The role of advance preparation in simultaneous interpreting

A comparison of professional interpreters and interpreting students

Stephanie Díaz-Galaz, Presentación Padilla* and María Teresa Bajo*
Pontificia Universidad Católica de Valparaíso / *University of Granada

Current comprehension models recognize the role of prior topic-specific knowledge in the processing of general and specialized discourse (e.g. Gernsbacher 1990; Johnson-Laird 1983; Kintsch 1988). In interpreting, there is widespread consensus that interpreters work better when they prepare in advance. However, research on how preparation affects interpreting has encountered such methodological challenges as high variability and the need for appropriately sensitive measures and tasks (Gile 2005). This article reports an experimental study to assess the effect of advance preparation on simultaneous interpreting of specialized speeches, comparing seven professional interpreters and sixteen interpreting students. All participants did two simultaneous interpretations, into Spanish (their ‘A’ language) from English, of presentations from scientific congresses: one with preparation materials provided half an hour beforehand, the other without preparation. Each source text contained both ‘neutral’ and ‘difficult’ speech segments (the three types of difficulty being terminology, syntactic complexity and lack of redundancy). Dependent variables were accuracy of interpretation and length of ear-voice span (EVS), the rationale being that longer EVS probably reflects processing difficulties. The results show that both groups worked significantly better after advance preparation, this being reflected both in accuracy and in ability to maintain a shorter EVS. Interaction between preparation and type of difficulty was also examined.

Keywords: simultaneous interpreting, advance preparation, expertise, specialized speeches
Introduction

The role of prior knowledge in comprehension

Comprehension is one of the complex cognitive processes that co-occur in simultaneous interpreting (for a review, see Christoffels & de Groot 2005). Currently accepted comprehension models, such as the Mental Models Theory (Johnson-Laird 1983), the Construction-Integration Model (Kintsch 1988) and the Structure Building Framework (Gernsbacher 1990), describe comprehension as a dynamic process that encompasses micro- and macroprocessing operations aiming at the construction of an internal (linguistic or abstract) representation of a particular event. Microprocessing includes low-level operations allowing the recognition and decoding of linguistic information explicitly provided in the text, while macroprocessing entails high-level strategies to integrate textual information with prior knowledge and with details of the communicative situation (Ericsson & Kintsch 1995; Kintsch 1998; van Dijk & Kintsch 1983).

In the comprehension of scientific discourse, there is a great deal of empirical evidence supporting the key role of prior topic-specific knowledge in discourse processing and related measures of performance (for a review, see McNamara et al. 2007). This issue has been investigated by comparing individuals with different degrees of knowledge about a certain domain, discipline or theme, such as sports (Chiesi et al. 1979; Kendeou et al. 2003), physics (Chi et al. 1981; Kendeou et al. 2007) or Star Wars (Means & Voss 1985). Research based on this paradigm has made it possible to determine, for example, that prior topic-specific knowledge was a strong predictor of text comprehension even in participants who were not experts in the domain concerned (Alexander et al.1994). More recent studies show that such knowledge contributes to the enrichment of comprehension processes in a number of ways. These include, at the microprocessing level, interaction with text structure and verbal ability; and, at the macroprocessing level, the management of meta-cognitive inferencing strategies (McNamara & O’Reilly 2009; Ozuru et al. 2009).

Comprehension is a fundamental component of most cognitive models representing translation and interpreting (Gerver 1975; Gile 2009; Moser 1978; Seleskovitch 1976; Setton 1999). According to these accounts, not only is translation equivalence established at the lexical/semantic level, but functional equivalence too must be attained in order to produce an acceptable target text. Translators and interpreters must therefore continuously integrate the information presented in successive sentences into the text/speech and create a representation of the situation in which this occurs, so as to reformulate the information in the target language (Zwaan & Radvansky 1998).
Several interpreting scholars have applied the tenets of these comprehension models to interpreting. Robin Setton (1999) takes the mental model theory of Johnson-Laird (1983) as a basis for his pragmatic-cognitive model of simultaneous interpreting (SI), since the dynamic nature of the mental structures thus identified would enable interpreters to address such requirements as the need to maintain cohesion throughout the conference. Mackintosh (1985) suggests applying Kintsch’s model (1988) of comprehension to consecutive and simultaneous interpreting. She proposes that many interpreting errors can be attributed to failures in macroprocessing operations such as deletion of irrelevant propositions or the construction of macro-propositions from the source speech.

Against this background, there is clearly much scope for further research into the role of topic-specific knowledge in SI.

*Previous empirical studies of preparation in simultaneous interpreting*

Preparation of interpreting assignments is a typical feature of professional practice among interpreters (AIIC 2004, 2006, 2009; Diriker 2004; Gile 2002). As mentioned above, both interpreting students and professional interpreters usually devote time before an interpreting exercise or assignment to acquiring a general idea of the topics to be covered, familiarizing themselves with the specific terminology and other relevant information. The final product of this preparation is usually a glossary that contains specialized terms and their target language equivalents (Donovan 2001; Martin 2002; Moser-Mercer 1992; Seleskivitch 1978). Several authors have hypothesized that this acquisition of prior topic-specific knowledge could have a positive effect on SI, by facilitating the anticipation and prediction of information (de Groot 2011; Moser 1978; Seleskivitch 1976).

