Searching to define expertise in interpreting

Moser-Mercer, Barbara; Frauenfelder, Ulrich Hans; Casado, Beatriz; Künzli, Alexander

How to cite

MOSER-MERCER, Barbara et al. Searching to define expertise in interpreting. In: Language processing and simultaneous interpreting: Interdisciplinary perspectives. Englund Dimitrova, B. & Hyltenstam, K. (Ed.). [s.l.] : John Benjamins Publishing Company, 2000. p. 107–131.

This publication URL: https://archive-ouverte.unige.ch/unige:83933

Searching to define expertise in interpreting

Barbara Moser-Mercer
Ecole de Traduction et d'Interprétation
Uli H. Frauenfelder
Faculté de psychologie et des sciences de l'éducation
Beatriz Casado
Alexander Künzli
Ecole de Traduction et d'Interprétation
Université de Genève

Introduction

Developing expertise in interpreting requires the integration of a large number of subs-skills and sub-processes of language processing, of which most occur more or less simultaneously: processes of language comprehension - auditory recognition, syntactic and semantic processing - (Marslen-Wilson & Tyler, 1980), activating two language systems at a time - source and target language systems - (Paradis, 1993; Grosjean, 1996), and language production - syntactic and semantic processing, grammatical and phonological encoding - (Levelt, 1989; Butterworth, 1980). In addition, interpreters need to engage in a certain amount of anticipation with regard to the incoming message, as they can hardly ever wait to process an entire sentence before starting with their interpretation, which in turn requires the interpreter to monitor his performance for semantic and grammatical correctness. Taken together these subskills make up a complex cognitive skill, for which - so it seems - not everyone appears to have the requisite aptitude.

This paper covers a series of experiments designed to isolate some of the aptitudes that seem to set expert interpreters apart from novices and as such might provide some preliminary guidance in the selection of interpreting candidates. When attempting to describe the reasons why certain individuals succeed at difficult tasks, whereas others fail, we are automatically led to studying the differences between these two groups in order to identify variables that might guide us in our search for the best aptitude test. We will thus first look at expertise research and how it contributes to the framework of this study, before discussing aptitude and proficiency and whether differences between expert interpreters and novices reflect differences in processing. This will be followed by a summary of each of the experiments carried out as part of the project and a discussion of the findings and their relevance to the ongoing discussion on aptitude testing in interpreting.

Expert-novice paradigm

Expertise research is concerned with understanding, in some detail, the knowledge, reasoning, and skills of experts in a variety of domains, with developing methodological tools to elicit such knowledge, and with describing similarities among experts in different domains. An expert is generally considered to be someone who has attained a high level of performance in a given domain as a result of years of experience. A novice is usually defined as someone who has little or no experience in a particular domain. Both categories allow for some degree of variation and Klein and Hoffman (1993:206) have argued for finer distinctions along the continuum from novice to expert. In translation and interpreting it is often acknowledged that the student obtaining his final diploma can call himself an expert with some degree of justification, but that years of experience in the field are still required for him to become a full-fledged professional.

Differences between experts and novices

The discussion that follows shows that differences between experts and novices have been found to relate to chunking of information, to reasoning, to speed of processing, to individuals' knowledge base and its organization.

Studies in expertise go back to the classic research of De Groot (1965, 1966) and Chase and Simon (1973) who were interested in discovering differences in the skills of chess players. De Groot (1965) concluded that while there were no obvious differences in the breadth or depth of search of possible moves on the board, the experts had a different representation of the problems they were setting out to solve. De Groot (1966) and Chase and Simon (1973) concluded that experienced chess players perceive relations between chess pieces as "chunks" and are thus able to remember a larger number of pieces, i.e. a larger number of valid board configurations. Anderson (1995) states that experts can recognize chunks in problems which are patterns of elements that repeat across problems.

Other pioneering work was concerned with experts' and novices' approaches to solving physics problems and was carried out by Simon and Simon (1978) and Larkin, McDermott, Simon, and Simon (1980a, 1980b). According to their studies experts and novices differed in several ways: Experts were able to reason forward and use the given facts to generate the equations they used; they solved the problems more quickly, and apparently generated some physical representation of the problem.

It appears that from early work in the field of expertise research there is evidence that expert-novice differences exist with respect to reasoning, knowledge structures, and the ways in which information is processed (Anderson, 1995). Künzli and Moser-Mercer (1995) concluded in light of their empirical research on the translation process and on sight translation (Moser-Mercer, in press) that expert translators' and interpreters' knowledge base appears to be somewhat differently organized (for further corroboration see also Foley and Hart, 1992).

With respect to factual knowledge expert translators and interpreters display better organization with more associative connections and more domain connections. With respect to expert translators and interpreters' semantic knowledge one sees experts' semantic interpretation almost always being tied to the context of a speech or a text, whereas novices' semantic interpretations are often entirely unrelated to the context. If we analyze experts' schematic knowledge we can see that they have built up schemata for different types of speeches, negotiating situations, texts, paragraphs, etc. Into which they embed individual utterances/sentences. Novices, however, tend to treat each utterance/sentence in a more isolated manner and fail to establish discourse links. At the level of strategic knowledge experts tend to proceed from known to unknown information, Novice translators and interpreters more often focus on the unknown and then easily get stuck. Experts thus use more global plans, whereas novices tend to favor low-level - microcontextual - plans (see also Mayer, 1992).

The acquisition of skill

While the above relates to differences between established experts and novices, it is also useful to shed some light on just how novices become experts and acquire the specialized skills needed to perform well on task.

