Proceedings of the 52nd Conference of the American Translators Association. New York: ATA, 2011.

COGNITIVE THEORY OF SIMULTANEOUS INTERPRETING AND TRAINING

Erik Camayd-Freixas, Ph.D.
Florida International University

Abstract: This is an original theory of simultaneous interpretation (SI) based on relevant principles of cognitive psychology, linguistics, and introspection by advanced professional practitioners about the mental processes and techniques utilized during actual performance of the craft. The objectives of the theory are: 1) to describe the different tasks involved in the process of SI; 2) to isolate each task in order to target the corresponding skill during focused training; 3) to describe the flow of tasks into a seamless SI process and the correct techniques that help to optimize performance; and 4) to lay the grounds for devising training methods and skills-building exercises for advanced SI performance. Thus far existing SI skills-building exercises, including shadowing, dual tasking, and paraphrasing, tend to focus exclusively on intra-lingual tasks, without actual inter-lingual translation. As such they are designed for the beginner. In contrast, the present cognitive theory of SI focuses on advanced techniques and optimization training methods.

1. INTRODUCTION

Simultaneous interpretation is arguably the most demanding of the three modes of interpreting. This is due in part to the fact that in SI the interpreter does not control the flow of communication, but must adapt to the pace of each speaker, without interrupting the proceedings. If anything is missed, there is generally no opportunity to ask for a repetition or clarification. The interpreter must be thoughtfully following the ideas in the speech, the "thread" of the conversation, as it were, and must be prepared to extrapolate meaning from context, in order to bridge or gloss over the spots that were unclear. The second source of difficulty is of course the fact that everything happens very fast: the interpreter must thoughtfully follow the fast-paced original speech, while at the same time formulating and delivering an equivalent TL version, and monitoring his/her own output for errors.

In this article we will examine the various elements and skills involved in this complex multitasking process we call simultaneous interpretation. We will do so step by step, considering each task and the techniques and exercises that reinforce it, as part of a methodical skills-building approach designed to develop confidence through knowledge of fundamentals in theory and steady improvement in practice. Given the challenges involved in SI, skills improvement is expected to be slow but steady. For this reason, it is wise to begin SI study and practice early on in your training program.

2. PRINCIPLES AND TECHNIQUES

2.1 Focus on Meaning not Words

The most basic principle applicable to all modes of interpretation is that, for the most part, interpreters do not translate the words, but rather reconstruct and restate the meaning. If we were to actually translate the words and phrases in a message, the process would take too long, and we would inevitably get stuck every time a source-language (SL) word in unclear or a target-language (TL) word evades us. So instead, interpreters listen and understand the SL message; they mostly let go of the original words and keep only the meaning; they make the meaning their own or "become" the speaker; and then they quickly restate the meaning in their own words in the TL, as though expressing their own ideas. Since there are usually various ways to express the same idea, they are not likely to get stuck looking for a particular word or for the perfect translation. When SL words are understood, they are abstracted into pure a-lingual concepts, ideas, and images, called *schemas*. A *schema* is an abstract (wordless) conceptual representation of an object, event, or situation, including any chronology, logical sequence, and causal relationship, in short, a mental image. There may be, however, in the original message, certain words that we cannot discard and that we have to actually translate; we shall call them *terms*. Yet, *terms* are the exception, not the rule. For example, in the following statement –"My dad drives a stick shift. So I had to borrow my friend's car, which is an automatic, so I could go take the test to get my driver's license"- the only *terms* to be translated are "stick shift" and "driver's license." Everything else is *schemas* to be reconstructed and restated. Thus, the basic principle of interpreting is: ***restate schemas, translate terms***.

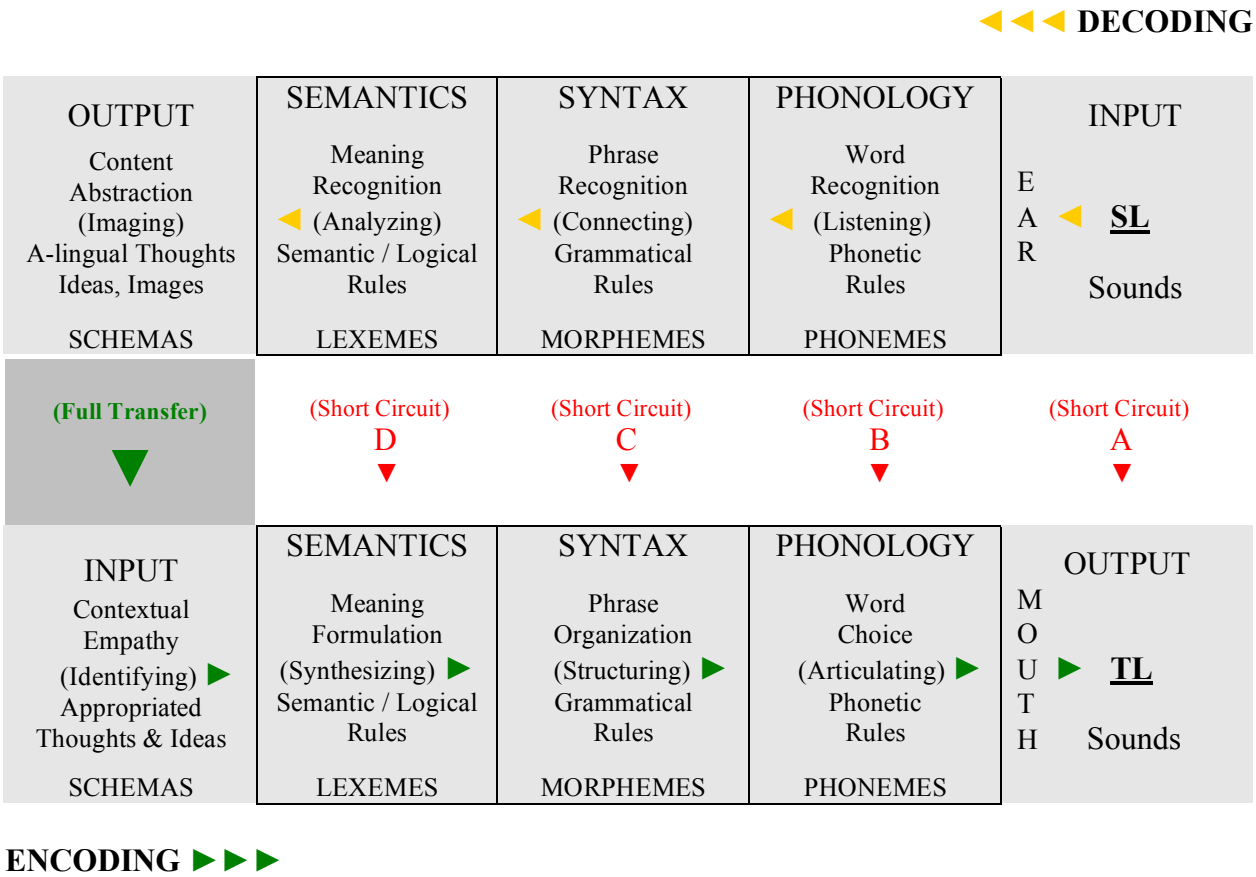

Fig. 1 Jensen Transformational-Generative Model of Interpretation Processing

In an effort to rigorously describe the complex mental processes that take place during interpretation, cognitive scientists have developed various theoretical models, which are for the most part too abstract to be of any real use to interpreters and trainers. However, my colleague Dr. John B. Jensen, a linguist at Florida International University and a simultaneous conference interpreter, has developed an illustrative model based on Noam Chomsky's transformational-generative grammar. I have adapted it to the present discussion on SI by adding tasks and functions at each stage that correlate to the SI process discussed later.

The Jensen Model correctly emphasizes the importance of processing the source message fully; leaving behind the original words and syntax, in order to abstract the meaning all the way back to pure (a-lingual) thoughts, ideas, and images (schemas); making that abstracted meaning one's own (empathic analysis); and then re-dressing it idiomatically in the syntax, words, and expressions that are natural to the target language. If the process is short circuited at any stage, the end product will not be an equivalent or natural-sounding TL version of the original message, resulting in a faulty interpretation.

2.2 The SI Process Step by Step

Having established the basic principle of restating meaning and translating only the terms, let us now consider the step-by-step process of simultaneous interpretation. It is with very specific training objectives in mind that we divide what is essentially a fluid process into steps or, rather, into isolated tasks. The first of these objectives is diagnostic: to help the trainee identify those specific SI tasks in which he/she has trouble. The second objective is remedial: to device practice exercises that specifically concentrate on the tasks that need improvement. The third objective is developmental: to learn how to isolate each task and exercise it until it becomes effortless, second nature, or mindless and "automatic." This last objective —making as much of the process as possible virtually automatic— is the key to mastery in simultaneous interpretation. Finally, let us not forget that SI is fluid (non-linear) multitasking. So working on an isolated task is not enough. We must learn which tasks tend to interfere with each other, and we must practice them in combination, until all the component tasks are integrated into a relaxed seamless performance.

Fluid listening

Active listening is the foundation of all interpreting, because we cannot interpret what we have not heard, understood, analyzed, and abstracted. So listening not only comes first chronologically speaking, but it is also the first priority in SI multitasking. That is, listening is the one task that must be fluidly maintained, even as attention is momentarily diverted to one of the other SI tasks. Whereas in consecutive interpretation one has the luxury to do nothing but concentrate quietly, intensely, and almost exclusively on the task of active listening, in SI there are several other competing tasks demanding our attention. In particular, the task of formulating the TL version interferes mentally, while our own voice as we interpret interferes aurally with the incoming message. Therefore our attention must be fluid and flexible: we may tune in more or less as we attend to other tasks, but we must never tune out. Here, good volume on our headsets is extremely helpful for deep message reception, as well as for long-lasting retention when we are trailing behind the speaker (see "acoustic memory" below). Headset volume should be as loud as

comfort and conditions allow. Another important aspect of fluid listening is to continuously follow the unfolding meaning of the entire speech. Following the "thread" of meaning is of vital importance for maintaining SI speed and accuracy. Novice interpreters tend to focus only on each immediate phrase, such that at the end of the speech they can report little about the ideas expressed. Instead of fluidly listening and following the line of meaning as it unfolds, the novice focuses too much on translating words and phrases, at the expense of accuracy, speed, and endurance. Finally, fluid listening means listening in a state of *relaxed concentration*, letting the words flow through you while you follow the meaning, without allowing your attention to get fixed, to linger on any one spot, or to dwell on any particular word or idea.