Some empirical studies have operationalized the construct of prior topic knowledge as “reference information” or “background information”, provided in the form of a speaker’s script or notes. However, research on this topic has so far not given very conclusive results. In Anderson’s study (1979, 1994), professional interpreters working into their ‘B’ language had advance access to a script, to a summary, or to no prior information about the source speech. The author observed no significant effect of the access to reference materials on the informativeness and intelligibility of target speeches. One reason for this, the author explains, might be the limited number of participants and the variability in their performance. As Gile (2005) points out, the dependent variables measured in this study might have not been sensitive enough to capture the effect of studying the related materials prior to interpreting.

In similar studies, Lamberger-Felber (2001, 2003) measured the effect on accuracy of preparation with a script. Three experimental conditions were compared:
one group of participants had a script and time to prepare; a second group also had a script, but no time to prepare; and a control group had no script at all. The two groups that had access to the script showed a higher percentage of correctly interpreted figures and names, and a lower percentage of errors and omissions.

On the other hand, Alonso Bacigalupe (1999) found the opposite trend in a study of eight interpreting students. The author measured interpreting errors in SI in four experimental conditions: advance preparation with a script; advance preparation with no script; availability of the script in the booth, but not beforehand; and no preparation or script. Contrary to expectations, the students who did not have the possibility of advance preparation or access to the script were the ones that made fewest errors.

However, preparation and interpreting actually entail different cognitive processing requirements according to whether they are done with or without a script. Recent studies have shown that interpreting with text presents an increased level of difficulty, and that the need to manage visual input from the text can affect the interpreter’s fluency and accuracy (Agrifoglio 2004; Lamberger-Felber & Schneider 2009; Shreve et al. 2010).

While the effect of background information on performance has proved difficult to observe, empirical data related specifically to the interpreting process (rather than the product) only partially support the assumption that prior knowledge facilitates the translation process. Griffin (1995) used a within-subject design in a study of ten professional translators. All participants performed the same word-translation task in two conditions: with related background information, and with unrelated background information. The researcher measured production times, i.e. the latency between a word’s initial appearance on screen and the moment when each participant began to type its translation on a computer keyboard. While the accuracy and acceptability of translated words were significantly higher when participants had access to relevant background information, their production times were longer. Bajo and Macizo (2009) found similar results: in their study, production times were longer when professional translators had previous access to a summary of the source text in two sight translation tasks: simultaneous (with ‘real-time’ production of an oral target text, while reading the source text) and consecutive (with production of the target text after reading the source text).

However, word translation and sight translation in both these conditions probably entail different cognitive requirements and processes from simultaneous interpreting. As explained above, for example, Agrifoglio (2004) found that one added difficulty of sight translation is the interference of visual input. Moreover, the unrelated information condition in Griffin’s study does not replicate a situation of non-preparation, since analyzing and discarding irrelevant information
do not imply the same cognitive processes as when no information whatsoever is available.

To explore the role of prior topic knowledge in interpreting, it should be studied in a task which is representative not only of real-life simultaneous interpreting (Ericsson 2000/2001), but also of the conditions commonly experienced in actual practice with regard to availability of materials for preparation.

The role of expertise in simultaneous interpreting

The development of expertise in simultaneous interpreting has received a great deal of attention from researchers who have explored the cognitive processes in translation and interpreting. While the study of differences in working memory capacity has produced mixed results (Chincotta & Underwood 1998; Christoffers et al. 2006; Köpke & Nespoulos 2006; Liu 2001; Padilla et al. 1995; Signorelli et al. 2012), studies that have focused on other memory and executive tasks have identified two areas in which experienced interpreters are at an advantage over those of limited experience: lexical and semantic processing (Bajo et al. 2000; Padilla et al. 2005), and overall cognitive processing (Yudes et al. 2012).

Studies comparing experienced and inexperienced interpreters in representative tasks, such as the simultaneous interpreting of real speeches, also show that skilled (professional) interpreters perform significantly better than inexperienced (student) interpreters: this is seen in the overall quality and accuracy of interpreting, as well as in the strategies used to overcome source text difficulties (Sunnari 1995; Tiselius & Jenset 2011); in the ability to distinguish between more important and less important information (Liu 2001; Liu et al. 2004); and in the management of syntactic density and complexity (Hild 2011).

Liu (2001; Liu et al. 2004) measured the accuracy of main and secondary ideas in simultaneous interpreting by experienced interpreters (with at least two years of professional experience), advanced students (two years of interpreter training) and novice students (one year of interpreter training). Experienced interpreters did best in terms of overall accuracy. They also interpreted more main ideas than inexperienced interpreters, whose interpretation included main and secondary ideas to the same degree. These results showed that skilled interpreters had developed an ability to distinguish relevant from irrelevant information. The study also manipulated the difficulty of critical sentences, measured by their degree of readability. Although both groups showed poorer accuracy with difficult critical sentences, the experienced interpreters still did better than the students in this respect.

A more recent study by Hild (2011) supports the hypothesis that experienced interpreters have less difficulty than inexperienced interpreters in managing the
linguistic complexity of source speeches. In this study of eight professional interpreters and eight interpreting students, Hild measured the accuracy of simultaneous interpreting from Bulgarian to English in relation to four parameters of linguistic difficulty: sentence density, syntactic complexity, type of clause, and type of subordination. Accuracy significantly worsened in relation to these difficulties, especially for the students. In addition, the performance of experienced interpreters was unaffected by the level of syntactic complexity, the type of clause or the type of subordination. This would suggest that the role of high-level macroprocessing operations increases with experience.

