Meta
Journal des traducteurs
Translators' Journal

META

English into Korean Simultaneous Interpretation of Academy Awards Ceremony Through Open Captions on TV

Taehyung Lee

Volume 56, numéro 1, mars 2011

URI : https://id.erudit.org/iderudit/1003514ar
DOI : https://doi.org/10.7202/1003514ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Lee, T. (2011). English into Korean Simultaneous Interpretation of Academy Awards Ceremony Through Open Captions on TV. *Meta*, *56*(1), 145–161. https://doi.org/10.7202/1003514ar

Résumé de l'article

Dans le présent article, nous avons mesuré, à l'aide d'un logiciel de montage vidéo, les paramètres temporels de la retransmission en direct de la cérémonie des Oscars. Celle-ci faisait appel, en même temps, à l'interprétation simultanée et à un sous-titrage visible en coréen. Les résultats ont montré que les sous-titres en coréen apparaissaient à l'écran 7,24 secondes après le début du commentaire original en anglais et qu'ils persistaient pendant 7,52 secondes après la fin de ce dernier. Ces valeurs sont statistiquement plus longues que celles qui étaient observables lors d'autres retransmissions en direct accompagnées uniquement d'une interprétation simultanée, ainsi que dans d'autres programmes de télévision utilisant des sous-titres simultanés. Certaines omissions ont été mises en évidence : d'une part, des phrases relativement courtes, dans l'interprétation simultanée ; d'autre part, les phrases venant après des décalages longs. En dépit des décalages longs, des questionnaires administrés à des spectateurs ont montré que ces derniers préféraient les sous-titres simultanés à l'interprétation simultanée seule, sans doute parce qu'ils peuvent ainsi entendre les commentaires originaux sans être perturbés par la voix de l'interprète.

English into Korean Simultaneous Interpretation of Academy Awards Ceremony Through Open Captions on TV

TAEHYUNG LEE*
Hanyang University, Ansan, Korea
tlee@hanyang.ac.kr

RÉSUMÉ

Dans le présent article, nous avons mesuré, à l'aide d'un logiciel de montage vidéo, les paramètres temporels de la retransmission en direct de la cérémonie des Oscars. Celle-ci faisait appel, en même temps, à l'interprétation simultanée et à un sous-titrage visible en coréen. Les résultats ont montré que les sous-titres en coréen apparaissaient à l'écran 7,24 secondes après le début du commentaire original en anglais et qu'ils persistaient pendant 7,52 secondes après la fin de ce dernier. Ces valeurs sont statistiquement plus longues que celles qui étaient observables lors d'autres retransmissions en direct accompagnées uniquement d'une interprétation simultanée, ainsi que dans d'autres programmes de télévision utilisant des sous-titres simultanés. Certaines omissions ont été mises en évidence : d'une part, des phrases relativement courtes, dans l'interprétation simultanée ; d'autre part, les phrases venant après des décalages longs. En dépit des décalages longs, des questionnaires administrés à des spectateurs ont montré que ces derniers préféraient les sous-titres simultanés à l'interprétation simultanée seule, sans doute parce qu'ils peuvent ainsi entendre les commentaires originaux sans être perturbés par la voix de l'interprète.

ABSTRACT

This article used video-editing software to explore the temporal aspects of the live coverage of the Academy Awards Ceremonies, which employed simultaneous interpretation (SI) and open Korean captions at the same time. The results showed that the Korean captions appeared 7.24 seconds after the beginning of the original sentences and remained on the screen 7.52 after the end of original. These figures were statistically longer than other live coverage of the same events with SI alone and other TV programs carrying live captions. It was also found that relatively short original sentences were omitted in SI and the long EVS (ear-voice-span) also left the sentences that followed uninterpreted. In spite of the long time lag, questionnaires indicated that viewers preferred SI through open captions to SI alone presumably because they could listen to the original voices of entertainers without being disturbed by the voice-over of SI.

MOTS-CLÉS/KEYWORDS

coréen, interprétation simultanée, sous-titre, décalage
Korean, simultaneous interpretation, caption, ear-voice span

1. Introduction

Since SI plays a pivotal role in international conferences, its use has extended to live TV broadcasting in Korea. In this context, many Korean TV companies now air major international events with English into Korean SI. The latest such event was the live coverage of US president Obama's inaugural ceremony by five Korean TV

networks at the same time. In the midst of these developments, a cable TV company in Korea has launched a daring attempt to cover the Academy Awards live with SI from English into Korean. In spite of the intrinsic difficulties of SI, some Korean viewers complained they could not hear the original voice of entertainers due to the SI voice over, which agrees Kurz's (1990) findings. Korean viewers wrote in their blogs that "I could not hear the original voice nor SI" (2002) or "The SI with high tone overlapping original was disturbing" (2003).

An edited version of the same ceremonies, on the other hand, was re-aired on the same day or some days later through Korean open captions. Korean viewers reacted favorably to this mode of captions in the re-aired program and wanted the caption mode introduced to live coverage too. "On Sunday morning, the edited version with Korean caption was very good. I hope this mode is being used in live coverage too" (2002).

The fact that the re-aired programs exclusively relied on captions indicates that TV companies regard captions to be superior to live SI. In this case, the captions were not the transcription of SI; rather, they were the final product of meticulous translation of the original remarks. Synchrony, which was impossible for SI through captions, might have been another positive factor that drew favorable feedback from Korean viewers.

In 2002, clearly influenced by the feedback from Korean viewers, the company began to include SI through captions for live coverage of the ceremony. The acceptance speech was aired with SI through open Korean captions while simultaneously interpreting the rest of the awards ceremony. This was an attempt to satisfy both audiences who want the original source, and those who could not follow the English original. In 2006, the use of SI through captions was extended to the master of ceremony, all of whose remarks were covered by SI through captions. Thus, no voice of interpreters was heard on the TV. This unique mode of live captioning substitutes interlingual captioning, which requires several steps including translation, proofreading and editing with SI, which is a real time language transfer.

