Assessing the Simultaneous Interpreting Outputs of Trainee Interpreters in Iraqi Departments of Translation

Research. Zakariya Yaseen Musa
Assistant Professor Jasim Khalifah Sultan Al-Maryani
University of Basra / College of Arts

Abstract:

Trainee interpreters at Departments of Translation, University of Basrah in Iraq commit significant errors that have a negative impact on the transference of speakers' messages. These errors have not yet been systematically and comprehensively examined. The present study tries to investigate this particular area. To assess the outputs, he study was based on an eclectic model that was based on two related frameworks, namely Riccardi's (1999) and Na'ja and Abu-Mighnim's (2012). This model helps to diagnose discontinuities in terms of meaning, disfluencies, and syntactic errors in the outputs of a randomly selected sample, which consists of twenty-four fourth-year students in Department of Translation, college of Arts, University of Basrah. They were invited to take a 3 and half minutes simultaneous interpreting test. The material of this test was two culturally-related political speeches with a delivery rate of around 145 words per minute (145 WPM). The errors in the outputs were quantitatively and qualitatively analysed. To check the validity of the analysis, an independent jury of specialized tutors was asked to evaluate the same outputs. The results obtained from the analysis and the jury's evaluation prove that the outputs were full of errors that impacted the transferred meaning. Neither the trainees nor their tutors were aware of these errors. Therefore, a revision of and an improvement for the followed training programmes should be carried on to improve the status quo of Iraqi trainee interpreters graduating from Departments of Translation.

Trainee interpreters Iraqi Universities Error Analysis Jury's Evaluation

تقييم مخرجات الترجمة الفورية للمتدربين في أقسام الترجمة العراقية

الباحث الأستاذ المساعد الدكتور

زكريا ياسين موسى جاسم خليفة سلطان المرياني

جامعة البصرة/ كلية الآداب

الملخص:-

يرتكب المترجمون الشفويون الذين يتدربون في أقسام الترجمة في العراق أخطاءً كثيرة تؤثر سلباً على نقل رسائل المتحدثين. لكن لم تتم دراسة هذه الأخطاء على نحو منهجي وشامل حتى الآن. لذا أتت الدراسة الحالية لتستقصي هذا الجانب في مخرجات الطلبة بكل تفاصيلها. ولتقويم مخرجات الترجمة الشفوية للطلبة المتدربين؛ تبنت الدراسة انموذجًا انتقائيًا يتكون من خلال الجمع بين إطارين متر ابطين يعودان ل ريكاردي (1999) و نعجة و ابو مغنم (2012). يساعد هذا الانموذج على تشخيص الانقطاعات في المعنى، والاختلالات والأخطاء النحوية في مخرجات عينة عشو ائية تتكون من أربعة وعشرين طالباً في السنة الر ابعة في قسم الترجمة/ جامعة البصرةً. أجرى الباحثان اختبار ترجمة فورية استغرق حوالي 3 دقائق ونصف. وتمثلت مادة هذا الاختبار في خطابين سياسيين ذوا ارتباط ثقافي تبلغ سرعة السرد فيهما حوالي (145 كلمة في الدقيقة). وقام الباحثان بتحليل مخرجات المتدربين كماً ونوعاً. وللتحقق من صحة التحليل، طلب الباحثان من لجنة مستقلة من التدريسيين المتخصصين تقييم عين المخرجات. و أتت النتائج التي استحصلها الباحثان من خلال التحليل فضلا عن تقييم اللجنة المنتدبة بان مخرجات عينة المتدربين مليئة بالأخطاء التي كان لها الأثر على المعنى المنقول بين اللغتين. وتجدر الأشارة في هذا السياق ان المتدربين ومعلميهم يجهلون وجود هذه الأخطاء. مما يستدعي إجراء مراجعة وتحسين للبرامج التدريبية المتبعة للأرتقاء بالمستوى العلمي الحالي

الكلمات الاساسية: **المترجمين الفوريين المتدربين في الجامعات العراقية**. **تقييم لجنة التحكيم**

1. Introduction

The study tackles the assessment of trainee interpreters' simultaneous interpreting outputs. The motivation behind such a topic relies on the fact that there is neither a systematic nor comprehensive research that has investigated this area in Iraqi Departments of Translation. Moreover, previous assessment studies overlooked interpreting and solely focus on written translations. Three significant phenomena would be assessed in this study; errors at the intertextual level, errors in the fluency of delivery, and errors at the intratextual level. These phenomena are investigated according to an adapted eclectic model of error analysis that is based on Riccardi's (1999) and Naj'a & Abu-Mighnim's (2012) models as well as the evaluation of a selected jury. Also, an SPSS statistical process is set to measure the convergence between the two investigations.

2. Literature Review

Assessment in translation depended in its early stages on linguistic theories. With the establishment of translation studies as an independent discipline, the tools used and procedures followed have sprung from this new field. In this respect, translation focused on multiple areas of study, especially translation criticism and translation training which were the two areas most investigated at the end of the previous century and the beginning of this new one. As perceived, the discipline has been well established, with Holmes's map as a key framework. But to date, there is no consensus on the specific criteria to evaluate acceptable translations. Hence, workable criteria for assessment have not been set yet. There are many valuable writings on translation assessment that are remarkably trustworthy for many scholars and translators. Most of these are mentioned below as they provide a solid theoretical background for our study. Different text types indeed adhere to different assessment criteria. For instance, Juliana House (1977) and Christina Nord (1993) study the written text, while Hatim and Mason (1997),

Setton (1999), and Riccardi (1999) focus on oral utterances. But citing some of them before offering the study's model seems significant.

Most translation assessment studies focus either the process of translation (translating) or the product (translation). One of the remarkable works is that of the pioneer Juliana House (1977). She attributes the initial idea of her model to Davy and Crystal (1969). In their book *Investigating English Style* (1969:66), Davy and Crystal classify a three-dimensional model for linguistic assessment. These dimensions cover the situational constraints of text analysis. As referred to above, House (1977) adapted, modified, and condensed Davey and Crystal's model into two sets of dimensions; A: language user and B: language use. The categories covered by language users are geographical origin, social class, and time, while language use is determined by medium, participation, social role relationship, social attitude, and province. This categorization reflects the fact that language user and language use are relevant for translation quality assessment. In other words, semantic, pragmatic and textual equivalents are essential replacements between the original and target texts (House, 2015: 28).

Nord tackles translating phenomena from a functionalist perspective. She demonstrates different criteria for the assessment of translation and interpreting. The purposefulness of translation and sponsors' goals are the main concerns of her study. Different modes of translation and interpreting are assessed as per this perspective including translation training, literary translation, and simultaneous interpreting. Nord (1997:40) underlines how functionalists understand translation training. They make use of the communicative function as a guide for training. Discussing error in translation, Nord (1997:31) emphasizes that Skopos is interested in measuring effectiveness on the target text audience. In other words, the assessment should be viewed more likely from the addressees rather than from the addressers. In this case, the mediator (i.e. the translator) shall take into account that, in some instances, what would be translated may not have the same function if it is retranslated to ST again, because the TT function may differ from that of ST.

Melis and Albir (2001) reveal that translations of sacred and literary texts have been heavily examined by scholars and researchers, while other areas are rarely visited. Hence, their study, as they anticipate, attempts to bridge the gap by focusing on two other areas to be assessed, namely, the translations of professional translators and those of trainee translators. Examining published translations as sacred and literary texts, Melis and Albir (2001:274) demonstrate many proposals that serve as starting points for assessment and analysis, such as Vinay and Darbelnet's technical procedures, Nida's dynamic equivalence, House's situational dimensions, Hatim and Mason's contextual dimension, Nord's functionalism and so on. The second area assessed in Melis and Albir's study is the texts translated by professional translators. This assessment provides feedback on whether the translator in question is qualified to occupy certain vacancies or membership that requires specific proficiency. In this assessment, technical, legal, scientific, commercial texts are the materials to be translated. It is generally adopted by translation agencies and international companies and is used as a mechanism for finding translators. Another important topic assessed is the method of teaching translators. Significantly, Melis and Alber (2001:278) adopt Melis's (1997) assessment sheet (see table 2.1). The criteria shown in the table refer to the type of assessment followed in each case (literary and sacred texts, professional translations texts, and trainee translators' texts).

Table (2.1) Assessment Criteria of Translation (Melis, 1997)

	PUBLISHED TRANSLATION	PROFESSIONAL TRANSLATION	TRANSLATION TEACHING
Object	Translation of literary and sacred texts	Translator competence	Students translator competence study plans programmes
Type	Product assessment Qualitative assessment Quantitative Assessment	Product assessment Quantitative assessment Procedure assessment	Product assessment Process assessment Qualitative assessment
Function	Summative	Summative Formative	Diagnostic Formative Summative
Aim	Informative Advertising Speculative Pedagogical	Economic-Professional Speculative	Academic Pedagogical Speculative
Means	Evaluation criteria	Non-literary translation Evaluation criteria Correcting scales Grading scales, tests, etc.	Translation evaluation criteria Correcting criteria Grading scales, tests. Exercises, questionnaires, etc.

Wu (2010) investigates the validity and reliability of the test in simultaneous interpreting. Many examiners have been invited to participate in this study. The high number of examiners is employed to inspect the test validity and measure its extent. Wu has conducted a three-minute test, then he presented the recordings to the examiners. Four methods of assessment have been employed by thirty examiners. Not all examiners have a background in teaching interpreting. Some are only translators with a limited interpreting experience. Positively, the results reveal that, in general, the examiners are inconsistent in their evaluation. They have adopted two methods of evaluation; the examiners with less interpreting experience use the transcripts of the recordings as the starting point, but the examiners with an interpreting experience depend on the recordings only. The study reveals that the former group shows much consistency in the judgments than those who are specialized in interpreting teaching.

Altman (1994:25) describes the different approaches to error in interpreting and other related disciplines. She perceives that error analysis in simultaneous interpreting is more difficult than error analysis in language teaching. At the beginning of her study, and as the experiment requires ideal interpreting to the text in question, Altman attempts to define ideal translation as a cornerstone. She assumes that there is a possibility to perceive a perfect translation and use it as a standard rendering. She (1994:37) concludes that her study is directed towards grouping and classifying common errors and mistakes in simultaneous interpreting. She also suggests that languages other than English and French, if involved, could lead to new facts and findings. The other significant benefit of this study is that it could help simultaneous interpreting trainers to adjust or modify their curriculum. Overall, the study is seen as a prediction of errors.

AL-Kilabi (2015) is among the pioneers of testing trainee interpreters' aptitude in simultaneous interpreting. He creatively suggests different testing criteria to measure the different skills and abilities necessary to take and design tests. He (2015:2) discusses many test types. The multi-choice method is the first one which is used to measure listening comprehension. The second method is back counting. It is used to figure out the participants' distraction resistance when they count in reverse way. The third method focusses on the oral cloze test. The participants are

given two texts with two different languages that contain many blanks. They are required to fill these blanks after listening to the subject-related recording. The fourth method is related to sight translation. A text is distributed to the participants; they are required to transpose it into language B orally. The memory test is the fifth method employed. It is designed to measure the participants' memory limitations. The last method is confined to simultaneous interpreting, where the participants are required to start interpreting a recorded segment simultaneously.

Moreover, AL-Kilabi (2015:12) proposes four test methods that are, according to him, much reliable to investigate simultaneous interpreting aptitude among trainee students, these are; sight translation in Arabic, shadowing in English, shadowing in Arabic, and listening comprehension. So, he dismisses the other test methods because of their unreliability; namely, oral cloze in Arabic, oral cloze in English, memory tests, and sight translation in English. Additionally, He (2015: 14) praises shadowing exercises because, as he believes, they cognitively involve measuring the listening and speech production which are the core of simultaneous interpreting. On the other hand, he criticizes many Arab interpreting and language centres because they ignore the role of listening comprehension exercises. He suggests that such a task should be taken into consideration more than often.

3. The Suggested Model of the Study

Regardless of the volume and difficulty of any academic study, the most important and critical decision that any researchers can take at the beginning of his/her project is the adoption of a certain model. As the literature reveals, the model is the cornerstone of any research. Critical readers, including examiners, are interested in finding the motives behind adopting a particular scheme and whether or not it has helped to arrive at concert conclusions. Within this study, the researchers were so keen to choose so carefully. Finally, the decision was to adopt an eclectic model, a comprehensive one that can account for Iraqi trainee students' interpreting errors. The researchers believe that this model will help to answer many of the offered research questions and validate the hypotheses. This model exclusively deals with error analysis in simultaneous interpreting. Because of limits of time and space, there are only three main types of errors the

study tries to explore in this context; intertextual, fluency, and intratextual errors. Within this area, discontinuities in terms of meaning (intertextual errors), disfluencies (errors of fluency), and linguistic errors (intratextual Errors) can be detected, diagnosed, and classified in terms of Ricarrdi's (1999) and Na'ja and Abu Mighnim's (2012) models.