More recently, Tiselius (2013) also found significant differences between experienced and inexperienced interpreters in a between-groups study: experienced participants (interpreters with more than 25 years’ experience) encountered fewer processing problems, and had more strategies to solve them, than interpreters with little or no experience.

Overall, research to date on the development of expertise in simultaneous interpreting shows that skilled interpreters not only make fewer errors and produce faster responses than inexperienced interpreters, but are also more efficient in the management of their cognitive resources (for a recent review, see Liu 2008). We therefore believe that it is important to continue exploring these differences, especially in specific tests such as interpreting with or without advance preparation of relevant materials.

This study

Against this background, the aim of the present study was to investigate the role of advance preparation in simultaneous interpreting. Specifically, the study aims at exploring (1) the role of advance preparation in overcoming difficult features of the source speech; and (2) the relationship between the effects of advance preparation and interpreting experience.1

Our first specific hypothesis on the role of prior topic-specific knowledge in discourse comprehension was that advance preparation would be reflected in better processing and performance, to be measured on the basis of ear-voice span

1. A pilot study conducted with interpreting students was previously published by Diaz-Galaz (2011). In the present study, we took part of the data from the pilot study and completed the sample with a few more interpreting students and a new group of experienced interpreters. While the aim of the pilot study was essentially to validate our methodology, in the present experiment we wanted to explore the effect of preparation as a function of experience in simultaneous interpreting.
(EVS) and accuracy with difficult features like scientific terminology, complex syntax and lack of redundancy. In comparing professional interpreters and interpreting students, we expected the former to benefit more from preparation.

Methodology

Study design

The experiment involved independent variables in relation to four parameters: (1) preparation vs. non-preparation; (2) level of difficulty of source speech segments (neutral vs. difficult); (3) type of problem in difficult segments (terminology vs. syntactic complexity vs. low redundancy); (4) experience (professional interpreters vs. interpreting students). Three of the resulting comparisons (nos. 1–3) were within-subject, while the fourth (no. 4) was between-subject.

Participants

Twenty-three subjects participated in the study: sixteen students about to complete a two-year undergraduate programme in English-Spanish conference interpreting, and seven professional interpreters who, at the time of the study, had been working regularly on the institutional and private markets in Spain for an average of 11.13 years (min. 4 years — max. 20 years). In the year when the study was conducted (2011), the interpreters had worked an average of 43.83 days (min. 25 days — max. 88 days) in the booth, including an average of 15.33 days at scientific and technical conferences (min. 5 — max. 44). The mean age of professional interpreters at the time of participation in the study was 37.14 years (SD = 6.74). Since proficiency in a second language and working memory capacity are both predictors of L2 listening comprehension performance (for a recent review, see Bloomfield et al. 2010), they were measured in order to establish a profile of participants.

Proficiency in L2

Participants were asked to fill out a language history questionnaire with a self-assessment of their fluency in Spanish (L1, i.e. the ‘A’ language) and English (L2, i.e. a ‘B’ or ‘C’ language), rating their skills in reading, listening, speaking and writing on a scale from 1 to 10. The students gave their general skills in L1 and L2 ratings of 9.31 (SD = 0.65) and 7.63 (SD = 0.82) respectively; for the professional interpreters, the level of self-assessed proficiency in L1 and L2 was 9.85 (SD = 0.19) and 8.9 (SD = 0.95) respectively. Comparison by ANOVA showed significant between-
groups differences in both L1 ($F(1, 21) = 4.635; \text{MSE} = 0.312; \ p = 0.043$) and L2 ($F(1, 21) = 11.89; \text{MSE} = 0.735; \ p = 0.024$).

**Working memory span**
Participants were also asked to perform the Spanish version of the reading span test (Daneman & Carpenter 1980), to assess their working memory span. In this test, sets of sentences are shown on a computer screen and participants are instructed to read each sentence aloud and to recall the last word of each sentence at the end of the set. The number of sentences in the set is increased gradually from two to six. The size of the largest set of sentences in which all last words were recalled correctly represents the participant’s memory span. Participants with scores of 3.5 or higher are usually considered to have a high memory span (Miyake et al. 1994). In this study, the mean scores for the students and the professional interpreters were 3.46 ($\text{SD} = 0.59$) and 4.2 ($\text{SD} = 0.75$) respectively. In this case too, an analysis of variance showed a significant difference between the two groups ($F(1, 21) = 8.91; \text{MSE} = 0.365; \ p = 0.007$).

**Background knowledge**
After completing the interpretation, participants were asked to fill out a questionnaire on their degree of familiarity with the source speech topic, as well as with the ideas and terminology it contained. This assessment was to be expressed on a scale from 1 (“very low level of knowledge”) to 5 (“very high level of knowledge”). The mean scores for familiarity with the topic were 1.88 ($\text{SD} = 0.89$) and 2.21 ($\text{SD} = 0.76$) for students and interpreters respectively; familiarity with the ideas expressed was rated as 2.09 ($\text{SD} = 0.95$) and 2.07 ($\text{SD} = 0.61$) respectively; while prior knowledge of the terminology used was rated as 1.97 ($\text{SD} = 0.90$) and 2.64 ($\text{SD} = 0.99$) respectively. $T$-tests showed no significant differences between the groups in terms of familiarity with the topic ($t = -0.88; \ p = 0.388$), ideas ($t = 0.06; \ p = 0.96$) or terminology ($t = -1.60; \ p = 0.124$).