When watching live SI through open captions, Korean viewers are orally exposed to the original English speech and then visually to incoming open Korean captions. In other words, viewers may have to carry out complex tasks of comprehending incoming oral messages and visual captions simultaneously. This is because the captions used in this program under examination are open captions, not closed captions. Unlike closed caption (CC) which needs a decoder in the TV set to see them, open caption is a part of the signal and needs no device to read them, and viewers cannot turn them off on their own (den Boer 2001). In this respect, Bartrina and Espasa (2005) noted that a main feature of audiovisual text is its being received through two channels, acoustic and visual. Since the viewers are native speakers of Korean, it can be assumed that reading Korean open captions will need less processing capacity than listening to English, which is a foreign language.

Although this kind of broadcasting with SI through captions to cover an awards ceremony was once tried in the Netherlands (Viaggio 2001), it was a first in Korea, and therefore requires academic analysis of this newly emerging mode. In this context, this article examined the temporal aspects of English into Korean SI through Korean open captions using a video-editing software. Based on the collected data, interpreters' information processing during this unique circumstance was also

explored. In addition, questionnaires were distributed to gather viewers' feedback. The findings will contribute to the academic development of SI studies as well as practical application of this innovative method.

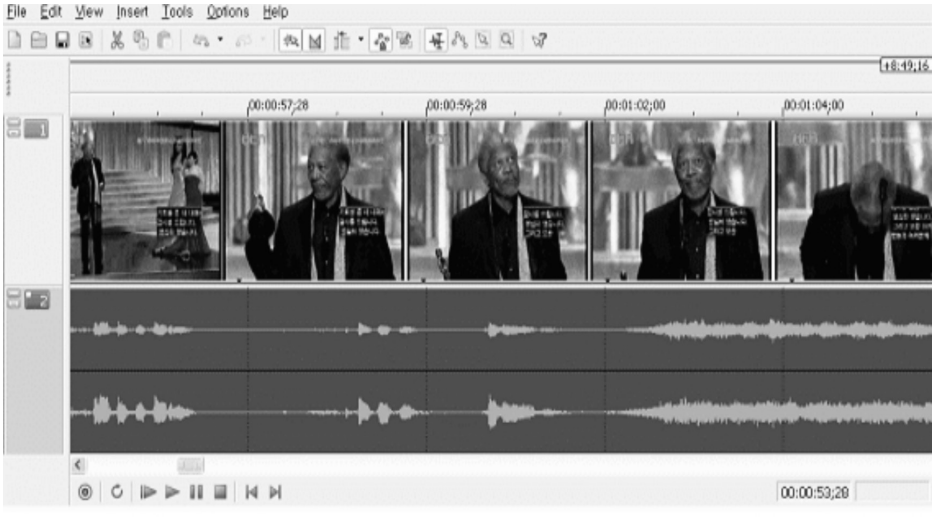
2. Materials and Procedures

The Academy Awards Ceremony broadcast in 2005, which employed SI through open captions, was chosen for this study. The results were compared with the temporal factors of Academy Awards Ceremonies in 1997 and 1999, which employed only English into Korean SI. The acceptance speeches of main actors, actresses, supporting actors and actresses were analyzed for this study. This kind of real-life data provided more reliable results than would studies via laboratory experiments with student interpreters.

The first phase of the study was to record the ceremony on videotapes. The recorded signal was transmitted to a PC equipped with picture capture board so as to digitalize and save the audio and video signal. The captured signals were analyzed using a video-editing software as seen in figure 1.

FIGURE 1

Measuring time span between audio signal and captions



The picture of the ceremony with open Korean captions was positioned in the upper window and the waveform of the English speech was located in the lower part of the screen. Using this software, it was possible to transcribe the original English speech and Korean open captions easily since portions of each could be repeated several times. It was also possible to measure the time span between the beginning of the speech and the beginning of the Korean captions on the screen at 1 millisecond level.

3. English into Korean media SI

3.1. Media SI

In general, there are substantial differences between general conference interpreting and live SI on TV in which interpreters are located in a studio receiving incoming video and audio signal through a satellite. Mack (2001) contended that TV interpreting demands a much more complex and articulate setting than does conference interpreting. Kurz (1990) also argued that interpreters, while confronted with all the difficulties encountered during "ordinary" conferences, also face particular challenges and time constraints.

One such constraint is that interpreters do not meet the speakers in advance so they cannot share background knowledge with them. This common ground coupled with incoming sound signals plays a pivotal role in understanding messages from the speakers. The context will also help interpreters grasp ideas better when the incoming message is not clear enough. Without this, the anticipation of interpreters is considerably jeopardized and they are much less able to handle unexpected change of direction in the speakers' discourse. Isolation from the speakers at the conference site also poses problems to interpreters for live television SI since they are geographically separated from the speakers. The only channel interpreters have access to is a monitor in front of them, and the flow of information is one-sided. Isolation from the speakers also makes it difficult for interpreters to see the conference room. Visual information including non-vocal signals from speaker and delegate is important for conference interpreting and the visual information provided by TV monitors is considered insufficient by many interpreters (Bühler 1985). Therefore, interpreters have no way to check the reaction from the audience. Technically, it is very difficult to fix sound quality problems in the speakers' message since, in most cases, they are transmitted via satellite. Sometimes the interpreters may lose the signal or experience noise from the sender. Furthermore, choosing Korean expressions to satisfy all general viewers during SI on TV would levy an extra processing load on interpreters. At typical international conferences, the audiences are experts in the given field, and the use of source language jargon during SI is allowed and sometimes recommended. In live media SI, however, this is not the case, and interpreters should choose the most general Korean equivalents to accommodate audiences of different ages and backgrounds. The interpreter must therefore strive to adapt his own verbalization for maximum intelligibility and acceptability by the presumed audience (Viaggio 2001). In addition to the difficulty of finding Korean expressions, the number of people listening to media SI is another psychological burden to interpreters. For ordinary international conferences, the audience in the room cannot exceed thousands, but it can be in the millions for live media SI on TV. The fact that any small mistake in SI is being exposed to an audience of millions on a real-time basis would be an unbearable burden for interpreters. Mack (2001) asserted that the awareness that one's work is accessible to a huge public can have a psychological impact on the interpreter.

