Meta
Journal des traducteurs
Translators' Journal

Studying Style in Simultaneous Interpretation

Sane M. Yagi

Citer cet article

Résumé de l'article

Divers outils ont été développés pour faciliter l'analyse quantitative du style d'interprétation, un sujet qui n'a été que vaguement traité jusqu'à maintenant. Ces outils peuvent permettre la recherche sur des questions telles que "Comment procède l'interprète pour partager le message original reçu ?", "Jusqu'à quel point reflète-t-il l'orateur original?" et "Jusqu'à quel point pratique-t-il une modification? ". De plus, un instrument adaptif de surveillance est mis au point pour faciliter la représentation graphique du développement linéaire d'un discours et de son interprétation équivalente simultanée. Non seulement nous permet-il d'évaluer la convergence et la divergence entre les deux discours, mais aussi et surtout de porter un jugement sur le tempo d'un interprète en caractérisant la périodicité (courte ou longue) de son discours, ainsi que sur son sang-froid ou son trouble en décrivant le degré de cohérence et d'aisance dans le discours.

Studying Style in Simultaneous Interpretation

SANE M. YAGI
Sultan Qaboos University, Oman

RÉSUMÉ

Divers outils ont été développés pour faciliter l'analyse quantitative du style d'interprétation, un sujet qui n'a été que vaguement traité jusqu'à maintenant. Ces outils peuvent permettre la recherche sur des questions telles que « Comment procède l'interprète pour partager le message original reçu ? », « Jusqu'à quel point reflète-t-il l'orateur original ? » et « Jusqu'à quel point pratique-t-il une modification ? ». De plus, un instrument adaptif de surveillance est mis au point pour faciliter la représentation graphique du développement linéaire d'un discours et de son interprétation équivalente simultanée. Non seulement nous permet-il d'évaluer la convergence et la divergence entre les deux discours, mais aussi et surtout de porter un jugement sur le tempo d'un interprète en caractérisant la périodicité (courte ou longue) de son discours, ainsi que sur son sang-froid ou son trouble en décrivant le degré de cohérence et d'aisance dans le discours.

ABSTRACT

Several tools are developed to facilitate the quantitative analysis of interpretation style, a matter that has hitherto been discussed only in vague terms. These tools can allow the investigation of questions such as: How does an interpreter divide up a source language input, to what extent does he mirror a source language speaker, and to what degree does he practise reformulation? Furthermore, an adaptive monitoring instrument is devised to facilitate the graphic representation of the linear developments of a source language discourse and its simultaneous interpretation equivalent. Not only does it allow the assessment of convergence and divergence between the two discourses, but this also permits commenting on an interpreter's tempo by characterising the narrow and broad periodicity within his discourse, and on his composure and tribulation by describing his consistency and fluency in the discourse.

Introduction

Style in simultaneous interpretation (SI) is one of the fundamental aspects of the interpretation performance. It concerns those features that constitute the method of interpreting rather than the substance of the information being rendered. Therefore, discussing it will inevitably lead to a consideration of interpretation techniques, strategies, and cognitive tasks.

The question of what constitutes good interpretation touches the very essence of SI style. Since there is a lack of consensus among theorists and practitioners alike as to the norms of SI and the quality criteria that interpreters need to observe, it is a foregone conclusion that there are no steadfast stylistic norms and criteria. In fact, it is not wide of the mark to say that there is no reliable method for assessing SI quality

or style. Pöchhacker (1994: 235) acknowledges this fact saying, "We seem to know what the [SI] product should be like, but we are less sure about a method for establishing what a particular product is like in a given situation. Quite obviously, researchers, teachers, and trainees need to have a method for looking at the product."

The International Association of Conference Interpreters defines quality in conference interpretation as "that elusive something which everyone recognises but no one can successfully define" (AIIC 1982). In the same vein, SI style is an elusive concept that interpretation users can sense but cannot verbalise. They can pass general value judgements about whether an interpretation performance is up to standard or not, but then they can neither agree in their priorities, tastes, and comprehension, nor in what they hold as standard. Confirming that quality assessors tend to use different yardsticks, Seleskovitch (1986) and Gile (1991a) have concluded that interpretation users are liable to give misleading evaluations because they have different expectations.

As was evident in the 1994 Turku conference on interpreting, researchers have been attempting to identify "just what it is that makes for excellence in [SI]" (Shlesinger 1997: 123). Participants in a workshop on simultaneous interpretation quality in that conference discussed the issue from three perspectives: market perspective, research perspective, and didactic perspective. They were critical that subjective criteria tend to be variable, so they expressed the need for identifying objective criteria by which quality can be assessed.

Due to the inconsistencies of quality perception among those whose discourse is the subject of interpretation as well as among those for whom the interpretation is made, there is an increasingly louder plea for more objective methods of quality assessment. There is a clear desire to pin down that which makes an interpretation performance better than another and to develop more tangible criteria for assessing the evasive aspects of quality. The need for objective methods of assessment was accentuated by several studies, among them that conducted by Gile (1991b), which found only a weak correlation between 'satisfactory quality' as perceived by a given speaker and the fidelity, linguistic acceptability, clarity and/or terminological accuracy of the translator's output.

As to the development of objective methods of SI assessment, Cartellieri (1983: 213) suggests that we find quantitative features that may eventually develop into qualitative criteria. Pöchhacker (1994: 234) advocates this idea, suggesting that the question then arises as to "how we should best go about defining and analysing the text produced by the interpreter as an 'objective', that is, physical reality. What are the textualised parameters and variables underlying judgements of quality in simultaneous interpreting, and how can they be measured and quantified in a corpus of texts?"

Quantitative Aspects of SI Style

One can deal with the simultaneous interpreted discourse on many levels: pragmatic or semantic, intra- or inter-textual, qualitative or quantitative, etc. In Pöchhacker's (1994: 238) words, the interpreted discourse is a "multi-faceted whole within a communicative situation."

Pöchhacker (1994: 236) asserts that "In simultaneous interpreting, the text as such is [...] a multi-parametric semiotic whole, which, in its full complexity, often

defies description," so he proposes "a text model with constituents in both the auditive and the visual channels, on a 'verbal-paraverbal-kinesic continuum.'" In his model of the SI 'audio-visual text', Pöchhacker suggests that "one can derive a number of textual features or parameters, such as slips and structure shifts in verbal production, voiced hesitation markers, peculiarities of voice quality and articulation, the use of (pictorial or verbal) visual information (such as slides), as well as prosodic and/or paraverbal features."

To these textual constituents, Pöchhacker (1994: 236) adds quantitative stylistic aspects of the interpreted discourse as manifested in "temporal phenomena like speed, pausing, and rhythmical pattern, which are often dominant in shaping the overall impression of a spoken text." It is these phenomena, and hesitation, that determine the audience's perceived degree of fluency and perhaps competence in an interpreted discourse. Other quantifiable aspects of performance affect the interpreter's time management ability and, to some degree, accuracy. For instance, delay and discourse chunking play an important role in how interpreters manage their time, and that may indirectly affect whether they resort to Gile's (1995) "law of least effort" and practise non-tactful omission.

The temporal phenomena Pöchhacker talks about have also been recognised as quantifiable gauges of SI performance by earlier researchers such as Goldman-Eisler (1968), Barik (1969), Gerver (1969), Chernov (1969), etc. The statistical tools developed in the present paper will be used to study only two quantitative aspects of SI style: fluency and chunking, which indirectly touch on a third aspect: lag. We need to be mindful, however, that discussing SI quality and style within a quantitative framework lays no exclusive claim to objectivity. A linguistic analysis of interpreted discourse may prove to be more reliable in quality assessment; yet if it is possible to correlate SI quantitative features with some content-based criteria, then quantitative analysis will indeed be indispensable. It will be used for substantiating and strengthening any judgement based on the linguistic content of an interpretation. Let us for now consider these stylistic quantitative aspects: fluency, chunking, and lag.

Fluency

A conference speaker and audience who do not speak the language of one another can only evaluate the simultaneously interpreted discourse by its form. They assess the performance of an interpreter by the fluency and nativelikedness in their TL discourse. Kopczynski (1994) conducted a survey among conference speakers and attendees to identify what they viewed as elements that contribute to quality in a simultaneously interpreted discourse. He found that both groups ranked fluency and style third on their list of priorities after content and terminological precision. Skilled interpreters (e.g., Jones 1998: 130) warn novices that "they should not make artificial pauses in the middle of a sentence because they are thinking of what to say next or are waiting for extra input from the speaker." They observe that audiences sometimes expect the interpreter to "keep up a continuous flow of sound in the booth" worrying about missing out on part of the SL discourse (Jones 1998: 128). They stress that "the constant objective of the interpreter is to provide a correct translation of the original in a form that sounds as natural and as authentic as possible in the target language: the audience should not feel they are listening to a translation" (Jones 1998: 90).

Not only does an interpreted discourse need to be fluent to earn the SI practitioner approval from their partners in the communication process, the speaker and audience, but also they need to imitate the tempo and intensity of the speaker's voice according to Kopczynski (1994). He found out that the majority of his questionnaire respondents had considered important that the interpreter assume a ghost role, ie, imitating the speaker. Although the validity of this conclusion is doubtful, as Kopczynski himself indicated, it points to the importance that interpretation users place on the method of TL discourse delivery. Because of the clear relevance of fluency to perceived interpreter competence, it is sound to consider it an aspect of SI style. Fluency is immediately relevant to the method of SI delivery; it represents the fluidness and smoothness of SI delivery. Therefore, it should not be controversial to consider fluency an aspect of interpretation style.

We can study fluency quantitatively if we succeed in identifying the elements that contribute to a seemingly effortless, fluid, and smooth interpretation. There is no doubt that false-start and hesitation ridden interpretation is non-fluent. And so is an interpretation with incomplete sentences, long-drawn-out delays, and a large volume of inactivity. Since these can be readily identified, fluency ought to be quantifiable. For instance, an interpreted discourse which consists of 40% pausing and has 10 false-starts, 15 hesitations, 13 incomplete sentences, and 7 instances of extended delays, is certainly less fluent than one that has 30% pausing, 5 false-starts, 5 hesitations, 3 incomplete sentences, and only 4 instances of extended delays. All of these properties of fluency are quantifiable; therefore, it should not be contentious to claim that fluency itself is one aspect of SI style that is also quantifiable.

