, this finding can be applied in interpreter training in scenarios using emotionally laden content. Future interpreters should be informed as to the problematic nature of interpreting affective language and be provided with more practice to deal with the challenge.

Funding

This work was funded by the National Science Centre, Poland (grant number 2020/39/D/HS2/00326).

Data availability

The data and the stimuli are available at:
https://doi.org/10.17605/OSF.IO/9ZRP8

Acknowledgements

I would like to thank Katarzyna Jankowiak and Łukasz D. Kaczmarek for their comments on the study design.

References

Altman, J. (1994). Error analysis in the teaching of simultaneous interpretation: A pilot study. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp. 25–38). John Benjamins.

Barik, H. C. (1971). A description of various types of omissions, additions, and errors of translation encountered in simultaneous interpretation. *Meta*, *16*(4), 199–210. https://doi.org/10.7202/001972ar

Bartłomiejczyk, M. (2010). Effects of short intensive practice on interpreter trainees' performance. In D. Gile, G. Hansen, & N. K. Pokorn (Eds.), *Why translation studies matters* (pp. 183–194). John Benjamins. https://doi.org/10.1075/btl.88.16bar

Chiang, Y.-N. (2010). Foreign language anxiety and student interpreters' learning outcomes: Implications for the theory and measurement of interpretation learning anxiety. *Meta*, *55*(3), 589–601. https://doi.org/10.7202/045079ar

De Bot, K. (2000). Simultaneous interpreting as language production. In B. Englund Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting: Interdisciplinary perspectives* (pp. 65–88). John Benjamins.

Déjean Le Féal, K. (1990). Some thoughts on the evaluation of simultaneous interpretation. In D. Bowen & M. Bowen (Eds.), *Interpreting – yesterday, today, and tomorrow: American Translators Association Scholarly Monograph Series, Vol. IV* (pp. 154–160). State University of New York at Binghamton.

Doherty, S., MacIntyre, A., & Wyne, T. (2010). 'How does it feel for *you*?': The emotional impact and specific challenges of mental health interpreting. *Mental Health Review Journal*, *15*(3), 31–44. https://doi.org/10.5042/mhrj.2010.0657

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221–233. https://doi.org/10.1037/h0057532

García, A. M., Mikulan, E., & Ibáñez, A. (2016). A neuroscientific toolkit for translation studies. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 21–46). John Benjamins. https://doi.org/10.1075/btl.128.02gar

Gieshoff, A. C. & Albl-Mikasa, M. (2024). Interpreting accuracy revisited: A refined approach to interpreting performance analysis. *Perspectives: Studies in Translation Theory and Practice*, 32(2), 210–228. https://doi.org/10.1080/0907676X.2022.2088296

Gruszczyński, W., & Ogrodniczuk, M. (2016). *Jasnopis czyli mierzenie zrozumiałości polskich tekstów użytkowych* [Jasnopis: Measuring the readability of Polish functional texts]. Aspra.

Hamidi, M. & Pöchhacker, F. (2007). Simultaneous consecutive interpreting: A new technique put to the test. *Meta*, *52*(2), 276–289. https://doi.org/10.7202/016070ar

Han, C. (2015). Investigating rater severity/leniency in interpreter performance testing: A multifaceted Rasch measurement approach. *Interpreting*, *17*(2), 255–283. https://doi.org/10.1075/intp.17.2.05han

Hansen, G. (2006). Time pressure in translation teaching and translation studies. In S. Öztürk Kasar (Ed.), *Interdisciplinarity on translation: Proceedings of the 2nd*

International Colloquium on Translation organised by Yildiz Technical University, Volume II (pp. 71–80). Les Éditions Isis.

Hyönä, J., Tommola, J. & Alaja, A.-M. (1995). Pupil dilation as a measure of processing load in simultaneous interpretation and other language tasks. *Quarterly Journal of Experimental Psychology*, *48*(3), 598–612. https://doi.org/10.1080/14640749508401407

Kimovska, S. K. & Cvetkoski, V. (2021). The effect of emotions on translation performance. *Research in Language*, *19*(2), 169–186. https://doi.org/10.18778/1731-7533.19.2.05

Kopczyński, A. (1980). *Conference interpreting: Some linguistic and communicative problems*. Wydawnictwo Naukowe UAM.

Korpal, P. (2017). *Linguistic and psychological indicators of stress in simultaneous interpreting*. Wydawnictwo Naukowe UAM.

Korpal, P. & Jasielska, A. (2019). Investigating interpreters' empathy: Are emotions in simultaneous interpreting contagious? *Target*, *31*(1), 2–24. https://doi.org/10.1075/target.17123.kor

Korpal, P., Jankowiak, K., & Kaczmarek, Ł. D. (2025). Emotional content affects interpreters: Insights from psychophysiological, behavioral, and self-report measures. *Interpreting* (online first). https://doi.org/10.1075/intp.00125.kor

Korpal, P., Jankowiak, K., & Kaczmarek, Ł. D. (2026). The psychophysiology of interpreting emotional language in the audiovisual modality. *Perspectives: Studies in Translation Theory and Practice* (online first). https://doi.org/10.1080/0907676X.2025.2587594

Kurz, I. (1993). Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter*, *5*, 13–21.