3.2. SI of Academy Awards Ceremony

The Academy Awards Ceremony is a ritual where nominees gather and winners are announced just before delivering their acceptance speeches. The ceremony includes

considerable amount of cultural content very different from fact-oriented international conferences typically dealing with such topics as politics or trade. Straniero (2003) pointed out that conference interpreting may be considered mainly as an interprofessional discourse, i.e., the interpreter works with a discourse produced by professionals. Conversely, media interpreting is mainly a professional-lay discourse, i.e., a discourse which addresses an undifferentiated mass audience within an entertainment logic. These unique characteristics therefore make interpreters vulnerable to the conditions in which they are working.

As is the case with other awards ceremonies, final winners are announced only at the last minute, so the interpreters cannot fully predict them or know who the next speakers will be. The nature of the acceptance speech is such that most of the winners are excited or nervous. Their pronunciation may be unclear due to excitement or hard to decode if they are non-native speakers of English. The interpreter is forced to devote much processing capacity to the Listening and Analysis Effort, and therefore slows down production (Gile 1995). Straniero (2003) also cited the accent of non-native speakers as a factor accounting for poor performance of media SI. Furthermore, the winners usually spoke quickly to thank as many people as they could within the limited time. Moreover, the time span is too short for interpreters to understand the main idea of any acceptance speech. This can be challenging for interpreters given that the interpreted version should cause the same reaction the speakers intended. Shlesinger (2003) said unpredictable or loosely contextualized items such as names and numbers increase the density of the discourse and place an exceptional burden on the interpreter. Even worse was the overlapping music indicating the speaker had exceeded the alloted time for their speech. This overlapping music seriously interfered with the understanding of the speakers. As Gerver (1974) pointed out, background noise weakens the monitoring capacity of interpreters, clearly creating an unfavorable situation for interpreters. Clapping from the audience also marred the understanding of incoming messages on the part of interpreters in some cases. Interpreters were also forced to wrap up their SI before loud applause began since the noise was so high that Korean SI was barely audible with the clapping.

Another issue we should consider is the viewers' attitude. As mentioned above, the general viewers of the program are not the audience attending international conferences in their special field. Therefore, they do not have enough knowledge about the way SI is processed and might not distinguish dubbing from SI. They might expect actor-like voice quality and speed in dubbed programs from SI. In this regard, Mack (2001) pointed out that most listeners (viewers) are unaware of the constraints of simultaneous language transfer and are therefore unprepared for active, cooperative listening as conference audiences generally are. He added that listeners expect a voice quality comparable to that of professional TV announcers, with a very short time-lag and close synchrony of text segments in source and target text. The varied needs of viewers were another intrinsic difficulty for interpreters in this type of media SI. Some viewers wish to listen to the original voice of their favorite entertainers without being interrupted by Korean SI. In this case, interpreters immediately receive negative criticism regardless of the quality of their SI.

Next, there can be technical difficulties in the combination of SI and captions. For one thing, interpreters in this mode would not have the extra processing capacity to consider readability for the viewers. Generally, translation for subtitling under-

goes strict proofreading to ensure accuracy and smooth flow. Even though linguistic corrections must be made to ensure the final quality of the captioning (Bartrina and Espasa 2005), neither interpreters nor stenographers have the capacity to edit this high speed input. Neither will have enough time or capacity to correct mistakes.

In spite of these limitations, there are some unique factors that can facilitate SI of the ceremony to a certain degree. First, the content of all the speeches tend to be similar to each other, expressing appreciation and gratitude by naming people the speakers wish to thank. This is quite different from SI for live news or political debate that include a wide range of topics. Another factor is that lists of nominees are announced in advance and interpreters have ample opportunity to prepare themselves for the SI of this event. With these lists, interpreters can study actors or actresses, movies, and related parts. Since accessing the released films is not difficult nowadays, watching the nominated films to become acquainted with the way the actors speak is an important part of the preparation. This accumulated background knowledge will help interpreters grasp the proper nouns, including names, movie titles during SI, thereby helping to improve accuracy.

3.3. Producing Open Captions

According to the Korean companies that produced the captions for this occasion, captioning involved the following steps. First, four stenographers, in two teams of two, received input from simultaneous interpreters. Next, the first two stenographers produced captions, taking three-second turns each, while the remaining two stenographers proofread the captions. They used their own uniquely developed computerized Korean stenograph for this job. The proofread captions were wired to the encoder in the broadcasting TV company. The encoder merges the captions with the picture before airing it. Then viewers can see the open captions with picture and sound on their TV at home.

This process clearly demonstrates that all the steps are processed on a real-time basis, which means the routine quality checking of any subtitle translation might not be applied. This might leave the possibility of poorer translation than normal subtitle translations. Synchronization, which is one of the most important considerations in subtitling, would be very difficult to achieve in this mode since so many parties are involved.

4. Results

4.1. English and Korean Syllables

First of all, each sentence of the original English was compared with its corresponding Korean captioning to analyze the quality of SI. This analysis was based on the total meaning-content, not just on the number of syllables. Then the quantitative analysis was executed beginning with the number of syllables, the EVS. The average number of English and interpreted Korean syllables in each sentence showed a high correlation ($r=0.684$, $p<0.001$) which reaffirms the fact that interpreters usually follow the path made by the speakers. One noteworthy thing is that not all of the English sentences were simultaneously interpreted into Korean. To be more specific, 61.4%