Chunking

A fundamental aspect of SI style that affects the interpreter's ability to cope with the seemingly unending flow of SL discourse is chunking. This is a coping strategy that interpreters use to divide up TL long stretches of discourse into chunks of manageable size. Gile (1995: 196) advocates chunking as a strategy that "can save short-term memory capacity requirements by unloading information from memory faster." Similarly, Jones (1998) urges SI trainees to use a technique based on chunking that he coined the 'salami technique.' It involves slicing up long sentences into a number of shorter ones. He says, "The salami technique is particularly useful when working from languages that have a natural tendency to long, complicated sentences, particularly those that can have Russian doll-like structures, with one subordinate clause fitting in another one, which in turn fits into a main clause (such as the so-called Schachtelsätze in German)" (Jones 1998: 102).

Chunking is a quantifiable stylistic feature of SI performance. By comparing the number of bursts in the SL discourse with those in TL, one can get an index that tells the magnitude of difference between the speaker's and interpreter's methods of chunking. The value of such an index lies in that it can enable researchers to make a statement about the interpreter's reformulation strategy; when the interpreter's chunks are more numerous than the speaker's, this is evidence that they practised the 'salami technique' and fragmented SL chunks. The magnitude of the difference indicates the extent of fragmentation. If, on the other hand, the interpreter's chunks were fewer than the speaker's, then clearly one would think the interpreter reformulated

the SL discourse by joining small chunks into bigger ones. The combination of such conclusions enable researchers to make an objective statement about SI style, a statement that is bounded only by the language-cultures of SL and TL.

Lag

Another important aspect of SI style is the way interpreters cope with the simultaneous tasks of listening and speaking, decoding and encoding, and all ensuing cognitive functions. Their coping strategy can be indirectly indexed by the duration of time between the onset of an SL chunk and the onset of its TL translation. This time span is often referred to as 'lag', 'ear-voice-span', or 'delay.'

Gile explains that simultaneous interpretation involves four types of mental efforts: listening, memory, production, and coordination. Successful performance necessitates that the total cognitive processing resources available to the interpreter match or exceed the cognitive requirements of SI tasks. When there is a deficiency in resources for any of the four mental efforts, "either the execution of a task is delayed, which may lead to heavier cognitive load on the processing of the next chunks, and ultimately, to failure sequences… or the task is not executed" (Gile 1997: 200). Gile is suggesting here that extended lag can cognitively overburden the interpreter because expending more of the finite mental resources on the listening effort will deplete those available for other efforts. Consequently, memory operations and production and coordination efforts will suffer, resulting in incidents of failure or in abandoning the task. "Regulation of the ear-voice-span… can be assumed to aim at optimising the balance between short-term memory load and speech production requirements. The further an interpreter lags behind the speaker, the clearer the understanding of his or her message, hence the easier its reformulation but the heavier the burden on working memory" (Gile 1997: 207).

Lag is an inherent aspect of SI performance; the interpreter needs to listen to what the speaker has to say before they can commence their interpretation. It tends to be variable in duration depending on some SL and TL variables (eg, speech delivery rate, information density, redundancy, word order, syntactic characteristics, etc). The average lag duration varies because of several factors that relate to the language combination, discourse type, information density, idiosyncratic preferences, etc. Nevertheless, many researchers attempted to measure the average delay for certain language combinations: Oléron and Nanpon (1964) and Barik (1969) found it to be two to four seconds long, Lederer (1978) three to six seconds, Treisman (1965) four to five words, Gerver (1972) 5.7 words. Goldman-Eisler (1972) found that the minimum lag varies from one language to another. Seasoned interpretation practitioners are less concerned with lag measurements. Jones (1998:83), for example, advises learners to commence rendering SL "once they have enough material from the speaker to finish their own interpreted sentence" trying to keep a more or less constant lag.

Since lag is the time difference between the onsets of an SL burst and its TL translation, it can be measured in linguistic units or in units of time. Thus, it is possible to convert this fundamental aspect of SI performance into an index in terms of which SI style can be discussed. With the knowledge of lag durations and the observation of an interpreter's pattern of listening and speaking, the analyst can

comment on their way of coping with the continuous SL flow and their handling of the simultaneous cognitive tasks of interpretation. The analyst can witness how an excessive delay, for example, can result in disrupting the interpreter's tempo and can lead to missing out on some SL chunks.

Linear Discourse Development

As the speaker and interpreter operate in the same time frame engaging in the delivery of identical information content, the first dealing and the second fielding, comparisons between their performances are inevitable. How similarly do they handle time? How different are their rhythms? How fluent one is compared to the other? Whose performance is most consistent? How synchronous are they with one another? Interpreters do not readily accept any comparison with SL speakers because they feel that the TL medium and world-view exert a great deal of influence over their performance and the way they render SL content into TL. Valid as this concern may be, the interpretation audience does make such comparisons anyway. If no sound comes through their headphones while they clearly hear the patter of the speaker, they will wonder why they are not getting any interpretation.

An inspection of the linear development of the SL and TL discourses, provided its aim is not assessing the interpreter's performance against that of the speaker, can reveal enlightening information about the interpreter's fluency, consistency, synchronicity, rhythm, and time-handling patterns. If we look at the speaker's and interpreter's systematic arrangements of pauses and speech bursts, we can learn about their schemes of discourse chunking, their degrees of fluency, and the interpreter's lag patterns. With knowledge of these, it becomes possible to have a quantitatively-backed discussion of SI style. It will be feasible to reflect on the interpreter's tempo or rhythm of activity, the degree of stability or consistency in their discourse, and their fluency. Consequently, when we know how stable and fluent an interpretation is, we can indirectly reflect on the interpreter's composure and tribulation. Let us explain how all of this is possible.

Interpreters are bombarded by a seemingly ceaseless flow of information; their task is to render it into the target language but at no point in time are they given the floor. Assuming the ghost role, they have to listen to the SL discourse and at the same time murmur its meaning into the headphones of the audience. For them to be able to do these crafty simultaneous tasks, they will have to expeditiously learn the speaker's style of information parcelling and develop a strategy of their own for dealing with SL parcels. If the information comes out in chunks that are so large that they can overburden interpreters cognitively, then the logical thing to do is to divide them up into smaller chunks. Gile (1995: 195) confirms this saying, "When faced with potential overload of memory… interpreters may choose to reformulate speech chunks earlier than they would normally do, sometimes before they have a full picture of what the speaker wants to say." If, however, the SL chunks are small and numerous, interpreters are likely to combine them into larger but manageable chunks. In other words, interpreters aim to reduce the demand on their cognitive processing resources by re-parcelling the SL information in such a way that they can successfully manage the listening, cognitive processing, and production phases.

Chunking is an adaptive dynamic process. Sometimes interpreters choose to di-

vide up a stretch of SL discourse and parcel it out into relatively smaller chunks, but at other times they may choose to combine small SL chunks into larger TL units. The decision is, in either case, determined by various SL and TL linguistic and extra-linguistic factors, but the ultimate goal is to cope cognitively with the steady flow of SL speech. Thus, to continuously adapt chunking to the interpreter's current situation is an exercise of skill of the highest order.

The number, size, and timing of chunks as well as the total speech volume in the TL discourse are of great value to the researcher. They are aspects relating to the chunking strategy that can offer quantitative tangible revelation about the style in a piece of discourse. Take the number of chunks as an example. If we compare the overall number of chunks in TL to that in SL and find them to be significantly different, then we can conclude that the interpreter practised a degree of reformulation by either fusion or fragmentation depending on which discourse had the greatest number of chunks. Similar information can also be obtained if we compare the average sizes of chunks. If we compare the amounts of speech or pause in SL and TL discourses, on the other hand, we can learn about the degree of interpreter fluency in relation to the speaker's. An interpreter who is bombarded by a large volume of SL speech cannot possibly produce a small volume of TL speech; if their pause time is disproportional to the speaker's, the interpretation audience will automatically assume that they are missing out. Jones (1998: 78), a seasoned practitioner, confirms this in relation to the beginning of an interpretation and the end. He says, "If a speaker begins and the interpreter says absolutely nothing, be it only for a few seconds… the participants listening to the interpreter may become very nervous, turn round and make signs at the interpreters' booth, or even interrupt proceedings to complain there is no interpretation." Furthermore, Jones (1998: 87) advises that "As soon as the interpreter senses the end is in sight they should accelerate their own output a little in order to finish as soon as possible after the speaker." The audience will also suspect that they are missing out on parts of the SL discourse whenever they experience exceptionally long pauses at any point in the interpretation. Therefore, interpreter fluency implies that SL speech volume must not be substantially larger than the speaker's.

Monitoring the speech or pause volume in the interpretation and comparing it with its counterpart in the SL discourse at regular intervals can actually chart the linear development of the two discourses. If we calculate the volumes of speech or pause at regular intervals, we can tell how fluent the interpreter is in relation to the speaker, and by the fluctuation of their degrees of fluency we can depict the cycles of discourse progression. Spoken discourse is usually punctuated by pauses (see Chafe 1988); the pause counterpart of a comma, for example, is shorter than that of a full stop, and a pause at the end of a sentence is shorter than that at the end of a paragraph, etc. Hence, the rhythm of interpretation, or its tempo, can evidently be portrayed in the pattern that the cycles produce. So spoken language equivalents of sentences, paragraphs, and sections can be identified with the cycles within the pattern.

Speech or pause volumes can also be enlightening in terms of interpreter composure and tribulation. When we observe the variability of these volumes at the monitored intervals, we can tell that interpreters were in control if their volume variability was comparable to that in the SL discourse. If it was significantly larger, how-

ever, the implication is that the interpreter vacillated dramatically between fluency and dysfluency. They were not in control, they were either lacking in equanimity and composure or encountering difficulties in comprehension or reformulation; for an interpreter in full control would have maintained a more or less stable degree of fluency. Jones (1998: 87) confirms this when he comments about the delay between the speaker and interpreter, saying: "The distance from the speaker should remain more or less constant." He further adds, "But on the other hand, some variation will always be necessary, depending on the speaker's rhythm, style, content, and also depending on specific syntactic difficulties." Although Jones' comments pertain to the interpretation lag, they are equally applicable to fluency variation. Let us ponder a hypothetical case in which two interpreters rendered four episodes of 60 seconds each. The first adopted a consecutive interpretation-like strategy delivering the 1st and 3rd episodes fluently and paused extensively throughout the 2nd and 4th episodes, while the second interpreter delivered all four episodes fluently. There is no doubt that the fluency variability in the first interpreter's discourse would be substantially higher than that in the second interpreter's. Therefore, it would be fair to claim that the second interpreter was more fluent and, hence, had less tribulation than the first had.