Anderson (1995) describes three stages of skill acquisition: the cognitive stage, the associative stage, and the autonomous stage. During the first stage, novices develop a declarative encoding of the skill, i.e. they commit to memory a set of facts that are relevant to the skill. During the associative stage, novices gradually detect errors and eliminate them and strengthen the connections among the various elements required for successful performance of a skill. During the autonomous stage, novices' procedures become more and more automatic (Hoffman, 1996). Once the novice has converted verbal or declarative knowledge into procedural knowledge, the learning of a skill nears completion. In translation and interpreting novices still need to engage in tactical learning whereby they learn specific rules for solving specific problems, such as how to convert particular syntactic constructions in the incoming message to matching constructions in the outgoing language (Moser-Mercer, in press). This tactical knowledge then becomes increasingly well organized and the novice develops a set of strategies designed to optimally solve the problems he encounters. As novices become more and more expert in a domain, they improve their ability to store problem information in long-term memory and to retrieve it (Anderson, 1995).

Thus, learning to interpret cannot merely be equated to automating the largest possible number of underlying processes. One hypothesis would be that shifting from consciously controlled to more automatic processing entails a change in processing efficiency which would explain the resulting increase in processing speed, without necessarily changing the underlying process itself. Another hypothesis would be that this shift results in a more or less dramatic restructuring of the process itself.

McLaughlin (1995) points out there is a constant modification of organizational structures. This process of restructuring is necessary as the novice begins to shift from consciously controlled to more automatic processes. In second language learning research gains in automaticity are thought to be more characteristic of early stages of learning another language, whereas restructuring is seen to occur more in later stages (McLaughlin, 1995). There may be an interesting parallel in the acquisition of the interpreting skill where initial gains might be attributable to increased automation of lower-level processes, whereas the development of true expertise at a higher level requires in addition the restructuring of higher-level processing. In language learning restructuring appears to provide an explanation for U-shape developmental functions: performance declines as more complex internal representations replace less complex ones, and increases again as skill becomes expertise (McLaughlin, 1995). Such examples abound in interpreting pedagogy, where candidates thought to possess sufficient aptitude for interpreting begin to plateau, without being able to reach the required level in the prescribed time. Since the leveling-off occurs at a fairly advanced stage, there is usually not enough time for the student to recapture a level of performance already achieved and to improve on it. Here is were issues of aptitude become inextricably tied to issues of developing expertise, and are difficult to tease apart.

Proficiency and aptitude

The training of interpreters is concerned with developing proficiency in individuals who have been selected for a particular training course based on their aptitude for the skill. Thus proficiency relates closely to the kind of expertise described in the section on differences between experts and novices, whereas aptitude relates to being suitable for a specific purpose, having a natural tendency or capacity to pursue the study of a particular skill.

Aptitude testing in interpreting is thus concerned with looking for parameters that set the capable individual apart from the rest. In much of the research on aptitude testing in interpreting (Gerver, Longley, Long and Lambert, 1989, Moser-Mercer, 1984;1994, Lambert, 1991, Longley, 1989, Gingiani, 1990) a number of parameters are suggested as important for success of a trainee interpreter: High proficiency in the active and passive languages and cultures, ability to grasp rapidly and convey the essential meaning of the original, ability to project information with confidence, good voice, wide general knowledge, speed of acquisition of new information, and the ability to work as a member of a team (Lambert, 1991).

In one longitudinal study Moser-Mercer (1984) was able to show that students' performed significantly better on a number of tests designed to measure some of the above parameters, and that this correlated well with performance on course final examinations.

Another study (Gerver, Longley, Long and Lambert, 1989) looked more closely at candidates' ability to grasp rapidly, and to convey, the meaning of spoken discourse. This study used the following tests: 1. Memory for text (based on Kintsch, 1974); 2. Logical memory test, chosen from the Wechsler Memory Scale (Wechsler, 1945); 3. Cloze test (based on Stubbs and Tucker, 1974); 4. Error detection test; 5. Factor-referenced cognitive tests, chosen and adapted from the group of factor-referenced tests developed by the Educational Testing Service (Eckstrom et al., 1976), among them tests for associational fluency, expressional fluency, and verbal comprehension; 6. Nufferno Speed test (Furneaux, 1956) to measure the effect of speed stress on performance of an intellectual task. Means and standard deviations of scores on these tests, together with means and standard deviations of the ratings of jury members in the final consecutive and simultaneous examinations were compared. Those who passed the course had higher scores on all of the above tests than those who failed, and these differences were significant for memory for text, logical memory, cloze and error detection, and synonyms. Some tests were more highly correlated with results in consecutive (memory for text, logical memory) than with those in simultaneous interpreting, which correlated better with the tests of error detection, cloze and synonyms, all requiring speed and accuracy of both perceptual and productive linguistic processes.

Looking at these results and comparing them to the findings of Anderson (1982, 1983), and Newell and Rosenbloom (1981), according to which performance of cognitive skills improves approximately as a power function of practice, we are left with the task of teasing apart the acquisition effect (learning effect) from the cognitive ability effect (by studying cognitive correlates or cognitive components). Already Carroll (1981) argued that the tasks contained in aptitude tests are similar to the processes described in information-processing accounts of cognitive functioning and he speculated that variability in students' aptitude for language learning may reflect differences in their ability to store and retrieve information from short-term memory.

While we argued above that automation is a necessary stage on the road to becoming an expert, as automated processes make less use of precious working memory capacity, Anderson (1983) suggested that automation increases the size of subtasks which in turn increase memory load. McLaughlin (1995) takes this reasoning one step further to conclude that increased automation may at some point be responsible for U-shaped functions of learning: Integrating large subtasks makes heavy demands on working memory and has an adverse effect on subsequent performance.