To the researchers' knowledge, there is no complementary research on error analysis for simultaneous interpreting at Iraqi universities. Previous studies on the assessment of simultaneous interpreting outputs focus only on a narrow area of errors committed. As far as the Iraqi context is concerned, there is a gap in research to be filled. This encourages the researchers to try to fill this gap and try to measure Iraqi trainee interpreters' outputs. Hence, an eclectic model formulated by two different authors will be adopted. The combination of these two models is helpful, as it can investigate the wider area of errors committed by trainee students. Although it does not provide a holistic approach to error analysis of SI in Iraq, it tries to cover the most common errors in the trainees' outputs.

As hinted above, two models have eclectically emerged here. The first model is Riccardi's (1999). It is a two part model that is designed to measure errors at different levels; the intertextual level (omission, addition, and substitution) and the disfluency level, which consists of two main categories (unfilled and filled pauses). Unfilled pauses are the clear silences calculated by the examiner per second. Filled pauses are of four subcategories; hesitation, repetition, correction, and a false start. The second model borrowed and modified here is Na'ja and Abu Mighnim's (2012). Although the original model consists of six subcategories, definiteness and indefiniteness, and verb tense subcategories are disregarded. Then, the vernacular subcategory has been added to measure the interference of formality and informality of Arabic. As the intratextual is assigned to measure syntactic errors, The modified version then consists of five subcategories: (wrong use of gender), (singular, dual, and plural), (vernacularism), (inflection) and (pronouns). At this particular point, it is very important to review

and explain in the subsequent subsections the components of the two models composing the adopted eclectic one outlined in Diagram (3.1).

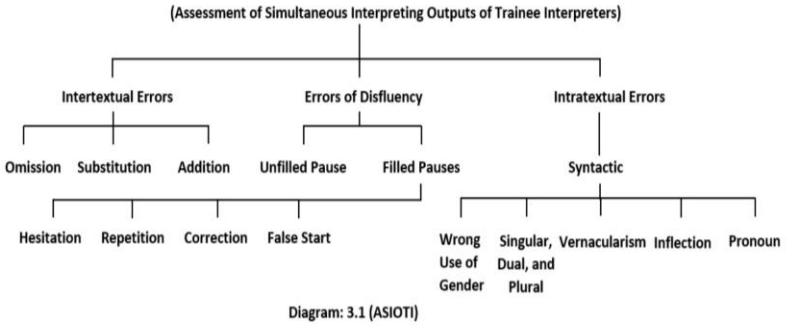

Diagram: 3.1 (ASIOTI)

3.1 Alesendra Riccardi's (1999) Model

Riccardi's (1999) model is an error analysis one. It investigates two main parts; discontinuities in terms of meaning (intertextual errors) and disfluencies (fluency errors). Diagram 3.2 outlines discontinuities in term of meaning (errors at the intertextual level):

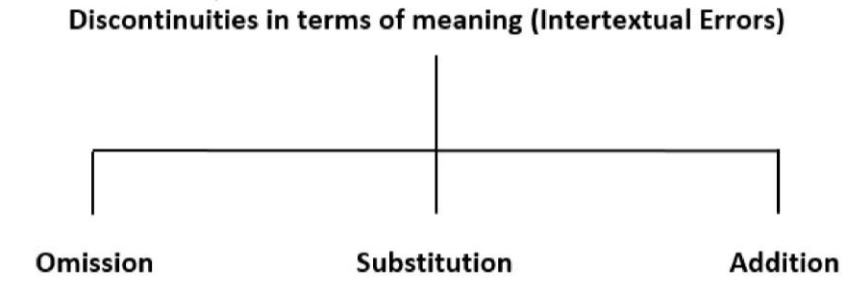

Diagram (3.2) Riccardi's (1999) Discontinuities in terms of meaning

3.1.1 Intertextual Errors

The errors at the intertextual level involve omission, substitution, and addition. The fourth category of error "logical-time sequence" is overlooked in this study as it is embodied within the substitution and addition categories. Generally speaking, intertextual errors are the departures made by the translator or interpreter. As there is mutual interference between the English and Arabic linguistic systems, the error monitoring is adopted to examine such departures.

Logically, when the term error (as meaning destruction) is mentioned in the assessment of translation/interpreting, it ought to be synchronized, at least, with phenomena such as omission, substitution, and addition. This synchronization is not arbitrary as it is unreasonable to assess errors isolated from each other unless there is a purpose or a limitation mentioned in advance. In this respect, Vancura (2017:13) offers a significant distinction between assessing errors in written translation and interpreting. She stresses that the criteria for assessing translation should differ from the ones used in assessing interpreting outputs. She assumes that on many occasions, minor impact errors in interpreting could be gross errors in translation. The interpreting literature reviewed in the previous chapter, however, has shown that emphasis should be on three main aspects; omission, substitution, and addition. Errors like an omission, substitution, and addition fall into two different assessing criteria in translation. They are either manipulated positively by the translator/interpreter or negatively. However, the current study doesn't look at these items from a positive perspective, as it focuses on errors only. Regardless of the many strategies that consider omitted, added, or substituted items as references for good TT comprehension, but in interpreting, particularly in simultaneous mode, they are more often seen as meaning destructive (Riccardi, 1999& 2001; Pio, 2003).

3.1. 2 Disfluencies (Fluency Errors)

This part of Riccardi's (1999) model is set to measure disfluencies in simultaneous interpreting. It examines various kinds of pauses and hesitations interpreters experience while rendering into the target language. Setton (1990:245) assures the absence of previous studies that tackle hesitation and fluency. However, he (as citing Culter, 1987:28) sheds light on some studies focusing on speech production, asserting that any monolingual speech is characterized by two articulating powers; the prosody and rhythm, both are monitored by the speaker adjustably. Moreover, a description of the many articulating phenomena was made by Setton as well. In one of the experiments he has conducted, Setton (1990:245) classifies several hesitations and pauses phenomena such as " *a) silent and filled pauses of varying length; b) vowel lengthening (drawl) and final consonant; c) stress, loudness and pitch variations; d) the use of lexical items as verbal fillers; and e) various solecisms and boggle effects".* In English language literature, Osgood et al. (1959:24) define four phenomena of hesitation and

disfluency: repetition, false start, filled pause, and unfilled pause. The study of Osgood et al was concluded from spontaneous utterances deducted from speeches of English native people. The average delivery rate of those utterances was 151 words per minute (Osgood et al., 1959:34). Osgood's study is considered the cornerstone and starting point for many other interrelated studies, especially for fluency in learning/teaching a foreign language and simultaneous interpreting.

Hesitation and pauses in simultaneous interpreting have been recently approached by different scholars such as Tissi, 2000; Riccardi, 1999, 2001 and 2005; Cecot, 2000; Pio, 2003; Rennert, 2010; Mankauskienė, 2018; and EL- Zawawy, 2019. A remarkable definition of error in fluency might be that of Rennert (2010:104):

a prosodic feature of speech that can be viewed as a function of several temporal variables. It is the complex interaction of pauses, audible breathing, hesitations, vowel, and consonant lengthening, false starts, repairs, repetitions, and speech rate that creates the impression of fluency or a lack thereof.

The illustration made by Rennart reviews the occasions where pauses take place. As the definition implies, pauses occur when there is an overlap with other pauses generated through hesitation, repetition, reconstruction, and interruption. Several studies adopt Alessandra Riccardi's (1999) model to examine errors in fluency. The most relevant to the present study is that of Pio (2003). Pio adopts Riccardi's model to measure interpreters' fluency. Her empirical study was conducted on different speech rates.

Disfluency, as a part of the model, is divided into two categories: unfilled and filled pauses. Unfilled pauses involve any clear silence that lasts more than three seconds. Filled pauses are of four types: vocalized hesitation, repetition, correction (reconstructing), and false starts. Logically, these types of errors occur more often when there are dissimilarities between the two linguistic systems in question. This can be noticed if the interpreter uses many pauses than usual, probably because of the non-correspondences between the processed input and the would-be output. Diagram (3.3) surveys the categories and subcategories of disfluency.

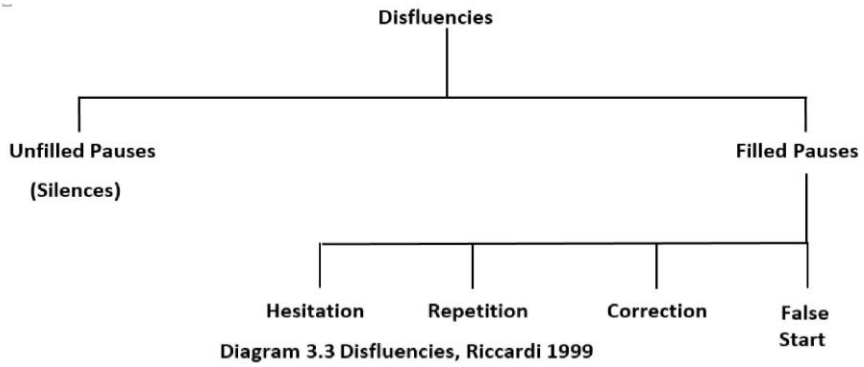

Diagram 3.3 Disfluencies, Riccardi 1999

3.2 Na'ja and Abu-Mighnim's (2012) Model

The second model that is emerged in the present study with Riccardi's (1999) is Na'ja and Abu-Mighnim's (2012). This model is set to examine the intratextual errors committed by trainee students. It is employed to cover different syntactic errors. Originally, the model is set to investigate foreign students' errors where Arabic is language B, while the tokens of the current study are students in the departments of translation at Iraqi universities, whose Arabic is language A. The authors have positively confirmed to the researchers the validity of this usage via several emails. However, as mentioned earlier, this model covers various syntactic errors. Diagram (3.5) summarizes the kinds of intratextual errors suggested by Na'ja and Abu-Mighnim (2012).

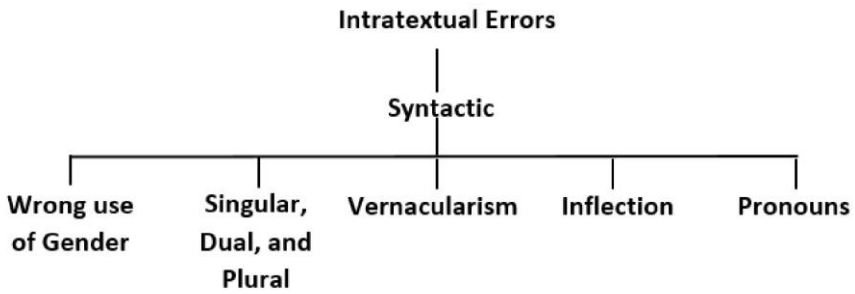

Diagram: 3.4 Syntactic Errors (Na'ja and Abu-Mighnim) Model 2012

3.2.1 Intratextual Errors

Vancura (2017) emphasizes both 'intertextual' and 'intratextual' errors are important in simultaneous interpreting. Regardless of how these two terms are employed in her study, in this context, the term intertextual refers to the measurement or comparison made between the two linguistic systems in question. On the other hand, intratextual is normally concerned with one linguistic system which could be that of either the output or the input. In either case, the parser

does not look for interrelated errors, but he sticks only to one linguistic system and attempts to detect errors pointed out there. The cornerstone of intertextual errors are the syntactic ones, which are classified into errors in the wrong use of gender; singular, dual and plural; vernacularism; inflection; and pronouns.

4. The Experiment

4.1 Introduction

Throughout direct involvement in academia, it has been noticed that there is a gap that is overlooked by researchers or tutors working at departments of translation in Iraqi universities. This gap is embodied in the errors that advanced Iraqi students of translation commit in simultaneous interpreting during lectures as well as at final exams. Therefore, the researchers decide to investigate the roots of this gap as well as try to suggest workable remedies that can improve students' performance. The data is elicited from the fourth year students of translation at Iraqi Universities. Simultaneous interpreting as a pedagogical item set by the Ministry of Higher Education and Scientific Research (MHESRDRD: 2016, p.178) is taught to students of that particular year. This fact enables the researchers to outline the study thoroughly. In the following, the model suggested above will be tested and assessed.

4.2 Pre-Assessment Conditions

Investigating outputs involves interdisciplinary processes. The researchers have to verify several related issues before choosing this topic, (to name but few), logistic materials, availability of resources, information about the subject matter, targeted audience, validity, space-temporal settings, beneficiaries, and so on. All of these remarkable phenomena have to be taken into consideration before kicking off the project. However, everything looks possible to go on in this study. Hopefully, none of these issues could undermine the project as the researchers are ready to change or amend the study plans.