of all original English sentences were successfully interpreted into Korean while the remaining 38.6% were omitted. This is in line with the finding of den Boer (2001: 171) who asked, "Can we really afford to leave so much of a source text untranslated?" Since language is redundant in nature, the number of sentences and syllables cannot be a conclusive criterion for the analysis of omissions; syllables or phrases can be omitted in the interpreted text without necessarily hurting the quality. Indeed, we can find many interpreters who produce a satisfactory target version in simple and neat form by omitting parts of the original. Nevertheless, this number can approximate the relative fidelity between two languages. Straniero (2003) reported an incorrect SI for a Formula One press conference by counting the number of original sentences. Mention should also be made here that these omissions could not be attributed to interpreters' inability; rather, they indicate that interpreters' available processing capacity was saturated at these moments (Gile 2008). Since SI is characterized by an almost permanent temporal overlap of language comprehension and language-production processes (Moser-Mercer 1997), there are inherent limitations to the capacity of the interpreter, no matter how expert and versatile (Massaro and Shlesinger 1997). One of the inherent limitations is the relative number of syllables in English into Korean SI. The total number of English syllables was 1421 and that of simultaneously interpreted Korean was 1497, although 38.6% of the original sentences were omitted. When we calculate the relative ratio, it is 105%, which suggests the interpreters already produced 105% of English syllables. It is in line with the finding of Lee (2000), who reported that Korean interpreters uttered 102% syllables of English speakers. Therefore, we can hypothetically calculate the number of syllables needed if the interpreters in this study were to omit no sentences in their SI. Since they covered only 61.4% of original sentences, we obtain the hypothetical number from the following formula:

$$\text{Number X} = 1497 \times 100/61.4 = 2438$$

This calculation shows that interpreters in this study should have uttered 2438 Korean syllables to leave no sentences omitted in their SI. The 2438 syllables are 172% greater than the original. In the above-mentioned study, Lee (2000) reported that Korean dubbing used 146% of original English speech and Korean translation used 159% more Korean syllables than English original speech. The figures 146% to 159% from the dubbing and translation demonstrate that it is the number interpreters need for the most accurate conversion from English into Korean. Dubbing undergoes translation of the original speech without a severe time constraint by translators, followed by rehearsal and recording; written translations are the same in terms of time factors. Considering these findings, we can easily notice that producing Korean syllables 172% of the original English is impossible to expect from interpreters engaging in English into Korean SI, considering the complex multi-processing of the job.

Apart from the issues of the number of syllables in the case of English into Korean SI, working memory can also provide theoretical support for the inherent limitations of the capacity of the interpreters. Cowan (2000/2001) explains that working memory is conceived as an activated portion of long-term memory, and it would be correct to assume that there are important limitations in human processing. Christoffels, de Groot *et al.* (2003) also argue that when memory tasks are administered in a second language (L2) the results are often found to differ from performance

in the native language (L1). Padilla, Bajo *et al.* (2005) also found articulatory suppression was not clearly present when L1 words were presented, while it was shown when L2 words were presented. Since interpreters in this study listened to their L2, it can be assumed that their processing capacity is more vulnerable than when listening to L1. This indicates that familiarity with a language plays an important role in the interpreters' ability to comprehend and produce.

Gile's Effort Model (1995) can also provide an explanation for this phenomenon. If the total available capacity of an interpreter at a certain point is insufficient to handle listening, memory and production, problems may arise. In this case, listening to the exceptionally fast and unclear speeches of the entertainers might have needed more capacity than the interpreters could provide. Then the interpreter's capacity was overloaded and omissions might occur. He also asserted that "the effects of interpretation constraints on production are stronger in simultaneous than consecutive." Lambert (1988) also argued that, under SI conditions, conflicting activities prevent the interpreter from processing material as deeply as under consecutive conditions. Therefore, no matter how expert and versatile the interpreters are, when it comes to the composite skill of SI, there are inherent limitations in the capacity of the interpreter (Massaro and Shlesinger 1997). Christoffels and de Groot (2004) pointed out that both the simultaneity of comprehension and production and the transformation component are sources of cognitive complexity in SI. Consequently, even the most successful professionals cannot sustain this level of performance if extreme conditions prevail for too long (Moser-Mercer 2000/2001). Here, one might ask if the stenographers might have lost some of the interpreted version during the caption producing process. Since den Boer (2001) asserts that stenographers can produce up to 1000 letters a minute, it becomes clear that no words were omitted by the stenographers once SI was produced by the interpreters.

4.2. Omissions

When the length of omitted and interpreted English sentences was compared, the length of interpreted sentences (14 syllables) was statistically longer than the omitted sentences (10 syllables, $p=0.000$). This is in line with Massaro and Shlesinger's (1997) assertion that the less information one has about the text at any given point, the more tenuous one's comprehension of it.

There can be several ways to interpret the reasons for this. The first is a temporal factor. Since the average word-per-minute (wpm) of the speeches was 138 wpm, it amounts to about 193 syllables. When we calculate the hypothetical speed of the incoming English message, 14 syllables will come in to interpreters within 4.3 seconds while the shorter sentences of 10 syllables will take 3.1 seconds. Since the average length of EVS, a time-lag between the moment an incoming message is perceived by an interpreter and the moment the interpreter produces his converted version of the segment, in English into Korean SI was found to be 3 seconds (Lee 2002), the above mentioned 3.1 seconds seems not much longer than average EVS in English into Korean SI. It means that these short sentences might be comprehended under extreme multi-processing while delivering the converted message of the previous sentence. In other words, it is highly possible that those short sentences, due to the extreme multi-processing, would not be given enough necessary processing capacity by the

interpreter. This is because conference interpreters usually begin to deliver their converted message quickly once they understand even part of incoming message. This supposition partly explains why the omitted sentences were shorter than interpreted ones.

Another way of supporting this logic would be the concept of "tail-to-tail span (TTS)." Unlike EVS, TTS means the time-lag between the end of a speaker's sentence and that of interpreter's converted sentence (Lee 2003). This is the span of time the interpreter is lagging behind the speaker in winding up delivery of the interpreters' rendition. It was found that TTS increased when Korean interpreters uttered relatively more syllables than the speakers; long EVS induced long TTS and the reverse was also true (Lee 2003). He also found that long TTS hurt the quality of the sentence being interpreted likely because longer TTS indicates that the interpreter needs more time and processing capacity for the sentence in question. Therefore, when an interpreter stays with a certain sentence longer than allowed, either due to the problem of understanding or finding Korean equivalents, he will not have enough time and processing capacity for the next incoming message. This will be especially true when the next incoming message is incidentally short enough to be in the TTS of the previous sentence. In this case, the interpreter might spend his entire processing capacity on the previous sentence, reserving a little or no capacity for the next short sentence. Under these circumstances, the next sentence might be omitted in SI.