This, for example, underlies our reasoning for choosing reading and listening span tests as part of the test battery (not reported on in this paper), as it has been claimed that part of individual differences in verbal ability can be accounted for by differences in working-memory capacity. Comprehension deteriorates when working memory is not adequately managed, i.e. if the processing demands of the text exceed working memory resources (Gernsbacher, 1995).

The Geneva project on aptitude testing

The long-term goal of our project is two-fold. It aims at gaining a better understanding of both simultaneous interpreting and of the processes involved in the acquisition of this skill. This project is first and foremost embedded in an interdisciplinary context[1]; expertise research provides the general paradigm for the study of the interpreting process from an ecological perspective. This research produces hypotheses that are then further tested in a controlled environment. Thus

---

[1]This project is embedded in an even larger interdisiciplinary project, initiated by a team of experts from a variety of departments (linguistics, philosophy, medicine, psychology) at the University of Geneva (*Projet plurifacultaire: Langage et Communication*).

controlled experimentation in language processing is in no way incompatible with the more applied methods of expertise research.

We aim to build on theoretical models that have been verified in part through methods of applied and basic research and to isolate those language sub-processes in which interpreters may differ from non-interpreters. In so doing, we seek to determine why some of these processes are only acquired with great difficulty and/or for which a special type of aptitude may be needed and how such aptitude can be identified through testing.

One strategy for identifying specific aptitudes involves making experimental comparisons between highly skilled professional interpreters and naive subjects. These comparisons can usefully be situated along some kind of continuum. At one extreme, the contrast between the performance of the two populations is based on situations directly involving simultaneous interpretation or some close approximation. Alternatively, these comparisons can be based upon standard psycholinguistic tasks (or other types of cognitive/memory tasks) not involving actual interpretation performance, but reflecting language processes that are presumably an integral but not distinctive part of interpretation activity. Given the difficulty associated with asking naive subjects to engage in SI, the former approach is somewhat problematic and, to our knowledge, not commonly exploited. However, tasks such as complex shadowing (DarÚ, 1989) have been developed to approximate some of the complexity and components of SI. The latter type of studies is more common and makes it possible to establish whether or not the differences between the two populations are restricted to some psycholinguistic skills or strategies that are specific to the interpretation process.

In what follows we will present some experiments that are situated along the continuum described above. Indeed, the choice of tests in this study, shadowing, delayed auditory feedback effects, verbal fluency, and reading and listening span (to be reported on in a future article), is designed to cover a number of aspects of the interpreting process. Shadowing experiments are certainly not new in the literature on interpreting (see Lambert, 1992) and have provided interesting insights into interpreters' ability to divide attention between different tasks. Verbal fluency provides information on individuals' efficiency to retrieve verbal information and to produce linguistic output in a limited period of time. Delayed auditory feedback experiments give us some clues as to how individuals overcome processing limitations and more specifically, how experts and novices manage self-monitoring.

Shadowing pilot study

Shadowing, or simultaneous repetition, is a verbal task in which a subject must repeat a spoken signal as quickly and as accurately as possible. This task has been used in different forms by psycholinguists and cognitive psychologists to study various aspects of language (Marslen-Wilson, 1985) and more generally, information processing (Cherry, 1953).

The variants of the shadowing task can be characterized according to the nature of the to-be-shadowed message and the nature of the dependent variable being measured. With respect to the first dimension, it is useful to mention the size (syllables and isolated words versus continuous texts) and the properties (undistorted versus with mispronunciations, rapid versus slow speech, or native versus second language, etc.) of the shadowed message. With respect to the second dimension, the number of errors and the ability of the subjects to comprehend the passages being shadowed were initially measured. Later, with advances in speech-handling technology, the actual temporal delay between the original and the shadowed messages has also been computed. Depending upon the particular research question being addressed, the emphasis was placed on the speed or the accuracy of the shadowing responses.

For the present purposes, it is important to identify the similarities and differences between the processing elicited in shadowing and simultaneous interpretation. At first glance, the two activities seem quite similar since both involve speaking and listening at the same time, in one case in the same language and in the other, in different languages. Clearly, both require the distribution of attention across these two activities and the retention of information in memory. However, SI is certainly more complex and more demanding given the additional processing required. Treisman (1965), among others, has reported results that confirm this greater complexity as revealed by longer latencies and more errors in SI than in shadowing.

In our pilot study conducted with shadowing (Milzow & Wiesenhütter, 1995), we compared the performance of professional interpreters and students.

Method

Subjects

A group of five students at the Ecole de Traduction et d'Interprétation of the University of Geneva, beginning their first semester in interpretation, and a group of five professional interpreters were asked to shadow a text in French and one in English. All subjects were native speakers of French. The interpreters all had between five and ten years of professional experience. Amongst the professionals, there were three women and two men.

Material

Two newspaper articles were selected, one in French and the other in English. Both were approximately 250 words long and of roughly equal difficulty. Both dealt with the subject of political economics. The vocabulary and the contents were deemed accessible to all subjects. The texts were well-structured and had few numbers and names. Both texts were read by native speakers who had previously familiarized themselves with their contents. The reading speed was between 120 and 150 words a minute which corresponds to the average verbal fluency of an interpreter.

Procedure

The subjects were placed in interpreting booths and heard the texts through headphones. They were instructed to repeat the incoming message as quickly and accurately as possible into a microphone placed directly in front of them. Their responses were recorded on a cassette recorder for further analysis.

The text in French was presented first and then the text in English. The first paragraph of the French text and the first two paragraphs of the English text were used as a warm-up exercise and thus were not included in the results. After shadowing the text in one language, subjects were asked to answer several questions testing their comprehension of the passage. The subjects did not expect the questions in French but were aware that they would be answering questions after shadowing the English text.