In normal SI stylistic analysis, however, comparisons are not made between interpreters' discourses directly. At first, each interpreter's discourse is compared with the speaker's, then the interpreters' parameters of interest will be compared against one another; ie, their performances are at first normalised by the speaker's, then the comparison is made. This way the speaker's performance is used as a benchmark. If the speaker's fluency variability is high, it will be reflected in both interpreters' variability. If one interpreter's fluency variability is significantly higher, however, we will be justified in thinking that that interpreter must have encountered some comprehension or reformulation hardship.

Thus, the quantitative aspects of SI style that we are considering rest primarily on the temporal characteristics of the SL and TL speech signals and the latter's linear discourse development is based on the quantifiable speech or pause volumes in the two discourses. Although quantitative measures do not directly reflect information content, we have seen how they can indirectly shed light on the elusive fluency, tempo, and tribulation. We, in fact, believe that these quantitative aspects can also be used to indirectly assess the quality of an interpretation performance, but we cannot talk about this here until it is experimentally established.

Tools for Stylistic Analysis

In this section, we will study the performances of two groups of interpreters: professionals who have been in the business for a minimum of 10 years, and novices who have received no formal training in SI and have never practised it. We will compare their stylistic parameters with those of shadowers, who simply mimic an SL signal as it is being delivered. Shadowing is a simultaneous task like SI but it suffers no language transfer hindrance; ie, shadowers' handling of the simultaneously presented SL discourse is quantitatively ideal as far as translators are concerned. Therefore, interpreters' performance is a quantitative approximation of shadowers; professionals come closer to it than novices. Because the three groups are so distinct, their performances

will be quantitatively well defined. As a result, the reliability of the analysis tools we have developed can be tested easily: If they produce statistically significant differences between the three groups of subjects, then they shall be accepted as reliable tools of stylistic analysis. For this purpose, a 10 minute long Arabic speech has been used as a stimulus. Three conference interpreters and three novices have been asked to translate it into English, and six university students have shadowed it. We will confine our comments here to two identical 120-second long extracts taken from the SL, interpreted, and shadowed discourses.

Below we will first discuss some theoretical stylistic analysis formulations, then we will study the extracts using the developed tools to see if there are any significant stylistic differences between the three subject groups. The analysis formulations are achieved by comparing the performance parameters of the subjects with those of the speaker. As they operate in the same time frame, and since the subjects are dependent on the speaker for input information, the way the speaker manages time is somehow reflected in the way the subjects manage their time. Barik (1973: 266) discovered that "the temporal characteristics of the translation generally co-vary with their counterparts in the original version." Our analysis formulations are based on a comparison between SL and TL total time spent in pausing, their average duration of pauses, and their total number of pauses. The formulations could have been equally efficiently expressed in terms of speaking parameters since pause and speech are symmetric.

Goldman-Eisler (1968: 31) considers pausing to be "an attribute of spontaneity in the creation of new verbal constructions… and [a phenomenon] concomitant with the cognitive processes" associated with spontaneous speech. In her investigation of the relationship between complex cognitive information and pausing, she discovered that "individual differences pointed to a factor of characteristic disposition to pausing, but extra time invested over and above the individual's characteristic pausing time seems to determine the intellectual quality of verbal statements as between individuals" (Goldman-Eisler 1968: 69). Similarly, Barik (1975) observed that low speech rates are typical of less qualified interpreters, and Yaghi (1994) found that TL non-tactical omission (i.e., failure to translate) is correlated, on the one hand, with the interpreter's speech rate, and on the other, with the SL speech volume. So pausing is critical in the analysis of spontaneous speech and is even of greater importance to the study of SI strategies. As explained in the previous section, differences in pausing patterns can reveal the pattern of chunking in a TL discourse and that in turn can expose the degree of mirroring or reformulation in it.

Chunking

Although the intention is not to evaluate the interpreter's performance against that of the speaker, a comparison needs to be made between their chunking methods in order to contextualize any statement we make about the interpreter's style. Such a comparison would enable the investigator to say that the interpreter behaved in a certain way when the SL discourse had some particular attributes. The comparison would reveal the degree of convergence or divergence between the two discourses, and that degree is indicative of the extent to which TL mirrors[1] SL chunking style. If an SL signal were to be reproduced in the TL track, which is a virtual impossibility, then we would expect the differences between SL/TL chunking parameters to be zero,

we would expect SL and TL numbers of speech bursts, mean burst durations, and speech volumes to be identical. Since speech and pause are symmetric, this hypothesis can be formulated with reference to speaker and translator pausing as follows:

Model 1: $\quad \# P_1 - \# P_2 \approx 0$

Model 2: $\quad \Sigma P_1 - \Sigma P_2 \approx 0$

Model 3: $\quad \mu P_1 - \mu P_2 \approx 0$

Model 4: $\quad \dfrac{v_2}{v_1} \approx 1$

where P_1 stands for speaker pauses and P_2 translator pauses, # is number of pauses, Σ is sum of all pauses in a discourse, μ is mean duration of pauses, v_1 is variance of speaker pauses, v_2 is variance of interpreter pauses, and $\approx$ is approximately.

For faithful mirroring to take place, the first three models will need to yield a value of zero and the last a value of 1. The first model merely depicts whether the interpreted discourse has been parcelled out in as many chunks as the source discourse, and hence, whether the interpreter has paused as many times as the speaker did. This, however, is not sufficient to claim that mirroring has taken place since the interpreter's overall volume of pausing may have been different; it is possible for the interpreter to have longer or shorter pause times than those of the speaker, but to manage to pause as often. The second model, therefore, ensures that SL and TL pause volumes are similar. The third one makes certain that the numbers of parcels as well as their sizes are taken into account; ie, that the interpreter pauses as often and for the same average duration as the speaker does. The last model ascertains that the interpreter and speaker are similar in their variation of pause length; ie, that the switching between long and short pauses is matched. With the four models together, we can make a statement about the chunking strategies of translators and can gauge the degree of reformulation[2] in the TL discourse by the magnitude of difference between speaker and translator parameters. In the first three models, values smaller than zero indicate that the interpreted discourse is: (1) more fragmented (since it has a larger number of pauses), (2) has a larger share of inactivity (since it has a larger overall pausing time), and (3) its SL bursts are uttered after a long delay (since it has a larger mean duration of pauses). Values larger than zero, on the other hand, indicate that the interpreter practised reformulation by fusing SL bursts together, that they exhibited more speech activity than the speaker did, and that they generally spoke after a short delay. Model 4 calculates the ratio of interpreter to speaker variation, thereby yielding 1 when their pause variances are similar. Values larger than 1 indicate that the interpreter varied their pause durations more than the speaker did.

To illustrate the utility of these models, we have calculated them for the shadowers and professional and novice interpreters, and plotted their values for each subject in the next four figures. On the axes are the translator and shadower codes: 0-6 for novice translators, 7-12 for professional translators, and 13-24 for shadowers. On the axes are the values of the models: numbers of pauses, sums of pausing time in centiseconds (cs), means of pause durations in centiseconds as well, and the ratio of interpreter to speaker pause variation.

Figure 1 shows the values produced by the first model; ie, the differences between SL and TL numbers of pauses.

Figure 1
Model 1: Segmentation
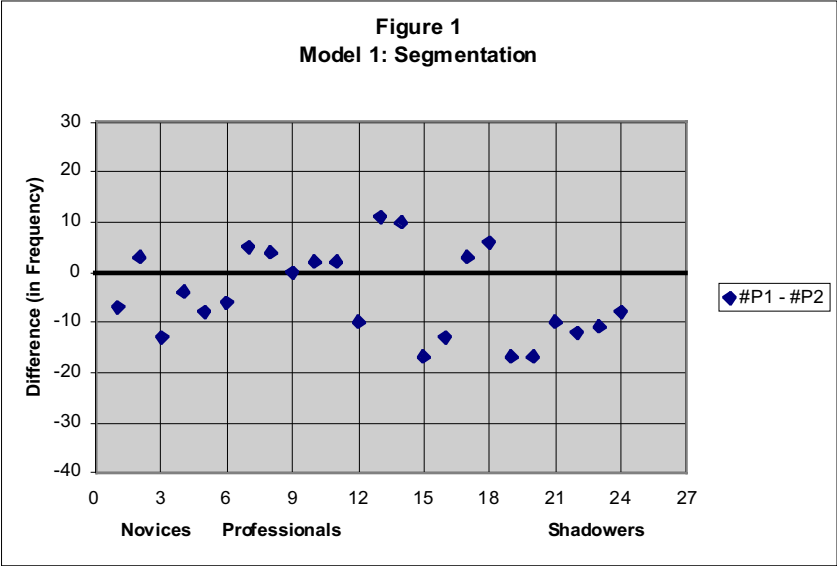
Notice that the majority of novices' values are lower than the majority of professionals' values and that the latter are lower than those of some shadowers but higher than others. Nevertheless, there is no clear distinction between professionals, novices, and shadowers in terms of how frequently they pause. In other words, the 1st model fails to establish that the three subject groups have different chunking strategies. When one factor ANOVA is carried out on the values plotted here, it corroborates this observation revealing that these values cannot set the subject groups apart in terms of proficiency [$F(2,23)=1.338$, $p=0.2838$]. Comparing between the groups' means shows the mean SL/TL difference in the number of pauses to be (0.5) for the professionals, (–5.8) for the novices, and (–6.3) for the shadowers. These means do set apart professionals from novices and professionals from shadowers but fail to establish significant differences between novices' and shadowers' means. This is the specific reason why Model 1 has failed in this instance.

Despite this finding, Model 1 offers useful information about the degree of reformulation practised by the subjects. Professionals paused almost as frequently as the speaker did; hence, they did little reformulation. Novices, on the other hand, paused more frequently, which means that their reformulation caused the TL discourse to be more fragmented than the SL discourse. Shadowers are two groups: one was more like professionals, the other more like novices. To assess chunking style fully, the other models need to be calculated; they will corroborate this conclusion and capture that which this model has failed to capture.