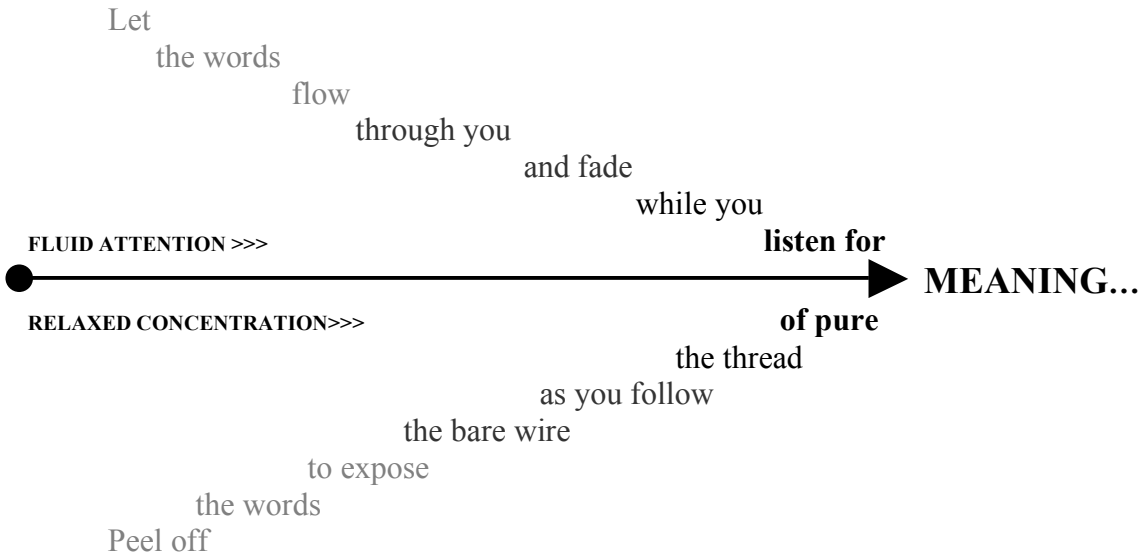

Fig. 2 Following The Thread of Meaning without Fixating on Words

Empathic analysis

Analysis is the process of: 1) letting go of the words in order to abstract the meaning (a-lingual images, ideas, thoughts); and then, 2) relating that meaning to our previous knowledge and experience, as stored in our long-term memory. In fact, when we relate, accommodate, and classify the new information within the context and categories of our previous knowledge and experience, we are said to have "understood." Our understanding of the message becomes deeper, and our interpretation more convincing, if we do not put up any resistance to the new information or to the values and assumptions that may be implicit in the incoming message, whether we agree with it or not. We should rather internalize the message within our knowledge base; take ownership of the ideas, make them our own; and identify with and "become" each speaker, with empathy for their situation and way of thinking, in order to render the message faithfully, not only in content, but also in spirit, tone, and intent. The interpreter should be transparent and allow the personality and state of mind of each speaker to shine through. "Empathic analysis," then, is an acceptance and ownership over the incoming ideas as we relate them to our previous knowledge, assumptions, attitudes, beliefs, and value system. The result is a role playing where the interpreter fully takes the place of and represents each speaker, leaving his

or her own values and convictions aside. Yet, this is only role-playing. In the background, the interpreter's conscience remains detached, even skeptical, and engages rather in "critical analysis." Thereby, the interpreter identifies points of disagreement and contention, but is aware of setting them aside for the sake of the role-playing and faithful interpretation. So, in fact, the interpreter plays the role of both listener and speaker, and while the role of listener comes first chronologically in the interpreting process, its critical conscience remains suppressed and uninvolved in the background as a neutral, non-influential observer, whereas the role of speaker takes over and becomes dominant through empathy, moving outward to the forefront to be publicly and convincingly expressed. Thus, we can now refine the basic principle of interpreting as: ***become each speaker; restate schemas, translate terms***.

Segment recognition

An SI segment is the shortest possible unit of clear, unambiguous meaning, which contains enough information to be interpreted unequivocally. When you listen analytically, you will recognize each segment as it becomes complete, as soon as doubts and ambiguities are cleared by additional incoming information. For example, if you hear "The chair called...," the information is still ambiguous; but if you hear "The chair called for a meeting...," then it becomes clear. Further, you will know that the segment is complete when you are able to abstract it into pure, a-lingual meaning. Instantly, a new mental image or schema is configured, which corresponds to your previous knowledge. A match is recognized, something clicks, you feel the "Aha!" experience, and you know you have "understood." Now you are ready to formulate and render the segment. All of this happens more or less automatically, intuitively. In fact, the functions of listening-analyzing-understanding-segmenting make up a single inseparable sequence. They occur in unison, without additional effort or deliberation.

In general, SI segments should be as short as possible, so that the interpreter may render each one quickly and move on fluidly without falling behind. Let's continue our previous example and see how we might make the segments shorter: "The chair called for a meeting of the steering committee." We had said that "The chair called..." is still ambiguous and "...for a meeting" clarifies and completes the meaning. So here we have two segments: "The chair called for a meeting | of the steering committee." However, if we were thoughtfully following the meaning of the entire speech as it unfolds, we might have previous contextual information to be able to recognize shorter segments as clear and complete. That is, if we were really paying attention, we might know that, here, "chair" means "director" and not a place to sit! We might also be able to predict that "called for" will be followed by "meeting." Moreover, we may be able to anticipate "steering committee" if it was mentioned earlier in the speech. Thus, following the meaning allows for contextual prediction, which enables us to make SI segments shorter. Short SI segments pose a distinct advantage for the interpreter, who can gain valuable time by immediately starting to formulate a rendition, thereby making it easier to keep apace of the speaker. We may thus render our previous example in shorter segments: *"El presidente | convocó | a una reunión | del comité directivo."*

Having said this, it is important to note also that short segments can be "chunked" together into one larger segment or compound unit of meaning. So our example, "The chair called for a meeting of the steering committee," can in fact also be considered one segment, a single idea. Therefore, a segment is not a rigid but a fluid quantity, whose length can be strategically

managed: expanded or contracted. This is important because, if shorter segments maximize our speed, longer segments maximize our memory and help us follow the thread of meaning in order to catch up when we have fallen behind, by taking *shortcuts*. It is much like taking long strides as opposed to quick little steps. Just like a runner adjusts his stride to negotiate the contour of the terrain, so must the interpreter regulate the length of segments according to the speed, density, and syntax of the incoming message (see "managing the décalage" below). This flexibility is possible because SI segments do not depend on words but on ideas, and ideas are fluid: they can be combined or divided as needed, in order to strategically cope with the demands of either memory or speed. This is yet another reason to focus on meaning, not words. The number of words in a message may be fixed and inflexible, but its ideas are fluid. Units of meaning are flexible: we segment them, not according to fixed rules, but at our convenience, simply to make our interpretation task easier.

Managing the décalage

Décalage is the amount of time (measured in ideas or segments) that the simultaneous interpreter lags behind the original speaker. In general, this is the time it takes the interpreter to process, formulate, and deliver each segment. Understandably, by then, the original speaker has moved on. Speakers are, and should be, oblivious to the interpreter and mindful only of their own speech. Typically, they will not wait for the interpreter. It is the SI interpreter's job to adapt to the speaker, and not the other way around. So, if a *décalage* is naturally induced by the time it takes to render each segment, one may be led to assume that it is a fixed quantity, always one segment long, in other words, that the interpreter would always want to lag exactly one segment behind the speaker, but this is far from true. Both the segments and the *décalage* are elastic, and their length can be managed. Managing segment length and *décalage* is indeed a fundamental strategy for successful simultaneous interpretation.

Obviously, a fast and dense speech will force the interpreter to lag further behind. We may call this "forced" *décalage*, in the sense that it cannot be deliberately managed or adjusted. The minimum quantity of forced *décalage* is always one segment, since one must at least listen to a full segment before it can be interpreted. Forced and managed *décalage* are in inverse proportion to each other: the more the *décalage* is induced by force, the less it can be managed. This is often the predicament of novice SI interpreters, who are forced by speed or density into a longer *décalage* than they can handle, and thus routinely fall behind and lose the thread of meaning. This is due to a combination of reasons, which can be addressed by proper training: 1) they lack the speed to minimize the *décalage* because a) they get stuck on translating words instead of restating meaning, or b) they lack practice in shortening segments and rendering them quickly and fluidly, so as to *heel* close to the speaker; 2) they lack the capacity for long *décalage* due to poor memory strategies and faulty segmentation techniques, because a) they focus too much on words rather than following the thread of meaning, or b) they lack practice in expanding the segments for enhancing memory and in abbreviating ideas through shortcuts as a technique for catching up. In this regard, training has a two-pronged approach: to build up the focus and discipline to stay just one segment behind the speaker for extended periods, and to push the capacity for handling a longer *décalage* without losing the thread of meaning; in other words, to minimize and to maximize the *décalage* by shortening and expanding segments. Through practice, experienced interpreters are able to approach the minimum forced *décalage* of one

segment, almost at will, and also have plenty of reserve capacity and catching-up ability to feel confident in increasing the *décalage* by choice, as a matter of strategy and technique.

In general, it is advisable to keep the *décalage* to a minimum and follow the speaker closely. That way if there is an increase in speed or *density* that forces a longer lag, one has enough reserve capacity or "enough room" to move back without falling too far behind to be able to recover. However, there are risks and drawbacks in attempting to maintain the *décalage* at a steady minimum. First, it is a stressful struggle that results in early fatigue. Second, it increases the potential for errors, self-corrections, and false starts, as a result of cutting segments too short. Third, staying too close to the speaker often leads to choppy, stop-and-go or staccato delivery, as well as excessive voice tension and an unpleasant hurried tone, all of which the listener perceives. Such excessive *heeling*, which is common when an interpreter is afraid to fall behind, only serves to project insecurity. On the other hand, good interpreting often requires the opposite strategy: to increase the *décalage*. This technique called *queuing* (literally, standing in line) consists in simply stopping to interpret for two or three seconds in order to just listen. It is used precisely when a segment of speech is particularly dense or fast; when there has been a distraction or lapse of attention; when low volume or sound quality makes it difficult to follow the speaker; or in any other circumstance where the interpreter begins to lose the thread of meaning. *Queuing* as a technique involves: stopping to listen; picking up the thread of meaning; and catching up, not so much by speaking fast (increasing delivery speed), but by using *shortcuts* (abbreviating ideas and eliminating redundancies). *Queuing*, thus, involves several skills that should be practiced. It is used routinely and effectively by experienced interpreters, who feel secure in their ability to handle and recover from a longer *décalage*.

In sum, flexibility is crucial, because the capacity to strategically manage (increase or decrease) the *décalage* is the key for not falling behind. An experienced interpreter, in fact, sits back and finds a comfortable *décalage*, neither too short nor too long (because both ends are stressful). This comfort zone is also borne by the interpreter's confidence in his or her speed and memory, so as to minimize the *décalage* if the need arises (as when the speaker speeds up) or to tolerate a longer lag without losing the thread of meaning. From this comfort zone, the interpreter instinctively adjusts the *décalage* in either direction, *heeling* or *queuing* as the situation requires, in response to speed, density, syntax, and other conditions. Such instinctual sense is developed with experience. Yet, the comfort zone should never be too far from minimum *décalage*, in order to maintain adequate reserve capacity at the back end. In the last analysis, the reason why we manage *décalage* and segment length is simply because we must make whatever adjustments are necessary in order to stay with the thread of meaning, at all costs. If we lose the thread, we have no choice but to skip over some information, make up a transition, pick up the thread again and keep going. Thus, we may further refine the basic principle of interpreting as: ***never let go of the thread of meaning; become each speaker; restate schemas, translate terms***.