Results

The performance of the two groups of subjects was evaluated on the basis of three different measures:

1. The temporal delay (in ms.) between words in the original and the repeated text

We selected 20 words from each passage for measuring the shadowing delays. These words were distributed throughout the text and individual sentences. They were also located with equal frequency at the beginning, middle and at the end of clauses. Finally, they began with phonemes whose onsets were relatively easy to locate with the speech analysis program. The production of both the original text and the subjects' responses were digitized with the Sound Designer II software on a Macintosh computer. The waveforms corresponding to both words were displayed simultaneously on a computer screen to measure the temporal interval separating the original and shadowed message .

2. The type and number of production errors made in the shadowed productions.

The categories established by Carey (1971) and Kraushaar & Lambert (1987) were adopted. We noted both the type of mistake (substitutions, omissions, etc.) and the level affected (syllable, word, or syntactic constituents).

3. The performance on the comprehension test.

The subjects received 5 comprehension questions. In order to answer these questions correctly, the subjects had to have understood the passages. A maximum score of 9 points was possible on both tests.

The results obtained for the first two measures are summarized in Figure1A and Figure 1B.

Insert Figure 1A and 1B about here

The interpreters did not shadow the French text more efficiently than the students in terms of the temporal delay in their shadowing responses (the difference of 102 ms is not significant). Moreover, students shadowed the English text more efficiently (delay and errors) than the interpreters. The interpreters made more errors than students in both languages. There were no significant differences in the comprehension test scores between the two groups.

Discussion

Initially, these results are quite surprising in light of the greater experience of the interpreters in dividing their attention between the activities of listening and speaking. In fact, one might have predicted that the interpreters would perform at the level of the "close" shadowers observed by Marslen-Wilson (1985). These subjects had the ability to shadow the incoming message with delays of only 250-300 ms., or approximately a one-syllable lag behind the original message. The performance of neither group tested in the present study even came close to this remarkable behavior.

It can be argued that despite the superficial similarities between shadowing and SI, some basic incompatibilities between these two tasks make it difficult for interpreters to perform efficiently in the former. In particular, these professionals habitually function on the basis of larger processing units in interpretation than those required in shadowing. This could well make it difficult for them to repeat immediately as is required by the shadowing task. Such an explanation based upon the working skills/habits of interpreters gains some support from the error data which show differences in the types of errors made by the two groups of subjects. Students as a whole proved to be rather faithful shadowers, since they made only minor mistakes such as distortions, repetitions, and corrections. Interpreters, however, made greater use of substitution, producing nonetheless sentences that were grammatically correct, and did not alter the meaning of the sentences. These substitutions resemble what Lambert (1992) called "sophisticated corrections". One might wonder whether at some level they were treating the shadowing task as if it were simultaneous interpretation, paraphrasing the input in the same language. We will return to this phenomenon in our general discussion.

Delayed auditory feedback pilot study

Speakers are constantly monitoring their output for correctness. In the case of simultaneous interpretation, the problem of self-monitoring raises some interesting questions. Since the comprehension system is occupied by the processing of the source language input during SI, we can ask what happens to monitoring activity that is presumably based upon the comprehension system. Can interpreters still self-monitor or do they rely less heavily on this feedback than normal speakers do? Little is known about this.

To answer this question we compared the language behavior of interpreters and students under conditions in which self-monitoring was made more difficult. The aim of the pilot experiment to be reported here (Grassi & Mazzoleni, 1996) was to determine how a task involving delayed auditory feedback (DAF) can affect the comparative ability of accomplished interpreters and students beginning their studies in interpreting to self-monitor their output. DAF involves introducing electronically a delay between the moment the speaker produces a speech sequence and the moment he hears this output. It is well-established that submitting subjects to such DAF conditions degrades their production performance considerably. The effect of DAF is observed on different properties of speech, including speech rate, the tone and intensity of the utterances, and the number of errors (omissions, substitutions, repetition of syllables and phonemes) in their pronunciation. However, not every individual is affected to the same extent by DAF. For example, men, young children and bilinguals speaking their non-dominant language are far more disturbed by DAF than are women, older children, and subjects speaking in their native language (Howell & Powell, 1983). More generally, it appears that individuals whose verbal fluency is low are more affected by DAF than those who feature high verbal fluency.

Previous comparisons in DAF experiments by Fabbro and Darò (1995) between interpreting students and students from other disciplines showed that the former were less sensitive to DAF effects. These authors proposed a verbal fluency task under NAF (normal auditory feedback) and DAF conditions to twelve interpretation students and 12 control subjects. The DAF conditions involved different delays, including those of 150, 200 and 250 ms. They noted a significant decrease in verbal fluency as well as a significant increase in terms of the number of errors by control group subjects under all three DAF conditions. Interpreting students did not show any significant difference under any of the three DAF conditions, either in their mother tongue or in the other language. Interpreters' high verbal fluency as well as their capacity to pay less attention to their own linguistic production appear to make them more resistant to interfering effects of DAF.

In the present study, students in interpretation were once again compared with professional interpreters in a reading task.

Method

Subjects

A group of six students in interpretation at the Ecole de Traduction et d'Interprétation of the University of Geneva and six professional interpreters were tested.

Materials

Two biographical texts were constructed for French and English. These texts were of comparable length (315 and 324 words, respectively) and of similar overall difficulty. In addition, a short practice passage was prepared for the two languages to familiarize the subjects with the task.

Procedure

Subjects were first asked to read the French text. Three students read the first half of the text in the NAF condition and the second half in the DAF condition. The three other students read the first half in the DAF condition and the remaining half in the NAF condition. The same procedure was used for the English text. The professional interpreters were also tested in the same fashion with both language texts.