Since the two groups of participants differed significantly in terms of their L1 and L2 skills and working memory span, these were taken into account as covariates in a between-groups ANCOVA of the dependent variables on which the interpretations were assessed.

**Experimental material**

**Selection of speeches**
Each group was required to do two simultaneous interpretations, from English into Spanish: one with prior preparation, the other without. With a view to ensuring that the two source speeches differed in topic but not in terms of difficulty,
they were selected by a panel of five interpreting lecturers from the Department of Translation and Interpreting at the University of Granada. The selection was made from six speech transcripts which were accessed on the Internet, all of them taken from scientific seminars. Rating was on a scale from 1 (“very easy”) to 7 (“very difficult”), the idea being to assess the probable level of difficulty for a final year interpreting student without specific knowledge of the topic. The two speeches selected for the experiment were the ones that were most similar in terms of difficulty. A full description and transcription of the speeches can be found in Díaz-Galaz (2012).

Comparison by ANOVA showed no statistical difference in the ratings provided by the five judges \(F(4, 25) = 1.627; \text{MSE} = 1.255; p = 0.2\), with two pairs of texts showing a significant correlation \(r = 0.93; p = 0.012\). The pair actually selected was the one with the higher degree of concordance among judges \(K = 1\), the topics of the two speeches being “The genetics of schizophrenia and bipolar disorder” (Sklar 2008) and “Neuroscience of visual perception” (Ramachandran 2003).

Identification of neutral and difficult segments
The judges selecting the texts were asked to mark the terms or segments that they considered would cause the interpreting students difficulty, and also to indicate why they considered each of these items difficult.

For each experimental speech, the judges identified 46 ‘difficult’ points. Without any manipulation of the original script, the items concerned fell into three categories: (i) terminology; (ii) syntactic complexity; (iii) presence of non-redundant items (such as proper names, figures and acronyms). Texts were then segmented into smaller units (phrases or clauses) around these difficult elements. For each of the two speeches, the frequency of segments in these three categories was as follows: terminology, 16 occurrences; syntactic complexity, 18; non-redundant items, 12. Segments which the judges did not mark as difficult were classified as ‘neutral’, indicating that they were not expected to pose a particular challenge in terms of general knowledge, linguistic competence or interpreting proficiency. Examples are “It is a great honor for me to be here today”, “How do we know this?” or “The answer is obvious”. There were 34 of these in each experimental text.

The following excerpts from the experimental speeches show examples of segments classed as difficult (in square brackets and bold type) and as neutral (in italics):

Speech 1
Today psychiatrists are stuck with a distinction that started over a hundred years ago, when the [psychiatrist Emil Kraeplin] was able to define or identify a
constellation of symptoms that he called [dementia praecox], what we would now call Schizophrenia. He identified the [thought disorder] and the [delusions] and [hallucinations], and something that he called [manic-depressive insanity], that we would now call [bipolar disorder] or more commonly [manic-depressive disease].

Speech 2
Well, the most striking evidence comes from patients with tiny lesions that damage just [V4], the color area, or just [MT], the motion area. So, for example, when V4 is damaged on both sides of the brain, you end up with a syndrome called [cortical color blindness] or [achromatopsia]. Patients with [cortical achromatopsia] see the world in shades of grey, like a black and white movie, but they have no problem reading a newspaper or recognizing people's faces or seeing the direction of movement.

Table 1 shows the number of words, paragraphs and segments in each of the two texts. In addition, the Coh-Metrix 3.0 program (McNamara et al. 2013) was used to analyze such features as the mean length of each speech segment, lexical diversity and lexical density. Lexical diversity (type-token ratio) is a measure of the number of new versus repeated words in a text; lexical density indicates the proportion of content words in relation to the total number of words in the text.

<table>
<thead>
<tr>
<th></th>
<th>Speech 1</th>
<th>Speech 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of words</td>
<td>1121</td>
<td>1311</td>
</tr>
<tr>
<td>Number of paragraphs</td>
<td>18</td>
<td>17</td>
</tr>
<tr>
<td>Number of segments</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Length of segments (mean number of words per segment) (SD)</td>
<td>13.84 (7.33)</td>
<td>15.99 (8.1)</td>
</tr>
<tr>
<td>Lexical diversity (type-token ratio, all words)</td>
<td>0.382</td>
<td>0.344</td>
</tr>
<tr>
<td>Lexical density (content words/total)</td>
<td>0.49</td>
<td>0.48</td>
</tr>
</tbody>
</table>

Speed of delivery
A number of studies show the effects of input rate on simultaneous interpreting, the optimal speed probably being in the range of 100–120 words per minute (wpm) (Barik 1973; Gerver 1969/2002; Liu 2001; Shlesinger 2003). Research on L2 listening comprehension shows that the speech rate to which the listener is exposed interacts with other factors, such as proficiency in L2, the speaker’s accent and text length, the combination of which tends to affect subjective perception of difficulty and the level of comprehension (Bloomfield et al. 2010).
Since the interpreters and interpreting students in our study differed in working memory capacity, L2 proficiency and interpreting expertise, it was necessary to avoid the confounding effect of either too fast or too slow a speed of delivery. With this in mind, the experimental speeches were recorded by a female native speaker of English and the audio files were digitally manipulated to achieve two versions of each speech, with differing speeds of delivery: 100 wpm for the students, and 120 wpm for the interpreters.