Figure 2 plots the volume of inactivity in TL as expressed in Model 2.
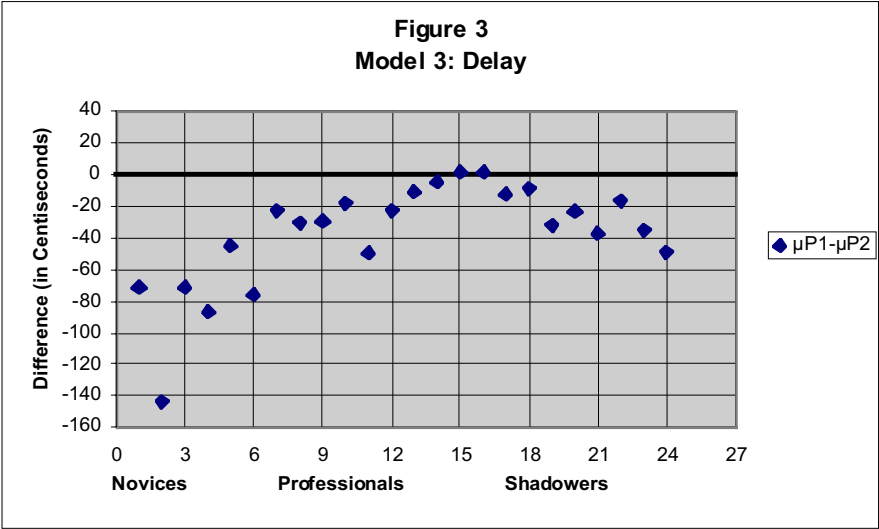
Figure 2
Model 2: Speech Volume
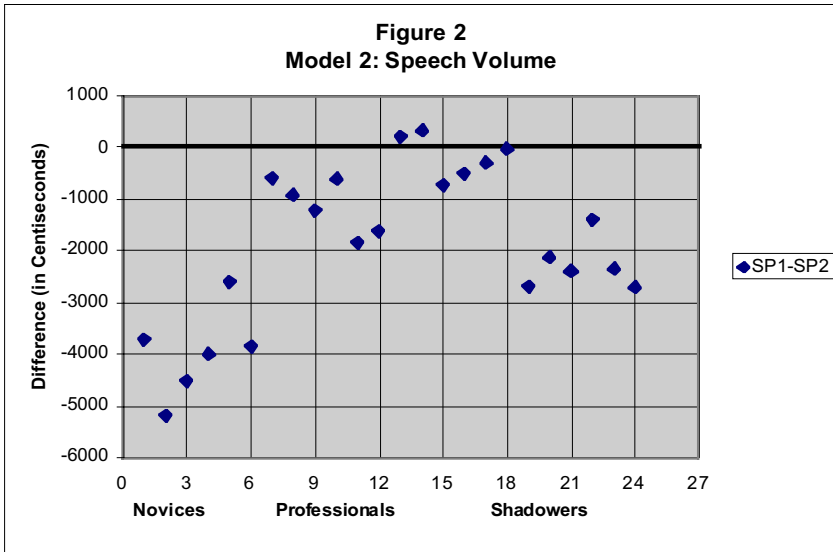
Notice that the shadowers are two groups; one faithfully mirrored the SL discourse (with the average of $\Sigma\, P_1 - \Sigma\, P_2$ being –1.72 seconds, the smallest) but the other exhibited a measure of autonomy having a larger volume of inactivity (with the average of $\Sigma\, P_1 - \Sigma\, P_2$ being –22.79 seconds). There are two possibilities for the difference between the two groups: either one had high enough speech rates to render SL in a shorter time and as a result had time left over, or the other group added material of their own to the SL discourse. On closer inspection, we have found the first possibility to be the case. The two groups had different speech rates, so those who spoke faster had a larger share of inactivity.

Professionals, on the other hand, had a volume of inactivity between the two groups ($\Sigma\, P_1 - \Sigma\, P_2$ being –11.31 seconds, on average). This implies that they neither slavishly mirrored SL nor were completely free to reformulate TL.

Novices behaved differently from either professionals or shadowers. Notice how they have the largest volume of inactivity and have exceeded SL levels of inactivity by 39.74 seconds on average. We definitely cannot attribute such a large volume of inactivity to the novices' high speech rate, since unlike shadowing, interpretation involves the time consuming skill of language transfer. If we consider the novices' degree of non-tactical omission, the extent of translation failure, we will find the reason for this large volume of inactivity. The average rate of omission for novices in the piece of discourse being considered is 47%. Clearly then, failing to render a number of SL bursts is responsible for the great volume of inactivity in novices' TL discourses.

When one factor ANOVA is applied to all the values produced by model 2, it confirms our observations [$F(2,23) = 17.946$, $p = 0.0001$]. Therefore, we can conclude that the difference between SL and TL volumes of inactivity as expressed in $\Sigma\, P_1 - \Sigma\, P_2$ is capable of discriminating between the subject groups in terms of proficiency. That is, while Model 1 revealed the degree of fragmentation in the TL discourse, Model 2 told us the degree of inactivity in the TL relative to the SL discourse, which is the inverse of speech volume.

Figure 3 plots mean pause durations as expressed in Model 3.

Figure 3
Model 3: Delay

As expected, the graph here is similar to that in Figure 2 since volume is a factor in the calculation of mean. Notice how one group of shadowers have mean pauses that closely resemble SL mean pauses (the average difference between their means being merely –5.6 cs), whilst the other group have mean pauses that are on average 32 cs longer. Professionals' mean pauses are longer than the SL speaker's but by a difference comparable to that between the mean pauses of the SL speaker and the second group of shadowers (29 cs). Novices' pauses, on the other hand, are the longest of all. The difference between their mean pause durations and the speaker's is about 82 cs. Once again, amateurs paused longer because of the high degree of omission that they exhibited in their translation.

One factor ANOVA applied to the values of Model 3 bears out these results [$F(2,23) = 19.416$, $p = 0.0001$]. This indicates that the differences between SL and TL mean pauses as expressed in $\mu P_1 - \mu P_2$ are successful in discriminating between the three subject groups in proficiency. We can, therefore, conclude that shadowers, professionals, and novices have different spans of delay, with the shadowers having the shortest and novices having the longest delay.

Figure 4 below depicts the difference in speech length variation between TL and SL discourses. It shows subject-to-speaker burst variation ratios. When Model 4 has a value of approximately 1, it reveals that the interpreter and speaker had a similar pattern of oscillation. When the value is larger than 1, the interpreter is said to have maintained more speech burst oscillation; but when the value is smaller than 1, the speaker is said to have varied their burst length more than the interpreter did.

Figure 4
Model 4: Variance
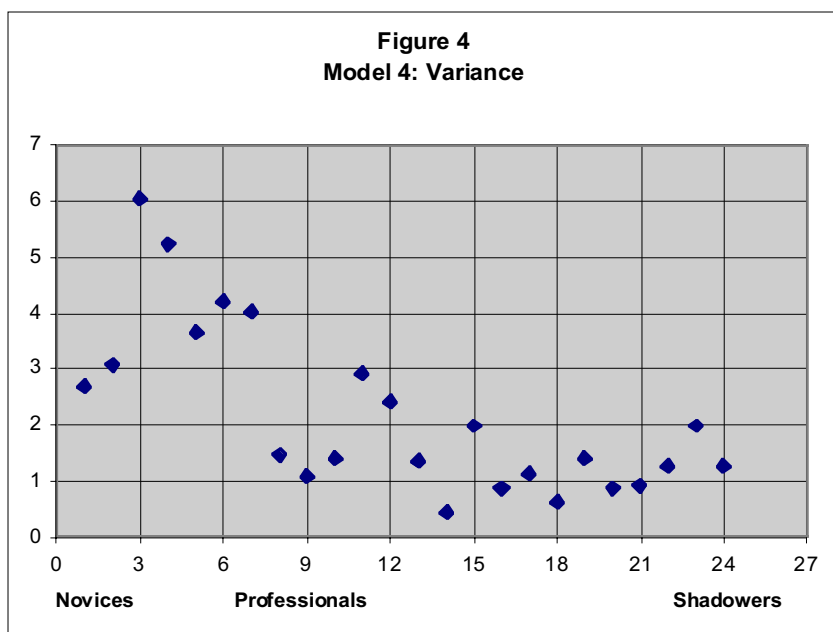
The three groups of subjects varied their speech burst durations in different ways: amateurs oscillated between long and short bursts a lot more than the others did; their average variance ratio being 4.14. This means that their burst variation was four times that of the speaker. Shadowers exhibited more stability when they attempted to mirror SL speech variation (their average variance ratio was merely 1.17; ie, their variation was similar to that of the speaker). Some professional interpreters, on the other hand, seem to have attempted to reflect the same level of speech burst variation that the speaker maintained, whilst others gave themselves more liberty. On average, however, professionals had a significantly higher level of burst length variation than the speaker (2.21). ANOVA confirms that the three groups had distinct patterns of burst variation [$F(2,23)=15.809$, $p=0.0001$]. The highest homogeneity of burst length was in shadowing; thereby indicating the highest degree of composure and stability. On the other hand, the lowest level of homogeneity was exhibited by the novices; thus reflecting a higher level of tribulation and a substantial lack of stability in performance.

Now the four models combined give us a fairly good idea about an interpreter's pausing strategy. The four together can help us substantiate any generalisation we may make about interpreters' chunking style and the extent of mirroring or reformulation in a TL discourse; we can talk about chunking in terms of the degree of fragmentation or fusion, the scale of inactivity, and the duration of delay. In our example above, for instance, professionals practised a lot of reformulation, pausing less often than the speaker, had a share of inactivity smaller than that in the SL discourse, and were slower than the SL speaker in producing subsequent utterances. They also maintained a level of speech burst variation that is substantially higher than that maintained by the speaker. Novices, on the other hand, were more slavish, taking pauses almost as often as the speaker, were slower in producing subsequent speech

chunks, and surpassed the SL speaker by a larger amount of inactivity. Their speech bursts tended to oscillate between long and short at a level four times higher than that of the speaker and two times higher than that of the professional interpreters.

Adaptive Monitoring

Another method for studying SI style is to monitor the volume of speech in the TL discourse and to compare it adaptively with the volume of speech in the SL discourse. Since the amount of speech in a translation is a crucial quantitative aspect of performance, as we have seen so far, indexing it at regular intervals can function as an adaptive gauge of linear discourse development. Whether or not such speech is contextually relevant cannot be detected by this quantitative method. This, however, should not be considered a defect since the monitoring method is designed for the study of SI style rather than content. The main purpose is to enable us to comment on chunking, rhythm of activity, and fluency rather than on the accuracy of interpretation.

This adaptive monitoring instrument is developed around the ratio of speech in the SL and TL tracks. The concept is mathematically simple: we need to compute the ratio of speech at fixed intervals in both speaker's and translator's tracks. Then we take the obtained interval speech ratio indices and compare them across the SL-TL tracks. Once that is done, we can move a step further and compute the cumulative indices for the two tracks separately, then compare the interpreter's against the speaker's. When graphed, the cumulative indices will portray three things: (1) periodicity or cycles depicting the rhythm of activity or the tempo in a discourse; (2) stability and consistency which will reveal the speaker's or interpreter's degree of composure and tribulation; and (3) the degree of fluency within SL/TL discourses. Let us explain how this is possible.