Subjects were asked to put on headphones through which the delayed or unaltered messages were heard and which prevented the spoken productions from arriving directly in the ear. Different research indicates that a delay of 200 ms disturbs subjects the most in their production.[2] This was the delay that was chosen in the present experiment. The productions of the subjects were recorded with a cassette player for subsequent analysis.

Results

A number of different measures were obtained to compare the performance of the two group on the texts in the two languages and the two feedback conditions. First, we computed the global reading times for the entire passage. These results are shown in the Figure 2:

Insert Figure 2 about here

[2]Howell and Powell (1983) proposed that this temporal delay of 200 ms is maximally disruptive since it matches the average time needed for the production of a syllable.

Figure 2 reveals that the detrimental effect of DAF was significantly greater for the students than for the professional interpreters for the French texts. There was a non-significant trend in the opposite direction for the English text.

Figure 3 shows the number of words read in a period of 50 seconds from the onset of the passage. This analysis shows a significant effect of DAF in both languages for students but not for professional interpreters.

Insert Figure 3 about here

Figure 4 shows the results of the durational measures for the individual words selected, an interaction between the effect of DAF and the languages in the two populations studied. In French, the students but not the interpreters, were slowed significantly, whereas in English both subject groups were significantly slowed in their reading rates.

Insert Figure 4 about here

Finally, Figure 5 shows the average number of errors for the entire read text. The results show a much greater number of errors and dysfluencies for the students due to the DAF condition.

Insert Figure 5 about here

Overall, interpreters showed smaller differences between their performance in normal and the DAF conditions. More specifically, in terms of all of the measures for French, the differences were significant for the students but not for the interpreters. The situation is somewhat less clear in the case of the English texts. The performance of both groups differs significantly according to analytical statistics.

Discussion & conclusion

Interpreters made fewer mistakes in DAF and had greater verbal fluency than did students both in English and French. The fact that interpreters, due to the acquired professional experience, are able to pay less attention to their own output might explain their better performance in DAF conditions. Interpreters are hence less affected by this delay.

According to Fabbro and Darò (1995), interpreters concentrate on the incoming message more than they do on their own output. The fact that they are more resistant to DAF may be because they are less dependent on their own feedback than students are.

Verbal fluency pilot study

The final pilot study to be presented here (Casado and Jimenez, 1996) uses the verbal fluency task. *Verbal fluency* refers to the speed and the ease with which a speaker can generate words and sentences. This term covers a wide range of language abilities usually associated with eloquence, garrulousness, wit, and, articulateness. Fillmore (1979) identified several dimensions defining how fluent individuals can produce language. One type of fluency is simply "the ability to talk at length with few pauses", i.e. the ability to fill time with talk. Disk jockeys or sports announcers are included within this category. A second kind of fluency is the "ability to talk in coherent, reasoned and semantically dense sentences." In this case the speaker expresses himself concisely without filling the discourse with semantically empty material. The third type of fluency, according to Fillmore (1979), refers to the "ability to be creative and imaginative; to express ideas in a novel way." This kind of speaker appears to have available a large range of alternative linguistic ways of responding to a situation and to choose the most appropriate one.

Research has shown that verbal fluency is influenced by gender, personality variables, and differential socialization patterns. Two types of explanations have been advanced to account for differences in fluency (Bock, 1996). The first appeals to <u>mental skill</u>, that is, the effects of differentially efficient processing mechanisms attained either through practice or inherently present. The other explanation appeals to differences in mental energy, as expressed by attention or working memory in the production of speech. These two perspectives are clearly complementary.

It is far from obvious how to test for verbal fluency. A pioneer in the study of verbal ability is certainly L.L. Thurstone, whose theory of primary mental abilities (Thurstone, 1938) has been very influential in American differential psychology. He assessed word fluency by testing the subjects' ability to generate words beginning with a specific target letter in a limited amount of time. Subsequently, different uses of the verbal fluency task have been made, in particular, in the field of aphasiology where such tests are commonly included in test batteries.

In this section we will examine how various verbal fluency tasks can be applied to the comparison between professional interpreters and students. Several subtasks (Casado & Jimenez, 1996; Benzoni & Cazajeux, 1996) were prepared with the objective of studying different linguistic levels of lexical processing: the semantic, morphological, orthographic and phonological levels.

Method

Subjects

Five first-year students at the Ecole de Traduction et d'Interprétation of the University of Geneva and five professional interpreters participated in the experiment. All subjects were French native speakers.

Material

The verbal fluency experiment consisted of the following subtasks: semantic task, free association task, spelling task, morphological task and a phonological task. In each subtask, subjects received the instruction to produce the maximum number of words in one minute on the basis of a specified cue. The nature of the cue varied across subtasks. For the semantic subtask, subjects were given four different semantic categories including: colors, animals, fruits and communication. In the free association task, subjects were asked to produce all the words that they associated with cat, table, rain and strike. For the morphological task, both prefixes (a-, dis-, re-, pré-) and suffixes (-isme -ance, -ure, -ité) were given as cues.

For the orthographic task, the subjects received four different letters: C, P, U and I. The phonological task included both high (dé, mar, ca, cor) and low frequency syllables (fu, bru, sta, ri). Finally, we also included an articulatory fluency task to establish the maximum articulatory rates that the subjects were capable of. Here subjects were required to repeat pairs of words 10 times as quickly as possible.

Procedure

All tasks were administered to subjects in interpreting booths at the Ecole de Traduction et d'Interprétation. All students took part in the experiment simultaneously, whereas the interpreters were tested individually. For each task, subjects were first given an example and then the cue. The subjects' linguistic productions were recorded on cassettes and evaluated by two independent judges according to well defined criteria.