4.3 Source of Data

The gap that the researchers want to bridge is related to the source of data. Many studies focus on assessment and error analysis of simultaneous interpreting (such as Cecot, 2000; Pio, 2003; Rennert, 2010; Vancura, 2017 ; Mankauskienė, and 2018 EL-Zawawy, 2019). But the fact is that data analyzed in these studies are rendered either by professional interpreters or by trainee students of different countries, not including Iraq. The direction of data collection is the

bone of this study. The outputs of fourth-year students of the Department of Translation will be used. It is believed that new facts can enrich the translation field in Iraq, as well as to stand on the proficiency of the targeted group.

4.4 Summative Assessment

Saldanha and O' Brien (2013:97) define summative assessment as knowledge gained until the end of course or term. In the same tunnel, Sawyer (2004:107) asserts that summative assessment can be adopted for many purposes; it can be manipulated in study projects and course examinations to determine learning level or teaching effectiveness. Hence, summative assessment is chosen as a primary tool in conducting this research. It is designed to test trainee students at the end of the course or final semester as Gipps (1997:17) asserts. This type of assessment is adopted by many educational programmes. It could enable the students not only to stand on their final scores but also to discover their level of proficiency. As for tutors, on the other hand, they can realize the effectiveness of the syllabus set and methods of teaching applied.

4.5 Validity and Reliability of Assessment

Sawyer (2001:107) demonstrates some test specifications that likely recommended in designing a test. Although he does not set such a format as a standard for all tests, he includes some components as a basis for a valid and reliable one. Gipps (1994:58) defines validity as "the extent to which a test measures what it was designed to measure". Saldanha and O'Brien (2013:28) state that when the findings of a researcher are real and related to the way others do in the social atmosphere, then it is a valid study. Per these definitions, the researchers attempt to validate the study by measuring the items that have been diagnosed earlier. The assessment, in general, is a valid theory: many studies use an assessment environment to target the same goals and purposes.

Reliability in research assessment is a crucial factor that researchers should pay attention to. Sawyer (2013:101) speaks of the consistency of the results when replicating a study by other examiners or researchers for the whole assessment. However, the following description of the test specification is attributed to Sawyer (2001:107-108).

4.5.1 Purpose and Role of the Test

The test is designed to provide data through which errors are diagnosed, classified, and described. Besides, there is a jury whose task is to use test content as outputs for evaluation.

Generally, the test is the core of this study as, without it, there is no possibility of examining the research hypotheses or find answers to research questions.

4.5.2 **Participants**

Fourth-year students (the academic year 2018-2019) are the participants of this study. They have been chosen because they study simultaneous interpreting as a compulsory course. This choice was made following the pedagogical agenda programmed by the Ministry of Higher Education and Scientific Research in Iraq (MHESRDRD, 2016:178). The trainee students (will interchangeably be called participants) are all from the University of Basra, College of Arts, Department of Translation. They have voluntarily agreed to take part in this study and their summative assessment could be used for research purposes. Participants were chosen randomly. Age and gender are disregarded. All participants are from morning studies. They have not been awarded (or promised) any credit for taking the test. Participants' identity is not shown; neither on transcript sheets nor recording files. This is due to their verbal request. Participants have been coded as A1, B1, C1, D1, E1, F1, G1, H1, I1, J1, K1, and L1 for the first speech, whereas A2, B2, C2, D2, E2, F2, G2, H2, I2, J2, K2, and L2 coding is used for the second speech.

4.5.3 Length and Speed of Speech

Sawyer (2001:204-205) surveys some shared opinions about text lengths prepared as a test for assessing simultaneous interpreting in English and some other languages. He states that the experiment he works on sets test length between 1208-4280 words (minimum 10 minutes, maximum 35 minutes). Nevertheless, each study has its data requirements. Wu (2010:102) uses three minutes as a test in his Ph.D. dissertation. He justifies that many examiners do not object to test length as long the test achieves the purpose of the study. He adds that some studies need to investigate some items such as "sustaining ability" and "gritting power", in these cases; the lengthy test is required to provide sufficient data. In the current study, neither the researchers nor the jury members requested a lengthy test. This fact was negotiated (verbally) with many tutors of simultaneous interpreting in four different Iraqi universities: they all agree that the suggested test length (which is about three and a half minutes) is sufficient in terms of the research limits and requirements.

On the other hand, there are various arguments on the speech rate. Mankauskiene (2018:19) identifies that the speech rate average used in her experiment is 156 words per minute (wpm). But Setton (1999:30) uses 100-120 wpm and 150-200 wpm, as the former is a normal speech

rate, while the latter average is high speech rate. Osgood (1959:34) shows that the average speech rate of English language native speakers is 152 wpm. Gile (2009:111) hypothesizes that the speech rate in interpreting, in general, is between 100 and 200 wpm. Therefore, it seems permissible to choose a speech that rates between 100 and 200 wpm, as long there is no consensus to determine a fixed speech rate among researchers and various studies in the past decades.

4.5.4 Test Situation and Setting

The test takes place in the interpreting lab of the Department of Translation, College of Arts, University of Basra. The researchers' request for using this facility was verbally approved by the person in charge of the Department. The lab infrastructure is complete. 12 desktop computers are ready to be occupied by 12 participants. A control computer wired to all other computers administrates the given orders. The control computer enables the examiner to show audio, video, or picture files on participants' screens. The volume of speech shall be adjusted by knobs there and here.

The 12 computers are situated on desks adhere to real booth conditions. Headsets are linked to computers: they enable participants to listen and speak through them as well. The software programme installed in the control computer manages all orders to participants' computers, but to start recording, each participant must click on the red recording bottom shown on his/her screen. As soon as this bottom is clicked, the software starts voice recording. This recording is normally saved in an output folder when terminating. For input presentation, the examiner uses the main computer to turn on whatever he needs to show or present, but before this is done, the examiner should make sure that the main computer screen is shown on targeted computers. After the recording is finished, participants are asked to terminate the recording, hence, there is a pause icon that must be clicked "as explained above". This process saves the outputs in the computers' internal storage.

4.5.5 Characteristics of Speeches

It is illogical to ask a mechanic to fix a TV or a driver to fly a jet. The same is true in the present context. The curriculum set by the Ministry of Higher Education and Scientific Research indicates teaching students at departments of translation different translation and interpreting materials, simultaneous interpreting is one of these. The curriculum (see MHESRDRD, 2016:178) suggests addressing specific pedagogical items and areas while teaching

simultaneous interpreting. The researchers stick to this fact. The chosen items must be adhered to what participants usually have in classrooms. Although this limits the scope of data selection, it is fair from the participants' perspective, as they are tested the items they have studied throughout the course.

The researchers have suggested five speeches as a test material. These have been presented to simultaneous interpreting tutors at the University of Wassit, University of Mustansirya, and AL-Iraqiya University. Later on, all tutors communicated with the researchers via social media to discuss the proposed choices. After that, the tutors agreed on two of the suggested speeches, i.e. they did not have an objection if any of these speeches are used as test material. Hence; both of these were chosen as the results obtained may enrich the study.

The first approved speech belongs to Angelina Jolie and delivered in the United Nations. It covers the suffering of Syrian refugees and the conflict started there in the last decade. The second speech belongs to the former president of the United States Barak Obama. It declares an official military withdrawal from Iraq and the future of the Iraqi security forces. Table (4.1) presents some details on the two speeches:

Speech of	Event	Date of Speech	Subject Matter and text type	Flow rate of Speech	Identity of Speaker	Length (by minute)	Type of presentation
Angelina Jolie	United Nations Security Council	24/Apr/2015	Humanitarian (Syrian Refugees)/Informative	147 wpm Total: 506 words	American	3.27	Audio
Barak Obama	Oval Office, White House	31/Aug/2010	Political (US Army Withdrawing from Iraq)/Informative	142 wpm Total: 495 words	American	3.30	Audio

Both speeches are uttered by American native speakers. Jolie's speech is a request to help the Syrian people and refugees. The speech bears a message to the members of the United Nations regarding the Syrian crisis, i.e. the miserable situation and mass deaths there. It surveys some disappointing stories about people losing their family members and images of destruction all over the country. The second speech addresses the American people, Iraqi people, and the whole world. It

declares an official military plan for withdrawal from Iraq. The speech represents a healing theme. The president appreciates the sacrifices of his forces as well as their families. Also, he supports the Iraqi people and their government. He also speaks about the future of Iraqi security forces and their mission to secure their country.

Generally speaking, both speakers seem very familiar to the participants. Angelina Jolie is an American actress whose movies are usually broadcasted to the Middle East with no restrictions. Barack Obama is the former President of the United States who used to address this region from time to time. Moreover, the vocabulary employed in both speeches is not complex, because it is taken from the humanitarian and political sectors with which participants are acquainted. It is expected that the participants are equipped with all required information and encyclopedic knowledge of such contexts.

Technically speaking, both speeches are under 150 WPM, each speech counts around 500 words. The flow rate of speakers seems very close in either speech. In Jolie's speech, the flow rate is 147 WPM, while it is 142 WPM in Obama's speech. Though there are few metaphoric expressions, the speeches are written in ordinary language which is potentially taught in media translation, a compulsory course the trainees study at the time of the test.

4.5.6 Directionality and Mode of Interpreting

The assessment of this study is neither comparative nor contrastive. It helps to figure out common errors committed by Iraqi trainee students. It is completely an error analysis study. The directionality is from English, i.e. language 2 (Lang. 2) into Arabic, i.e. language 1 (Lang. 1). The study believes that trainee students' Lang. 1 proficiency has never been tested before so the directionality enables them to test this hypostudy.

Both speeches are written to be read. But for research purposes, they have been accounted as spontaneous speech said lively by speakers. For this reason, each group of participants has been informed briefly about the speaker as well as given some hints on the theme. No dictionaries or other resources would be consulted.

4.5.7 Test Administration

The researchers themselves administered the test. they ensured that all resources are available on test day. they informed the trainee interpreters a couple of weeks in advance. Because of the students' busy schedule as well as the summative test environments, the test was appointed shortly before the final examinations, where the students have completed their

semesters which is recommended by many scholars such as Gipps (1997:17), Sawyer (2004:107), and Saldanha and O'Brien (2013:97).

Before the students occupy booth seats, they have signed a letter to use their outputs for research purposes. They have seated randomly on the specified PCs. No malfunction has been reported when turning on the PCs. Each trainee has set the headset properly. A sample recording was played, so the participants could adjust the volume as per their desire. After a while, a brief of the speech was given. One minute of warming up was played to get the participants into the interpreting mood. The warming up sample is not part of the test material. Later on, the participants have been informed to be ready; Angelina Jolie's speech was played. After 03:27 minutes, the speech finished. The participants terminated the recording and left the lab. By doing this, the output is saved automatically in each PC. All outputs are saved successfully. A memory stick was used to transfer the recordings instantly into the researchers' PC. The second group was then invited to occupy the seats. The same procedures have been followed as in the first test. But the speech played belongs to Barack Obama. No credit has been given to the participants through this process.

4.6 Transcription

After all, recordings are uploaded from the lab. They have been coded randomly and saved for analysis and evaluation. Two-word formats files have been initiated, one for those who interpreted Angelina Jolie's speech, and the other for those who interpreted Barak Obama's speech. The process of transcribing the outputs for each participant started. It was done manually, i.e. no specific software programme was used. All outputs were typed including hesitations, stops, silences, and so on. The typed speeches were divided into 17 sentences. Each sentence was highlighted by a certain colour for easiness while conducting quantitative analysis. The first speech consists of 154 utterances, while the second consists of 196 utterances. The difference in the number of utterances does not make any variation as there are mean readings that will be calculated. See Table (4.2) for the abbreviations used.

Table (4.2) Abbreviations Used in the Transcription Process

Abbreviation	Full Expression
UFP 00S	Unfilled Pause 00 Second

FPH	Filled Pause Hesitation
FPR	Filled Pause Repetition
FPC	Filled Pause Correction
FPF	Filled Pause False-Start

4.7 Transcription of Source speeches

The transcripts of the source speeches used are obtained from two sources; Angelina Jolie's was taken from the United Nations website and Obama's was taken from CBC News' website. To date, both speeches are still available on these websites.