All four of the recordings to be interpreted were analyzed using Praat©, version 5.3.77 (Boersma & Weenink 2014), a package for processing data on quantitative parameters such as speech rate and pausing time. Table 2 shows that phonation time (total speaking time, excluding pauses), speech rate (measured in syllables per second, including pauses) and articulation rate (in syllables per second, excluding pauses) remained broadly similar in all four recordings, thus making them comparable for the purposes of our study.

Table 2. Speed-of-delivery measures for the two experimental speeches

<table>
<thead>
<tr>
<th></th>
<th>Speech 1</th>
<th></th>
<th>Speech 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>100 wpm</td>
<td>120 wpm</td>
<td>100 wpm</td>
<td>120 wpm</td>
</tr>
<tr>
<td>Number of syllables</td>
<td>2630</td>
<td>2381</td>
<td>2493</td>
<td>2244</td>
</tr>
<tr>
<td>Duration (seconds)</td>
<td>696.9</td>
<td>600.02</td>
<td>745.2</td>
<td>600.02</td>
</tr>
<tr>
<td>Phonation time (speaking time, excluding pauses; seconds)</td>
<td>597.2</td>
<td>517.9</td>
<td>575.3</td>
<td>471.18</td>
</tr>
<tr>
<td>Articulation rate (syllables per second, excluding pauses)</td>
<td>4.40</td>
<td>4.60</td>
<td>4.33</td>
<td>4.76</td>
</tr>
<tr>
<td>Speech rate (syllables per second, including pauses)</td>
<td>3.78</td>
<td>3.97</td>
<td>3.35</td>
<td>3.74</td>
</tr>
</tbody>
</table>

Preparation materials

For each speech, the preparation materials consisted of a 250-word summary; a biosketch of the speaker (about 80 words); nine slides based on the speech; a conference programme including the speech (for communicative context); and an English-Spanish glossary containing 30 specialized terms, in some cases with additional information about the term or concept. Our purpose in preparing these items was to recreate the type of research and reading done by a professional interpreter before an actual conference (for example, see Abril & Ortiz 1998; Donovan 2001; Gile 2009; Moser-Mercer 1992).
Procedure

The experimental sessions were conducted at the University of Granada, in two different facilities: in the School of Translation interpreting laboratory (for the students), and in the Experimental Psychology Department behavioral science laboratories (for the professional interpreters). All participants completed their interpretations individually, with no audience. The speeches were presented as audio recordings. Participants interpreted one speech after 30 minutes’ preparation based on the materials above, and one without any prior preparation or knowledge of the topic. Participants were not allowed to access the Internet, or to consult any materials other than those provided. To ensure that speech content would not act as a confounding variable in the comparison of the two experimental conditions (preparation vs. no preparation), the order of the preparation conditions and the presentation of the speeches was counterbalanced. A 15-minute break was allowed between the two interpretations.

Dependent variables

The variables measured in this study were EVS and accuracy of interpretation. Research on cognitive processing has traditionally used mental chronometry measures, such as latencies and reaction times, to infer mental operations and measure cognitive demand (see Posner 1978). In simultaneous interpreting research, this latency has traditionally been measured by EVS. Research suggests that this may be a reliable measure of cognitive processing in simultaneous interpreting tasks (Timarová et al. 2011). EVS was measured at the beginning of the 80 source speech segments, the theoretical overall total of EVS measurements from the entire sample in the two experimental conditions being 3,680.

In the event, some source speech segments were omitted in the target speech. For the purposes of this study, this was considered to be the case only when there was no trace whatsoever in the target speech of a given source text segment. For the 220 segments (7.14%) in this category, EVS values could not be calculated. In addition, EVS values of more than two standard deviations from the mean were considered as outliers and were excluded from the analysis (n=471, i.e. 15.3% of total observations). This meant that EVS was actually measured on 2,989 segments.

Scoring

Three criteria were assessed by one of the authors, in order to obtain a multivariate measure of the interpreters’ accuracy: (i) use of vocabulary and terminology;
(ii) content; (iii) absence of syntactic interference from the source language. Other features of delivery, such as fluency, intonation and voice quality, were not assessed.

The interpretation of each of the 80 source speech segments was rated on the above criteria, on a scale from 1 to 3. For terminology and content, the rating was as follows: 1 = omission; 2 = some inaccuracy; 3 = correct/appropriate. For interference, the scoring covered a range from clear interference (1) to none (3). Overall accuracy was defined as the mean value of the scores for terminology, content and control of interference. For purposes of validation, two independent judges (PhD students from the Department of Translation and Interpreting at the University of Granada, with practical and theoretical training in translation and professional experience as English-Spanish translators) assessed interpretation samples from four participants (16% of the total sample). Comparison between their assessment and ours showed that Kendall’s coefficient of concordance was significant ($W=0.92, p<0.05$).