Here, a word might be said about *shortcuts*, that all-important technique that makes it possible to increase the *décalage* and still be able to recover and catch up. Shortcuts involve abbreviating ideas by eliminating redundancies and expressing the same meaning with fewer words. This does not mean that information is omitted. In other words, abbreviating does not mean summarizing. If any information must be sacrificed, a distinction must be drawn between substantive and accessory meaning. Accessory meaning includes emphasis, reiteration,

redundancies, clarifications, and certain modifiers and descriptors, which may be eliminated without sacrificing the essential information or substantive message. If any substantive meaning, however, is eliminated, accuracy is compromised and the rendition may become unacceptable. Thus, interpreters should practice distinguishing between substantive and accessory meaning in different utterances, eliminating redundancies, and expressing ideas accurately but in their simplest and most economical form. The following example presents various levels of abbreviation of the same ideas, where shortcuts become progressively drastic. Notice to what extent ideas may be acceptably abbreviated without sacrificing substantive meaning. Remember that the only reason to take shortcuts is to keep apace and thereby safeguard the integrity and completeness of the substantive message.

1-	Another new initiative | that we are working on right now | is the renewal of those contracts | which are scheduled to expire | at the end of this year.	**[5 segments]**
2-	Another new initiative | that we are working on now | is the renewal of contracts that expire | at the end of this year.	**[4 segments]**
3-	We are also working on renewing contracts | that expire at year's end.	**[2 segments]**
4-	We're also working | on year-end contract renewal.	**[2 segments]**
5-	We're renewing year-end contracts as well.	**[1 segment]**
6-	We're renewing contracts too. (Information sacrificed: "year-end")	**[1 segment]**

Table 1 Shortcut Strategy: Abbreviating the Idea by Eliminating Redundancies

This illustrates just how powerful shortcuts can be as a catching-up technique. Here, even five segments are "chunked" into one, but this requires *queuing*, increasing the *décalage* in order to listen to various segments and to have them available to chunk together. This is not possible if you are *heeling* too close to the speaker. As the *décalage* increases, shortcuts become possible (and even necessary). In reality, therefore, shortcuts are a technique for speed; sometimes, you get there faster by taking long steady strides than quick short steps. Shortcuts work by eliminating redundancies in order to say the same thing with fewer words. With practice, and the confidence to allow for a comfortable *décalage*, shortcuts allow us to cover plenty of ground with a steady, elegant pace, while keeping our delivery from sounding hurried or hectic.

Many trainers advocate eliminating redundancies constantly and routinely, particularly in conference interpreting. Redundancies can and should be eliminated in SI to save time, without omitting any substantive information. How much of the information is redundant (and therefore expendable) depends on the immediate context: whether the information was mentioned earlier in the speech, yet recently enough to still be a fresh antecedent. If so, there is no need to repeat it. This is, however, a decision-making process, which can be taxing. There is a price to be paid in

terms of stress. When you interpret everything just like you hear it, you can function more or less mindlessly, on "automatic pilot" so to speak. But when, on top of the interpreting task, you are also making decisions as to whether information is redundant and how to glean out the real substance of the message, the level of concentration can get exhausting after a while. So eliminating redundancies is really a technique for catching up, not an activity that can be sustained for very long. Still, interpreters should practice it intensively in their training, so that it may become more automatic, efficient, and sustainable; that is, so that they can increase their capacity.

We have arrived at a fundamental aspect of SI performance and training: minimizing and maximizing *décalage* (that is, the techniques of *heeling* and *queuing*). Let's review: 1) minimizing and maximizing *décalage* is inseparable from minimizing and maximizing segments; 2) both activities require intense concentration, which quickly becomes fatiguing; 3) therefore, a key objective of SI training is to promote efficiency (for speed) and relaxed focus (for endurance) in *décalage* and segmentation management, through intensive and sustained practice in the technical elements of *heeling* and *queuing*; 4) both activities lead to certain accuracy risks and performance disadvantages, so the interpreter should find a suitable middle, or comfort zone, as the default setting from which to adjust backward and forward; 5) through practice and experience, the interpreter should develop an instinctive sense for this adjustment or regulation, in the strategic and efficient management of segmentation and *décalage*; and 6) this instinctive regulation is a conditioned response to: a) the variable "terrain" of the incoming message, and b) the interpreter's own state of concentration and energy reserves. Thus, an experienced simultaneous interpreter goes with the flow, negotiating changes in message speed, density, syntax, and sound conditions, while budgeting his own concentration and energy resources. From the default comfort zone, the interpreter aggressively *heels* when the speaker speeds up, and *queues* when the message increases in density or difficulty. In addition, the interpreter constantly regulates the *décalage* along this continuum, *heeling* or *queuing* more or less aggressively, and *coasting* or resting periodically in the comfort zone, according to his current energy level and taking advantage of any restful opportunities to catch a second wind and come back with a burst of concentration. Since *heeling* and *queuing* are stressful in different ways, and yet both are techniques for speed, which ultimately serve the same purpose, the interpreter may rest from one by moving to the other across a period of *coasting*, or simply alternate among them to prevent fatigue.

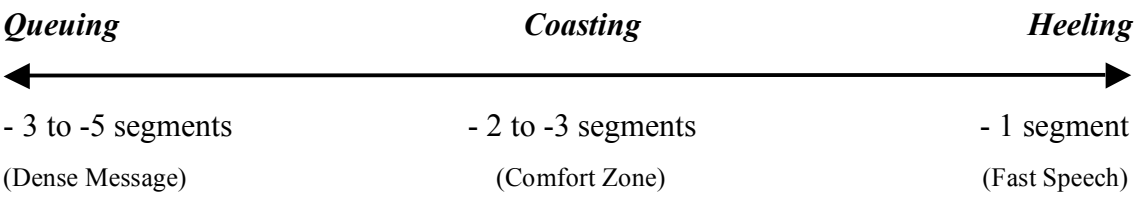

Fig. 3 The Décalage Continuum

Typically, an interpreter will begin a session by heeling aggressively at minimum *décalage*. Yet, after a couple of minutes of intense heeling, the interpreter will be warmed up, will get used to the speaker and the topic, will establish a confident pace, and will relax into the comfort zone of

a moderate *décalage*. From then on, the interpreter will flexibly regulate his stride with variable segmentation and elastic *décalage*, negotiating the contours of the message and budgeting energy efficiently. This can be maintained for 20-40 minutes of continuous SI, depending on message speed, density, and conditioning, before fatigue sets in. What actually happens with fatigue and diminished concentration is that the interpreter loses the ability to comfortably regulate the *décalage*, which becomes entirely forced. Like the novice, the fatigued interpreter struggles to keep up and hold on to the thread of meaning, reacting forcibly at the mercy of the message, with no room to maneuver, the result of depleted capacity. The interpreter should recognize such early signs of fatigue and either ask for a break if working alone or switch with a fresh colleague, before performance begins to falter and accuracy is compromised.

1	**2**	**3**
Listen and Process Segment # 1	**Chunk together Segments # 1 & 2**	Listen and Process Segment # *n*...
		Deliver as one Segments # 1&2

Table 2 Moderate Décalage with Variable Segmentation

Formulation, comparison, adjustment, delivery, and self-monitoring

Here is another inseparable sequence in the SI process. We have completed the "reception" phase (decoding the incoming SL message) and are now entering the "production" phase (encoding the outgoing TL message). We have abstracted the source message into pure thought or conceptual *schemas*, except for perhaps a few *terms* that we have actually translated mentally. Now, to produce the interpreted message, we first formulate it as we speak, by putting together the schemas and terms into expressions that are idiomatic and natural to the target language, exactly as though we were expressing our own thoughts, in our own words. Thus, interpreting involves making the message your own, and expressing it naturally as though they were your own ideas. Notice that when we speak our own mind and express our own ideas, we do not formulate exactly what we are going to say before we begin speaking. Rather, we know the general direction (schemas) in which we want to go in order to express ourselves, and we simply find the words as we go along. As we unfold our schemas in the process of verbalizing them, the words simply seem to "come to us" to fill in the skeletal structure (syntax) of entire phrases, since we naturally construct speech, not one word at a time, but one phrase or segment at a time. This is because our conceptual schemas are strings of organized thought, equivalent to a semantic structure, that is, a structure of meaning containing logical and temporal relationships, such as cause-and-effect, how-and-when, who-did-what-to-whom, etc. To each semantic structure, there correspond one or more grammatical phrase structures, such as subject-verb-object plus prepositional phrases, which we are conditioned to associate rather effortlessly during our own natural speech. By substituting different words into our finite syntactic repertoire of stock (yet flexible) phrase structures, we are able to express a seemingly infinite variety and combination of schemas.

The central lesson here is that the closer our interpretation process is to our natural speech, the more correct its form, the easier the task, and the better the product. In terms of formulation, the natural (and therefore correct) way is to formulate on the go, as you speak, finding the words as you go along. Once again, we may further refine the basic principle of interpreting as: ***never let go of the thread of meaning; become each speaker; restate schemas, translate terms, formulating as you go***.

An integral part of formulation is a barely conscious process whereby we instinctually compare our version to the original, and quickly make any necessary adjustments, before delivering our revised version. Here is where the *translator* consciously looks for perfection, while the *interpreter*, pressed for time and fluency, settles for acceptable equivalence. The interpreter must be comfortable with "close enough" because excessive comparison and adjustment results in hesitation, which compromises the quality of the delivery. Thus, formulation-comparison-adjustment is another inseparable SI sequence whose steps occur almost instantly and in unison, to the point that they are virtually indistinguishable from each other. Finally, when we deliver our interpretation, another more conscious and deliberate comparison takes places, between what we say and what we intended to say. That is, we must continuously monitor our delivery, because in SI it is very easy for an incoming word or idea to be substituted into our outgoing message. Monitoring, however, is a kind of passive listening that requires only minor concentration, since it does not involve any information processing (at this point the information has already been processed and delivered). Monitoring involves a sort of listening "with one ear" for recognition and back-match between execution and intention. With practice, interpreters learn to monitor themselves mindlessly, until a discrepancy is detected, which sounds to the experienced interpreter like a loud dissonant alarm, signaling that a quick self-correction is necessary. That is, monitoring requires little attention because it works on the homeostatic principle. In conclusion, monitoring must be considered an inseparable part of the delivery, a mere follow-through of the production process that began with the formulation phase. Yet, since we formulate as we go, that is, as we deliver, both sequences [formulation-comparison-adjustment] + [delivery & monitoring] are actually non-linear, intertwined, and ultimately inseparable parts of a single fluid movement. Formulation, comparison, adjustment, delivery, and monitoring are like phases of a single brushstroke.