The main point here is that the multiprocessing of SI including understanding, converting, and producing by interpreters is not a single loop. More than one set of multiprocessing tasks might be operating in the interpreters' processing depending on the length of incoming sentences. In other words, the Effort Model or working memory should encompass not only one loop of processing for incoming message but more than one sentence concurrently.

4.3. EVS and TTS

EVS is one of the most salient variables that shows interpreters' time management during SI. During this EVS, interpreters carry out concurrent processing of comprehending, converting, planning, uttering and other unknown processing. In this case, the time needed for producing open captions will be part of the EVS as well.

Computer assisted analysis of SI revealed that the average EVS in this data was 7.24 seconds.[1] This means that viewers could see the Korean captions appearing on the TV screen 7.24 seconds after hearing the beginning of the English sentence. This 7.24-second gap is much longer than average EVS of 3.13 seconds found in English into Korean SI in general (Lee 2006), even if we consider the time for caption production. To find the difference more clearly, this long EVS was compared with the EVS of SI for the Academy Awards in 1997 and 1999 with SI alone (2.50 sec). Statistical treatment showed that the EVS of SI through captions was longer than mere SI ($p=0.000$).

Since the EVS of SI alone in 1997 and 1999 was 2.50 seconds, we can safely assume that interpreters in this study used some 2-3 seconds for their SI while stenographers and proofreaders used the remaining 4-5 seconds before the captions were seen on the screen. Although the additional time to produce captions by the stenographers and proofreaders excessively lengthened the final EVS, the size of 7.24

second EVS seems to be much longer than normal EVS in SI. The main problem stemming from these exceptionally long EVS is that no synchrony between the original English dialogue and the Korean captions on the screen is achieved. Bartrina and Espasa (2005) asserted that one of the audiovisual text's main features is the synchrony between verbal and non-verbal messages, which is essential. This long EVS was already a problem in earlier experiments noted by den Boer (2001) who pointed out that:

> this kind of delay is inevitable in live subtitling and can be very disturbing to two groups of viewers. Those who are used to 'normal' subtitles will expect them to synchronize while viewers who have no experience with subtitles, but understand enough of the source text to be able to follow it, will be disturbed by the lack of synchronization (den Boer 2001: 171).

Franco and Araujo (2003) also argued that the fact that speech and image hardly synchronize is a problem. Whenever speech depended on images, like the events in this study, lack of synchronism impaired the reception of closed subtitles and the understanding of the contents.

Besides the synchrony issue, another finding concerning EVS was that the long EVS of a sentence induced omission of the next incoming sentence in SI. The length of previous EVS of interpreted English sentences averaged 6.67 seconds while that of omitted sentences was 8.28 seconds ($p=0.000$). This statistically significant difference indicates that when interpreters began to produce the converted version too late, the quality of the following sentences was often sacrificed. In this regard, Van Dam (1986) succinctly described that "when, for whatever reason, we fall behind the speaker, the amount of information backlog is proportional to the increasing distance between the speaker and the interpreter. In other words, the further behind we fall, the more information we must store in short-term memory." On the other hand, Massaro and Shlesinger (1997) contended that novice interpreters tend to favor short EVS, failing to make effective use of the text-schematic knowledge available for longer EVS. Although both assertions seem logical, one of the most important maxims in this regard will be how to keep the balance between the benefit of accumulating more information from long EVS, and cognitive breakdown in the extreme multi-processing stemming from the EVS. Since no interpreters engaging in SI would wait for the next incoming sentence without delivering the interpreted version for the current sentence, beginning too late means the second sentence pushes into interpreters' ears even before they finish delivering the first sentence. Thus the interpreter's capacity diminishes and the final quality suffers if unable to handle the demand. This is exactly what a long TTS illustrates in the previous section.

Regarding this, Daneman and Green (1986) claimed that speaking involves a highly complex and skillful coordinating of processing and storage requirements. Speakers must rely on other word selection heuristics that likely tax working memory. Christoffels, de Groot *et al.* (2003) asserted that if finding an appropriate word for a concept during SI takes a long time, it is likely that the interpreting process may break down due to the loss of valuable processing capacity and time. Cowan (2000/2001) said that if an item was not refreshed before it decayed from memory, it would be lost. Although many fundamental skills could be relevant, one of them is the control of attention. Speedy retrieval of equivalents will be one of the most important skills and, as mentioned earlier, failing to complete the current sentence as quickly as possible

and move on to the next incoming sentence will lead to overloading of processing capacity. Therefore, successful interpreters are those who manage to maintain the delicate balance between their cognitive strengths and weaknesses and who have developed coping skills (Moser-Mercer 2000/2001). Interpreters also should be able to distribute their limited information processing capacity optimally to several modules of processing.

To confirm the negative role of a long TTS, the span during which interpreters lag behind speakers in concluding a TL utterance, the average length of TTS of this data was examined. The results showed that the average length of TTS was 7.52 sec (SD 6.25). This means the Korean captions for a certain English sentence are on the screen until 7.52 seconds after the audio is gone. Here, once again, the length of TTS seems quite long since the viewers have to see the captions remaining on the TV screen 7.52 seconds after the segment has finished. The next incoming dialogue will proceed with the captions for the previous dialogue on the screen. This means that the synchrony was completely neglected. As repeatedly mentioned, the reason for this is the extra time needed to produce the captions after the completion of SI in case of caption TTS.

When the length of caption EVS (7.24 sec) and caption TTS (7.52 sec) of this data was compared, there was no statistically significant difference between them ($p=0.35$). To see the relative length of TTS, the 1997 and 1999 Academy Awards Ceremony TTS and caption TTS in this study were compared. As was expected, caption TTS (7.52 sec) was much longer than previous Academy Awards Ceremony TTS that employed SI only (4.70 sec, $p=0.000$).