**Results**

Descriptive and inferential statistical analyses were conducted on the study data. Student’s $t$-tests for dependent samples were used to assess the effect of preparation on the dependent variables (EVS and accuracy of target speeches). To test the significance of main effects and interactions among more than two dependent variables, repeated-measures analyses of variance were conducted. Statistical significance was established at $p<0.05$.

**Effects of advance preparation in interpreting students**

Table 3 shows the results for EVS (in milliseconds) and overall accuracy, according to whether the interpreting students prepared. EVS was significantly shorter ($t=−3.13; p=0.007$), and overall accuracy significantly higher ($t=5.819; p<0.001$), after preparation.

<table>
<thead>
<tr>
<th>Preparation</th>
<th>No preparation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2425.05 (349.96)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.48 (0.33)</td>
</tr>
</tbody>
</table>
Regarding the effect of advance preparation on difficult and neutral speech segments, Table 4 shows that, irrespective of preparation, EVS was consistently and significantly longer for difficult segments (ANOVA: $F(1, 15) = 17.397; MSE = 38302; p < 0.001$); the interaction between preparation and level of difficulty was not significant ($F < 1$). For accuracy too, scores differed significantly according to the level of difficulty, with lower accuracy in difficult segments (ANOVA: $F(1, 15) = 334.29; MSE = 0.006; p < 0.001$); in this case, the interaction between preparation and level of difficulty was significant ($F(1, 15) = 27.494; MSE = 0.016; p < 0.001$). Though the increase in the level of difficulty had a significant effect in both conditions, it was greater when the students did not prepare ($F(1, 15) = 185.891; MSE = 0.012; p < 0.001$).

**Table 4.** Interpreting students (n=16): mean EVS and overall accuracy in neutral and difficult segments, with and without prior preparation

<table>
<thead>
<tr>
<th></th>
<th>Neutral segments</th>
<th>Difficult segments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preparation Mean (SD)</td>
<td>No preparation Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2295.57 (340.31)</td>
<td>2449.50 (433.50)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.56 (0.31)</td>
<td>2.36 (0.27)</td>
</tr>
<tr>
<td></td>
<td>Preparation Mean (SD)</td>
<td>No preparation Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2534.83 (413.58)</td>
<td>2618.38 (371.64)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.37 (0.36)</td>
<td>1.84 (0.28)</td>
</tr>
</tbody>
</table>

Table 5 shows EVS and overall accuracy for difficult speech segments. While EVS was longer in segments with terminological and syntactic difficulties than in segments containing non-redundant information, these differences were not significant (ANOVA: $F(1, 15) = 2.447; MSE = 95634; p = 0.104$); in addition, there was no significant interaction between preparation and type of difficulty ($F < 1$). In other words, the students maintained a fairly steady EVS throughout the difficult segments in both experimental conditions.

**Table 5.** Interpreting students (n=16): mean EVS and overall accuracy in difficult segments, with and without prior preparation

<table>
<thead>
<tr>
<th></th>
<th>Specialized terminology</th>
<th>Complex syntax</th>
<th>Lack of redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prep. Mean (SD)</td>
<td>No prep. Mean (SD)</td>
<td>Prep. Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2621.56 (512.90)</td>
<td>2636.57 (446.80)</td>
<td>2663.10 (545.45)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.40 (0.35)</td>
<td>1.72 (0.35)</td>
<td>2.28 (0.37)</td>
</tr>
<tr>
<td></td>
<td>Prep. Mean (SD)</td>
<td>No prep. Mean (SD)</td>
<td>Prep. Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2392.35 (549.12)</td>
<td>2584.71 (439.55)</td>
<td>2392.35 (549.12)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.39 (0.41)</td>
<td>1.89 (0.30)</td>
<td>2.39 (0.41)</td>
</tr>
</tbody>
</table>

For accuracy too, there were no significant differences related to the type of difficulty (ANOVA: $F(1, 15) = 1.412; MSE=0.036; p=0.259$). On the other hand, the
effect of preparation was significant \( (F (1, 15) = 44.649; MSE = 0.146; p < 0.001) \), as was the interaction between preparation and type of difficulty \( (F (1, 15) = 4.365; MSE = 0.048; p = 0.027) \). In segments containing specialized terminology, accuracy was significantly lower in the absence of prior preparation \( (F (1, 15) = 9.51; MSE = 0.037; p = 0.008) \).

Three general trends thus emerge among the interpreting students: (i) preparation had a significant effect on both EVS (reduced) and accuracy (increased); (ii) EVS was significantly longer, and accuracy significantly lower, in difficult speech segments than elsewhere; (iii) the effect of preparation was consistent, showing no significant interaction with the level of difficulty.

**Effects of advance preparation in experienced interpreters**

Table 6 shows that preparation was associated with shorter mean EVS and greater accuracy: Student’s \( t \)-test showed that data for both these variables differed significantly according to whether the interpreters had prepared beforehand (EVS, \( t = -2.53, p = 0.045 \); accuracy, \( t = 3.085, p = 0.022 \)).

**Table 6.** Experienced interpreters \((n=7)\): mean EVS and overall accuracy, with and without prior preparation

<table>
<thead>
<tr>
<th>Preparation</th>
<th>No preparation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2411.01 (287.30)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.81 (0.15)</td>
</tr>
</tbody>
</table>

Table 7 shows that EVS and accuracy did not differ significantly between neutral and difficult segments (ANOVA, \( F < 1 \) for both variables); in addition, the interaction between preparation and level of difficulty was not significant \( (F < 1 \) for both variables).