4.8 The Unit of Translation/ Interpreting

The scope of the unit of translation is one of the most controversial issues in translation/ interpreting studies. Many arguments have been posed during the last three decades with no consensus on what composes this unit. For example, Baker (1992:11) describes the word as the smallest unit of translation. But Newmark (1988:31) and Munday (2012:86) criticize the usage of a word as the unit of translation. They support their criticism by saying that there are many words in various languages that have inadequate referents in the target text, i.e. the word in a language has, as Munday assumes, different semantic spaces in the target language. Besides, Newmark (1988: 55) prevails that:

The text cannot be the UT(i.e. unit of translation) in the 'narrow' sense defined by Vinay and Darbelnet. That would be chaos- The largest quantity of translation in a text is done at the level of the word, the lexical unit, the collocation, the group, the clause and the sentence - rarely the paragraph, never the text - probably in that order. The text can rather be described as the ultimate court of appeal; every stretch at every level of the translation has to conform to the unity of the text, its integrating properties, what Delisle calls its 'textual organicity\ if such exists (often it does not).

On the other hand, Aissi (1989:111) proposes that the unit of translation ranges from word-level all the way to sentence level. He claims that all components at the word, phrase, clause, or

sentence level can be segmented and taken as a unit of translation. However, in this study, the phrase is used as the smallest unit of translation/interpreting.

4.9 Translating the Transcripts of the Source Speeches

Translating the transcripts of the source speeches is done by a professional translator. He was briefed that the transcripts are of spontaneous speeches delivered by two native Americans. The recordings of the source speeches were also provided. After translating them, the translator has approved and signed each translated transcript.

4.10 Data Analysis

First of all, it is worth mentioning that the adoption of a mixed-method as an analysis tool is borrowed from previous studies on error assessments and classification. Therefore, it is useful here to explain the way numerals and statistics are shown in the tables, charts, or diagrams presented below. The very initial process is calculating the total errors of each participant. This is done by combining all errors each sub-type in question in one table. Below is an example of how the numbers, average, and mean reading are calculated:

Each speech is divided into 17 segments. Each segment consists of several utterances, the number of utterances that have been modified is shown in the cell next to each participant. The same counting process is repeated in the other sixteen segments. Then, all errors from the seventeen segments are combined in one main table. The second process is to obtain the average of each type of error. This is done by combining the total number of errors for all participants, then dividing this number by 12 (the number of participants), the number obtained is the general average which represents 0 degrees on the scale. The last process is to figure out the mean. To do so, the participant's total number of error is subtracted from the average. In this case, if the numeral is positive, it indicates that the participants' errors are less than the average, if negative, it indicates that the errors are higher than the average. For more detail, the equations below summarize the processes of obtaining the general average and mean readings:

$$\frac{total\ number\ of\ errors\ committed\ by\ all\ participants}{(the\ number\ of\ participants)} = General\ Average$$

Ex:

The total number of omissions in Angelina Jolie's speech is 916 errors. The total number of participants is 12, hence;

$$\frac{916}{12} = 76.33\ is\ the\ average$$

$$\text{Average} - participant's\ total\ number\ of\ error = Mean$$

Ex:

The average of errors attributed to an omission is 76.33. F1's total number of errors is 87, hence:

$$76.33 - 87 = -10.67\ is\ the\ mean\ of\ participant\ F1$$

Regardless of how difficult both speeches are, the participants have not committed that disparate number of errors. The total number of errors in Jolie's speech is 2105 errors, while it is 1968 errors in Obama's. The highest type of error committed in both speeches is omission with 43.52% in Jolie's and 47.21% in Obama's speech. On the other hand, the lowest type of error committed in Jolie's is the wrong use of pronouns which is only 0.24%, while in Obama's speech, the lowest is the error in singular, dual, and plural which constitutes 0.20%. Before demonstrating the tables, let's examine the percentages of errors in the table (4.35):

Table (4.35) Summary of the Errors Committed in Jolie's and Obama's Speeches

Type of Error	Jolie's Speech	Obama's Speech
Omission	43.52%	47.21%
Substitution	15.49%	14.84%
Addition	2.38%	1.32%
Unfilled Pause	11.97%	13.36%
Hesitation	9.93%	8.03%
Repetition	6.22%	4.22%
Correction	4.23%	5.49%
False Starts	1.00%	1.52%
Wrong Use of Gender	0.38%	0.30%
Singular, Dual and Plural	0.29%	0.20%

Vernacularism	1.33%	0.36%
Inflection	3.04%	2.74%
Pronouns	0.24%	0.41%

In Tables (4.36) and (4.37), the horizontal columns view the statistics for each participant when reading them within error types, while vertical columns view the statistics for each error type. Figures (4.39) and (4.40) show the highest and lowest error types committed within each speech.

Table (4.36) Summary of the Errors Committed in Angelina Jolie's Speech

Assessment of Trainee Students' Outputs													
Angelina Jolie's Speech													
	Intertextual Errors			Errors of Fluency					Intratextual Errors				
					Filled Pauses				Syntactic				
Part.	Omission	Substitution	Addition	Unfilled Pauses	Hesitation	Repetition	Correction	False Start	Wrong Use of Gender	Singular, Dual and Plural	Vernacularism	Inflection	Pronoun
A1	72	25	2	0	26	19	14	1	2	0	2	5	1
B1	95	33	8	32	26	8	5	1	1	1	3	8	1
C1	78	27	4	59	9	1	3	0	1	0	2	7	1
D1	53	16	1	0	26	17	9	6	1	2	1	6	0
E1	102	30	16	14	17	6	6	5	0	0	0	5	0
F1	87	18	0	51	2	4	3	0	0	1	5	2	0
G1	67	37	3	7	18	9	10	2	0	0	3	5	1
H1	81	37	1	27	19	17	9	1	1	0	2	5	1
I1	76	34	2	0	26	7	6	3	2	1	0	2	0
J1	56	30	2	4	23	16	8	2	0	0	4	6	0
K1	81	14	0	41	8	7	6	0	0	0	2	5	0
L1	68	25	11	17	9	20	10	0	0	1	4	8	0
Total	**916**	**326**	**50**	**252**	**209**	**131**	**89**	**21**	**8**	**6**	**28**	**64**	**5**
Grand Total of Errors													**2105**

Table (4.37) Summary of the Errors Committed in Barak Obama's Speech

Assessment of Trainee Students' Outputs													
Barak Obama's Speech													
	Intertextual Errors			Errors of Fluency					Intratextual Errors				
					Filled Pauses				Syntactic				
Part.	Omission	Substitution	Addition	Unfilled Pauses	Hesitation	Repetition	Correction	False Start	Wrong Use of Gender	singular, dual and plural	Vernacularism	Inflection	Pronoun
A2	72	27	3	12	7	3	9	0	0	0	1	3	1
B2	95	16	1	12	13	6	8	1	1	0	0	8	1
C2	96	16	4	59	5	7	10	2	2	1	0	5	2
D2	105	14	2	72	14	0	2	5	0	0	1	3	0
E2	78	35	4	0	13	1	9	7	0	0	0	3	0
F2	77	37	0	33	1	1	2	1	0	0	0	6	0
G2	55	31	0	0	5	12	14	1	0	0	2	7	1
H2	68	8	0	14	32	1	4	2	1	0	0	4	0
I2	95	18	4	41	11	3	8	4	0	0	0	6	1
J2	37	23	0	4	15	7	16	3	0	0	1	4	0
K2	65	38	4	0	18	31	19	1	2	3	2	3	2
L2	86	29	4	16	24	11	7	3	0	0	0	2	0
Total	**929**	**292**	**26**	**263**	**158**	**83**	**108**	**30**	**6**	**4**	**7**	**54**	**8**
Grand Total of Errors													**1968**

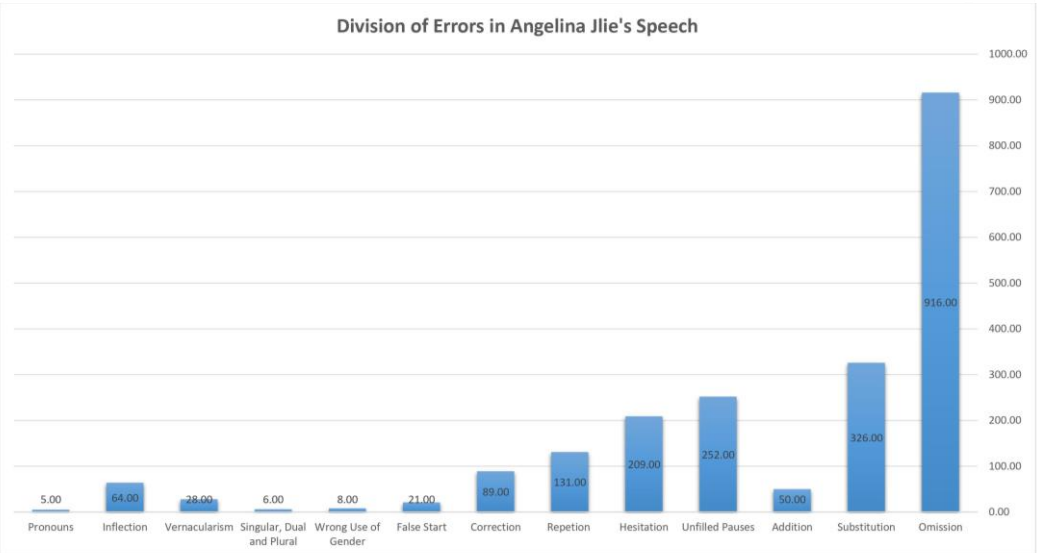

Figure (4.39) Division of Errors in Angelina Jolie's Speech

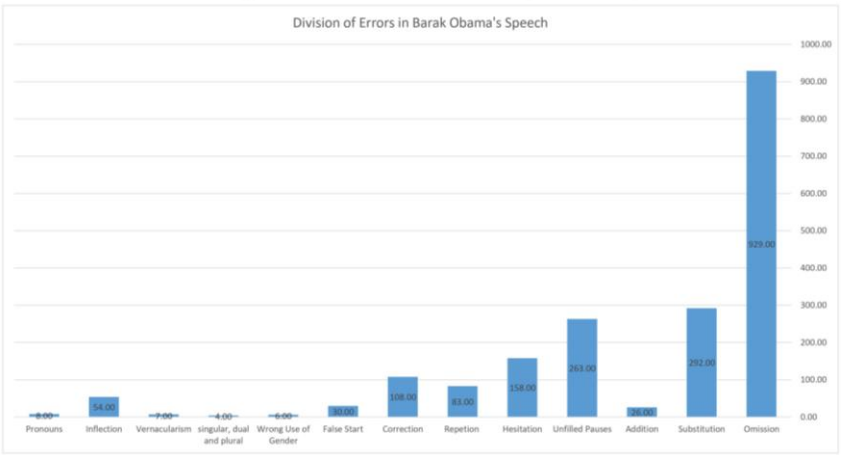

Figure (4.40) Division of Errors in Barak Obama's Speech

The examination of the results reveals that many indicators prove the reliability of the test. The most important indicator is that the percentages of committing errors are very close in the two speeches with only two variables. One of these variables is that in Jolie's speech, the percentage of committing repetition is higher than that of committing correction, while in Obama's speech, the reverse is true. Probably, whenever the speech is culturally related, the number of repetitions becomes lower, and correction becomes higher. The second variable concerns the intratextual analysis, errors in vernacularism in Jolie's speech is the second-highest error committed, while in Obama's speech, error in pronouns is the second. After examining where most vernacularism errors have been detected, the scanning reveals that Jolie's source speech includes specific problematic numbers that raised the total of errors. If these errors are excluded, the percentage of committing this error type becomes similar to those in Obama's speech.