Clean delivery

Delivery should be smooth and even-paced, emulating the quality of broadcast voice. In order to achieve smooth pacing, the interpreter should avoid: a) jerky spurts typical of too short a *décalage* (heeling); b) stop-and-go where SI actually becomes a sort of fast consecutive; and c) hurried slurs that result from catching up after queuing, without using shortcuts. The voice should be clear and expressive, with meaningful tone; yet diction should always be natural and never overacted, ultra-correct, or affected. To achieve appropriate expressiveness, the interpreter should "become" the speaker and purposefully direct the communication to the listener. That is, even if you are interpreting from an isolation booth, you should visualize a direct and purposeful face-to-face communication with the audience. This will make the communication meaningful and appropriate, as opposed to a self-involved soliloquy. Finally, each phrase or segment should be formulated and delivered fluidly and confidently, avoiding the hesitant tone, fillers, uneven pace, and insecure volume of one who is deciphering, pondering, or doubting.

2.3 Managing Selective Attention

During simultaneous interpretation, attention must be selectively directed to the various tasks in the *reception* and *production* sequences (listening-understanding-abstracting vs. formulating-delivering-monitoring). This is because human information processing can operate only on a single channel at a time. That is, human beings are unable to process two channels of information at once. Imagine that you are at a cocktail party and you try to listen to two conversations around you. You immediately realize that you can only get bits and pieces of each conversation as you listen to one or the other, but you cannot get them both together: if you concentrate on one, you sacrifice part of the other. However, what you can do —and become good at— is to shuttle back and forth between both, and then, when you come back to each conversation, fill in what you missed by extrapolating meaning from context, between what came before and what comes after.

Channel 1:	***a b c d . . . . h i j . . . . . o p q r s . . . v . . . . y . .***
Channel 2:	***. . . . . 5 6 7 . . . 11121314 . . . . . 2021..2324..26***

Table 3 Selective Attention Shuttling and Extrapolating Meaning

Whereas the previous example involves listening to two conversations (two input/reception channels), actual SI involves an attention split between listening and speaking (one input/reception channel and one output/production channel).

In fact, selective attention management is absolutely crucial to simultaneous interpretation. The key is to make the most efficient use possible of our single-channel attention capability, by knowing how much attention to devote to the listening side and how much to the speaking side, as well as knowing exactly when to switch from one to the other. (By now, you have probably realized that SI is not actually "simultaneous" after all.) Experienced interpreters will often recommend a general percentage of attention split between listening and speaking, which can go anywhere from 80/20 to 70/30 to 60/40 to 50/50 to 40/60 to 30/70 to 20/80. In a way, they are all correct in this diverse recommendation because, in fact, the attention split, like we saw earlier about the *décalage*, must be flexible and subject to continuous adjustment. According to the characteristics of the message, the interpreter has to shift or slide the center of attention to favor listening (e.g., while *queuing*) or speaking (e.g., while *heeling*). At the same time, one could reasonably postulate that for an "average" message (if there is such a thing) there must be an "average" attention split, a sort of optimum point or home base from which to adjust in either direction. Determining this theoretical "center" of attention is of interest for both training and performance. What we are looking for is not an exact percentage but a general tendency for favoring one side or the other and approximately how much. Yet, contrary to the earlier example, which involved listening equally to two conversations, the listening-speaking SI task is asymmetrical, and therefore the optimum point is not simply a 50/50 split. Given that the reception side (listening) involves rather passive mental functions compared to the intensely active production side (speaking), it is certain that the latter generally requires more attention. It is, then, reasonable to assume that the SI attention split tends to favor the production side (formulating-delivering-monitoring) by a margin of 40/60 to 30/70. That is, as a general average, roughly two thirds of the attention goes to producing the interpreted message, and only one third

to the listening side. This is reflected in the relative amount of time one devotes to attending the production of the outgoing message over the reception of the incoming message. Moreover, some of the tasks on each side require more concentration than others. Thus, as a general recommendation, the following chart illustrates the approximate distribution of attention across the different SI tasks.

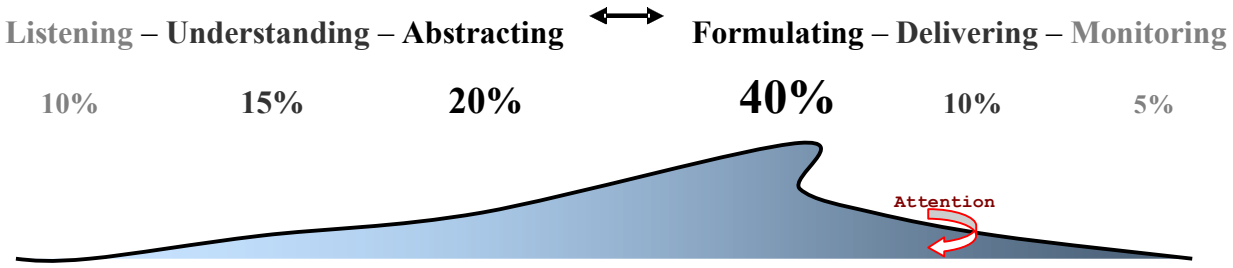

Fig. 4 Suggested Time-Distribution Flow of Selective Attention across SI Tasks

Notice that the SI concentration curve has the shape of a breaking wave, which grows in intensity up to the moment when the interpretation of a segment has been formulated. At that point the wave breaks, and then its momentum flows swiftly and easily through delivery and monitoring. Soon after formulation, selective attention returns to listening, and the process begins all over again for each new SI *segment*. To have an idea of the speed with which this process can take place, consider that human "reaction time" (the time it takes, for instance, between thinking of a phrase and beginning to say it) is approximately 200 milliseconds, that is, about two tenths (.2) of a second. Thus, an SI interpreter commonly handles 2-3 words per second. This is proportionate with two other important psycholinguistic findings: that the optimum listening speed for ease of comprehension is about 170 words per minute, while the optimum interpreting speed to avoid increased stress is about 120 wpm. The interpreter must make up the difference by managing resources strategically, using shortcuts, and speeding up at the cost of increased stress, inevitably fatiguing faster than the original speaker.

According to this "breaking wave" model, the interpreter focuses on formulating the speech and directing it purposefully to a listener or audience, while simply allowing the incoming message to become an "inner voice" that feeds ideas and tells the interpreter what to say. The process should be as natural as expressing one's own ideas, except that our inner voice, which normally feeds us thoughts, images, and schemas to verbalize, is replaced here by the ideas we have abstracted from the outer voice of the speaker. In a very real sense, we become the speaker, while the original speaker becomes a coach who whispers in our ear what we ought to think and say. Once we know what to say, that is, once we have formulated the segment, delivering it into words and monitoring the output are comparatively mindless functions that we can put "on automatic." That is, as soon as our delivery is under way and we can "see the road ahead" (that is, the words we are going to use), we can proceed to say them mindlessly, and monitor them "with one ear," while our primary attention shuttles back immediately to our listening function, in order to understand and abstract the meaning of the next segment.

At this point we might introduce the concepts of primary and secondary attention. Our visual field, for example, has a *focus* at the center and a *fringe* around the edge. The same can be said about our auditory field, about our other senses, and indeed about our mental attention as we think. In all cases, the focus receives primary, central or "active" attention while the fringe receives secondary, "passive" or homeostatic attention.

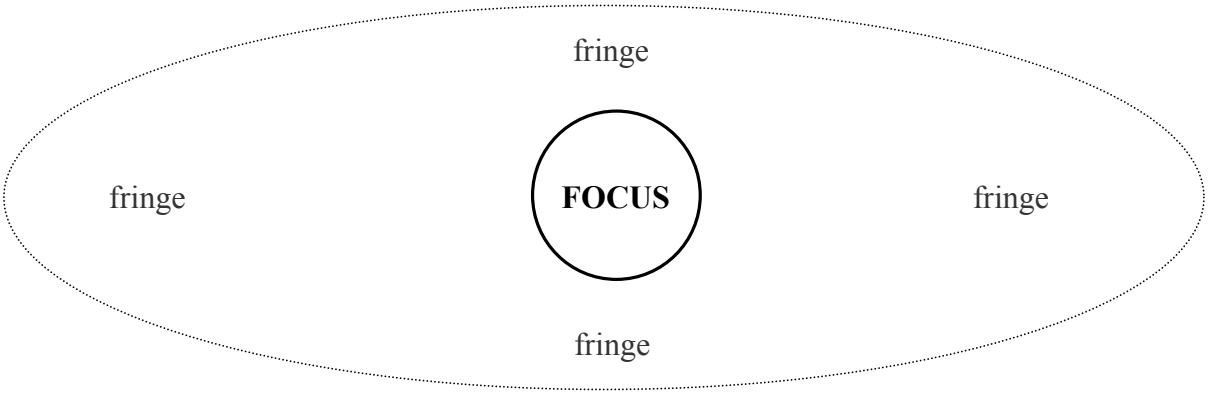

Fig. 5 Primary And Secondary Attention

When an object comes into our fringe or peripheral vision, we say we see it "out of the corner of our eye." Fringe or *peripheral attention* works according to the "stay-the-same" or *homeostatic principle*. That is, when everything in the fringe remains uniform or static, nothing stands out to call our attention. Everything is in equilibrium, and we are scarcely aware that we are actually monitoring the fringe. But when a change or something out of the ordinary that does not fit the status quo enters our fringe, it immediately captures our interest and we shift our focus or *central attention* to it for an instant, in order to cognize it. The same principle applies to our auditory field. Let us go back to our cocktail party example, and let's say that your central attention is focused on your conversation with a couple of friends in the middle of a crowded room. There is music and other conversations all around you, but to you that is just peripheral noise, which you are scarcely aware of monitoring. Suddenly, something stands out across the room: it could be a voice louder than the rest or, even without being louder, you recognize the voice of a friend you were expecting or whom you are surprised to see, or perhaps it is your own name you hear being mentioned amid all the noise. It captures your attention and you turn to look.

This distinction between central and peripheral attention, and the homeostatic principle on which the latter operates, are absolutely crucial for SI. In fact, what we call "monitoring" is simply the conscious use of your peripheral attention. If everything at the fringe remains normal, no action is necessary, but as soon as a discrepancy is detected, our focus needs to instantly shift to it in order to resolve it. Now, if we return to our SI tasks (listening-understanding-abstracting vs. formulating-delivering-monitoring), we see that the optimal stance is to keep our focus centered on formulating and briefly shuttling back to abstracting in order to grasp each new segment. All other functions are simply monitored "on automatic" (for homeostasis) by our peripheral attention, until a discrepancy is detected. Only then do we focus on these other functions for a quick troubleshooting before returning to the optimal center.