4.4. EVS and TTS of CNN Captions

In order to see the relative size of the EVS and TTS of the Academy Awards Ceremony, they were compared with the TTS and EVS of a CNN news program which employed no SI. It was found that the average EVS of CNN was 4.13 seconds, TTS was 3.32 and they were statistically different ($p=0.000$). Since the TTS is shorter than EVS, it can be fairly assumed that the caption producers for CNN were trying to shorten the time to make captions, although their captions appeared 4.13 seconds after the beginning of the original speech. This might have been possible since there was only one party, stenographers, involved.

When the EVS of CNN (4.13 sec) was compared with caption EVS (7.24 sec), caption EVS was statistically longer than the CNN case ($p=0.000$) as was expected. This may be due to the extra involvement of interpreters prior to the stenographers. Next, the TTS of both cases was examined, as in the above case, caption TTS (7.52 sec) was much longer than CNN TTS (3.32 sec, $p=0.000$) for the same reason mentioned earlier.

5. Users' Preferences

To gain users' feedback, two versions of the Academy Awards Ceremonies were shown to 79 students studying English at a university in Korea. The acceptance speeches with SI through captions that were analyzed in this study were one of the video clips; the other was the 1999 ceremony which only used SI. After watching the

two video clips consecutively, the students completed questionnaires. One limitation regarding the subjects is that they were not the general public, and not necessarily representative of most Korean viewers. Many Internet posts by Korean viewers, however, expressed similar observations so there was apparently little difference in opinion between the two groups.

Firstly, the subjects almost totally preferred SI through captions (in 2005) rather than the SI only mode (in 1999). Ninety-four percent of the subjects replied that they liked the SI through captions and only 4% wanted SI alone mode. This generally concurred with the opinion posted by Korean viewers on the Internet. Some of them expressed their satisfaction on changing SI alone mode to SI through captions: "The live coverage of Academy Awards Ceremony is getting better. I was much more comfortable as SI was not heard at all" (2007); "This year was better than last year. First of all, I feel pleased as SI was not heard" (2006). This suggests that the new mode is relatively well accepted by Korean viewers.

Secondly, the students were asked to choose one of the two modes for movies they prefer, namely, Korean dubbed movies or English captions with original English dialogue. Some 98% of the respondents favored caption movies rather than Korean-dubbed movies. One reason for this clear preference seems to be that the students like to hear the original voices of actors and actresses. Another important aspect would be that they could check their understanding of the oral input with the synchronized visual input, i.e., the Korean caption on the screen. Franco and Araujo (2003) pointed out that listening viewers are used to watching films with open subtitles and their comprehension is facilitated by captions.

The next item "I like the caption mode since I can listen to the original voice (2005)," reinforced the above-mentioned point. A total of 97% of the subjects answered that they like the caption mode since it gives them direct access to the original English voice undisturbed by Korean SI. This is one of the most tangible factors making caption mode more popular than SI. In fact, the disturbance coming from the overlapping of Korean SI with original voice was one of the main complaints by general Korean viewers watching the Academy Awards Ceremonies in Korea.

This point was reaffirmed when the students were asked if "Overlapping Korean SI and original English voice is confusing" (SI alone mode in 1999). Some 87% replied they were being disturbed by the voice over. This is understandable knowing that viewers wish to focus their attention on the entertainers' original voice. In this situation, adding the overlapping Korean SI with excited original speeches must have been unpleasant for Korean viewers.

When they were asked "I read Korean caption only when watching SI with captions," contrary to the earlier expectation, 87% of the subjects replied negatively. This means that they rely more on the incoming oral signal than on Korean captions on the screen. A major reason for this would be that the subjects were students studying English and their foreign language ability was strong enough to process the incoming English message. Another factor would be that, as mentioned above, captions and Korean SI version are not offered simultaneously. Instead there is a more than a seven second gap and subjects had no choice but to process incoming oral messages first rather than waiting 7 seconds to read the Korean captions. It is possible therefore that different groups with varying levels of English fluency may have answered that they depend mainly on reading incoming Korean captions.

The next question was, "Reading the Korean captions impaired the understanding of incoming English." Unlike SI, where only one mode of oral input is necessary, SI through open captions provides both visual and acoustic input for viewers. In spite of the seemingly complex information processing, 69% of the subjects reported having no difficulties, while 31% said yes. Besides the long EVS of captions, another possible reason is that the subjects totally ignored Korean captions on the screen while relying exclusively on listening.

The next statement in the questionnaire was, "It is irritating to see the Korean captions appear on the screen far after the English." Some 76% of the subjects admitted that the long time lag between the original and the captions was annoying. This means that although the subjects prefer SI through captions, they are not completely happy with the long time-lag. The same problem was also found by den Boer (2001), and should be a subject of further study in the future.

To see whether they appreciate the relatively short EVS of SI alone mode, the subjects were asked "It is good to listen to SI since it is real time processing." A marginally larger number (54%) replied they liked the concurrency of SI while 46% said no.

The next question was, "SI is good since I don't have to read the captions." More than half of the subjects (56%) rejected this as being a merit of SI that it requires no visual reading of the captions. Hence more than half of the students did not appreciate the advantage of SI as saving the extra effort to read the captions.

Next, subjects were asked if "The quality of captions was satisfactory." Some 83% replied they were satisfied with the quality of captions from SI through caption mode. Surprisingly, when the same question was asked regarding the quality of SI alone mode, only 57% replied they were satisfied with the quality of SI in this mode. SI and captions are the same product in terms of the process since they are both the results of SI. Therefore, we cannot assume that the quality of SI alone mode was inferior to the SI through captions. To examine this paradoxical result, a meaning-based qualitative analysis was carried out comparing all the original sentences with SI in SI alone mode (in 1999), and the original with the SI through captions mode (in 2005). A t-test was run and revealed that there was no statistical difference in terms of the quality of the two versions ($t=-0.049$, $p=0.961$). Considering that the subjects were not professional interpreters or linguists, we can presume they do not have the capacity to form a judgment about the concurrently incoming original and SI version on a real time basis. Therefore, it can be safely assumed that the main factor influencing them was how the subjects received the converted message, SI or captions. This again means that the subjects preferred captions where they can hear the original voice to SI that sacrifices the original voice.