SL and TL linear discourse development can be represented graphically in the form of either speech ratios per interval or cumulative speech ratios. If interval indices were to be put on the X axis, ratios on the Y axis, and the individual speech ratios were to be plotted per interval, one would be able to visualise the precise time when acceleration and deceleration occur in a discourse; ie, we can observe the fluency modulation. Accepting Chafe's (1988) observations about spoken language punctuation, the graph of interval speech ratios can be used to identify punctuation units (ie, periodicity) within a discourse. If an essay is divided up into sections, sections into paragraphs, paragraphs into sentences, and sentences into clauses and phrases, for example, then perhaps a speech can be viewed like an essay. Assuming that it is true that speakers do punctuate their discourse, we can view the SL and TL discourses in terms of units that are equivalent to those in written language. And since pausing is one of the punctuation devices that speakers use, the period when the speech ratio decreases is most likely a period where a pause is located. Therefore, the lowest points in an SL or TL graph where the speech ratios are at minimum will demarcate the biggest discourse units, the equivalent of sections in an essay. They represent intervals with the highest volumes of pausing and these are equivalent to the blank spaces and font- or paragraph-formatting that delineate sections in an essay. Intervals with relatively higher speech ratios will delimit the equivalent of paragraphs, and those with yet a higher speech ratio will delineate sentences. In other words, by

plotting the individual interval speech ratios, we can map the discourse development, identify its periodicity, and draw the rhythm of activity in a discourse. When the linear development of both SL and TL discourses are plotted on the same graph, we can see to what extent they match and we can identify where there is convergence and where there is divergence. If we desire to compare between the performances of two interpreters rendering a single piece of discourse, we can plot the speaker's and the two interpreters' linear developments on one graph, thereby having the opportunity to compare interpreters' performances in relation to the speaker's.

The linear discourse development in SL and TL can also be depicted in the form of cumulative speech ratio graphs, where each point on the graph tells the speech ratio in the discourse from the beginning up to the time that it represents. Because each point represents the speech ratio in all that precedes it, this cumulative speech ratio graph is less sensitive to occasional variation and more concerned with the overall pattern of performance. Therefore, it is excellent for depicting fluency and the degree of stability (consistency or variability, however one may view it). By computing the cumulative speech ratios for each interval in the discourse, then plotting them on a graph similar to the one described above, the investigator can visualise the discourse originator's fluency and stability of performance over time. Wherever on the Y-axis the line of points hovers is the level of fluency in the discourse. On the other hand, the degree of wavering in this line represents the amount of variability in performance, thus implying lack of stability.

Both SL and TL linear patterns of development can be plotted in this way and the investigator can compare them visually. Model 5 below is used to figure out statistically the difference between the performances of two people. These can be the speaker and interpreter, or two interpreters who render the same piece of discourse into the same language. We can depict at regular intervals the precise degrees and locations of convergence and divergence between their performances by simply deducting at every interval the interpreter's cumulative speech ratios from those of the speaker or the other interpreter. We can tell how well they compare in terms of degree of fluency and stability of activity. If we want to compare two interpreters, we can deduct at every interval Interpreter Y's cumulative speech ratios from those of X. The closer the difference between them to zero, the more similar they are. When the difference is negative, Interpreter Y is more fluent; when positive, Interpreter X is. The magnitude of difference represents the degree of divergence between their performances. In fact, delay of a certain type can be extrapolated from the comparative ratio graph line. We said that every point on the cumulative ratio line sums up the speech ratio in the relevant discourse up to the interval it stands for, and every point on the comparative ratio line represents the difference between corresponding points on the SL and TL cumulative ratio lines. Therefore, if the value of every point on the comparative ratio line were to be multiplied by the duration of the interval, we would obtain the SL-TL speech ratio difference per interval in centiseconds. In other words, every point on the comparative ratio line tells the offset, the phase shift, or the delay between speaker and interpreter performances.

To appreciate the value of adaptive monitoring, let us first discuss the mathematical model that makes it possible, then we will apply it to the discourses produced by our subjects. We will see the difference in performance between shadowers, professional interpreters, and novices.

Speech Monitoring Model

For purposes of calculating interval speech ratios, the two discourses subject to comparison need to be divided up into corresponding intervals of some duration. We have chosen to divide up the discourses into intervals of 500 cs each. We think this interval is appropriate because it roughly accommodates one sentence assuming that the natural speech rate is 3.5 syllables/second. The interval would hold 16 syllables, which makes out 8 average size words that consist of two syllables each. If other investigators wish to capture details in units smaller than a sentence, they may choose to divide up their discourses into still smaller intervals. Speech ratio is computed by measuring the duration of time within this interval that is occupied by speech, then dividing the sum duration by the length of interval. The speech ratios are calculated separately for the intervals within a piece of discourse, then they are plotted sequentially with the speech ratio of the first interval being first and the last being last. Speech ratios are calculated independently for the two discourses (say, SL and TL) but are plotted on the same graph for comparison purposes.

The adaptive monitoring model that uses cumulative speech ratios, on the other hand, is a little more complex. It may be expressed as a comparison between two interpreters or one interpreter and a speaker. It takes this form:

Model 5: $I^1(t) \approx I^1(t)$ which can be rewritten as:

$$\left(\frac{\sum\limits_{j=1}^{n} \sum\limits_{i=1}^{k_j} S_{ij}^1}{\sum\limits_{j=1}^{n} t_j} \right) - \left(\frac{\sum\limits_{j=1}^{n} \sum\limits_{i=1}^{k_j} S_{ij}^2}{\sum\limits_{j=1}^{n} t_j} \right) \approx 0$$

where is I^1 SL or Interpreter X's index, I^2 is TL or Interpreter Y's index, t is an interval of time whose duration is determined by the investigator, S^1 and P^1 are speech and pause durations in the first track (i.e., SL or Interpreter X's), and S^2 and P^2 are speech and pause durations in the second track (i.e., TL or Interpreter Y's), $\sum\limits_{j=1}^{n}$ is the sum of intervals $j=1$ to n, $\sum\limits_{i=1}^{k_j}$ is the sum of speech bursts $i = 1$ to k_j within interval j, and is the time duration of the jth interval.

Once the two tracks are divided up into corresponding time intervals of equal length (t), as stated above, the ratio of speech is calculated within each of these intervals producing interval speech ratios (r). Then, the cumulative ratios are computed for the tracks separately by adding the interval speech ratios one at a time and dividing by the number of added intervals, i.e.,

$(r_1 + r_2)/2, (r_1 + r_2 + r_3)/3, \dots (r_1 + r_2 = r_3 + \dots + r_n)/n$

where the subscript stands for the interval whose speech ratio (r) is added. The end value constitutes the cumulative ratio for the SL or TL track and that is what is termed in this model 'Speaker and Translator Indices', $I^1(t)$ and $I^2(t)$ respectively. Finally, the cumulative indices of SL and TL (or of the two compared interpreters, whichever the case might be) are compared for each interval by deducting the second from the first. Values close to zero indicate that the interpreter produces a volume of

speech equal to that in the SL discourse; negative values indicate that the interpreter produces more speech, but positive values that the speaker produces more.

Application

Plotting the SL/TL interval, cumulative, and comparative speech ratios provides a graphic representation of the development of the two discourses. From the shapes of the graphs, we can learn about their periodicity, stability, and fluency. Periodicity is manifested by a pattern of repetitive cycles in interval ratio graphs (as in Figures 5a and 5b) where the speech ratio ascends to a peak then descends to a dip, stability is expressed by the straightness of cumulative graphs, and fluency is portrayed by the level at which the cumulative graphs reside (as in Figure 5). To illustrate this, we will now study the performance of one shadower, one professional interpreter, and one novice interpreter in a 120-second extract from the stimulus SL discourse, but first let us consider the SL discourse. Figure 5a below depicts the development of SL discourse.

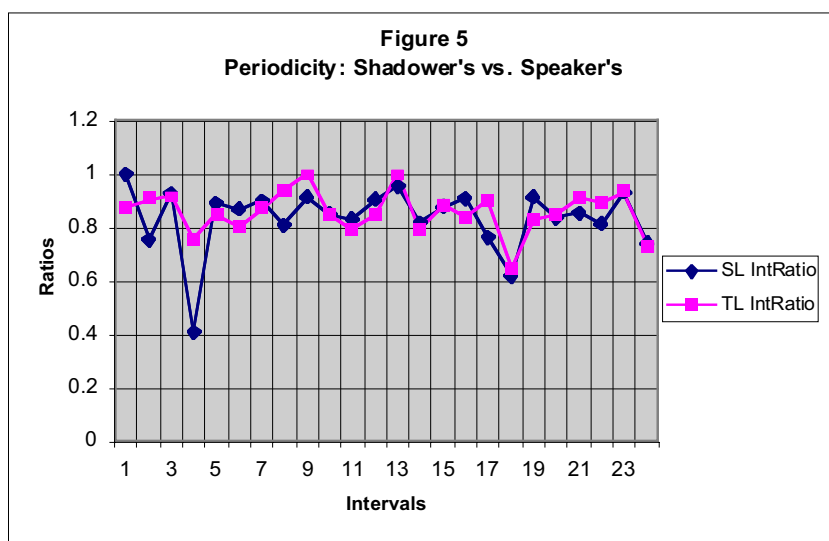

Figure 5a
SL Periodicity: A Closeup
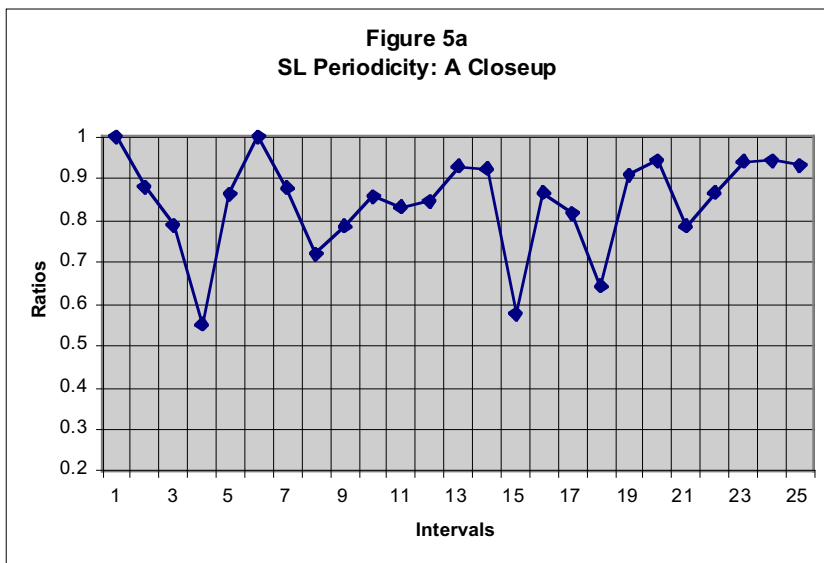
Notice that the speaker's graph here consists of cycles or periods that begin at a high ratio of speech and end at a relatively smaller ratio. Obviously, these cycles can be delineated differently according to whether we are looking for a detailed or a global pattern of cycles. One way of viewing the cycles in the SL extract under consideration is by thinking that the first cycle spans across 1st-4th intervals beginning at a 100% speech ratio but ending at 41%, while the second is less dramatic, spanning across 5th-8th intervals beginning at an 89% speech ratio and ending at 81%, the third spanning across 9th-11th at 92-83%, the fourth across 12th-14th at 90-82%, the fifth across 15th-18th at 88-62%, the sixth across 19th-22nd at 92-82%, and the seventh across 23rd-24th at 94-74%. If we want to study the global pattern, on the other hand, we may have to view this extract as consisting of three cycles: the first being between intervals 1 and 4, the second between intervals 5 and 18, while the third is between