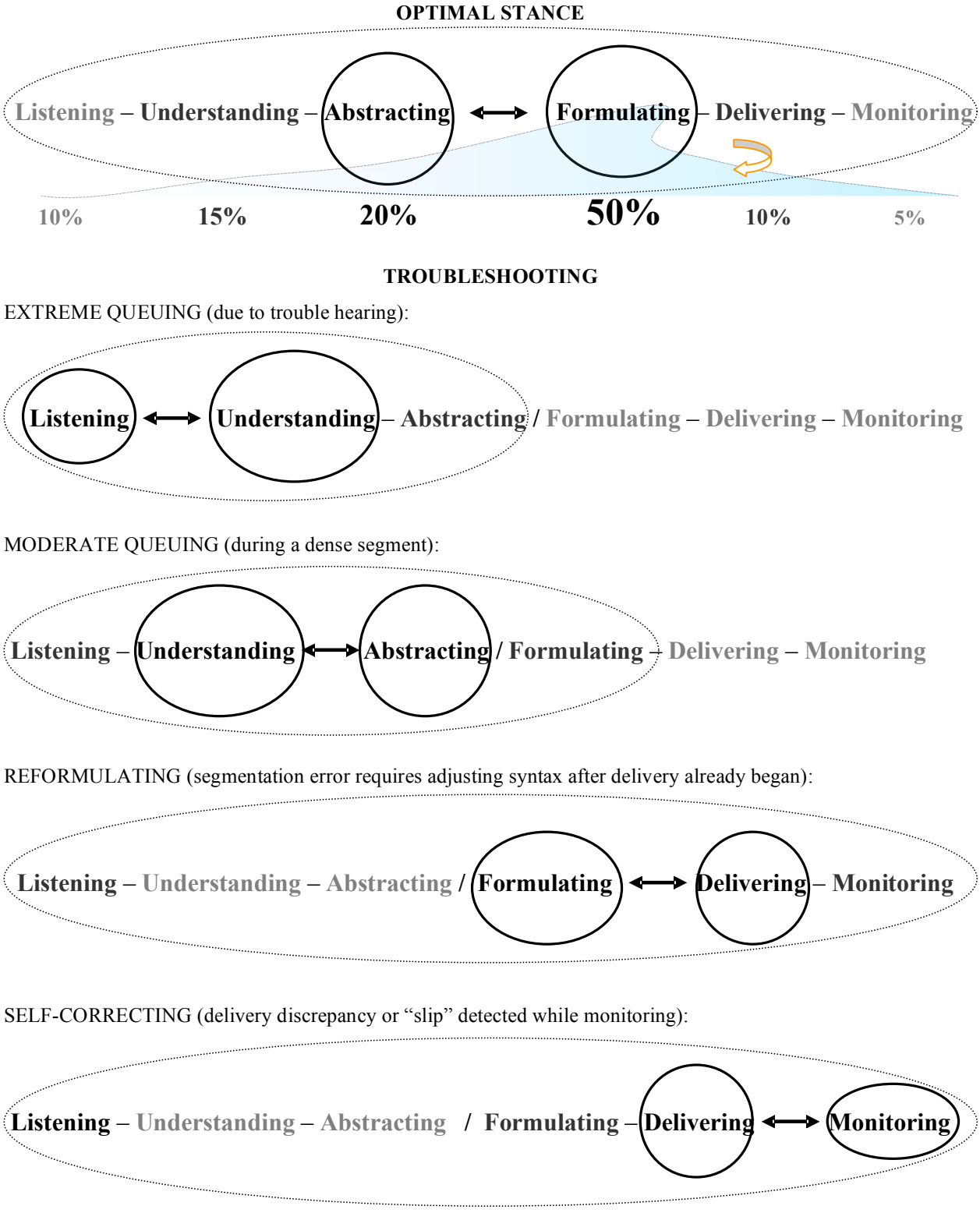

Fig. 6 Central and Peripheral Attention Shifting within SI Segments

Let us first look at each type of troubleshooting, before returning to our optimal stance:

During EXTREME QUEUING, you stop all speaking functions to concentrate 100% on listening and understanding, because you are having trouble hearing. It could be a simple audio problem (low volume or interference) or a momentary lapse of attention, which should be corrected immediately. If it is a recurring problem, it should be brought to the attention of the speaker, moderator, or sound technician. Consider the different listening problems that may arise and how to respond to them. In any case, whenever the listening function demands more than minimal or momentary attention, formulation (the most time-consuming function of all) is sacrificed, and the interpreter may fall behind and have to skip or gloss over one or more segments.

During MODERATE QUEUING, the interpreter can hear fine, but has difficulty understanding the message. Central attention is focused on understanding and then abstracting, with listening and formulating staying in the fringe or peripheral attention. Moderate queuing is more forgiving than extreme queuing, because formulating is not entirely ignored, and thus can be more quickly completed. Moderate queuing is in fact a common and necessary technique to deal with dense segments of the incoming message. "Dense" simply means difficult to understand. A portion of the message may be dense either because it is too fast, contains tightly condensed meaning, includes difficult terminology or complex syntax, or is not clearly articulated or pronounced. Often, the problem is merely momentary, but condensed meaning, difficult vocabulary, and complex syntax can be recurring and pervade an entire speech. In such challenging instances, in which attention time is consistently taken away from formulating, the interpreter may have to resort to frequent queuing, shortcuts, and occasional summarizing, so as not to fall behind, thereby simplifying the message while attempting not to compromise its substantive accuracy. Sustained queuing is exhausting, because our mental resources are thinly stretched between the stressful demands of both, understanding and formulating, as each function requires large amounts of concentration. It is like pulling a short blanket to cover your neck, only to expose your feet. Even proficient interpreters can become fatigued after only 15 minutes of this tug-of-war, and therefore should be timely relieved by a partner, request a few minutes to rest, or ask the speaker to slow down. Occasionally the problem with understanding may be persistent, such as a heavy accent or unclear articulation by the speaker. In such cases, the interpretation will suffer, and all the interpreter can do is his/her best. However, if the interpreter has trouble understanding a standard form of the source language, or has to devote more than minimal or momentary concentration to understanding, that interpreter may not be ready to interpret simultaneously from that language.

REFORMULATING is a corrective measure that must be undertaken when one begins to deliver a segment, only to realize that it cannot be completed grammatically by continuing the syntax sequence to which one was originally committed. In short, the syntax must be reformulated, either by stating the segment over (false start) or —a more elegant solution— by redirecting the syntax in mid sentence so as to make its ending work grammatically and meaningfully. In any case, this happens because of two types on interpreter error. The first is misunderstanding caused by mistaking, for instance, a noun for a verb, or by making an assumption of meaning that turns out to be incorrect (faulty prediction). The second and most common type of error is cutting a

segment too short. This may be due either to: a) excessive healing (following the speaker too closely without allowing him/her to finish the segment); b) the original segment being syntactically too long (hyperbaton); or c) marked differences in syntax between target and source languages (as is often the case between European and Asian languages). Preventing this requires queuing, because cutting a segment too short is in fact an error on the listening side or reception sequence, which results in formulation and delivery being started prematurely. Exercises in semantic prediction and syntax manipulation (discussed below) are useful in preventing the need for the reformulation, as well as making reformulation smoother and more effective when the need for it arises. The syntactical discrepancy that leads to reformulation is detected during *comparison* and *adjustment*, the intermediate steps that follow formulation and accompany delivery. Thus, during reformulation, central attention is shifted forward between formulating and delivering, while peripheral attention remains centered back on listening for the next incoming segment. Reformulation must be done quickly and, as soon as delivery is under way, central attention must instantly shift back to listening (so as not to lose the thread); then swiftly flow forward for another breaking wave. Like with any troubleshooting, the interpreter, after a quick correction, resumes coasting at the optimal attention stance.

SELF-CORRECTING is necessary when a discrepancy or "slip of the tongue" is detected during monitoring. Monitoring is a semi-conscious alertness to the automatic comparison between what we intended to say and what we actually heard ourselves say. It operates on the homeostatic principle under minimal peripheral attention. Yet, when a discrepancy between intention and word is detected, it stands out like an alarm. Slips are not uncommon in interpreting, given everything that is going through the interpreter's mind during such an intensive multitasking. Typically, a slip is a substitution of an intended word by an extraneous one that often sounds alike or is somewhat similar in some other manner, and appears to fit, thereby slipping through the vigilant comparison-adjustment pre-filter that should immediately precede delivery. It is always better to catch our mistake before we say it and avoid a self-correction, but inexperienced interpreters all too often relax this comparison vigilance. What is worse, they too often relax even monitoring itself, the final filter, and that causes mistakes go undetected and uncorrected. Although less frequently, this happens to experienced and vigilant interpreters too, because there is a point of vulnerability in the delivery process. When we have formulated and our delivery is under way, we shift central attention back to listening for the new segment, and leave the rest of the delivery on automatic. This is necessary in order to maximize the efficiency of our selective attention, but it leaves the rest of the delivery under peripheral attention only, and therefore vulnerable to substitutions and intrusions. Especially since at that very moment we are busy attending to the new incoming message, it is easy for one of the new terms to get formulated and substituted into our delivery of the previous segment. Interpreters should be aware of this tendency and vulnerability in order to guard against it. Yet, there are other potential sources of intrusion. Any thought that crosses our mind can potentially disrupt our concentration and interfere with our delivery. To begin with, interpreters should be thoughtfully following the topics and ideas of the speech, so they are actively engaged in thinking about the material they are interpreting. They must not let their minds wander, no matter how thought provoking the topic may be. Other sources of slips, however, come from extraneous thoughts that should be avoided. If the interpreter is preoccupied with something else, consciously or unconsciously, a proverbial "Freudian slip" may creep in. Whatever the cause, the self-correction must be quick and effective, ensuring an instant return to the optimal stance. Interpreters should practice quick

formulas for self-correction, such as "I mean…" or "Rather…" --TV news anchors are good models for this. Often, the self-correction can be rendered with a simple change in tone and rise in emphasis: "He was hurt… uh… He was not hurt in the accident." Unnecessary self-corrections, the product of insecurity, are to be avoided. Finally, do not let an error or self-correction itself become a distraction. Notice that in Figure 6 above no troubleshooting scenario ever leaves unattended the function of listening-for-meaning. It simply never leaves the corner of our eye. As Figure 2 also emphasizes, we must follow the thread of meaning and never let go.

Even under the best of conditions, some level of distraction is inevitable, but we must not allow it to disrupt our concentration or cause a lapse of attention, especially on the listening side. When there is a pause in the speech, you can relax, tend to your needs and distractions, and get ready to focus again when the speech resumes. Cognitive psychologists have documented the phenomenon of "attention fluctuation" –the fact that our attention fluctuates several times per minute. In Figure 4.10 below, concentrate on the center of the empty circle and try to ignore the X outside. Even if you do not look directly at it, count how many times you become aware of the X or your mental attention is momentarily drawn to it.