Before moving to the next section, which is about the Jury's evaluation, the participants are lined up according to the number of errors committed. Those with fewest errors are at the top, while the ones with the highest number of errors are at the bottom accordingly. Tables (4.38) and (4.39) are the standing of both participants in both groups:

Table (4.38) Participants' Lineup in Angelina Jolie's Speech

Code of Parts.	Total Errors	Lineup
D1	138	1
J1	151	2
I1	159	3
G1	162	4
K1	164	5
A1	169	6
F1	173	7
L1	173	8
C1	192	9
E1	201	10
H1	201	11

B1	222	12

Figure 4.41 shows the lineup of the participants in Angelina Jolie's speech:

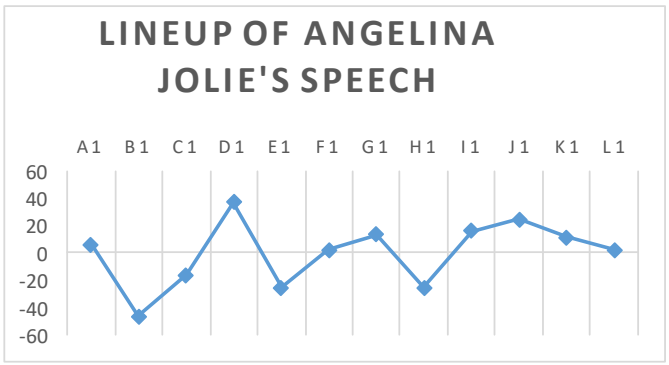

Figure (4.41) Participants Lineup in Angelina Jolie's Speech concerning Average Errors

Table 4.39 Participant's Lineup in Barak Obama's Speech

Code of Parts.	Total Errors	Lineup
J2	110	1
G2	128	2
H2	134	3
A2	138	4
E2	150	5
F2	158	6
B2	162	7
L2	182	8
K2	188	9
I2	191	10
C2	209	11
D2	218	12

Figure 4.42 shows the lineup of participants in Barak Obama's speech:

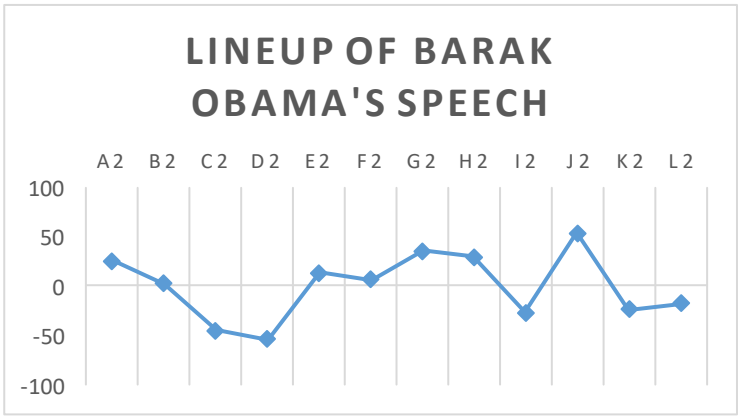

Figure (4.42) Participants Lineup in Barak Obama's Speech concerning Average Errors

4.11 Jury Evaluation

4.11.1 Introduction

Generally speaking, any jury evaluation is seen as having conclusive decisions when they are called for examining research findings. In most instances, the researchers call for such a jury to prevail something seems controversial. Or, they are called for evaluating specific results in the study to show whether the examiner was biased toward specific facets in the research or not. These two points are the most common ones where the jury is called about. Sawyer (2004:177) states that test validity and reliability could be undermined if there is no external examination of the results. In this study; the jury members are called where they would evaluate the participants' outputs and grade them as per the given scoring criteria. Then, the evaluation of each speech (Jolie's speech and Obama's speech) is juxtaposed with the results of the error analysis obtained.

4.11.2 Jury Composition

Sawyer (2004:182) states that the minimum size of the jury in interpreting the environment is constituted at least three members. He (2001:184) adds that instructors even with one year experience could participate in any evaluating team. The jury members in this study have, at least, four years of teaching experience. One of the jury members holds a Ph.D. in translation, he is the head of the department of translation in AL-Iraqiya University/Baghdad. The second one is a teacher at the University of Basra, college of Arts, Department of Translation. while the third jury member is an assistant lecturer at the Department of Translation, College of Arts, University

of Wassit. All three jury members agree that they adopt blind ratings (see Sawyer, 2004: 182) as long they work in three different cities. The researchers are the ones responsible for distributing hard copies of scoring sheets to jury members. They have briefly explained to each of them the purpose of the study, as well as the purpose of their evaluation and how significant it is to the research findings.

4.11.3 Scoring Sheet

According to the research requirements, the scoring sheet (form A) of Sawyer (2004:242) is adopted and broadened to adhere to the requirement of the current study. The scoring sheet consists of four main parts. They are posited in the below section separately.

4.11.3.1 Test, Testee, and Jury information

The top part of the scoring sheet includes test, testee, and evaluator information. It contains nine fields; (range of grading, mode of interpreting, direction, type of exam, the language, code of participants, speech no., code of grader, and the date of evaluation). Soon this is finished, the evaluator shifts to look at the criteria set in the scoring sheet.

4.11.3.2 Criteria for Scoring

Six criteria should each output be evaluated about; the first criterion is meaningfulness. The grader assesses to the extent to which the meaning is conveyed, does the meaning in the participant's rendering is the same/equivalent to the SS? The second criterion is accuracy. It tackles whether the participant has faithfulness to SS or not; does the participant render accurate information? The third criterion is clarity, it handles how clear the rendering of the participant is, it is also set to measure the presence of the participant as well. The fourth criterion concerns the coherence. As it is widely known, and as it has been experienced so far, the interpreters, in general, could repeat or correct utterances as required. Sometimes they interrupt specific utterances and start a new one incoherently. Thus, do all these phenomena affect the coherence of output? The fifth criterion is about intonation, sometimes the macro-linguistic phenomena are not controlled properly by the interpreter, how does the participant deal with such phenomena? The last criterion is about basic interpreting skills, do the participants employ such skills while interpreting, how did they do when facing challenges?

4.11.3.3Grading

The range of scoring is between (0-10). Many strategies would be employed by the evaluator to assess the participant's performance. The grader in this respect could assess how well the

participant has done the criterion in question. He might only rely on listening to the recording or take notes before type down the final score. However, as long this is the evaluator's decision, there would be no intervene made ever. The study is much concerned about the final scoring than those nuances. Final scores are grouped in one main table for each speech and presented in the research in the appropriate section.

4.11.3.4 Proofing the Scores

The last part of the scoring sheet is the signature. It is to confirm that this scoring sheet has been filled legally by the jury member specified by the letter of College of Arts, the University of Basra (numbered: 3/17/108 in 03-Feb-2020). Otherwise, the scoring sheet is considered fake and jury judgment is suspected.

4.12 Procedures

Blank scoring sheets along with recordings of participants have been distributed to each evaluator. The researchers had to travel for the three universities despite the long distances between a university and another, as long this task strengthens the research methodology. Each evaluator has been briefed about the whole study and the purpose of scoring. This has been done without giving hints about the results of the researchers. Because if given, it would draw an impression to the evaluator about participants' overall performance. This effects scoring in a way or another. The second step was to collect the draft scoring sheets from the three evaluators after the evaluation is finished. The scoring sheets have been combined in one sheet for each participant to save some space in the research volume. By doing this, the collecting of data is completed. Later on, all data have been computerized for demonstration.

4.13 Analysis of Evaluation

There are six criteria the participants have been evaluated for. As mentioned earlier, the grading is from (0-10), what is shown in the table is the total of the three jury marks. For instance, in the meaningfulness criterion, Jolie's speech; A1's total mark is (11), it is a combination of the marks of all three jury members; (jury (1): five marks, jury (2): four marks, and jury (3): two marks only).

4.13.1 Jury's Evaluation of Angelina Jolie's Speech

Table (0.00) demonstrates the total marks of the participants who interpreted Angelina Jolie's speech:

Table (4.46) Summary of Marks Granted in Angelina Jolie's Speech

Part.	Meaningfulness	Accuracy	Clarity	Coherence	Intonation	Basic Interpreting Skills	Final Score	Mean
A1	11	14	13	22	11	15	4.78	0.12
B1	9	13	10	12	16	12	4.00	-0.66
C1	7	13	17	11	20	15	4.61	-0.05
D1	13	19	11	22	12	16	5.17	0.50
E1	5	11	11	16	17	12	4.00	-0.66
F1	8	13	14	11	19	16	4.50	-0.16
G1	12	12	14	21	12	14	4.72	0.06
H1	7	17	14	17	13	14	4.56	-0.11
I1	13	13	11	19	16	15	4.83	0.17
J1	13	15	14	23	14	15	5.22	0.56
K1	7	17	16	16	18	14	4.89	0.23
L1	10	14	15	17	10	18	4.67	0.00

The table above shows that participant marks according to the jury evaluation. J1 is taking the lead with (5.22) marks. While the least mark (4.00) is the lowest, it is imputed to two participants B1 and E1. Figure (4.49) shows the total marks of this speech:

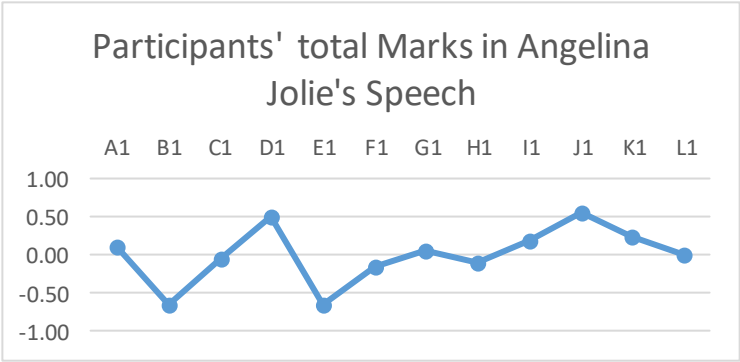

Figure (4.49) Total Marks in Angelina Jolie's Speech

4.13.2 Jury's Evaluation of Barak Obama's Speech

Table (0.00) demonstrates the total marks of the participants who interpreted Barak Obama's speech:

Table (4.46) Summary of Marks Granted in Barak Obama's Speech

Part.	Meaningfulness	Accuracy	Clarity	Coherence	Intonation	Basic Interpreting Skills	Final Score	Mean
A2	9	9	15	18	17	14	4.56	-0.14
B2	6	14	13	17	15	15	4.44	-0.25
C2	5	11	9	7	14	13	3.28	-1.42
D2	5	14	13	8	19	12	3.94	-0.75
E2	9	9	17	21	22	19	5.39	0.69
F2	9	9	17	15	19	15	4.67	-0.03
G2	14	12	19	19	11	17	5.11	0.42
H2	13	20	11	18	22	20	5.78	1.08
I2	7	13	17	13	18	14	4.56	-0.14
J2	20	19	21	22	18	23	6.83	2.14
K2	12	7	13	14	8	12	3.67	-1.03
L2	7	8	12	19	14	14	4.11	-0.58

The table above shows each participant's final score according to the jury evaluation. J2 is taking the lead with (6.83). While the least mark is (3.28), it is imputed to C2. However, figure (4.56) demonstrates the standing of participants in this speech:

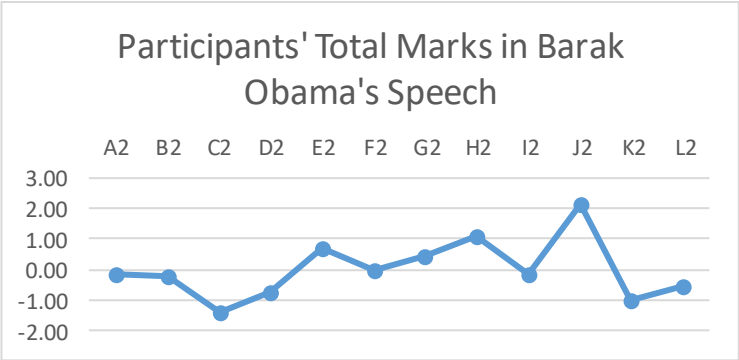

Figure (4.56) Total Marks in Barak Obama's Speech

4.14. Summary of Jury Evaluation

Within this point, all jury marks and final scores have been demonstrated, but two important statistics should be posited here. Table (4.54) shows the percentage of marks given for each criterion in each speech, while table (4.55) and table (4.56) are the lineups of participants as per their scores for each speech as well.