19 and 24. Had the SL extract been longer, we could have obtained an even larger
global pattern of cycles. For demonstration purposes, Figure 5b below plots the in-
terval speech ratios for a stretch of 365 seconds, 120 seconds of which precede our
extract and 120 follow it, making our extract feature between the 26th and 49th inter-
vals. The 365 seconds constitute more than half the SL discourse that was used as a
stimulus for the interpretation we will be analysing.


Figure 5b
SL Periodicity : A Broad View
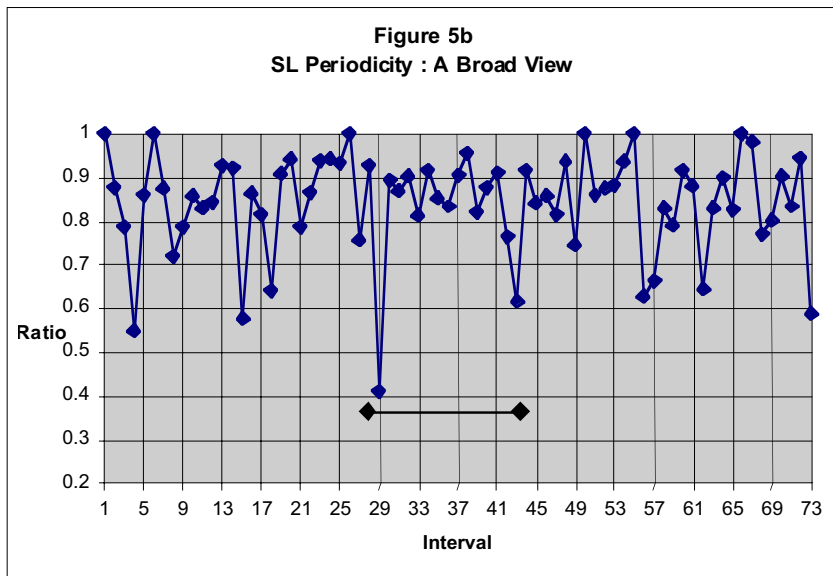
Here we can see that there are two large cycles in the six-minute extract, the first
extends over intervals 1-29, and the second over 30-73. The extract subject to our
analysis falls predominantly in the second global cycle but does span over the inter-
section between the two cycles; four intervals fall in the first and 19 in the second.
Had the entire SL passage been considered for analysis, even this global pattern of
cycles could have changed.

Our knowledge of discourse structures, however, tells us what to expect in terms
of periodicity. Phrases and clauses, roughly corresponding to Chafe's (1980) 'focus
clusters', are normally bound by pauses; they constitute the smallest cycle we can
have, so the interval comprising the terminal pause of these phrases or clauses will
have lower speech ratio than a non-pause interrupted stretch of discourse. Sentences
constitute bigger cycles and they are normally followed by longer pauses (Crystal
1969: 172; Chafe 1980), so when the terminal pause falls in an interval, it will make
it have a speech ratio lower than that of the interval at the end of a clause or phrase.
The spoken equivalent of a paragraph, episode in Chafe's (1980) terminology, on the
other hand, constitutes yet a larger cycle and it is usually followed by a very long
pause. The interval containing such pauses will have a speech ratio lower than that of
the interval terminating phrases, clauses, and sentences. The longer a discourse unit
is, the larger its cycle is.

Figure 5
Periodicity: Shadower's vs. Speaker's

Now if we compare the shadower's interval speech ratios (TL IntRatio) with the speaker's (as in Figure 5), we will find them to be quite similar and to be at about the same level (86 and 84% respectively). In the speaker's graph, however, the cycles are 3 to 4 intervals long; ie, 1500-2000 cs long since each interval spans to 500 cs, while the cycles in the shadower's graph are more variant: the first extends across the 1st - 4th intervals, the second 5th - 11th, the third 12th - 14th, the fourth 15th - 18th, and the fifth 19th - 24th. Clearly, the speaker's and shadower's patterns of cycles are not identical, yet superimposing the latter on the former has facilitated the comparison between their periodicities. Notice how the shadower rendered the first SL cycle in a comparable period that spans over four intervals, but did not pause as long as the speaker did. The second and third SL cycles, however, have been combined into one TL cycle; ie, the shadower practised reformulation by fusion. The fourth and fifth SL cycles have been replicated in TL, but not the sixth and seventh. These have been combined together in the fifth TL cycle. In other words, the shadower attempted to faithfully mirror the first SL cycle, reformulated the second and third, replicated the fourth and fifth, but reformulated the sixth and seventh SL cycles.

Figure 6 below presents a different aspect of the shadower's performance, the cumulative speech ratios, which indicate the overall development of SL and TL discourses.

Figure 6
Discourse Development : Shadower's vs Speaker's
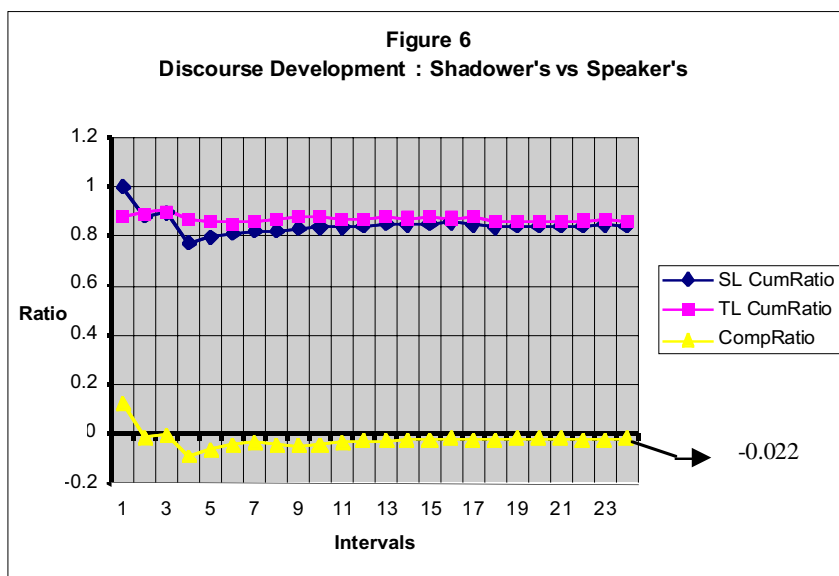
The three lines here give a global view: SL CumulativeRatio shows the development of SL in terms of the ratio of speech in the discourse, TL CumulativeRatio depicts the progression of TL shadowed discourse, and ComparativeRatio portrays the convergence and divergence between SL and TL in terms of cumulative speech ratios as calculated by the adaptive monitoring model, .

The line representing the cumulative speech ratios in the shadowed discourse is only very slightly higher than that of the SL speaker's. These graph lines communicate one message: there is very little difference between SL and TL discourses in terms of fluency; both the speaker and shadower spoke consistently between 84-87% of the time.

As explained in the previous section, the levels of SL and TL cumulative ratios indicate the degree of fluency in a discourse. The straightness of the graphs, on the other hand, signifies to what extent the discourse originators were consistent; ie, it marks their fluency variability. Maintaining level fluency indicates that the speaker is in control; therefore, a twisting graph shows turbulent performance where fluency comes only in bursts. The shadower's performance as depicted in the above graph points to levels of fluency and tribulation that match the SL speaker's.

The comparative graph line, at the bottom, hovers around zero, which signifies the degree of divergence between SL and TL in terms of fluency and offset. Thus, the fact that the line is barely below zero (the last value being -0.04) indicates that the shadower was a little more fluent and was slightly ahead of the speaker. They were also about 20 cs ahead of the speaker; this value is obtained by multiplying the comparative ratio average of –0.04 by the duration of our interval (ie, 500 cs).

To appreciate the information we have so far about the shadower's performance, we need to study the performances of a professional and a novice interpreter. Let us consider this graph:

Figure 7
Periodicity: Professional's vs Speaker's
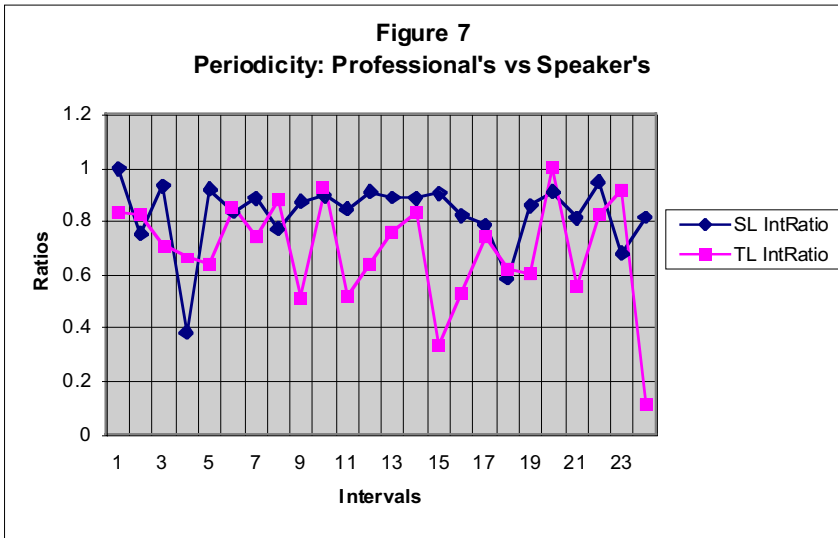
Periodicity in the professional's discourse appears to be more variant in length and in speech ratio than that in the shadower's. Notice in this figure that the TL discourse is comprised of about six cycles: the first ranging between 1st-5th intervals, the second between 6th-9th, the third between 10th-11th, the fourth between 12th-15th, the fifth between 16th-21st, and the sixth between 22nd-24th. Furthermore, the speech ratios within intervals range between 100% and 11% with the average being 69%.