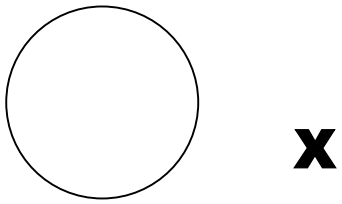

Fig. 7 Attention Fluctuation

With simple figures such as this you can practice increasing your attention span. You can also sharpen your peripheral attention: for example, look fixedly at the center of the X but concentrate mentally on the circle, or vice versa. Attention fluctuation is involuntary; so when we talk about "selective" attention, it is precisely to distinguish it from this involuntary kind. Thus, fluctuation of attention is normal and generally does not present a problem for the interpreter, as long as we maintain our focus and do not simply let our attention wander or linger on a distraction. Sources of distraction are all around us, and also within us. Some SI interpreters erroneously believe that closing their eyes will help them concentrate and avoid distractions. The fact is that visual contact with the speaker and the audience is essential for "tuning in" and "becoming" the speaker, and for purposefully directing our interpretation to the audience; in short, for keeping us focused on the job at hand. The fluctuation of attention caused by distractions in our visual field is harmless. In fact, it is a necessary mechanism that enables us to monitor our surroundings. Besides, some of the most troublesome distractions come from within: hunger, thirst, fatigue, and other discomforts. So interpreters should take care of such needs before their working session begins. Of course, the more conspicuous the distraction the harder it is to ignore. One of the most common and conspicuous is self-consciousness, worrying about how one is doing, which leads to self-doubts, and to mounting nerves. To combat this, interpreters should focus on the job and forget themselves, lose themselves in their work; ironically, attention fluctuation and the myriad distractions in the room actually help us direct attention away from ourselves.

Let us now return to our OPTIMAL STANCE to examine the role of memory in reinforcing SI selective attention management. There are two different forms of remembering: recognition and recall. Recognition is a rather passive, automatic, and nearly instantaneous function, which requires little concentration. Sounds are listened to and we recognize the words because they match sounds we have stored in memory; the meaning of those words is understood because it matches ideas in our memory; those ideas are put together and abstracted into schemas because they match schemas we have in memory, such as before-and-after, cause-and-effect, part-to-whole, and countless others. In fact, the entire reception sequence is based on effortless recognition, although each subsequent function, listening, understanding, and abstracting, is a bit more active and demanding than the previous one, and so they are assigned 10%, 15%, and 20% of our attention time, respectively, in our optimal stance. Recall, on the other hand, is a fully active, deliberate, and quite demanding function. When we produce or formulate language, we have to recall, that is, actively retrieve from memory the schemas we have received, the syntax to organize them, and the words to express them. This is why formulating and the beginning of delivery need to receive the bulk of our attention time for optimal SI efficiency.

There is, however, another aspect of memory that is very important for the listening function in SI. When we hear language, we need to pay attention to it (active listening) to be able to start processing the information. Through attention, we bring the message forth into our short-term memory (STM). Then, we relate personally to that message because it reminds us of previous knowledge, experiences, and schemas we retrieve from long-term memory (LTM), thereby enriching and reinforcing comprehension. But if we do not pay attention to the incoming sounds, they soon fade and are no longer available for processing. Yet, they do not fade immediately, but "linger in our ear" for a couple of seconds. This is because we have two preliminary forms of memory that operate even before we pay attention and start processing the message. First, we have a "sensory buffer" that works like this: a sound makes our eardrum vibrate and that vibration takes a bit of time before it fades; in addition, our auditory nerve continues to fire impulses to the brain for a short while, even when the original sound vibration is gone, and then those impulses also fade away. Second, we have what we might call an "acoustic memory": the brain registers the sound of an entire phrase or segment, and it remains available like an "echo" in our mind for a few seconds, even before we pay attention and start to process it. For the simultaneous interpreter, acoustic memory is an invaluable resource, because it gives us time to concentrate on formulating and starting to deliver the previous segment. Then, when we go back to listen to the new segment, its "echo" is still there, available to us. Thus, interpreters learn to tap into this resource, such that when we are in the optimal efficiency stance and we shift attention back to listening in order to grasp each new segment, we are no longer listening to the original voice but working off our acoustic memory of it. Without this, "simultaneous" interpreting would not be possible, because if we had to listen constantly to the speech itself, we could never switch attention away to concentrate on formulating.

What is the capacity of this acoustic memory? The answer is: only one segment. As the original speaker keeps talking, each new segment "overwrites" the previous one (which was already fading). So when our attention goes back to grasp it, it finds a strong new "echo" masking-over the fading previous one, such that it can no longer be accurately grasped. This means that when we are interpreting at optimal efficiency and fluency, we keep only one segment in our acoustic memory, that is, we are lagging behind the speaker by only one segment of *décalage*. When a

second segment begins to enter our acoustic memory before we have grasped the one already there, immediate queuing is required. In queuing we stop speaking in order to concentrate on listening. In so doing, we slightly sacrifice fluency of delivery and efficiency of attention, but this allows us to pick up the fading segment in acoustic memory and merge it with the new segment that we hear. Typically when we catch up or recover from queuing, we tend to heel or follow the speaker closely for a few seconds, until we feel secure and steady to ease up and slide back again to our optimal stance. This is because, on the listening side, both queuing and heeling work in a similar way: bypassing acoustic memory and listening to the original voice directly. It is easy, therefore, to transition from one to the other without any change in our listening stance. The only difference is that in queuing we hold back our delivery, while in heeling we let our speech go and follow so closely behind the original that we are virtually "stepping on the heels" of the speaker. Contrary to queuing and heeling, both of which work off the original voice, the optimal efficiency stance requires that we work off our acoustic memory (although even from there, like a driver on the rearview mirror, our attention takes quick "glimpses" at the incoming voice to "see" what is next and start tying-in the meaning of the new segment, such that it may flow, both semantically and grammatically, with our ongoing discourse).

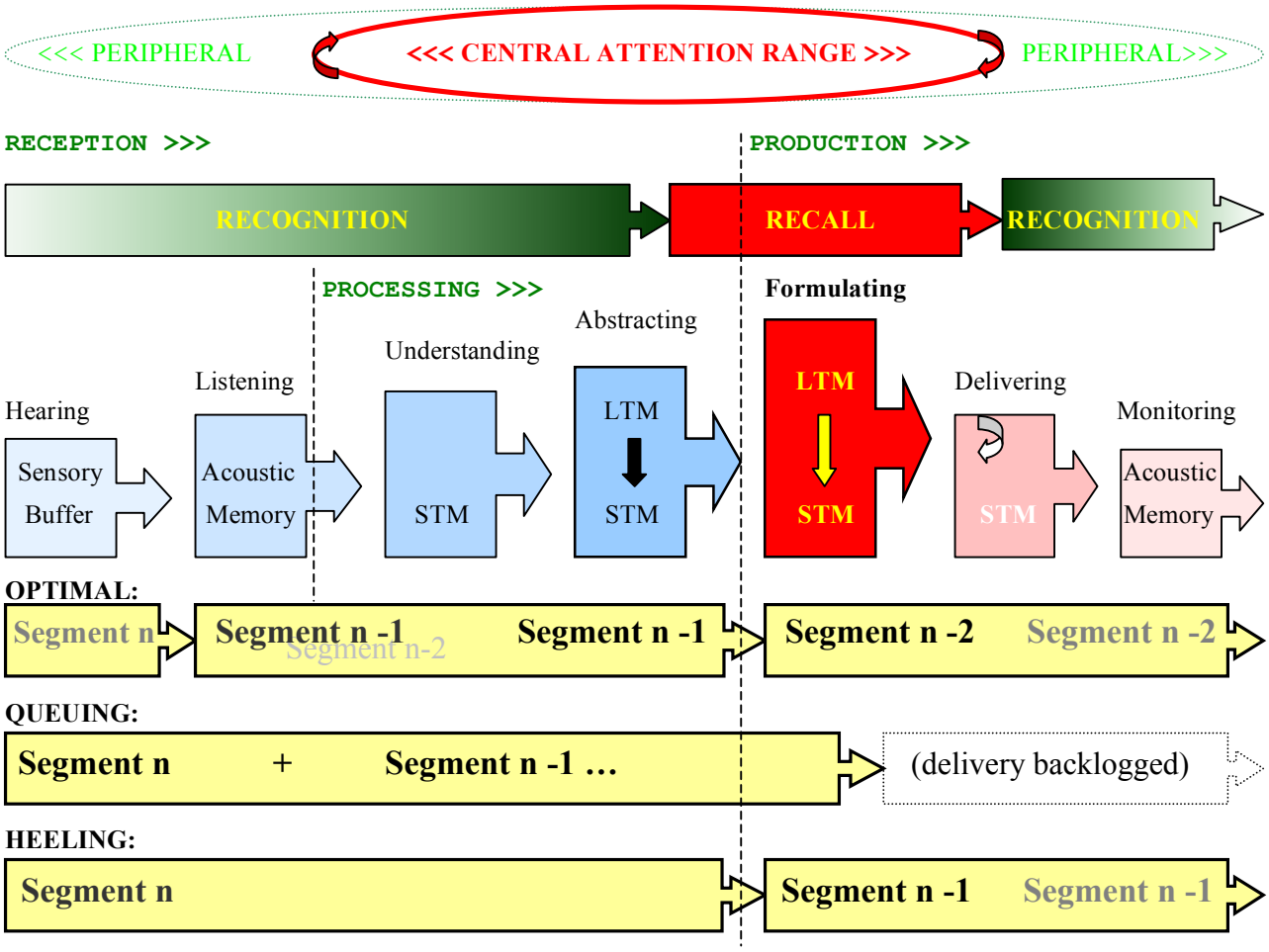

Fig. 8 The Flow of SI Segments through Memory and Attention

Figure 8 represents a still snapshot of the continuous listening-and-speaking process in simultaneous interpretation, a process that takes place at the speed of thought. Figure 8 illustrates the roles of memory and selective attention, as they relate to the flow of segments, where "n" is the new incoming segment, "n-1" is the previous segment flowing through acoustic memory in order to be processed, and "n-2" is the segment undergoing formulation and delivery.

Notice that in the OPTIMAL efficiency stance the interpreter is able to have three segments under the scope of attention: the one being delivered (n-2), the one being processed (n-1), and the new incoming segment (n). This is the result of efficient selective attention management, which makes for an easier, seamless flow of segments and a smooth, more fluent interpretation. Under acoustic memory, Segment n-1 has overwritten the fading n-2, which in turn has already been processed for formulation and delivery. As soon as delivery of n-2 is under way it is left in automatic mode, while central attention instantly shifts back to grasping and processing the next segment (n-1); meanwhile, the incoming segment (n) enters acoustic memory, becoming the new n-1 and displacing the old one (which by now has become the new n-2). In this manner, each new segment n becomes n-1 and then n-2 as it is processed forward and up, climbing the hill of increased activity and attention, until it gets over the hump of formulation (where the "wave" breaks), after which it is all easy coasting down the slope of delivery and monitoring. On the other hand, when QUEUING becomes necessary, acoustic memory is at least partially bypassed in order to grasp n directly off the voice and merge it or "chunk" it together with n-1, which is already in short-term memory (STM) and may itself be the product of a previous chunking. Through chunking, n and n-1 will be formulated together as one segment, resulting in a *shortcut*. As attention was focused almost exclusively on the reception side (with only some residual attention left over for formulation), all speaking has momentarily stopped, and delivery is becoming backlogged. The shortcut, then, helps the interpreter catch up. Finally in HEELING, acoustic memory is completely bypassed as segments are rapidly processed off the voice and quickly delivered, at the risk of cutting their meaning too short, and therefore engaging in choppy delivery, self-corrections, and/or false starts. Contrary to queuing, in which central attention shifts back almost exclusively to the reception side, heeling centers primary attention forward towards production and delivery, resulting in an often hurried tone.