Results

We computed the number of correct productions in each of the different subtasks. In Figure 6 we summarize the results for the different conditions (collapsed across the semantic categories).

Insert Figure 6 about here

Globally, we see that both groups of subjects produced the greatest number of words in the semantic fluency subtask and differed only marginally in the form-based (phonological and orthographic) and the morphological subcategories. In the latter subtasks, the average number of words produced in one minute is relatively low. More interesting for the present purposes is the comparison between the two different populations. Here we see only small differences in the scores for the populations, none of which reach significance.

Discussion

Unlike what we predicted, the present study showed no differences in the performance of the student and professional interpreting populations. This is, of course, not to say that there are no differences between individual subjects in the two different population types. Indeed, large differences can be found (up to a factor 2) between the student subjects for specific subtasks. This variability and the small number of subjects make it difficult to draw firm conclusions here. Clearly, similar experiments with more subjects are required.

Conclusions

In this study, we have compared the performance of novice and expert interpreters by testing them in a series of three separate pilot experiments. These were designed to tap into different parts of the language processing and language interpreting system. We obtained clear-cut differences between the two groups in the experiment involving the reading under delayed auditory feedback conditions. In contrast, interpreters' and students' performance did not differ in either the shadowing or the verbal fluency tasks. It is not possible to determine precisely why the performance of the two groups on the latter two tasks did not differ in the expected direction. However, these experiments tested only a limited number of subjects, making it difficult to generalize. Nonetheless, in what follows we will speculate on some reasons for the absence of differences.

In the shadowing task, we believe that interpreters' acquired processing strategies for interpreting made it difficult for them to adjust to the different requirements of a shadowing task. In particular, the much longer shadowing delays observed for the English text (the condition which would most closely resemble the usual interpreting condition, i.e. working from one's B- into one's A-language) could well reflect the difficulty interpreters had - in this condition in particular - in suppressing automated processing strategies and in adapting to changed processing requirements. In the verbal fluency task, interpreters did not perform better than the students in any of the subtasks (the fact that professional interpreters outperformed students on the semantic subtask is statistically not significant). This result is surprising and the phenomenon needs to be studied with a greater number of subjects.

The DAF experiment brought out a clear performance difference between novices and experts which could be explained by interpreters a) being less dependent on monitoring their own output; and/or b) their improved ability to process two inputs simultaneously and consciously allocating attention to one or the other (Lambert et al., 1995). Moreover, it is also not clear whether the observed differences are in part due to a variation in the total capacity of resources available, or to the efficiency with which certain cognitive processes are executed by interpreters. Further research is necessary to shed more light on the question of whether these performance differences reflect a change in processing efficiency on part of interpreters, who have managed to shift from consciously controlled to more automatic processing, or whether they reflect a more or less dramatic restructuring of the process itself (McLaughlin, 1995). We believe that our results, when taken in conjunction with other experimental evidence on improved DAF performance by advanced interpreting students (Fabbro & Darò, 1995), provide useful evidence on which to build for developing more refined measures that would allow us to answer some of the questions raised in this paragraph.

Finally, with a variety of experimental conditions, we have tested the global hypothesis according to which expert interpreters should outperform novice students on all experimental tasks. While this hypothesis had to be rejected, more specific hypotheses could be confirmed. Most important seems to be the fact that in comparing novices and experts in variety of language processing-related experimental conditions we found evidence of performance differences that will allow us to hone in on aspects of experts' and novices' behavior that may hold further clues to what sets the two apart.

References

Anderson, J.R. 1982. "Acquisition of cognitive skill." *Psychological Review* 89:369-406.

Anderson, J.R. 1983. *The architecture of cognition*. Cambridge. MA: Harvard University Press.

Anderson, J.R. 1995. *Cognitive Psychology and its Implications*. New York: W.H. Freeman.

Butterworth, B. 1980. *Language Production, Vol. I: Speech and Talk*. London: Academic Press.

Casado, B. and Jimenez, J. 1996. *Mental processes involved in conference interpretation: Verbal fluency*. Unpublished M.A. thesis, Ecole de Traduction et d'Interprétation, University of Geneva.

Carey, P.W. 1971. "Verbal retention after shadowing and after listening". *Perception and Psychophysics*, 9, 79-83.

Carroll, J.B. 1981. "Twenty-five years of research on foreign language aptitude." In K.C. Diller (ed), *Individual Differences and Universals in Language Learning Aptitude*. Rowley, MA: Newbury House.

Chase, W.G. and Simon, H.A. 1973. "Perception in chess." *Cognitive Psychology* 4:55-81.

Cherry, C. (1953). "Some experiments on the recognition of speech with one and two ears". *Journal of the Acoustic Society of America*, 2, 975-979.

Darò, V. 1989 "The role of memory and attention in simultaneous interpretation: A neurolinguistic approach". *Interpreters' Newsletter*, 2, 50-56.

DeGroot, A.D. 1965. *Thought and Choice in Chess.* The Hague: Mouton.

DeGroot, A.D. 1966. "Perception and memory versus thought: Some old ideas and recent findings." In B. Kleinmuntz (ed), *Problem Solving: Research, Method, and Theory*. New York: Wiley, 19-50.

Eckstrom, R.B., French, J.W., Harman, H.H. and Dermen, D. 1976. *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

Fabbro, F., & Darò, V. 1995. "Delayed auditory feedback in polyglot simultaneous interpreters." *Brain and Language,* 48(3), 309-319.

Foley, M. and Hart, A. 1992. "Expert-novice differences and knowledge elicitation." In R.R. Hoffman (ed), *The Psychology of Expertise. Congitive Research and Empirical AI.* New York: Springer, 233-244.