Speech ratio is considerably more variant in the professional's cycles than in either the speaker's or the shadower's. While the SL cycles had a speech ratio standard deviation of 0.12, the cycles in the shadower's discourse had an even smaller standard deviation of 0.8. The professional's speech ratio standard deviation, on the other hand, was 0.20. This significantly larger variability in the speech ratio within the professional's cycles is a direct consequence of the interpretation process. Interpreters take longer pauses than shadowers to decode then recode SL bursts into TL; hence, their speech ratio increases when offering their TL output and decreases when taking SL input. Larger speech ratio variability is inherent in the interpretation process despite the conscious attempt of interpreters to "provide a correct translation of the original in a form that sounds as natural and as authentic as possible" (Jones 1998: 90).

Figure 8 below depicts the professional interpreter's discourse development against the speaker's. Immediately after the first cycle, both TL and SL graphs maintain a somewhat even direction (at the 74% and 84% speech ratio levels respectively), although the interpreter's is slightly more winding. Any twisting in a cumulative ratio graph indicates lack of consistency and stability in the discourse. Therefore, we may claim that the professional interpreter's performance here was less steady than the SL speaker's, thus confirming Barik's (1973: 267) conclusion that "translated speech is less rhythmical than natural speech."

Figure 8
**Discourse Development: Professional's vs. Speaker's**
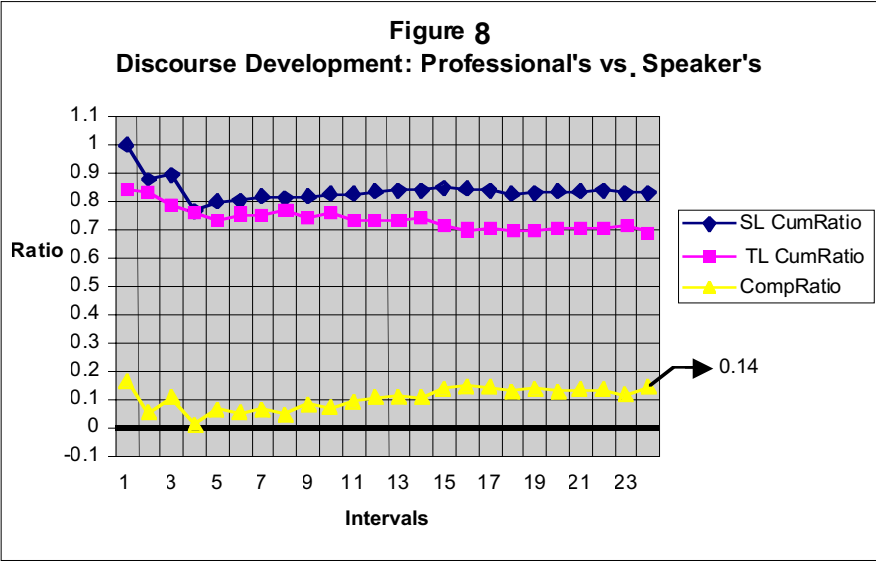
Fluctuation in the TL cumulative speech ratio graphs is to be expected since it is speakers rather than interpreters who are in charge of the overall discourse planning. It is the speakers who originate ideas and who parcel them out into spoken packages which form what we call here cycles; hence, they know how big each package will be, and because they are the discourse planners they can manage to parcel ideas in roughly even cycles. Interpreters, on the other hand, do not have definitive knowledge about how a current SL speech unit will end, let alone what the overall pattern of idea packaging is like. Gile (1995: 195) observes that when interpreters are faced with potential memory overload, as with an SL and TL that are syntactically very different, "[they] may choose to reformulate speech chunks earlier than they would normally do, sometimes before they have a full picture of what the speaker wants to say." Because interpreters lack long term planning and they handle each idea independently of the ones that subsequently follow it, their discourse periodicity is more variant. Furthermore, the fact that they also need to adapt some SL ideas to the needs of their TL audience constitutes yet another factor that may cause them to alter SL periodicity. Figure 7 above shows the degree of divergence in periodicity between TL and SL.

When SL and TL discourse developments are compared, the comparative ratio graph at the bottom of Figure 8 is produced. Here we see that the interpreter was generally less active than the speaker; ie, the amount of speech in his discourse was small. The graph also tells us that at the end of the first cycle, the interpreter was able to keep up with the speaker but as the speech progressed, interpreter's performance started lagging behind. At the fifteenth interval (the end of the fourth cycle), his performance stabilised and he managed to maintain speech ratio at a constant level (with an SL-TL difference of 14%). If we were to comment on this comparative ratio graph in one sentence, we would say that at the beginning the interpreter tried to match up with the speaker in fluency but as the discourse progressed he resigned himself to one level and maintained it. Consequently, his initial performance was

jerky but it eventually stabilised. In terms of the delay, the professional's performance was offset by 70 cs (0.14*500 cs).

The novice's performance as shown in Figure 9 is strikingly different from those of the shadower's or the professional interpreter's. Let us first identify the periods and ranges of speech ratios within the shadower's discourse: the first spans between the 1st-5th intervals at a 20%-0% speech ratio, the second between 6th-13th at 49%-0%, the third between 14th-19th at 27%-39%, and the fourth between 20th-24th at 72%-65%. Notice how low the speech ratios are at the tail of the first and second cycles. Neither the professional nor the shadower ever reached this low level of speech ratio. In fact, the novice spent 1500 cs at the tail end of the second cycle exhibiting no speech activity whatsoever, while the lowest level of speech activity that the speaker went down to was 41% at the end of the 1st cycle, the shadower 65%, and the professional 11%. The novice interpreter was evidently trying to match the speaker's periodicity (cf. 1st cycle) but found it difficult to listen and speak at the same time, so he adopted the consecutive strategy of pausing to listen for SL input then speaking the TL interpretation. This is why the third and fourth cycles had a relatively higher level of fluency.

Figure 9
Periodicity: Novice's vs Speaker's
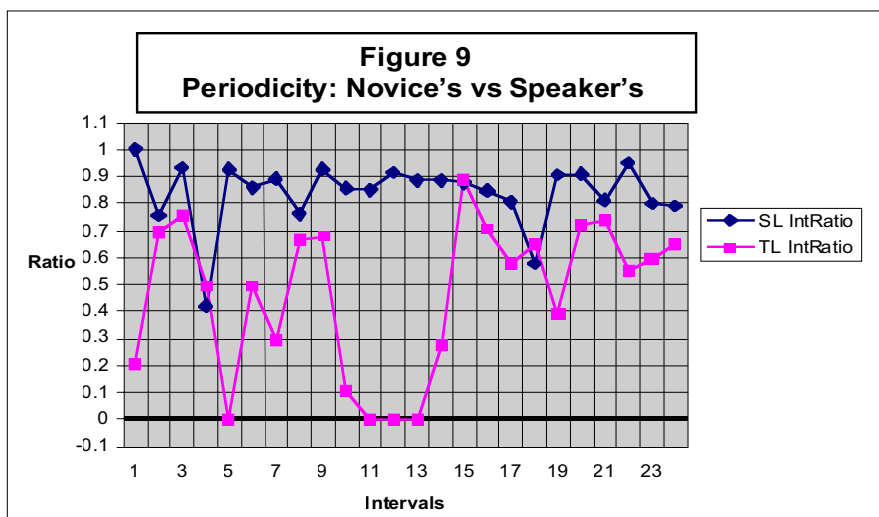
Notice that the shadower had significantly lower average speech ratio than those of the speaker, shadower, or professional interpreter. At 46% speech ratio, the novice spent more than half the discourse time in silence. His performance was also a lot jerkier than that of either the shadower or the professional interpreter. The standard deviation of his cycles' speech ratios stands at 0.28, which signifies how high variability was in his performance and consequently how lacking in consistency and stability.

A close inspection of Figure 10 below and of the transcript of the novice's discourse indicates that he had a consecutive-interpretation-like strategy. He listened long enough to make sense of an SL idea, then translated it without monitoring what the speaker said during their involvement in the transformation into TL. As a result, the novice had a low speech ratio but a very high rate of omission.

Figure 10
Discourse Development: Novice's vs. Speaker's
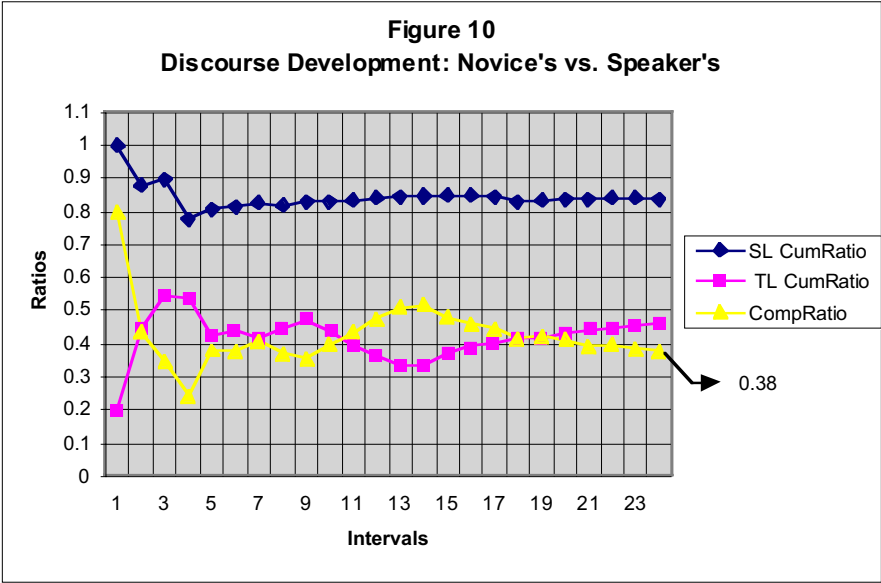
Notice the disparity between SL and TL cumulative ratio levels. The TL cumulative ratio graph is at a very low level, signifying a low level of fluency, and is a lot more winding than either the shadower's or the professional's. The novice's performance is clearly turbulent and lacking in consistency. The comparative ratio graph, on the other hand, depicts his effort at coping with the continuous flow of SL discourse. The lowest dip at the 4th interval, represents his best performance, the middle hump his worst, and the smooth ending his relative stability. In terms of delay, his offset is 190 cs (0.38*500cs), which is the largest of all subjects. Obviously, this is a very inexperienced interpreter but his performance illustrates the potential power of our model 5 as a tool for stylistic analysis.