The next question was "What will be the most appropriate form of interpretation for future Academy Awards Ceremonies?" Some 93% preferred captioned SI to SI alone mode. However, when they were asked, "What will be the most appropriate way of interpretation of future Academy Awards Ceremonies for general viewers?" their answer was quite different. The portion that preferred captions dropped sharply to 37%, and 63% said SI alone mode is good.

6. Conclusion

As Massaro and Shlesinger (1997) pointed out, one of the reasons why progress in the understanding of SI has been slow is that SI behavior is complex. In this context, the temporal data of EVS and TTS, the results of the computer-aided analysis, can be a small stepping-stone that might partly explain the process of SI. In-depth examination of SI data revealed that some of the original English sentences were not interpreted into Korean presumably due to the extreme multi-processing on the part of the interpreters. Statisticals showed that the length of omitted English sentences was shorter than interpreted ones. It was also found that EVS in this data was 7.24 seconds, which was statistically longer than the 2.5 second-long EVS of the previous coverage of the same events using only SI without captions. TTS was also fairly long, up to 7.52 seconds, and was also longer than that of the previous coverage through SI alone. Another important finding was that longer EVS often meant the next incoming sentences were omitted in SI.

Questionnaires indicate that SI through captions seems to be a sustainable way to televise the ceremonies. This merit of captions is reaffirmed by the fact that rebroadcasting of the ceremony usually employed only Korean captions that were undoubtedly retranslated and polished. Furthermore, the entire ceremony was aired through captions beginning in 2006. While the questionnaire subjects were mainly students studying English, the general public also reacted favorably to this new mode. This reaction seems mainly influenced by the fact that Korean viewers are accustomed to watching foreign movies with Korean open captions. Although the long time-lag between original dialogue and the appearance of captions was viewed as a problem, the desire to listen to the original voice of their favorite entertainers seems to compensate for the inconveniences stemming from the long EVS. Adapting captions in place of the interpreters' voice seems to solve the problem of the irritating overlapping of SI and original voice for Korean audiences. When SI was used until 2005, though for presenters only, viewers could not turn off either the SI or the original voice. With SI through caption mode, on the other hand, Korean viewers can listen to the original English remarks by not reading incoming Korean captions or those who cannot understand English can focus their attention on incoming Korean captions even though they are not synchronized.

In spite of favorable reaction from viewers, reducing the long time-lag between original dialogues and Korean captions, in reality, would not be an easy task for either interpreters or stenographers. It would be fair to assume that professional interpreters are already maintaining optimal EVS under this extreme multiprocessing of SI. Lee (2002) asserted that, at least for English into Korean SI, maintaining optimal EVS is extremely important since too short EVS might end up with a clumsy target language due to the lack of comprehensive understanding of source text, while a too long EVS will lower the quality of SI of the sentence as well as the incoming one too. For interpreters under this circumstance, reducing the EVS further to help following caption production might disrupt the whole processing and might result in poor performance in their SI. Lambert, Daró *et al.* (1995) also suggested that "the explicit request to consciously focalize one's attention on one of these SI tasks not only does not improve the interpreter's performance, but is even detrimental to general, or more natural production." TTS, on the other hand, seems to provide a little room for

improvement on the part of interpreters. As Kurz (2004) pointed out, since interpreters working for the media must endeavor to be very quick without 'hanging over' excessively after the speaker has finished, interpreters for this kind of ceremony should try to reduce TTS. However, it is very unlikely that stenographers would be able to reduce their time to produce captions since the current time span would be the result of the most efficient processing. With both parties already trying their utmost to reduce the time-lag, there will not be much room for improvement except possible reduction in the length of TTS. Therefore, on a long-term basis, if we can introduce an automated machine proofreading system and utilize an enhanced voice recognition technology in the future, the time for caption making could be cut down to a certain degree. This is highly possible when we consider the fact that researchers, including a Korean one, are actively involved in creating the automatic production of subtitles (Melro, Oliver *et al.* 2006; Daelemans, Hothker *et al.*; 2004; Yuh 2006).

Another possible technical recommendation we can consider at this point would be a Multichannel Television Sound (MTS). MTS is the method of sending an additional sound signal to an existing sound signal. Therefore, two audio channels, such as original English remarks and Korean SI, can be sent to TV viewers who can choose one of them. This method is similar to what Kurz (1990) reported with an Austrian case. Sending original English in one channel and Korean SI in another channel will clearly solve the problem of overlapping between SI and original dialogue. Lee (2000) suggested this mode several years ago to solve the problem of overlapping of SI and English remarks. A Korean viewer also preferred this mode: "Since it is live coverage, captions would be difficult to adopt, but I believe the time has come to consider MTS" (1997). Since some English dramas are already being aired with MTS, English and Korean dubbing, in Korea, adopting MTS for the Academy Awards Ceremony would not be a difficult job. Viewers could then choose one of them, greatly enhancing the possibility of satisfying both types of viewers.

One limitation of this study, however, is that the questionnaire subjects, as mentioned before, are students studying English. Although some general Korean viewers expressed the same reaction, we cannot assert that the results of this questionnaire represents the majority of Korean viewers watching this ceremony with SI through open Korean captions. Future work might therefore involve a broad sampling of general viewer impressions on the use of this new method, or comparing those of university students vs. general viewers. This might yield a more comprehensive evaluation of this new technology. If interpreters and related researchers in the broadcasting field successfully enhance the quality of SI through captions mode by resolving these limitations, this new technology will be a revolutionary tool of communication in the future. This is especially noteworthy because the huge size of general public TV viewing is incomparably greater than the numbers attending general international conferences. As asserted by den Boer (2001), academic researchers should provide serious input towards improving this new tool.

NOTES

* The writer expresses deep appreciation to the Monterey Institute of International Studies for providing an office and access to the library while this article was being written.

1. Although the term EVS does not fit exactly in this case, since we are talking about the time span between the beginning of original speech and the beginning of the captions on the screen, we will use EVS to describe the span.