3. CONTRASTIVE SYNTAX AND PREDICTION IN SI

3.1 Syntactic manipulation

It goes without saying that command of specialized terminology and a broad general vocabulary in both languages are essential in SI for efficient understanding and formulation. Building a personalized glossary is a good way to attain this command. However, many interpreters (and trainers) tend to focus exclusively on terminology, and fail to realize that the syntactical differences between languages are just as important in SI for effective *segmentation*. In fact, every interpreter's glossary should contain a section on contrastive syntax, with phrases whose construction or word order varies markedly from one language to the other. Contrastive syntax is the comparison of the rhetorical patterns typical to each language, and their equivalencies. For

example, the simple English phrase [“Upon arriving,”] is rendered very differently into Spanish using a different preposition + infinitive: [“Al llegar,”]. The phrase [“Being that it costs too much,”] may have several equivalents: [“Siendo que… Dado que… Puesto que… Debido a que… Ya que… + cuesta demasiado,”]. Usually it is best to choose the simplest and shortest equivalent: the English [“In order to…”] may be rendered into Spanish as [“Con el propósito de…”], [“A fin de…”], or simply [“Para…”]. These and other phrases and equivalencies may belong in the contrastive syntax section of an interpreter’s personalized glossary.

Another aspect of contrastive syntax involves sentence construction. Syntax is more or less flexible in every language. Word order and the order of clauses may be rearranged grammatically in various ways without altering meaning. In general during SI, it is best to follow the original sentence structure as closely as possible. This way the interpreter does not have to spend the extra time and concentration that complex rearrangement of syntax requires. However, this mirroring of the source language syntax may necessitate employing a target language syntax that, although correct, may be less common. For instance, the end of the previous sentence [“…that complex rearrangement of syntax requires.”] is more naturally rendered into idiomatic Spanish with the verb upfront [“…que **requiere** la reorganización compleja de la sintaxis.”], but it is also correct, though less common, to place the verb at the end and mirror the English syntax: [“…que la compleja reorganización de la sintaxis **requiere**.”]. It is always easier and faster to mirror the original syntax. An interpreter may also have to rearrange the syntax in mid sentence after beginning delivery, in order to make its conclusion work grammatically. For all of these reasons, interpreters should gain command and flexibility of syntax by doing syntactical manipulation exercises, that is, rearranging the syntax of sentences in every possible way.

There are yet other syntactical structures that are peculiar to one language, and thus are not easily mirrored in the target language. Some peculiar English structures to watch out for are the following:

Multiple Adjectives – In English, contrary to many other languages, adjectives are usually placed before the noun. Multiple adjectives will then present a challenge in SI. For example: “He is an honest, hard-working, and law-abiding **citizen**.” Until you hear the noun “citizen” you cannot close the segment and begin to interpret it, but by then you may have fallen behind in your SI. One possible solution is to mirror the syntax and adapt the conclusion resourcefully: “El es honrado, trabajador y respetuoso de la ley **como** ciudadano” [“He is honest, hard-working, and law-abiding **as a** citizen”].

Multiple Adverbs – Multiple adverbs present a similar challenge: “He wantonly, willfully, and maliciously **drove** into the crowd.” Until you get to the verb “drove” you cannot close the segment, and thus you risk falling behind. Again, one solution is to mirror the syntax, begin interpreting the segment before even knowing what the verb is going to be, and then adapt at the end to make it work grammatically: “**De manera** temeraria, voluntaria y maliciosa, condujo contra la multitud” [“In a wanton, willful, and malicious **manner**, he drove against the crown”].

Prepositions at the End – “This is the possibility I suggest we should worry **about**.” Again, you can wait for the verb + preposition [“worry about”], and risk falling behind, or you can begin your interpretation and adapt the syntax as you go: “Esta es la posibilidad **que** sugiero **nos debe**

preocupar" ["This is the possibility I suggest **should concern us**"]. The first part of the sentence ["This is the possibility…"] remains unchanged, and thus we can consider it a stand-alone segment and begin to interpreter it. We can in fact anticipate or predict an anomaly (in this case the preposition at the end) by paying attention to syntactical cues, like the fact that English often omits "that" ["This is the possibility (**that**) I suggest we should worry about"]. When we mirror the syntax we would supply these implicit connectors ["Esta es la posibilidad **que** sugiero…"]. Mirroring the SL syntax allows us to make our segments shorter and move on with our interpretation, gaining valuable time to adapt the predicate syntax without sacrificing fluency, grammar, or meaning ["This is the possibility | that I suggest | **should concern us**"]. As part of their syntax manipulation exercises, interpreters should practice anticipating and eliminating postponed prepositions.

Clauses of Purpose – "I need **for you to consider** this very carefully." Romance languages typically render such clauses in the subjunctive: "Necesito **que consideren** esto con mucho cuidado" ["I need **that you consider** this very carefully"]. Since the anomaly is not postponed, but is signaled early on by the construction "for you to," it should not cause any major delay in formulation, provided that the interpreter has the correct TL equivalent handy. This is yet another item for the contrastive syntax section of the glossary.

There are other peculiarities of English, such as the variable uses of the verb "would" or the gerund "-ing" ending. Every language, in fact, has its syntactical peculiarities, such as the ability in Spanish to begin a sentence with the verb: "**Inciden** en el costo del producto importado la materia prima, la fabricación, los fletes, los arbitrios, la gestión de mercadeo y la tasa cambiaria." Mirroring such syntax into English requires creativity and may result in a somewhat awkward, albeit correct and meaningful construction: "**Affecting** the cost of imports are the raw materials, manufacturing, freight, duty, marketing, and exchange rate."

3.2 Prediction

Prediction is the ability to anticipate the conclusion of a segment, as well as the general direction of the speech. It is a valuable technique that allows us to switch attention sooner from listening to formulating and delivering. There is of course a potential for error if the technique is not carried out properly. Prediction does not mean guessing. Rather, the way it works is that as soon as we anticipate the conclusion of a segment, we switch central attention to formulating, but continue to monitor the listening side with our peripheral attention under the homeostatic principle. If our prediction is verified, we proceed with our delivery, but if there is a discrepancy with our assumptions, our homeostatic alarm goes off and we adjust our formulation. In this way, there is no guessing and the potential for error is virtually eliminated.

Syntactic prediction

There are three main types of prediction in SI: syntactic, semantic, and pragmantic. *Syntactic prediction* is the ability to anticipate how a phrase or sentence will end. In turn, there are three main types of syntactic prediction: *formulaic*, *reiterative*, and *grammatical*. If, for example, an attorney says "Ladies…" during a closing argument, the interpreter goes straight to delivery,

expecting to hear a homeostatic verification of the formula: "…and gentlemen of the jury." If we hear "as a matter…" we automatically anticipate "…of fact."

Reiterations or repetitions also create an anticipation to be verified. If, for example, during a campaign speech a politician repeats a slogan or catchy phrase such as "…and that is not what our citizens deserve," by the third time it is enough for the interpreter to hear "…and that is not…" in order to proceed with delivery. All speech contains redundancies, reiterations, and repetitions, which create reasonable anticipations to be verified. If a speaker says, "Commercial tuna fishing employs large nets. The problem with this manner…," the interpreter can anticipate the reiteration: "…of catching tuna."

Grammatical prediction works by anticipating certain parts of speech or syntactical structures, and it can be quite accurate and specific. In the previous example, we can confidently extend the reiterative prediction grammatically: "The problem with this manner …of catching tuna …**is that**…" Grammatical prediction, however, can be general rather than specific. For instance, if we hear a subject and a transitive verb, we can expect a direct object: "He bought _____." Even if we do not yet know specifically what he bought, such partial predictions are still useful, because they allow us to begin our formulation.

Semantic prediction

Semantic prediction is the anticipation of meaning. When it is based on the immediate context, it can work together with syntactic anticipation to help determine the outcome of segments. Suppose that in the previous example the speech has to do with an automobile aficionado; the context may then lead you to expect the following development: "He bought *his first car*..." You may even anticipate and listen for the year, make and model.

Yet, semantic anticipation may reach well beyond the sentence, to the general intent and direction of the entire speech. In this case it is based on knowledge of the world, of language and culture, of the subject matter, and of the ongoing discussion. In the earlier example about commercial fishing, you might expect, from your knowledge of current issues, that the speaker will turn to the problem of dolphin getting caught on commercial tuna fishing nets and what ought to be done about it. The more we know about the subject matter, and the less new information we find in the message, the faster we can process it, and the smoother and more accurate the interpretation. Even general information, such as whether the speaker is for or against a particular issue, will help us make better sense of the argument and avoid misunderstanding. The argument will have a logical sequence that organizes the speech and helps us anticipate subsequent steps. This logical sequence is a sort of pragmatic "syntax" that operates beyond the sentence to organize the exposition of ideas throughout the entire speech.

Pragmatic prediction

When we combine syntactic and semantic prediction we have a pragmatic or rhetorical anticipation of the different styles of sentences, paragraphs, and speeches. We thus recognize and anticipate the expected patterns of the different expository modes: **narration**, **description**, **argumentation**, **comparison**, **contrast**, **exemplification**, **generalization**, **classification**,

analysis, etc. All forms of prediction, however, depend on the interpreter's ability and discipline to constantly follow the unfolding meaning of the speech. Whether a general prediction turns out to be fulfilled or not, is secondary. By simply following the flow of ideas and continuing to anticipate what is next, we remain engaged in thoughtful processing, immersed in the subject matter, and in tune with the speaker's point of view. That is correct interpreting.

4. INCREMENTAL TRAINING METHODS

4.1 Diagnostics

The first step in incremental training is to conduct some diagnostics. For this purpose, it is important that you record your practice. Now that you know the relevant criteria, you can evaluate yourself. Self-diagnosis through introspection is very important because only you know what goes on inside your mind while you are interpreting.

Measure your SI *performance base*, that is, your current level of proficiency in terms of *accuracy*, *speed*, and *endurance*. For this purpose, you may use recorded selections rated in words per minute, such as those available from the National Court Reporters Association. Then, practice incrementally, pushing yourself beyond your performance base to higher levels of speed and endurance. Record and listen to your own performance, and measure your accuracy. Note that your level of accuracy will improve over time with consistent practice and evaluation. Keep track of your progress.

Review the tasks and techniques discussed above, and determine your weaknesses, sources of inefficiency, and the differences between the way you have been interpreting and the procedures and techniques recommended here. You may have developed habits that lead to inefficient and diminished performance, or simply it may be time to try new methods.