Speech Variability Monitoring Model

If the graphic depiction of variation in the previous model is insufficient or there is a desire to quantify performance stability and consistency, another formula can be used. We can monitor the variation in speech or pause length within a piece of discourse by calculating the cumulative variance. As in the case of the Speech Monitoring Model, it is possible with such a formula to compare the stability in the performances of two interpreters or one interpreter and a speaker.

In Model 4, variance was calculated for the entire second track discourse and was divided by the total variance in the first track; thus it portrayed the ratio of variances. To monitor such a ratio, it is necessary to cumulatively calculate it for each interval. An adaptive formula similar to that in the Speech Monitoring Model can serve this purpose.

Model 4: which can be rewritten as:

$$\frac{\left(\dfrac{\displaystyle\sum_{j=1}^{n}\ \sum_{i=1}^{k_j}\left(S_{ij}^2 - \bar{S}^2\right)^2}{\displaystyle\sum_{j=1}^{n} k_j - 1}\right)}{\left(\dfrac{\displaystyle\sum_{j=1}^{n}\ \sum_{i=1}^{k_j}\left(S_{ij}^1 - \bar{S}^1\right)^1}{\displaystyle\sum_{j=1}^{n} k_j - 1}\right)} \approx 0$$

where v^1 and v^2 are the variances of speech bursts in Track1 (the Speaker's or Interpreter X's) and Track2 (the Interpreter's or Interpreter Y's) respectively, S^1 and S^2 are likewise speech bursts in Tracks 1 and 2, $\bar{S}^1$ and $\bar{S}^2$ are the average speech bursts in all previous intervals within Tracks 1 and 2, $\sum_{j=1}^{n}$ is the sum of intervals $j=1$ through n, $\sum_{i=1}^{k_j}$ is the sum of speech bursts $i=1$ through k within the jth interval, and k_j is the number of speech bursts in the jth interval.

As in the Speech Monitoring Model where the speech ratio was calculated for the intervals cumulatively, the speech burst variance is calculated here for the first interval in Track2 and divided by its counterpart in Track1. Then the speech burst variance is calculated for the 1st and 2nd intervals in Track2 and divided by the counterpart variance in Track1, and so on until the last interval in the discourse. Every time the speech burst variance in Track2 is calculated, we get a value on the degree of stability, consistency, or tribulation, however one may view it. And every time we divide that value by its counterpart in Track1, we get a ratio that indicates the degree of similarity between the two discourses in terms of stability, consistency, and tribulation. When this ratio is roughly equal to 1, the two discourses are said to have a similar level of stability, but when the ratio is greater than 1, the discourse whose ratio is the numerator will be considered to have a more turbulent performance. Thus, the ratio produced by the Speech Variability Monitoring Model is indicative of which interpreter was more stable and had less tribulation in their performance.

Conclusion

Through the models presented in this paper, it has been possible to explore some tangible and some elusive aspects of simultaneous interpretation. We can now discuss chunking, mirroring, and reformulation quantitatively and reflect on the overall development of an SI discourse. It is possible now to comment on an interpreter's tempo by describing their broad and narrow TL periodicity, and to learn about their composure and tribulation by describing the stability and fluency in their TL discourse.

Comparing the chunking patterns in our SL/TL discourses to study the extent of interpreters' mirroring has enabled us to make generalisations about the performances of those who participated in our experiment. We have seen how the professionals reformulated SL and endeavoured to avoid SL fragmentation, while novices failed to be in

control of reformulation; they divided up the SL discourse into manageable fragments. Barik (1973: 267) observed something similar when he concluded that "less qualified [interpreters]… tend to render their versions in more fragmented form."

The study of mirroring compared SL/TL volumes of activity and inspected the spans of delay in TL. Our professional subjects were distinctly more active in the interpretation process than novices and they appear to have produced only slightly less speech than that in the SL discourse. Furthermore, they rendered SL bursts after a substantially shorter delay than that of the novices.

Shadowers in our experiment form two groups: one avoided fragmenting SL bursts, matched the SL volume of speech, and delivered their bursts after a very short delay, while the other fragmented SL bursts as novices did but had a volume of speech substantially larger than novices' and their delay was similar to professionals.'

The adaptive monitoring instrument offered by model 5 facilitated a graphic representation of the linear development of the SI discourse, which made it possible to compare TL to SL discourse progression (development), thereby facilitating the assessment of convergence and divergence between them. The model allows us to quantify development by calculating interval, cumulative, and comparative speech ratios; together, these portray such elusive aspects of SI performance as periodicity, stability, and fluency. With these being quantified, interpretation investigators can talk about SI discourses with more confidence, since they can support their views with numerical evidence.

There is no doubt that the models discussed here are but a modest attempt at objectifying SI stylistic analysis. They are tools that can facilitate what has hitherto been discussed in vague terms. They are precise enough to allow the investigator to witness instances of reformulation and fragmentation, and to talk about how consistent, composed, and fluent an interpreter is. They can give researchers a mechanism for analysing interpreter style without having to resort to the content of their interpretation. Furthermore, they can offer interpreter trainees instantaneous feedback that may help them modify their style and improve their coping performance.

NOTES

1. 'Mirroring' is used here in the sense of 'imitating' or 'reflecting a picture.' It is not intended to imply, as well, exact correspondence.
2. 'Reformulation' is a superordinate term that covers both fusion (merging several SL bursts into one TL burst) and fragmentation (dividing up one SL burst into several TL bursts).

REFERENCES

AIIC (1982): *Practical Guide for Professional Interpreters*, Geneva, Association Internationale des Interprètes de Conférence.

BARIK, H. C. (1969): *A Study of Simultaneous Interpretation*, thesis, University of North Carolina at Chapel Hill.

—— (1973): "Simultaneous Interpretation: Temporal and Quantitative Data," *Language and Speech*, 16-3, pp. 237-270.

—— (1975): "Simultaneous Interpretation: Qualitative and Linguistic Data," *Language and Speech*, 18-3, pp. 272-297.

CARETLLIERI, C. (1983): "The Inescapable Dilemma: Quality and/or Quantity in Interpreting," *Babel*, 29, pp. 209-213.

CHAFE, W. (1988): "Punctuation and the Prosody of Written Language," *Written Communication*, 5, pp. 395-426.

—— (1980): "Some Reasons for Hesitating," *Temporal Variables in Speech: Studies in Honour of Frieda Goldman-Eisler* (Hans W. DECHERT and Manfred RAUPACH, eds.), The Hague, Mouton.

CHERNOV, G. V. (1969): "Linguistic Problems in the Compression of Speech in Simultaneous Interpretation," *Tetradi Perevodchika*, 6-1, pp. 52-65.

CRYSTAL, D. (1969): *Prosodic Systems and Intonation in English*, Cambridge, Cambridge University Press.

GERVER, D. (1969): "The Effects of Source Language Presentation Rate on the performance of Simultaneous Conference Interpreters," *Proceedings of the 2nd Louisville Conference on Rate and/or Frequency Controlled Speech* (E. FOULKE, ed.), Louisville (Kty), University of Louisville.

—— (1972): "Simultaneous and Consecutive Interpretation and Human Information Processing," *Research Report HR 566/1*, London, Social Science Research Council.

GILE, D. (1991a): "Methodological Aspects of Interpretation (and Translation) Research," *Target,* 3-2, pp. 153-174.

—— (1991b): "A Communication-Oriented Analysis of Quality in Non-Literary Translation and Interpretation," *Translation: Theory and Practice. Tension and Interdependence* (Mildred L. LARSON, ed.), Bringhamton (NY), State University of New York, pp. 188-200.

—— (1995): *Basic Concepts and Models for Interpreter and Translator Training*, Amsterdam/Philadelphia, John Benjamins.

—— (1997): "Conference Interpreting as a Cognitive Management Problem," *Cognitive Processes in Translation and Interpreting* (Joseph H. DANKS, Gregory M. SHREVE, Stephen B. FOUNTAIN, and Michael K. MCBEATH, eds.), Thousand Oaks (Calif.), Sage Publications, pp. 196-214.

GOLDMAN-EISLER, F. (1968): *Psycholinguistics: Experiments in Spontaneous Speech*, London, Academic Press.

—— (1972): "Segmentation of Input in Simultaneous Translation," *Journal of Psycholinguistic Research*, 1-2, pp. 127-140.

JONES, R. (1998): *Conference Interpreting Explained*, Manchester, St. Jerome Publishing.

KOPCZYNSKI, A. (1994): "Quality in Conference Interpreting: Some Pragmatic Problems," (Sylvie Lambert and Barbara Moser-Mercer, eds.), *Bridging the Gap: Empirical Research in Simultaneous Interpretation*, Amsterdam/Philadelphia, John Benjamins, pp. 87-100.

LEDERER, M. (1978): "Simultaneous Interpretation: Units of Meaning and Other Features," *Language Interpretation and Communication* (David GERVER and H. WALLACE SINAIKO, ed.), New York and London, Plenum Press, pp. 323-332.

OLÉRON, P. et H. NANPON (1964): «Recherches sur la traduction simultanée», *Journal de psychologie normale et pathologique*, 62-1, p. 73-94.

PÖCHHACKER, F. (1994): "Quality Assurance in Simultaneous Interpreting," *Teaching Translation and Interpreting 2: Insights, Aims, Visions; Papers from the Second Language International Conference* (Elsinore, Denmark, June 4-6, 1993), Amsterdam/Philadelphia, John Benjamins, pp. 233-242.

SELESKOVITCH, D. (1986): "Who Should Assess an Interpreter's Performance?," *Multilingua*, 5-4, p. 236.

SHLESINGER, M. (1997): "Quality in Simultaneous Interpreting," Y. GAMBIER, D. GILE, and C. TAYLOR (Eds), *Conference Interpreting: Current Trends in Research, Proceedings of the International Conference on Interpreting; What Do We Know and How? (Turku, August 25-27, 1994)*, Amsterdam/Philadelphia, John Benjamins, pp. 123-131.

TREISMAN, A.M. (1965): "The Effects of Redundancy and Familiarity on Translating and Repeating Back a Foreign and a Native Language," *British Journal of Psychology*, 56-3, pp. 369-379.