Furneaux, S. 1956. *Manual of Nufferno Speed Tests and Manual of Nufferno Level Tests.* London: Institute of Psychiatry.

Gathercole, S.E. and Baddeley, A.D. 1993. *Working Memory and Language*. Hillsdale,NJ: Lawrence Erlbaum Associates.

Gernsbacher, M.A. 1995. "The structure building framework: What it is, what it might also be, and why." In B.K. Britton and A.C. Graesser (eds), *Models of Text Understanding.* Hillsdale, NJ: Lawrence Erlbaum Associates, 289-311.

Gerver, D., Longley, P.E., Long, J. and Lambert, S. 1989. "Selection tests for trainee conference interpreters." *META, XXXIV*(4):724-735.

Gingiani, A. 1990. "Reliability of aptitude testing: A preliminary study." In L. Gran and C. Taylor (eds), *Aspects of Applied and Experimental Research in Conference Interpretation.* Udine: Campanotto, 42-53.

Grassi, S., & Mazzoleni, D. 1996. *Les tests d'aptitude pour l'interprétation de conférence: Le delayed auditory feedback.* Uunpublished M.A. thesis, Ecole de Traduction et d'Interprétation, University of Geneva.

Grosjean, F. 1996. “Living with two languages and two cultures.” In I. Parsnis (ed), *Cultural and Language Diversity and the Deaf Experience*. Cambridge/New York: Cambridge University Press, 20-37.

Hoffman, R.R. 1996. “How can expertise be defined? Implications of research from cognitive psychology.” In J. Fleck and R. Williams (eds), *Exploring Expertise*. Edinburgh: University of Edinburgh Press.

Howell, P., & Powell, D.J. 1983. “Delayed auditory feedback with delayed sounds in varying duration". *Perception and Psychophysics*, 42(2), 166-172.

Kintsch, W. 1974. *The Representation of Meaning in Memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Klein, G.A. and Hoffman, R.R. (1993). “Seeing the invisible: Perceptual-cognitive apsects of expertise.” In M. Rabinowitz (ed), *Cognitive Science Foundations of Instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates, 203-226.

Kraushaar, B., & Lambert, S 1987. “Shadowing proficiency by ear of input and type of bilinguality.” *Bulletin of the Canadian Association of Linguistics*, 9, 17-31.

Künzli, A. and Moser-Mercer, B. 1995. “Human strategies for translation and interpretation.” In L. Dreschler-Fischer and S. Pribbenow (eds), *KI-95 Activites: Workshops, Posters, Demos.* Bonn: Gesellschaft für Informatik e.V.

Lambert, S. 1991. “Aptitude testing for simultaneous interpretation at the University of Ottawa.” *META, XXXVI*(4):586-594.

Lambert, S. 1992. “Shadowing.” *META, XXXVII*(2):263-273.

Lambert, S., Darò, V. and Fabbro, F. 1995. “Focalized attention on input vs. output during simultaneous interpretation: Possibly a waste of effort!” *META* 40:39-46.

Larkin, J., McDermott, J., Simon, D. and Simon, H. 1980a. “Expert and novice performance in solving physics problems.” *Science* 208:1335-1342.

Larkin, J., McDermott, J., Simon, D. and Simon, H. 1980b. “Models of competence in solving physics problems.” *Cognitive Science* 4:317-345.

Levelt, W.J.M. 1989. *Speaking: From Intention to Articulation.* Cambridge, MA: MIT Press.

Longley, P. 1989. “Aptitude testing of applicants for a conference interpretation course.” In L. Gran and J. Dodds (eds), *The Theoretical and Practical Aspects of Teaching Conference Interpretation.* Udine: Campanotto, 105-108.

McLaughlin, B. 1995. “Aptitude from an information processing perspective.” *Language Testing*, :364-381.

Marslen-Wilson, W.D. 1985. "Speech shadowing and speech comprehension." *Speech Communication*, 4, 55-73.

Marslen-Wilson, W.D. and Tyler, L.K. 1980. "The temporal structure of spoken language understanding." *Cognition* 8(1), 1-71.

Mayer, R.E. 1992. *Thinking, Problem Solving, Cognition.* New York: W.H. Freeman.

Milzow, M., & Wiesenhütter, A. 1995. *"Les tests d'aptitude pour l'interprétation de conférence: le shadowing."* Unpublished M.A. thesis, Ecole de Traduction et d'Interprétation, University of Geneva.

Moser-Mercer, B. 1984. "Testing interpreting aptitude." In W. Wilss and G. Thome (eds), *Translation Theory and its Implementation in the Teaching of Translating and Interpreting.* Tübingen: Narr, 318-325.

Moser-Mercer, B. 1994. "Aptitude testing for conference interpreting: Why, when and how." In S. Lambert and B. Moser-Mercer (eds), *Bridging the Gap. Empirical Research in Simultaneous Interpretation.* Amsterdam: John Benjamins, 57-68.

Moser-Mercer, B. (in press). "Acquiring expertise in sight translation;" In S. Lambert (ed), *A Cognitive Approach to Interpreter Training.* Amsterdam: John Benjamins.

Moser-Mercer, B. (in press). "The expert-novice paradigm in interpreting research." In E. Fleischmann, W. Kutz and P. Schmitt (eds), *Translationsdidaktik. Beiträge der VI. Internationalen Konferenz zu Grundfragen der Übersetzungswissenschaft, Leipzig 11.-13.9.1996.* Tübingen: Narr.

Newell, A. and Rosenbloom, P.S. 1981. "Mechanisms of skill acquisition and the law of practice. In J.R. Anderson (ed), *Cognitive Skills and their Acquisition.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1-55.