Table (4.54) Percentage of Marks Granted in the Speeches of Angelina Jolie and Barak Obama

Criteria	Jolie's Speech	Obama's Speech
Meaningfulness	11%	11%
Accuracy	17&	14%
Clarity	16%	18%
Coherence	21%	19%
Intonation	18%	19%
Basic Interpreting Skills	17%	19%

As shown above, meaningfulness has the lowest percentage in both speeches 11%, while coherence is the highest in Julie's speech (21%). On the other hand; coherence, intonation, and basic interpreting skills are diagnosed as having the highest percentages in Obama's speech. Figure (4.57) and (4.58) shows the division of criteria concerning jury evaluation for both speeches:

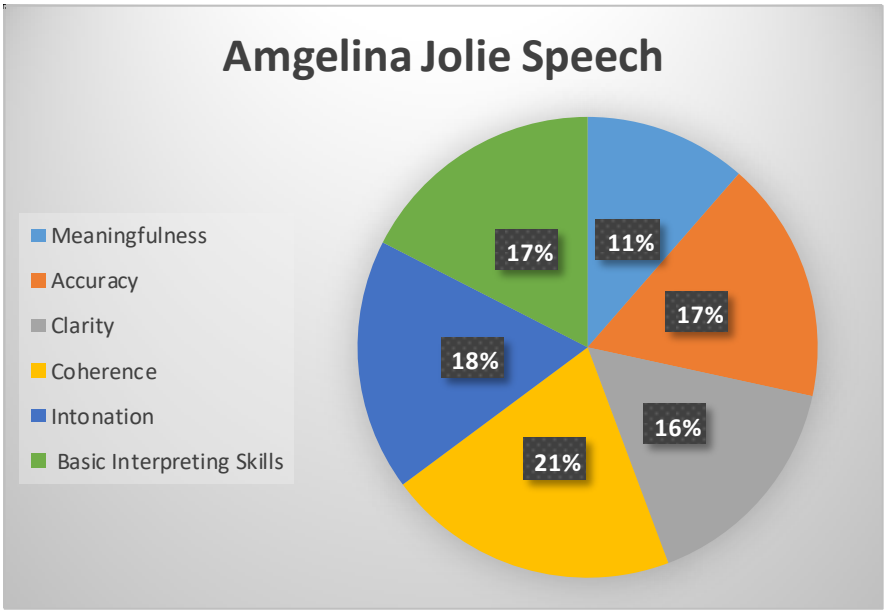

Figure (4.57) Division of Criteria in Angelina Jolie Speech

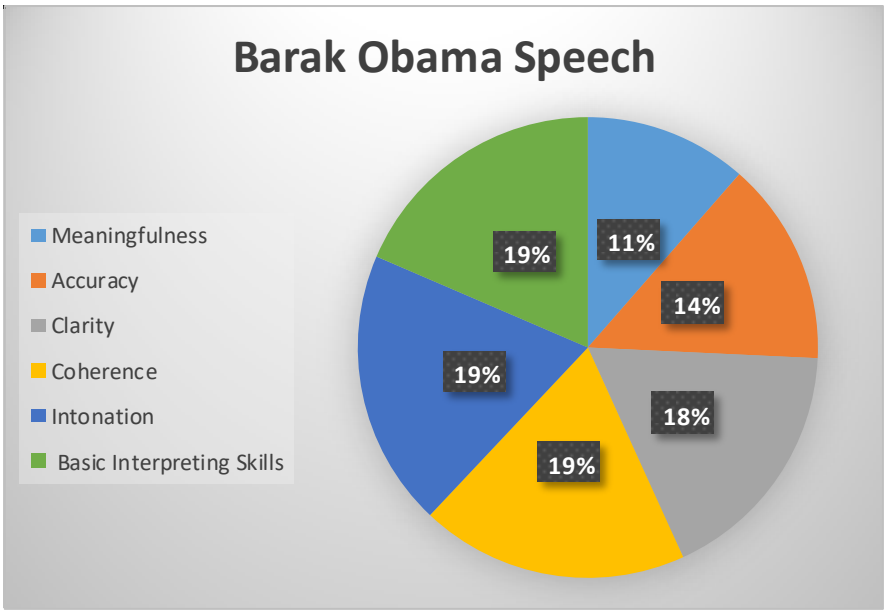

Figure (4.58) Division of Criteria in Barak Obama's Speech

Table (4.55) Participants' Lineup in Angelina Jolie's Speech According to Jury Evaluation

Code of Parts.	Final Score	Lineup
J1	5.22	1
D1	5.17	2
K1	4.89	3
I1	4.83	4
A1	4.78	5
G1	4.72	6
L1	4.67	7
C1	4.61	8

H1	4.56	9
F1	4.50	10
B1	4.00	11
E1	4.00	12

Table (4.56) Participants' Lineup in Barak Obama's Speech According to Jury Evaluation

Code of Parts.	Final Score	Lineup
J2	6.83	1
H2	5.78	2
E2	5.39	3
G2	5.11	4
F2	4.67	5
A2	4.56	6
I2	4.56	7
B2	4.44	8
L2	4.11	9
D2	3.94	10
K2	3.67	11
C2	3.28	12

These two lineups along with the last two figures are the gist. They simply reflect how the jury members, who are tutors in departments of translation, perceive the performance of their students when they interpret simultaneously.

4.15 Convergence of Error Analysis Results and Jury Evaluation Results

4.15.1 introduction

This part juxtaposes the results of the error analysis obtained along with the results of the jury evaluation. The purpose of this juxtaposition, first, is to check the closeness between the two investigations. Soon this is positively or negatively confirmed, it decides the extent to which the assessment is reliable. Subsequently, it helps to reveal many important facts and answers for the research hypotheses and questions. To do this so, the data obtained so far from error analysis and jury evaluation are inputted in the SPSS software program for measuring this convergence.

4.15.2 The Statistical Description

The idea of the SPSS statistical test is based on the calculation of the ratio of deviation of any average among the statistical distribution averages to the standard error. In this context, the SPSS test which is referred to as (t), was used to compare the two independent samples. Four conditions should exist for proper SPSS calculation: these conditions are:

- The variance must be measured quantitatively.
- The variance must follow the normal distribution.
- The observation must be independent.
- Samples must be chosen randomly.

The test (Shapiro- Wilk) has been used to examine if the observations have been equally distributed to the research samples. Table (4.57) displays that the variances follow the normal distribution as long the statistic of (Shapiro-Wilk) reached (0.904) at sig. level (0.026), this is seen ≤ as sig. level (0.05) with a degree of freedom (df=24) which refers to the acceptance of the hypostudy which reveals that variances follow the normal distribution.

Table (4.57) Shapiro-Wilk Test of Normality

Tests of Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
test	.904	24	.026
a. Lilliefors Significance Correction			

4.15.2.1 Convergence of Angelina Jolie's Speech

The hypostudy is described after the distribution of variances has been acknowledged. The hypostudy is:

- Null Hypostudy $A = \pi r^2$: there is sig. convergence between the results of the average of the researchers and that of the jury.
- Alternative hypostudy H_1: there is no sig. convergence between the results of the average of the researchers and that of the jury.

The employment of the appropriate hypotheses, reveals that (t) is (10.902) at sig. level (0.000) which is ≤ from (0.05). The degree of freedom is (22). There is no Mean Difference as long both of them are (1.73843) with Std. Error Difference = 0.15945. This result indicates the acceptance of the Null hypostudy with 95% as a Confidence Interval of the Difference. Table (4.56) demonstrates the results much clearly:

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
test	Researchers Assessment	10.902	22	.000	1.73843	.15945	-2.06911	1.40774
	Jury Evaluation	10.902	21.930	.000	1.73843	.15945	-2.06917	1.40768

4.15.2.2 Convergence of Barak Obama's Speech

The hypostudy is described after the distribution of variances has been acknowledged. The hypostudy is:

- Null Hypostudy H_0: there is sig. convergence between the results of the average of the researchers and that of the jury.
- alternative hypostudy H_1: there is no sig. convergence between the results of the average of the researchers and that of the jury.

The employment of the appropriate hypotheses, reveals that (t) is (6.028) at sig. level (0.000) which is ≤ from (0.05). The degree of freedom is (22). Also, there is no Mean Difference as long both of them are (1.96111) with Std. Error Difference = 0.32533. This result indicates the acceptance of the Null hypostudy with 95% as a Confidence Interval of the Difference. Table (4.59) demonstrates the results much clearly:

Table (4.59) Independent Samples Test of Barak Obama's Speech

		t-test for Equality of Means						
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
							Lower	Upper
test	Researchers Assessment	6.028	22	.000	1.96111	.32533	-2.63580	1.28642
	Jury Evaluation	6.028	17.611	.000	1.96111	.32533	-2.64568	1.27654

4.16 Discussion

Generally speaking, the experiment principally works to facilitate and produce a reliable test that would be the backbone of the assessment. For this reason, two speeches with very parallel characteristics have been tested. Sawyer (2004:101) Bachman (1990:174) state that retesting the samples of the same population on the parallel test produce reliable results. As has been demonstrated (see, table 4.37, 4.38 and figures 4.37, 4.38), the same population was tested with parallel materials and the same facts have been obtained. Additionally, the hierarchy of errors resulted is correlated in both speeches. On the other hand, the jury members have very convergent results to that obtained from the error analysis. These pieces of evidence reflect another facet of test reliability, as long the test is examined by two different groups of people

independently. Hence, test reliability is confirmed, and the study follows a very reliable test in these advancements. This leads, finally, to fair assessment.

Within the error analysis, the examinations reveal that the participants committed a tremendous number of errors in all three intertextual level categories, but with disparate amounts. The assessment shows that the task seems very problematic, the outputs produced by the participants are poor. The participants, as it could be noticed, don't produce very reliable outputs. There might be many factors that effected on their general performance (here, the researchers are not in a position to suggest specific instruction or remedies as long the goal of this study is assessing the trainee students while interpreting simultaneously). However, there are huge numbers of omission and substitution in the renderings, with a fewer number of additions. The hugest number of errors (omission and substation) effect on the meaning being conveyed. As it has been noticed, there is no perfect rendering at all. This fact might seem very disappointing to the people in charge of the training of the students, but that what it has been experimentally proved.

The highest and more problematic error that has been observed more likely in the disfluency level is the silent pauses, the participants have very lengthy pauses recorded. Perhaps, these pauses are committed due to the clear failure of; the difficulty of comprehension, searching for suitable equivalents, or appropriate structures in TL. Also, these pauses might be attributed to the time of being simultaneously interpreting. The study shows that the participants have lengthy pauses in the last minutes of the exam more often. This is an indication that the participants are not capable to interpret for long periods. Other pauses are also committed with the fewer amount, fillers like a', um, repetition, and correction have been recorded as well. Although these types of pauses are seen as useful strategies adopted by the interpreter to think about specific term or structure, or to correct something said erroneously, but they are still seen as time-wasting. The interpreter should logically realize that while the interpreter is under pressure and trying to find out new structures or terms, it must be reminded that the speaker is keeping producing utterances at the same time, hence, losing lots of information said by the speaker will be missed by the interpreter. The last category assessed in the disfluencies is a false start. The study demonstrates the many occasions where the participants interrupt their rendering, and start a new segment, then, the interrupted segment is incompletely left out.

Within the intratextual category, Although the participants interpreted into their A language, they have also committed errors. Subsequently, these errors don't represent that much of errors as if comparing them to the numbers of errors in the intertextual and errors of disfluency. Two syntactic errors seem so common as long they have been committed more than usual; these two syntactic errors are; inflection and vernacularism. These two errors have been committed more often when compared to the other three subcategories. The study reveals that inflection is not highly mastered by the tested population. There is a considerable number of errors committed in both speeches. The second-high type committed is the vernacularism. It is shown that the renderings include a high number of vernacular terms. Such errors indicate that the students need to speak the more formal language than they currently do. Other intratextual errors (Wrong use of gender), (singular, dual, and plural) and (pronouns) might be attributed to the pressure of simultaneity the participants undergo because they have not been commonly committed.

As far as the jury's evaluation is concerned, meaningfulness is granted the lowest scoring. While the criterion coherence has the highest evaluation in turn. In this case, the jury believes that the participants produce coherent, but, less meaningful outputs than the meaning implemented in the inputs. The other significant fact that has been revealed; the participants don't produce accurate renderings. Although this fact doesn't seem positive for the participants, it refers to the idea that they cope and follow the suitable strategies very well when they face difficulties. In this respect, the students seem equipped with good interpreting skills despite the low scoring in clarity. There is a reversible relationship between clarity and basic interpreting skills, the interpreters, in general, make some corrections and repetitions to produce more reliable interpreting. This is understood that the student employs the skills acquired properly. But when it comes to clarity, these corrections and repetitions are seen errors and time-wasting.

The other fact that has been uncovered by the jury is that the participants master the tone of the speaker very much. The intonation produced by them is close to that of the speaker as the scoring tells. Finally, the jury's evaluation provides indications to tackle information and judgement of two criteria (basic interpreting skills, and intonation) that couldn't be investigated by (ASIOTI). The role of the jury is essential for the accomplishment of this assessment.

5. Conclusion, Recommendation, and Suggestions for Further Research

5.1 Conclusion

Throughout the aforementioned assessment, it has been found that:

1. None of the trainee interpreters at the Departments of Translation in Iraq provide an error-free rendering. Consequently, the meaning that the source language utterances convey has been variably distorted in the renderings.
2. The errors are committed on three main levels; the intertextual, disfluency, and intratextual. On the intertextual level, there are very influential amounts of improper omissions, substitutions, and additions. The highest error on the disfluency level is the silent pauses while on the intratextual level, inflection (**التصريف**) and, to a limited extent, vernacularism (**العامية**) are the most problematic.
3. No successful interpreting strategy has been adopted by the trainee interpreters to decrease the number of errors committed. This could be attributed to the trainee's incompetence in English and Arabic as well as the poor training they underwent.
4. Neither the trainers nor the curriculum designers seem to be aware of the quality and quantity of these errors. This reveals the failure in the training strategic planning adopted by the Ministry of Higher Education and Scientific Research, exemplified by the Department of Translation/ University of Basrah.
5. As there has been no systematic nor comprehensive framework to assess trainee interpreters' outputs in the Arab world, nor there is transparency in terms of criteria used to assess interpreters' outputs or their performance in general, the eclectic model suggested by the present study provides workable and valid mechanisms through which errors and their possible causes could be empirically diagnosed. The dependence of assessment on both the outlined error analysis and the jury evaluation is insightful.