REFERENCES

Bartrina, Francesca and Espasa, Eva (2005): Audiovisual Translation. *In*: Martha Tennent, ed. *Training for the New Millennium Pedagogies for translation and interpreting.* Amsterdam: John Benjamins, 83-100.

Bühler, Hildegund (1985): Conference Interpreting: A Multichannel Communication Phenomenon. *Meta*. 30(1):49-54.

Christoffels, Ingrid and de Groot, Annette (2004): Components of Simultaneous Interpreting: Comparing Interpreting with Shadowing and Paraphrasing. *Bilingualism: Language and Cognition*. 7(3):227-240.

Christoffels, Ingrid, de Groot, Annette and Waldrop, Lourens (2003): Basic Skills in a Complex Task: A Graphical Model Relating Memory and Lexical Retrieval to SI. *Bilingualism: Language and Cognition*. 6(3):201-211.

Cowan, Nelson (2000/2001): Processing Limits of Selective Attention and Working Memory. *Interpreting*. 5(2):117-146.

Daelemans, Walter, Hothker, Anja and Tjong Kim Sang, Erik (2004): Automatic Sentence Simplification for Subtitling in Dutch and English. *In*: *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, 1045-1048.

Daneman, Meredyth and Green, Ian (1986): Individual Differences in Comprehending and Producing Words in Context. *Journal of Memory and Language*. 25(1):1-18.

den Boer, Corien (2001): Live Interlingual Subtitling. *In*: Yves Gambier and Henrik Gottlieb, eds. (*Multi)media Translation. Concept, Practice, and Research.* Amsterdam: John Benjamins, 167-172.

Franco, Eliana. C and Araujo, Vera Santiago (2003): Reading Television. *The Translator*. 9(2):249-267.

Gerver, David (1974): The Effects of Noise on the Performance of Simultaneous Interpreters: Accuracy of Performance. *Acta Psychologica*. 38(3):159-167.

Gile, Daniel (2008): Local Cognitive Load in Simultaneous Interpreting and its Implications for Empirical Research. *Forum*. 6(2):59-78.

Gile, Daniel (1995): *Basic Concepts and Models for Interpreter and Translator Training.* Amsterdam: John Benjamins.

Kurz, Ingrid (2004): Television as a Source of Information for the Deaf and Hearing Impaired: Captions and Sign Language on Austrian TV. *Meta*. 49(1):81-88.

Kurz, Ingrid (1990): Overcoming Language Barriers in European Television. *In*: David Bowen and Margareta Bowen, eds. *Interpreting-Yesterday, Today, and Tomorrow*. New York: State University of New York at Binghamton, 168-175.

Lambert, Sylvie, Daró, Valeria and Fabbro, Franco (1995): Focalized Attention on Input vs. Output during Simultaneous Interpretation: Possibly a Waste of Effort. *Meta*. 40(1):39-46.

Lambert, Sylvie (1988): Information Processing among Conference Interpreters. *Meta*. 33(3):377-387.

Lee, Taehyung (2006): A Comparison of Simultaneous Interpretation and Delayed Simultaneous Interpretation from English into Korean. *Meta*. 51(2):202-214.

Lee, Taehyung (2003): Tail-to-Tail Span: A New Variable in Conference Interpreting Research. *Forum*. 1:41-62.

Lee, Taehyung (2002): Ear Voice Span in English into Korean Simultaneous Interpretation. *Meta*. 44(4):596-606.

Lee, Taehyung (2000): Temporal Aspects of Live Simultaneous Interpretation on TV: A Case of Academy Award Ceremonies. *Conference Interpretation and Translation*. 2:85-110.

Mack, Gabriele (2001): Conference Interpreters on the Air: Live Simultaneous Interpretation on Italian Television. *In*: Yves Gambier and Henrik Gottlieb, eds. (*Multi)media Translation. Concept, Practice, and Research.* Amsterdam: John Benjamins, 126-132.

Massaro, Dominic and Shlesinger, Miriam (1997): Informational Processing and a Computational Approach to the Study of Simultaneous Interpretation. *Interpreting*. 2(1/2):13-53.

MELERO, Maite, OLIVER, Antoni and BADIA, Toni (2006): Automatic Multilingual Subtitling in the eTITLE Project. *In*: *Proceedings of Translating and the Computer 28*. London: Aslib.

MOSER-MERCER, Barbara (1997): Methodological Issues in Interpreting Research: An Introduction to the Ascona Workshops. *Interpreting*. 2(1/2):1-11.

MOSER-MERCER, Barbara (2000/2001): Simultaneous Interpreting: Cognitive Potential and Limitations. *Interpreting*. 5(2):83-94.

PADILLA, Francisca, BAJO, Maria Teresa and MACIZO, Pedro (2005): Articulatory Suppression in Language Interpretation: Working Memory Capacity, Dual Tasking and Word Knowledge. *Bilingualism: Language and Cognition*. 8(3):207-219.

SHLESINGER, Miriam (2003): Effects of Presentation Rate on Working Memory in Simultaneous Interpretation. *The Interpreters' Newsletter*. 12:37-49.

STRANIERO, Sergio Francesco (2003): Norms and Quality in Media Interpreting: the Case of Formula One Press Conferences. *The Interpreters' Newsletter*. 12:135-174.

VAN DAM, Ine (1986): Strategy of Simultaneous Interpretation: A Methodology for the Training of Simultaneous Interpreter. *In*: Manfred KUMMRER, ed. *Building Bridges*: *Proceedings of the 27th Annual Conference of the American Translators Association*. Medford: Learned Information, 441-456.

VIAGGIO, Sergio (2001): Simultaneous Interpreting for Television and Other Media: Translation Doubly Constrained. *In*: Yves Gambier and Henrik GOTTLIEB, eds. (*Multi)media Translation. Concept, Practice, and Research*. Amsterdam: John Benjamins, 23-33.

YUH, Sanghwa (2006): *Multilingual Machine Translation of Closed Captions for Digital Television with Dynamic Dictionary Adaption*. Doctoral dissertation, unpublished. Seoul: Graduate School of the Sogang University.