Paradis, M. 1993. "Linguistic, psycholinguistic and neurolinguistic aspects of 'interference' in bilingual speakers: The Activation Threshold Hypothesis." *International Journal of Psycholinguistics* 9:133-145.

Simon, D.P. and Simon, H.A. 1978. "Individual differences in solving physics problems." In R. Siegler (ed), *Children's Thinking: What Develops? Hillsdale*, NJ: Lawrence Erlbaum Associates, 325-348.

Stubbs, J.B. and Tucker, G.R. 1974. "The Cloze Test as a measure of English proficiency." *Modern language Journal* 58:239-241.

Treisman, A. 1965. "The effects of redundancy and familiarity on translation and repeating back a foreign and native language." *Birtish Journal of Psychology* 56:369-379.

Wechsler, D. 1945. "A standardised memory scale for clinical use." *Journal of Psychology* 19:87-95.

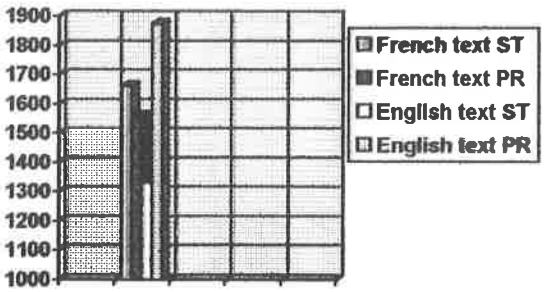

Fig. 1a: Delay (in ms) for the two language texts
and two groups of subjects (**ST**udents/**PR**ofessionals)

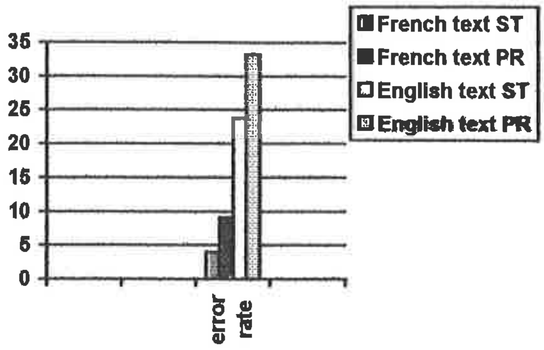

Fig. 1b: Shadowing errors per language and group of subjects (STudents/PRofessionals)

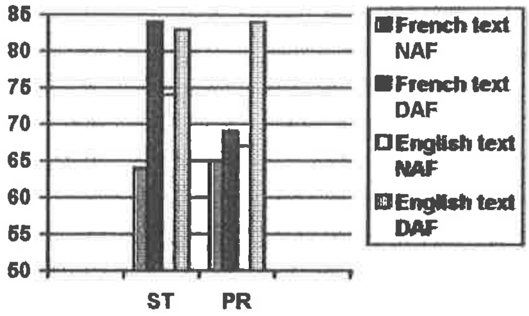

Figure 2: Average total reading time in seconds for different subject groups (STudents/PRofessionals) and conditions (No Auditory Feedback/Delayed Auditory Feedback)

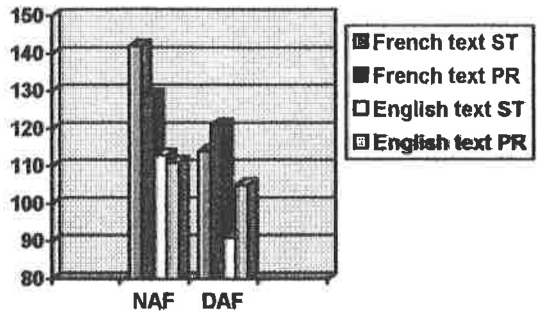

Fig. 3: Average number of words read during the first
50 second period for different groups and conditions

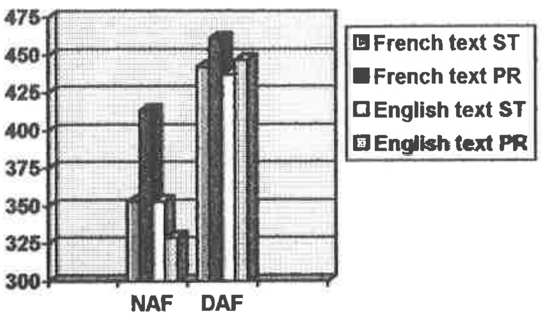

Fig. 4: Average duration (in ms) of the 14 selected words for different conditions and groups

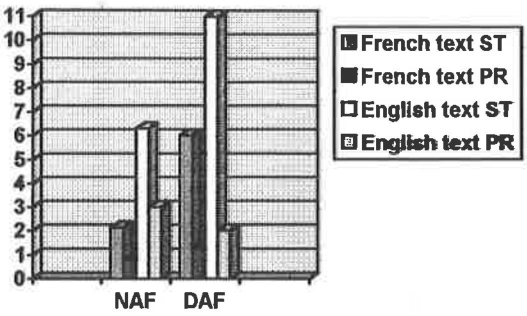

Fig. 5: Average number of errors per text and group

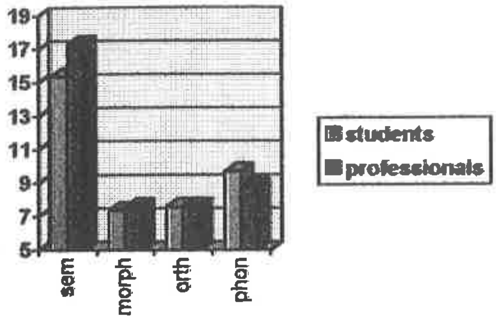

Fig. 6: Average number of words produced by professional interpreters and students in the various verbal fluency subtasks (**sem**antic, **morph**ological, **orth**ographic, **phon**ological).