**Table 7.** Experienced interpreters \((n=7)\): mean EVS and overall accuracy in neutral and difficult segments, with and without prior preparation

<table>
<thead>
<tr>
<th>Neutral segments</th>
<th>Difficult segments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preparation</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2378.01 (372.85)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.84 (0.14)</td>
</tr>
</tbody>
</table>

Trends for EVS and overall accuracy in difficult speech segments can be seen in Table 8. EVS was significantly shorter in segments containing non-redundant
information such as figures and acronyms ($F (2, 12) = 20.204; MSE = 27179; p < 0.001$); the interaction between preparation and type of difficulty was not significant ($F < 1$). In terms of accuracy, results were significant neither for differences related to the type of difficulty nor for the interaction between preparation and type of difficulty (for both these analyses, $F < 1$).

The interpreters’ preparation was therefore efficient, helping them achieve accuracy even in difficult segments. As for the students, mean EVS was significantly shorter following prior preparation. Overall trends for the professional interpreters thus overlap only to a limited extent with those for the students. There were two similarities: (i) the professionals’ EVS and accuracy were significantly related to preparation; (ii) the effect of preparation was consistent, showing no significant interaction with the level of difficulty. On the other hand, the professionals’ EVS and accuracy showed no significant differences between neutral and difficult speech segments.

<table>
<thead>
<tr>
<th>Specialized terminology</th>
<th>Complex syntax</th>
<th>Lack of redundancy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prep. Mean (SD)</td>
<td>No prep. Mean (SD)</td>
<td>Prep. Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2603.88 (314.73)</td>
<td>2790.61 (420.42)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.78 (0.24)</td>
<td>2.50 (0.37)</td>
</tr>
</tbody>
</table>

**Interpreting expertise**

Differences between the two groups of participants were tested by ANCOVA. Since the groups differed in terms of working memory capacity and L2 (English) skills, these two variables were factored in as covariates. Table 9 shows that once this had been done, the experienced interpreters were seen to have been more accurate than the students ($F (1, 39) = 12.227; MSE = 0.079; p = 0.001$). This main effect of expertise remained significant after conducting a factorial ANOVA (i.e. not controlling for working memory capacity and L2 skills) ($F (1, 42) = 27.332; MSE = 2.078; p = 0.001$).
Table 9. Experienced and inexperienced interpreters: mean EVS and overall accuracy, with and without prior preparation

<table>
<thead>
<tr>
<th></th>
<th>Inexperienced interpreters</th>
<th>Experienced interpreters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preparation Mean (SD)</td>
<td>No preparation Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Preparation Mean (SD)</td>
</tr>
<tr>
<td>EVS (ms)</td>
<td>2425.05 (349.96)</td>
<td>2540.70 (379.70)</td>
</tr>
<tr>
<td>Overall accuracy</td>
<td>2.48 (0.33)</td>
<td>2.10 (0.26)</td>
</tr>
</tbody>
</table>

EVS in the two groups was also tested by ANCOVA. In this case, the difference between the groups was not significant ($F<1$). Comparison by factorial ANOVA also showed no significant difference between them ($F<1$). In addition, there was no significant interaction between preparation and group ($F<1$). In other words, regardless of the differences in participants’ interpreting experience, L2 competence and working memory capacity, both groups maintained similar EVS values.

Discussion

The results of this study reinforce the idea that advance preparation supports the SI process, resulting in shorter EVS and greater accuracy: for the students, this was seen even in difficult segments with specialized terminology, complex syntactic structure and non-redundant information; for the professional interpreters, segments with non-redundant elements were managed better after preparation.

How does advance preparation facilitate the interpreting process? Previous research shows that the ability to simultaneously comprehend and produce speech could be related to the ability to use lexical knowledge efficiently in support of these tasks: in a study by Padilla et al. (2005), professional interpreters avoided the articulatory suppression effect when dealing with L1 words, but not when processing non-words. The authors explain that efficiency might be related to the role of familiarity and to the ability to access language-specific long-term knowledge during the simultaneous interpreting process, either via the episodic buffer (Baddeley 2000) or through domain-specific knowledge structures such as a long-term working memory (Ericsson & Kintsch 1995). Indeed, when preparing for an interpreting assignment, special attention is given to preparing glossaries of specific terms and their target language equivalents. According to Gile’s Gravitational Model of language availability (2009), advance preparation activates and increases the frequency or familiarity of relevant source and target language information in the long-term memory. The items concerned thus remain available for prompt access and retrieval during SI. Advance preparation should therefore facilitate the
balancing of comprehension and production, arguably one of the greatest sources of difficulty in simultaneous interpreting (Christoffels & de Groot 2005; de Groot 2011).

These data are also consistent with a processing capacity account of simultaneous interpreting such as the Effort Model (Gile 2009). According to this model, the different cognitive operations involved in SI must share limited attentional resources. If any of these operations take(s) too much of this limited capacity, the other operations will not have sufficient resources to carry on and the entire process will stall. In the present study, lack of preparation increased processing capacity requirements, this being reflected in longer EVS and lower accuracy for both groups. Longer EVS places increasing strain on the overall interpreting process, making it hard for the interpreter to maintain comprehension and production at a good enough level (Gile 2009).