5.2 Recommendations

The study comes up with the following recommendations:

1. The tutors of simultaneous interpreting are invited to design their curricula based on the errors they outline relying on the suggested model.
2. A diagnostic assessment should be conducted at the beginning of the academic term or semester. This assessment should be designed to diagnose the flow rate of participants in English and Arabic. The average flow rate in either language is helpful to design the curricula.

3. The trainee interpreters should undergo further in-house training, such as shadowing, to improve their interpreting skills.
4. Trainee interpreters must be transparently made aware of the criteria their trainers use in assessing their performance; and the criteria must be inspired by the real-life professional practice of interpreters.

5.3 Suggestions for Further Research

Many gaps have not been bridged yet. These include:

1. Assessment of trainee simultaneous interpreters' into B language outputs.
2. Assessment of Iraqi professional simultaneous interpreters' outputs while rendering into their A or B.
3. Measurement of the lag time consumed by Arab interpreters when they interpret into their A or B language.
4. Investigate trainee interpreters' lengthy pauses in simultaneous interpreting.

Bibliography:

Aaron, Abdul-Salam (1988) ***Book of Sibawayh Abu Bishr Amr ibn Qanbar.*** **Cairo, AL-Khanji Library.**

Abdalla, Mohamed. (2018) ***The Influence of Translation on the Arabic Language: English Idioms in Arabic Satellite TV Stations.*** **Newcastle, Cambridge Scholar Publishing.**

Abdulazeez, Abdulazeem (2015) ***Analysis of written errors in Gender of third-level students at the Institute of Teaching Arabic to non-Arabic Speakers at the Islamic University of Madinah.*** **Unpublished MA Thesis. Islamic University of Madinah. Available from: https://www.academia.edu/attachments/36503823/download_file?st=MTU4MzUzNjQzMiwzNy4yMzYuMjMuMTMsOTQwNjIwMDI%3D&s=swp-toolbar. {Accessed: 23-Sep-2019}.**

Abdullah, Mahmood (1999) ***The Effects of Definite and Indefinite Phenomena in the Linguistic Context.*** **Unpublished MA Thesis. University of Al–Albeit, Available from: http://mohamedrabeea.net/library/pdf/ba372ae7-d0e6-4e66-930e-1385f98cafa1.pdf. {accessed 28-March-2019}**

Abu-Chacra, Faruk (2007) ***Arabic: An Essential Grammar.*** **London, Routledge Taylor and Francis Group.**

Aissi, Layachi (1987) ***"An Analytical Study of the Process of Translation: with Special Reference to English / Arabic".*** **Unpublished PhD Dissertation, University of Salford, Available from: https://dokumen.tips/documents/english-arabic-study.html. {Accessed: 25-Apr-2019}**

AL-Amayra E. and AL-Amayra H. (2013) Comments on Pronouns: A Comparative Study. ***IUG Journal of Humanities Research.*** **21 (1), 43-72.**

AL-Jarf, Reima (2007) SVO Word Order Errors in English-Arabic Translation. *Meta: Journal des traducteurs*, 52 (2), 299-308.

AlKilabi, Ahmed (2015) Assessing University Students' Aptitude for Simultaneous Interpretation. In: Izwaini, Sattar (ed.) *Papers in Translation Studies*. London, Cambridge Scholar Publishing.

Al-Qinai, Jamal (2010) Training Tools for Translators and Interpreters. *Pan-Pacific of Applied Linguistics*, 14 (2), 121-139.

AL-Shinbiry, Hamid (2004) *Phonetic System of Arabic Language*: Applied and Descriptive Study. Cairo, Arabic Language Center.

AL-Sulaiman, Misbah, and Alsinjari Ahmed (2018) Problems of Translating Definite and Indefinite Articles from English into Arabic. *International Journal of English Literature and Social Sciences (IJELS)*. 3 (6), 1076-1131.

Altman, Janet (1994) Error Analysis in the Teaching of Simultaneous Interpretation: A Pilot Study. In: Lambert S. and Moser-Mercer B. (eds.) *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Amsterdam, John Benjamins Publishing Company, pp. 25-38

Amin, Aras (2017) The Frequency of Morpho-Syntactic Errors by Kurdish EFL Learners. *Journal of Humanities and Social Science*, 22 (3), 68-70.

Amin, Karimnia. (2011) Waddington's Model of Translation Quality Assessment: A Critical Inquiry. *Elixir Ling and Trans.*, 40 (1), 5219-5224.

Ammar, Mahmood (2001) Feminine and Masculine in Arabic: Modern Usage. *The Journal of Jordan Academy for Arabic*. 61 (1), 105-134.

Angelelli, Claudia, and Jacobson, Holly (eds.) (2001) *Testing and Assessment in Translation and Interpreting Studies*. Kent, John Benjamins Publishing Company.

Armstrong, Scott (1985) *Long Range Forecasting*, New York: Wiley.

Aydin, Feriduddin (1997) *Tenses in Arabic*. Istanbul, Al-Ibar Publishing.

Aziz, Yowell (1989) *A Contrastive Grammar of English and Arabic*. Mosul, Ministry of Higher Education and Scientific Research, University of Mosul.

Bachman, Lyle (1990) *Fundamental Considerations in Language Testing*, New York, Oxford University Press.

Baker, Mona, and Saldanha, Gabriela (eds.) (1998) *Routledge Encyclopedia of Translation Studies*, 2nd edition, New York, Routledge Taylor and Francis Group.

Baker, Mona. (1992) *In Other Words: A Coursebook on Translation*. London, Routledge Taylor and Francis Group.

Barik, Henri (1971) A Description of Various Types of Omissions, Additions, and Errors of Translation Encountered in Simultaneous Interpretation. *Meta*: *Journal des traducteurs*, 16 (4), 199-210.

Besien, Fred V. (1999) Anticipation in Simultaneous Interpretation *Meta: Journal des traducteurs*, 44 (2), 250-259.

Brenes, Cesar (2017) Observing Structures' Syntactic Errors and the Perception Towards Writing in the Composition Course. *Kanina*, 1 (1), 109-130.

Carrove, Mariona (1999) *Towards a Theory of Translation Pedagogy*. Unpublished PhD Dissertation, University of Lleida. Available from: https://www.tdx.cat/bitstream/handle/10803/8115/Tmsc1de3.pdf?sequence=1&isAllowed=y. {Accessed: 01-April-2019}

CBS News (2010) *Obama's Iraq Speech: Text*. Available from: https://www.cbsnews.com/news/obamas-iraq-speech-text/. {Accessed: 07-July-2019}.

Chen, Zhongmei (2010) Simultaneous Interpreting: Principles and Training." *Journal of Language Teaching and Research*. 1 (5), 714-716.

Christoffels, Ingrid and Groot, Annette (2004) Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition*. 7 (3) 227-240.

Cokely, Dennis (1986) The Effects of Lag Time on Interpreter Errors. *Sign Language Studies*. 53 (1), 341-375.

Corder, S. P. (1981) *Error Analysis and Interlanguage*. Oxford, Oxford University Press.

Crystal David, and Davy Derek (1969) *Investigating English Style*. New York, Longman Group UK Limited.

Dewl, Haru (2015) *Comparing Two Translation Assessment Models: Correlating Student Revision and Perspectives.* Unpublished PhD Dissertation. Kent State University. Available from: https://etd.ohiolink.edu/!etd.send_file?accession=kent1448504394&disposition=attachment. {Accessed: 28-Oct-2019}

Diaz-Galaz, Stephanie, et al. (2015) The Role of Advance Preparation in Simultaneous Interpreting: A Comparison of Professional Interpreters and Interpreting Students." *Interpreting*, 17 (1), 1-25.

Ding, Yan (2017) Using Propositional Analysis to Assess Interpreting Quality. *International Journal of Interpreter Education,* 9 (1), 17-39.

Drugan, Joanna (2013) *Quality in Professional Translation: Assessment and Improvement*. London, Bloomsbury.

Duez, Danielle (1982) Silent and Non-Silent Pauses in Three Speech Styles. *Language and Speech,* 25 (1) 11-28.

Erzurum, Ozum (2019) Interediting: Revealing the Daul Role of Staff Interpreters on Turkish News Channels. *Translogos,* 2 (1), 68-84.

Esmail, Whaj. (2017) Applying Errors Analysis in Teaching Translation. *AL-Ma'amoon College Journal*, 29 (1), 175-189.

Fablo, Caterina (1998) Analyse Des Erreurs En Interprétation Simultanée. *The Interpreters' Newsletter.* In: Gran, Luara, and Snelling David (eds.) Scuola Superiore di Lingue Moderne, 101-115.

Fengling, Liu (2017) A Comparative Study of Nida and Newmark's Translation Theories. *International Journal of Liberal Arts and Social Science*. 8 (8), 31-39.

Fugen, Christian, et al (2007) Simultaneous Translation of Lectures and Speeches. *Mach Translat*. 21 (1), 209-252.

Gabrea, M. and O'Shaughnessy, D. (2000) Detection of Filled Pauses In Spontaneous Conversational Speech. In: *6th International Conference on Spoken Language Processing*. Beijing. INTERSPEECH.

Gernove, Ghalley (2004) *Inference and Anticipation in Simultaneous Interpreting*. Amsterdam, John Benjamins Publishing Company.

Gerver, David, and H. Wallace Sinaiko (eds.) (1978) *Language Interpretation and Communication*. New York, Plenum Press.

Gerver, David, et al. (1989) Selection Tests for Trainee Conference Interpreters" *Meta: Journal des traducteurs.* 34 (4), 724-735.

Gile, Daniel (1999) Testing the Effort Models' Tightrope Hypothesis in Simultaneous Interpreting: A Contribution. *Hermes, journal of Linguistics*, (23) 153-172.

Gile, Daniel (2018) Simultaneous Interpreting. In: Chan, Sin-wai (ed.) An Encyclopedia of Practical Translation and Interpreting. Hong Kong, The Chinese University Press, pp 531-561.

Gile, Daniel. (2009) *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam, John Benjamins Publishing Company.

Gipps, Carolina (1994) *Beyond Testing: Towards a Theory of Educational Assessment*. London, The Falmer Press.

Goldman-Eisler, Frieda. "A Comparative Study of Two Hesitation Phenomena." *Language and Speech*, vol. 4, no.1, 1961, pp. 26-18.

Gregorio, De Gregoris (2014) The limits of Expectations vs. Assessment Questionnaire-based Surveys on Simultaneous Interpreting Quality: The Need for a Gestaltic Model of Perception. *International Journal of Translation*, (16), 57-87.

Hatim, Basil, and Ian Mason (1997) *The Translator as Communicator*. London, Routledge.

House, Juliana (1977) A Model for Assessing Translation Quality. *Meta Journal des traducteurs*. 22 (2) 103-109.

House, Juliana (2015) *Translation Quality Assessment: Past and Present*. London, Routledge Taylor and Francis Group.

Husein, Sahya (2014) *Error Analysis on Machine Translation: The Study on Indonesian Translation into Arabi Using Google Translate and the Effect on its Meaning*. Unpublished MA Thesis. University of Maulana Malik Ibrahim Malang. Available from: http://etheses.uin-malang.ac.id/8273/1/12720060.pdf {Accessed 28-June-2019}.

Imbarek, Louisa (2010) *Students' Errors in Translating from Arabic into French: Fourth Year as Sample.* Unpublished MA Thesis. Oran University. Available from: https://theses.univ-oran1.dz/document/THA2434.pdf. {Accessed: 21-Aug-2019}.

Isfoor, Mohammed (2007) The Impact of Translation on Arabic. *The Journal of AL-Sharjah University for Legalization and Humanitarian Sciences,* 4 (2), 195-216.

James, Carl (1980) *Contrastive Analysis*. London, Longman.

Jasim, Jasim (2010) Error Analysis Theory of Traditional Arabic. *The Journal of the Jordan Academy of Arabic*. 79 (1), 151-210.

Juma'a, Imad (2006) *Grammar of Arabic: Syntax and simple Morphology*. Jordan, Dar AL-Nafa'is Publishing Company.