Kurz, I. & Färber, B. (2003). Anticipation in German-English simultaneous interpreting. *FORUM. Revue internationale d'interprétation et de traduction / International Journal of Interpretation and Translation*, *1*(2), 123–150 https://doi.org/10.1075/forum.1.2.06kur

Leanza, Y., Pointurier, S., & Duchesne, L. (2025). Interpreting is an emotional effort: Public service interpreters' coping strategies in emotionally charged situations. *Interpreting and Society: An Interdisciplinary Journal*, *5(1)*, 3–26. https://doi.org/10.1177/27523810241310909

Lee, S.-B. (2015). Developing an analytic scale for assessing undergraduate students' consecutive interpreting performances. *Interpreting*, *17*(2), 226–254. https://doi.org/10.1075/intp.17.2.04lee

Lehr, C. (2011, June). *The happier, the better? Exploring the impact of positive and negative emotions on performance in translation* [Paper presentation]. International Symposium for Young Researchers in Translation, Interpreting and Intercultural Studies, Barcelona, Spain.

Lehr, C. (2013, December). *Influences of emotion on cognitive processing in translation: A framework and some empirical evidence* [Paper presentation]. International Online Workshop on Affective factors in Translation Process Research: To Feel or Not to Feel? That is the Question. Birmingham, UK.

Lehr, C. & Hvelplund, K. T. (2020). Emotional experts: Influences of emotion on the allocation of cognitive resources during translation. In R. Muñoz Martín & S. Halverson (Eds.), *Multilingual mediated communication and cognition* (pp. 44–68). Routledge.

Lemhöfer, K. & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Liu, M., Schallert D. L., & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting*, *6*(1), 19–42. https://doi.org/10.1075/intp.6.1.04liu

Mehus, C. J. & Becher, E. H. (2016). Secondary traumatic stress, burnout, and compassion satisfaction in a sample of spoken-language interpreters. *Traumatology*, *22*(4), 249–254. https://doi.org/10.1037/trm0000023

Moser-Mercer, B., Künzli, A., & Korac, M. (1998). Prolonged turns in interpreting. Effects on quality, physiological and psychological stress (pilot study). *Interpreting*, *3*(1), 47–64. https://doi.org/10.1075/intp.3.1.03mos

Ndongo-Keller, J. (2015). Vicarious trauma and stress management. In H. Mikkelson & R. Jourdenais (Eds.), *The Routledge handbook of interpreting* (pp. 337–351). Routledge.

Rojo López, A. M. (2017). The role of emotions. In J. W. Schwieter & A. Ferreira (Eds.), *The handbook of translation and cognition* (pp. 369–385). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119241485.ch20

Rojo López, A. M. & Ramos Caro, M. (2016). Can emotion stir translation skill? Defining the impact of positive and negative emotions on translation performance. In R. Muñoz Martín (Ed.), *Reembedding translation process research* (pp. 107–130). John Benjamins. https://doi.org/10.1075/btl.128.06roj

Rojo López, A. M., Cifuentes Férez, P., & Espín López, L. (2021). The influence of time pressure on translation trainees' performance: Testing the relationship between self-esteem, salivary cortisol and subjective stress response. *PLoS ONE*, *16*(9), e0257727. https://doi.org/10.1371/journal.pone.0257727

Roziner, I. & Shlesinger, M. (2010). Much ado about something remote: Stress and performance in remote interpreting. *Interpreting*, *12*(2), 214–247. https://doi.org/10.1075/intp.12.2.05roz

Spinolo, N., Olalla-Soler, C., & Muñoz Martín, R. (2022). Finding a way into an interpreter's heart: Methodological considerations on heart-rate variability building on an exploratory study. *The Interpreters' Newsletter*, *27*, 63–87. https://doi.org/10.13137/2421-714X/34392

Timarová, Š., Čeňková, I., Meylaerts, R., Hertog, E., Szmalec, A., & Duyck, W. (2014). Simultaneous interpreting and working memory executive control. *Interpreting*, *16*(2), 139–168. https://doi.org/10.1075/intp.16.2.01tim

Tommola, J. & Helevä, M. (1998). Language direction and source text complexity. Effects on trainee performance in simultaneous interpreting. In L. Bowker, M. Cronin, D. Kenny, & J. Pearson (Eds.), *Unity in diversity? Current trends in translation studies* (pp. 177–186). St. Jerome.

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in Psychology*, *8*, 1454. https://doi.org/10.3389/fpsyg.2017.01454

Vermeer, H. (1987). What does it mean to translate? *Indian Journal of Applied Linguistics*, 13, 25–33.

Whyatt, B. (2019). In search of directionality effects in the translation process and in the end product. *Translation, Cognition & Behavior 2*(1), 79–100. https://doi.org/10.1075/tcb.00020.why

Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology & Psychology*, *18*(5), 459–482. https://doi.org/10.1002/cne.920180503

Zwischenberger, C. (2010). Quality criteria in simultaneous interpreting: An international vs. a national view. *The Interpreters' Newsletter*, *15*, 127–142.