With regard to differences between the professional interpreters and students, the results indicate that both groups benefited greatly from prior access to related sources of information. The participants in our study differed in years of experience in SI, working memory capacity and knowledge of L2. The professionals interpreted more accurately, while EVS was comparable in the two groups, even when controlling for WM capacity and skills in L2. These results might suggest that greater accuracy is in part a function of experience and expertise, allowing faster access to (prior) knowledge and efficient use of cognitive resources for better performance in the limited time available.

Methodological implications for future research

One main conclusion of this study is that familiarity with the topic affects both processing (EVS) and performance (accuracy) in simultaneous interpretation. The work described above has methodological implications for future research, highlighting the need to control as far as possible for participants’ prior knowledge. One possible way of doing so is to select topics that are not commonly dealt with in class or in professional interpreting. This is probably easier achieved with interpreting students than with professionals, but gauging participants’ prior knowledge is not easy and relies largely on rather subjective methods of assessment. Control of this variable might be helped by adding a briefing or similar preparation to the experimental task, so that all participants have at least minimum prior knowledge of the topic(s) involved. This may make it possible to reduce performance variability in experimental designs, thus helping isolate the effects produced by the experimental manipulations and variables.

As in previous works that measured EVS (Bajo et al. 2000; Barik 1973; Christoffels et al. 2006), mean EVS in this study was about 2–3 seconds. While
interpreters may adapt EVS when dealing with problems in SI, the trend in this study was to lengthen or shorten it according to the difficulty of any given segment, but not according to the interpreter’s experience. Even when the effect of skills in L2 and working memory capacity was accounted for, no appreciable difference in EVS was found between professional interpreters and students. This may mean that: (i) as suggested by Timarová et al. (2011), EVS is a reliable indicator that varies according to local fluctuations in the difficulty of simultaneous interpreting; (ii) EVS may tend not to exceed a certain duration, reflecting a natural limit in a simultaneous interpreter’s processing capacity. This of course would need to be further explored in future research.

Another important methodological implication of our study is related to the source speech presentation rate. In this study, the speed of delivery was digitally manipulated to avoid the confounding effect of too high or too low a rate of delivery. Speech rate (including pauses) and articulatory rate (excluding pauses) nevertheless remained similar between the two groups. Admittedly, while speech rate in w.p.m. provides a rough indication of content expressed in a given amount of time, it does not account for variations in word length. Calculating speech rate and articulation rate in syllables per minute could provide a more sensitive measure of speech rhythms and pausing behaviour in discourse-length materials (Bloomfield et al. 2010) such as those used in interpreting research.

Conclusions

The findings of this study indicate that both inexperienced and experienced interpreters can benefit from advance preparation, usually done by identifying reliable sources of information, extracting relevant information from them, and drawing up a glossary. By the same token, interpreting students should find preparation of specific background knowledge useful before any form of interpreting practice or test on a given topic.

With regard to professional conditions, the findings of this study indicate that interpreters are fully justified in asking for access to conference materials in order to prepare for an assignment. Although tight professional schedules may make it very difficult to prepare at length for any given assignment, the professional interpreters in this study benefited significantly from fairly limited advance preparation. The habit of preparing in advance for interpreting assignments could also be evidence of deliberate practice leading to expert performance (Ericsson 2006). Deliberate practice tasks are specifically designed to improve performance,

2. The authors thank an anonymous reviewer for pointing this out.
with immediate feedback and, where appropriate, repetition. In the case of advance preparation, interpreters study conference materials with the specific aim of improving their performance, and can often obtain immediate feedback as to whether this was effective. In this way, they can refine their methods of preparation and gain insight into what helps them best, thus enabling them to improve performance on subsequent assignments.

Users of interpreting services, such as conference organizers, can learn from this study how important it is to provide interpreters with relevant information, like the topic of the conference and presentations and, if possible, other materials like speakers’ slides. Conference organizers, in particular, will surely not fail to appreciate that this is in their interest, helping set the scene for better interpretation and thus contributing to the success of the event.

References


The role of advance preparation in simultaneous interpreting

and models in interpreting and translation research: A tribute to Daniel Gile. Amsterdam: John Benjamins, 159–177. DOI: 10.7202/1011268ar


Authors’ address

Stephanie Díaz-Galaz
Instituto de Literatura y Ciencias del Lenguaje
Pontificia Universidad Católica de Valparaíso
Av. El Bosque 1290, Edificio C, Oficina 7–16, Campus Sausalito
Viña del Mar
Chile

stephanie.diaz@ucv.cl
ppadilla@ugr.es
mbajo@ugr.es
About the authors

Stephanie Díaz-Galaz is an Assistant Professor at Pontificia Universidad Católica de Valparaíso, Chile. Her research interests are cognitive processes in translation and interpreting, the acquisition of translation and interpreting skills, and the development of expertise in translation and interpreting.

Presentación Padilla is a Full Professor in the Department of Translation and Interpreting at the University of Granada, Spain. Her main fields of research are interpreter training and cognitive processes in interpreting.

Teresa Bajo is a Full Professor in the Department of Experimental Psychology at the University of Granada, Spain. She carries out research on inhibition in memory control and in language selection, cognitive processes in language translation and interpretation, and individual differences in working memory.