Khalifa, Ariyaf (2011) *Syntactic Structural Transformation Between Masculine and Feminine Gender in Similar Verses in The Holly Quran*. Unpublished MA Thesis. Middle-East University. Available from: http://mohamedrabeea.net/library/pdf/e6838d8b-c145-48f5-ba7f-e07b3df3ea4b.pdf. {Accessed: 24-Aug-2019}.

Khalil, Aziz (2010) *Contrastive Grammar of English and Arabic*. Amman, Jordan Book Center.

Khider, Mohammed (1982) *Simplified Inflection in Arabic Grammar*, Jordan, AL-Manar Library.

Korpal, Pawel (2012) Omission in Simultaneous Interpreting as a Deliberate Act. In: *Translation Research Projects 4*, Pym, Anthony and Orrego-Carmona (eds.) Tarragona, Intercultural Studies Group, pp. 103-111.

Krevelj, Stela, and Renata Geld (eds.) (2017) *UZRT 2016 Empirical Studies in Applied Linguistics*. Zagreb, FF Press.

Kumar, Chinta (2013) The Eclectic Method: Theory and Its Application to the Learning of English. *International Journal of Scientific and Research Publications*. 3 (6), pp.1-4.

Lambert, Sylvie, and Barbara Moser-Mercer, editors. *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. John Benjamin Publishing Company, 1994.

Lee, Zi-ying (2015) *The Reflection and Self-Assessment of Student Interpreters through Logbooks: A Case Study.* Unpublished PhD Dissertation. Heriot-Watt University. Available from: https://pdfs.semanticscholar.org/ee8f/b51af8cc8f11bdf3dc7cff57c6cf1a2072e5.pdf. {Accessed: 23-Oct-2019}

Lennon, Paul (2008) Contrastive Analysis, Error Analysis, Interlanguage1. *Bielefeld Introduction to Applied Linguistics*. Gramley S. and Gramley V. (eds.) Bielfeld, Aisthesis, pp. 51-60.

Li, Xiangdong (2015) Putting Interpreting Strategies in their Place: Justifications for Teaching Strategies in Interpreting Training. *Babel*. 61 (2) 170-192.

Limongi, Eliana (2000) *The Interdependence of Extratextual and Intratextual Factors in Translated Texts: A Sample Analysis of Ícaro Brasil, Varig's Bilingual in-Flight Magazine.* Unpublished MA Thesis,

University of Federal De Santa Catarina. Available from: https://core.ac.uk/download/pdf/30360021.pdf. {16-Dec-2019}.

Maclay, Howard, and Osgood Charles (1959) Hesitation Phenomena in Spontaneous English Speech." *Word.*, 15 (1), 19-44.

Mahasneh, Anjad (2015) *Translation Curriculum Evaluation in Jordan: A New Perspective*. Germany, LAP Lambert Academic Publishing.

Mankauskiene, Dalia (2018) *Problem Triggers in Simultaneous Interpreting from English into Lithuanian*. Unpublished PhD Dissertation, Vilnius University. Available from: https://www.academia.edu/36698338/Mankauskiene_summary_final. {Accessed: 02-Sep-2019}.

Marcel, Thelen (2008) Translation Quality Assessment or Quality Management and Quality Control of Translation. *Translation and Meaning*. 8 (1) 411-424.

Marshal, Bethan (2011) *Testing English: Formative and Summative Assessment Approaches to English Assessment.* London, Continuum.

Melis Nicole, and Amparo Albir (2001) Assessment in Translation Studies: Research Needs." *Meta Journal des traducteurs.* 46 (2), 272-287.

Michela Cecot (2011) Pauses in Simultaneous Interpretation: A Contrastive Analysis of Professional Interpreters' Performances. *The Interpreters' Newsletter.* 11 (1), 63-85.

Milcu, Marilena. (2012) Evaluation and Self-evaluation in Simultaneous Translation: Assessment Methods. *Procedia – Social and Behavioral Sciences*. 46 (1), 4253-4257.

Ministry of Higher Education and Scientific Research (2016) *English Language Curriculum Development and Updating Project in Iraqi Universities*. Baghdad, Ministry of Higher Education and Scientific Research: Department of Research and Development,

Mohammed, Hala (2017) Common English Syntactic and Pronunciation-related Errors Made by Trainee Simultaneous Interpreters. *AL-Ustath*. 1 (221) 79-106.

Munday, Jeremy (2012) *Evaluation in Translation: Critical Points of Translator Decision-Making*. London, Routledge Taylor and Francis Group.

Na'ja, Suha and Abu-Mighnim Jamila (2012) Analysis of Morphological Errors of Arabic Language Learners in light of the linguistic convergence. *Journal of Tikrit University for the Humanities*. 19 (10), 167-219.

Newmark, Peter (1988) *A Text Book of Translation*. New York, Shanghai Foreign Language Education Press.

Nima'a, Fau'ad (1973a*) Summary of Arabic Grammar.* Cairo, The Scientific Office for Authorship and Translation.

Nima'a, Fau'ad (1973b) *Summary of Arabic Grammar.* Cairo, The Scientific Office for Authorship and Translation.

Niska, Helge (1999) *Text Linguistic Models for the Study of Simultaneous Interpreting*. University of Stockholm. Available from: https://pdfs.semanticscholar.org/5df6/7a0f9e46c0cef69e15cc39b2e68b4207bb22.pdf?_ga=2.258718349.681757606.1583272993-1407134683.1562170250. {Accessed 15-Dec-2019}.

Nord, Christiane (1997) *Translating as a Purposeful Activity: Functionalist Approaches Explained*. Manchester, St Jerome Publishing.

Phelan, Mary (2001) *The Interpreter's Resource*. Clevedon, Multilingual Matters LTD.

Pio, Sonia (2003) The Relation Between St Delivery Rate and Quality in Simultaneous Interpretation. *The Interpreters' Newsletter*. 12 (1) 69-100.

Pochhacker, Franz (2016) *Introducing Interpreting Studies*. London, Routledge Taylor and Francis Group.

Pym, Anthony (1993) Christiane Nord. Text Analysis in Translation. Theory, Method, and Didactic Application of a Model for Translation-Oriented Text Analysis. Translated from the German by Christiane Nord and Penelope Sparrow" *Traduction, Terminologie et Redaction*, 6 (1), 184-190.

Pym, Anthony (2009) On Omission in Simultaneous Interpreting. Risk Analysis of a Hidden Effort. In: Hansen, Gyde and et al (eds.), *Efforts and Models in Interpreting and Translation Research.* Amsterdam, John Benjamins Publishing Company, pp. 83-105.

Reishaan, Abdul-Hussein, and Ali Ja'far (2008) Time, Tense, and the Verb Form in Arabic and English: A Contrastive Study. *AL-Qadisiya Arts and Educational Sciences Journal.* 7 (1-2), 99-112.

Renard, Lucie (2017) The Difference between Formative and Summative Assessments. *Bookwidgets Interactive Learning. Available from:* https://www.medford-lakes.k12.nj.us/ourpages/auto/2019/7/11/51422861/The%20Differences%20Between%20Formative%20and%20Summative%20Assessments%20by%20Lucie%20Renard%204_14_17.pdf. {Accessed: 21-Oct-2019}.

Rennert, Sylvi (2010) The Impact of Fluency on the Subjective Assessment of Interpreting Quality. In: Riccardi, A. and Kellet, C. (eds.) *The Interpreters' Newsletter,* pp. 101-115.

Riccardi, Alessandra (1999) Attuali metodi di valutazione presso la SSLMIT. *Quality Form 1997: Esperienze, Problemi, Prospettive*. 13 (1), 33-53.

Riccardi, Alessandra (2002a) Evaluation in Interpretation: Macrocriteria and Microcriteria. In: Hung, Eva (ed.) *Teaching Translation and Interpreting 4: Building Bridges*. Amsterdam, John Benjamins Publishing Company, pp. 114-126.

Riccardi, Alessandra (2002b) Interpreting Research: Descriptive Aspects and Methodological Proposals. In: Garzone, Giuliana and Viezzi Maurizio (eds.) *Interpreting in the 21st Century: Challenges and Opportunities*. Amsterdam, John Benjamins Publishing Company, pp.15-27.

Riccardi, Alessandra (2005) On the Evolution of Interpreting Strategies in Simultaneous Interpreting." *Meta: Journal des traducteurs*. 50 (2), 753-767.

Rustipa, Katharina (2011) Contrastive Analysis, Error Analysis, Interlanguage and the Implication to Language Teaching. *Jurnal Pengembangan Humaniora*. 11 (1), 16-22.

Sak-Wernicka, Jolanta (2017) *Blind People's Pragmatic Abilities*. Newcastle, Cambridge Scholars Publishing.

Saldanha, Gabriela and Sharon O'Brien (2013) *Research Methodologies in Translation Studies*. London, Routledge Taylor and Francis Group.

Salimi, Esmaeel, and Zohreh Shahrestani (2009) A Product Approach to Units of Translation in Iranian English-Persian Literary Translation. *Journal Language and Translation*. 10 (2), 129-152.

Sawant, G. Data "History of Translation "*Literary Endeavor*. 4 (1), 109-115,

Sawyer, David (2004) *Fundamental Aspects of Interpreter Education: Curriculum and Assessment*. Amestrdam, John Benjamin Publishing Company.

Seeber, Kilian (2001) Intonation and anticipation in simultaneous interpreting. *Cahiers de linguistique française*. 23 (1), 61-97.

Setton, Robin (1999) *Simultaneous Interpretation: A Cognitive-Pragmatic Analysis*. Amsterdam, John Benjamins Publishing Company.

Shamma, Tarek (2009) *Translation and the Manipulation of Difference*. London, Routledge Taylor & Francis Group.

Shamsan, Muayad, and Abdul-majeed Attayib (2015) Inflectional Morphology in Arabic and English: A Case Study. *International Journal of English Linguistics*. 5 (2), 139-150

The Obama White House (2010) The End of the Combat Mission in Iraq. *YouTube*. Available from: www.youtube.com/watch?v=yzO9LZzZoOk. {Accessed: 01-March-2019}

Tissi, Benedetta (2000) Silent Pauses and Disfluencies in Simultaneous Interpretation: A Descriptive Analysis. In: Riccardi, A and Viezzis M. (eds.) *The Interpreters' Newsletter,* pp. 103-127.

Toro, Cristina (2007) Translation Studies: An Overview. *Capa*. 2 (20), 9-42.

Toth, Andrea (2011) Speech disfluencies in Simultaneous Interpreting: A Mirror on Cognitive Processes. *Journal of Translation and Interpretation*. 5 (2), 32-31.

United Nations (2015) Angelina Jolie on the Middle East (Syria) – Security Council 7433rd Meeting. *YouTube*. Available from: www.youtube.com/watch?v=Pd2CdyOCvJc. {Accessed: 01-March-2019}

USA for UNHCR the UN Refugee Agency. (2015) *Angelina Jolie Addresses UN Security Council on Ongoing Refugee Crisis in Syria*. Available from: https://www.unrefugees.org/news/angelina-jolie-pitts-speech-to-the-un-security-council/ {Accessed 3-March-2019}

Vancura, Alma (2017) Speech characteristics as progress indicators in simultaneous interpreting by trainee interpreters. *Govor/Speech*. 34 (1), 3-32.

Viktorovich, Viktor, et al (2015) Scientific Methods in Social and Humanitarian Researches. *Indian Journal of Science and Technology*. 8 (S10), 1-9.

Volkova, Tatiana (2014) Translation Model, Translation Analysis, Translation Strategy: An Integrated Methodology. *Procedia-Social and Behavioral Sciences*. 154 (1), 301-304.

Waddington, Christopher (2001a) Different Methods of Evaluating Student Translation: The Question of Validity. *Meta Journal des traducteurs*. 46 (2), 311-325.

Waddington, Christopher (2001b) Should Translation Be Assessed Holistically or Through Error Analysis. *Hermes, Journal of Linguistics*. (26) 15-37.

Wu, Shao-Chuan (2010) *Assessing Simultaneous Interpreting*. Unpublished Ph.D. Dissertation, Newcastle University. Available from: https://pdfs.semanticscholar.org/76bf/72cd868afe434582737e3f7092f2e26b3fb2.pdf?_ga=2.214906135.1232538430.1583444173-1407134683.1562170250. {Accessed: 02-Apr-2019}.

Yudes, Carolina, et al (2013) Comprehension and Error Monitoring in Simultaneous Interpreting. *Applied Psycholinguistics*. 34 (1), 1039-1057.

Zehnalova, Jitka (2013) Tradition in Trends in Translation Quality Assessment. In: Jitka Zehnalova, et al (eds.). *Tradition and Trends in Trans-Language Communication*, pp. 41-58.