Needless to say, diagnostics and troubleshooting are not just for beginners: they constitute for the professional interpreter an on-going activity that is fully necessary for maintaining skills. As in other practical arts, when the ability to self-criticize is undeveloped, interpreting performance is limited; when it is abandoned, proficiency declines. Unfortunately, many interpreters never develop adequate self-diagnostic skills because of lack of formal training. Others in time become self-satisfied, rigid, and impervious to criticism: their performance declines without their acknowledging it.

An easy and effective way to self-diagnose is to try the different exercises in this section. Because these exercises address specific SI tasks and skills, you can readily see which of them need improvement. Then it is up to you to address them through concentrated practice. Even if you are well advanced, determine which exercises you find most challenging, and then isolate and practice the relevant skills until they become virtually effortless and second-nature. As you develop those isolated skills, do not forget to integrate them into a seamless performance, aiming for mastery of process. Regardless your level, you can always measure your progress against your performance base.

4.1 Exercises

The following exercises are presented in developmental order, focusing on increasing integration of skills. When applicable, practice should progress from intra-lingual to inter-lingual exercises, and from written to oral form.

Focusing on meaning, not words (paraphrasing) – Read (later listen to) sentences and paragraphs; practice restating the meaning in your own words, first in writing and then orally. Avoid using the original words and syntax. Do this first within each language, and then translating bi-directionally. Practice restating the meaning in various different ways. Finally, paraphrase entire speeches orally in both languages.

Restating schemas, translating terms - Read (later listen to) sentences and paragraphs; underline or identify only the terms. Then, in your own words, orally reconstruct the relationship between those terms, according to the original meaning expressed in the passage. Do this first within each language and then translating bi-directionally.

Segmentation – Take written texts in both languages and draw slashes with a pencil in order to segment the texts strategically the way you would during SI. Remember to use longer segments (*queuing*) for the denser portions of the text. After practicing with texts, listen to recordings in both languages, and segment mentally, according to the speed, density, and syntax of the incoming message. Do not interpret yet. For now, just listen and segment strategically.

Interpret silently – Practice interpreting silently in your mind, without speaking. That is, carry out the process up to the formulation phase only, without delivery. You can do this anytime with TV programs or other aural input. Make sure you practice in all your working languages. Stop to note down the spots that gave you trouble, so that later you can analyze the cause and devise a remedial strategy. Start at your performance base and push for speed.

Interpret for correct form – Practice interpreting for correct form, without worrying, for the time being, about correctness of content. Delivery should be smooth and even-paced, emulating the quality of broadcast voice. Maintain a moderate but flexible *décalage*, and avoid interpreting in jerky spurts, excessive stop-and-go, fast consecutive, or hurried slurs. Voice should be clear and expressive, with meaningful yet natural intonation. "Become" the speaker and envision your audience: direct your communication purposefully to your listeners. Formulate and deliver segment by segment fluidly and confidently, avoiding hesitation, fillers, and insecure volume or tone. Start at your performance base and push for speed.

Prediction – Do phrase completion exercises, trying to predict the ending. Read newspaper articles and other texts, and as you read try to anticipate how aspects of the argument or story will develop and what the final outcome will be. As you segment a text, notice how contextual knowledge of the subject increases cumulatively as you read on. This helps you predict better because there are fewer ambiguities as the passage progresses, which in turn enables you to recognize shorter segments and therefore increase your interpreting efficiency and speed.

Syntactic and semantic adaptation – Do syntax manipulation exercises in both languages, by changing the word order of sentences in every possible way, grammatically, without altering meaning. Watch for ambiguities that may be created as a result of changing the word order. First do these exercises in writing, and then orally. Aim for maximum flexibility and resourcefulness in syntax and sentence construction. Practice anticipating and eliminating postponed prepositions, by using alternate constructions. Also, do exercises completing sentence translations, and practice changing direction and adjusting the syntax in mid sentence, in order to adapt to unexpected developments in the incoming message, without need for self-corrections and false starts. This is important particularly during heeling. Finally, do exercises in extrapolating or bridging meaning from context and "filling-in the blanks." Read and listen to excerpts with missing or unintelligible portions, and supply or gloss over the meaning of the unclear parts. Finish sentences with unclear ending by logically extending the ongoing meaning.

Shortcuts – Following the model in Table 1, practice distinguishing between substantive and accessory meaning in different utterances, eliminating redundancies, and expressing the ideas accurately but in increasing levels of abbreviation, down to their simplest and most economical form. Determine to what extent ideas may be acceptably abbreviated without sacrificing substantive meaning. Remember that shortcuts are to be used only to the extent necessary to keep apace and thereby safeguard the integrity and completeness of the substantive message. Practice shortcuts with written materials in both languages, first in writing and then orally, first within each language and then translating. Then move on from written input to listening, and practice shortcuts by paraphrasing within the same language. Finally, practice interpreting with queuing and shortcuts. The added concentration needed to eliminate redundancies can be fatiguing; so practice it intensively for increasing periods in order to develop endurance.

Managing *décalage* – Practice simultaneous *shadowing* in both languages; that is, repeating what you hear, within each language, without translating. Start by closely following the original (heeling); then slowly increase the *décalage*. Shadowing in your weaker language at high speed, while heeling, is an excellent way to develop fluency and improve your pronunciation. Next, do shadowing while queuing: let the original speaker get one segment ahead of you; then, when the speaker begins to utter the second segment, you begin to shadow the first, and so on. Following, let the speaker get two segments ahead, and then three segments, while you shadow with extreme queuing and shortcuts. Finally, do actual SI in both languages, while minimizing and maximizing the *décalage*; that is, practice extreme heeling (using syntactic and semantic adaptation in mid segment in order to prevent self-corrections and false starts), and then practice extreme queuing (using shortcuts in order not to fall behind and lose the thread of meaning). Practice both heeling and queuing for extended periods in order to develop endurance. Then alternate frequently and at will between heeling and queuing, during the same SI session, in order to develop maximum control. Start at your performance base and push for speed.

Interpreting non-stop - Practice never stopping until the end. If you ever fall behind while queuing, skip and pick up the thread of meaning again on the next segment, without stopping. Try to bridge or gloss over the meaning without leaving any sentences unfinished, so that your listener does not notice that a segment was ever skipped. Continue interpreting at all costs, until you finish the entire session or passage. Start at your performance base and push for speed.

Selective Attention Management (multitasking) – While shadowing with *décalage*, as explained above, write down numbers from 1 to 100 on a piece of paper; then backwards from 100 to 1; then in multiples of 3; then write down the national anthem or another known poem or piece of prose; then an original letter. After each exercise, tell back the meaning of the story or piece you just shadowed. Having one input channel (for shadowing) and two output channels (for speaking and writing) forces you to manage your selective attention and get used to multitasking, in a way that approaches SI. Since you are only shadowing and not translating, the function of selective attention is isolated for the purpose of concentrated practice. Next, try two input channels and one output channel: listen to and shadow one selection, while you read an unrelated text at the same time. For example, you can shadow a tape while you read a newspaper article. Keep track of each piece separately, without mixing their meanings. At the end of the exercise, report back the story or argument behind each piece.

Avoiding translation short circuits – Practice written translation. Review Figure 1 and check all your written translation work for short circuits, such as false cognates and borrowings of terminology and syntax. Have your work checked for correctness of content by an instructor or peer. Do the same with all your recorded interpretation practice.

Comparison, monitoring, and adjustment – Read (and later listen to) sentences and paragraphs with their corresponding translation. Practice detecting translation mistakes or inaccuracies and quickly supplying a correct or improved version.

Endurance training – Place yourself in a state of relaxed concentration, forget technique and everything you have learned, clear your mind of all your troubles, and just interpret. Keep your relaxed focus and do not think about yourself and your doubts, limitations, or needs. Visualize yourself as a tireless interpreting machine that is incapable of error, and focus on the immediate interpreting task. Maintain this relaxed focus while you interpret non-stop for longer and longer periods, until you reach 45 minutes of effective and uninterrupted SI.

REFERENCES

Anderson, J. R. *Cognitive Psychology and Its Implications*. 6th Ed. New York: Worth Publishers, 2004.

Bock, K. "Toward a Cognitive Psychology of Syntax: Information Processing Contributions to Sentence Formulation." *Psychological Review* 89: 1 (1982) 1-47.

Cowan, N. "Processing Limits of Selective Attention and Working Memory: Potential Implications for Interpreting." *Interpreting* 5:2 (2000/01) 117–146.

Danks, J. et al. (Eds.) *Cognitive Processes in Translation and Interpreting.* Thousand Oaks, CA: Sage, 1997.

Gambier, Y., Gile, D. & Taylor, C. (Eds.) *Conference Interpreting: Current Trends in Research*. Amsterdam / Philadelphia: John Benjamins, 1997.

Gerver, D. & Sinaiko, H. (Eds.) *Language, Interpretation, and Communication*. New York: Plenum Press, 1978.

Gile, D. *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam / Philadelphia: John Benjamins, 1995.

Gran, L. & Dodds, J. (Eds.) *The Theoretical and Practical Aspects of Teaching Interpretation*. Udine: Campanotto, 1989.

Jesse, A. et al. "The Processing of Information from Multiple Sources in Simultaneous Interpreting." *Interpreting* 5:2 (2000/01) 95-115.

Lambert, S. "Shared Attention during Sight Translation, Sight Interpretation, and Simultaneous Interpretation." *Meta* 49:2 (2004) 294-306.

Lambert, S. & Moser-Mercer, B. (Eds.) *Bridging The Gap: Empirical Research in Simultaneous Interpretation*. Amsterdam / Philadelphia: John Benjamins, 1994.

Liu, M. "How Do Experts Interpret? Implications from Research in Interpreting Studies and Cognitive Science." In Gile, D. et al. (eds.) *Efforts and Models in Interpreting and Translation Research* (Amsterdam / Philadelphia: John Benjamins, 2008) 159-178.

Mizuno, A. "Process Model for Simultaneous Interpreting and Working Memory." *Meta* 50:2 (2005) 739-752.

Moser-Mercer, B. "Simultaneous Interpreting: Cognitive Potential and Limitations." *Interpreting* 5:2 (2000/01) 83-94.

Seleskovitch, D. & Lederer, M. *Pédagogie raisonnée de l'interprétation.* Traductologie 4. Paris: Didier Erudition, 1989.

Shreve, G. & Angelone, E. (Eds.) *Translation and Cognition: Recent Developments*. ATA Scholarly Monograph Series XV. Amsterdam / Philadelphia: John Benjamins, 2010.

Tirkkonen-Condit, S. & Jääskeläinen, R. (Eds.) *Tapping and Mapping the Processes of Translation and Interpreting*. Amsterdam / Philadelphia: John Benjamins, 2000.

Wei, L. "Positive Transfer: A Neuropsychological Understanding of Interpreting and the Implications for Interpreter Training." *Translation Journal* 6:3 (2002).