NUMBERS IN SIMULTANEOUS INTERPRETATION

Cristina Mazza
Freelance Conference Interpreter

1. Introduction
1.1 Numbers

In everyday practice, conference interpreters report a certain difficulty dealing with numbers. On first analysis, it might seem surprising that numbers cause so many problems. Indeed, since their main characteristic is univocity of meaning (Alessandrini 1990), their interpretation should be relatively easy.

In most modern languages, numbers are represented in writing both by their *number name*, for example "five" in English, and by *numerals*, that is special symbols for representing numbers visually, for example "5" (Hurford 1987). While the Arabic code of numerals is almost universal, number names are specific to each language, and are used to reproduce numbers both orally and in writing. Since numerals must represent a potentially infinite set of quantities, the system developed makes it possible to create virtually all numbers from a small set of basic words and fixed syntactic rules, each providing a single piece of information on the number. The main constituents are called *base digits* (0...9 in the decimal system), which are used to create another two lexical classes: *teens* (eleven, twelve,…nineteen) and *tens* (twenty,…ninety). A fourth set of items, called multipliers (e.g. hundred), is used to mark the magnitude of the base digits.

Although this system allows the representation of all numbers, it sometimes requires a long string of items to reproduce a large quantity. For example, English uses two words to say "35", "thirty five", while "35,035" needs as many as six terms, "thirty five thousand and thirty five". The difficulty experienced by interpreters could therefore be a memory problem, due to the need to remember many words for expression of a single concept.

1.2 Memory, numbers and simultaneous interpretation

Memory is a complex cognitive activity, whose functions are the *encoding* (comprehension), *storage*, and *retrieval* of input information. These tasks are carried out by two different kinds of memory, *short-term memory* (STM) and *long-term memory* (LTM). Information which must be stored for a short time (1.5-2 seconds) is processed in the STM, while information stored for a long time (from minutes to years) is managed by the LTM (Atkinson 1999). One of the main functions of STM is to hold and manipulate information while it is used for some cognitive activities (storing, but also reasoning, calculating, etc.). In this case, it is said that STM acts as a Working Memory (Baddeley 1990).

The working memory (WM) is simply the *temporary storage of information that is being processed* (Baddeley 1987: 34). The many experiments conducted over the last decades (Miller 1956; Craik and Lockhart 1972; Baddeley 1987; Harrington 1992) have highlighted that the WM also has an important role in linguistically related complex cognitive tasks, and in the planning and organization of tasks.

However, WM also has several limits. The most striking characteristic is perhaps its very limited capacity in terms of the number of items it can store at the same time (the *memory span*). In his milestone article, *The Magical Number 7*, Miller maintains that on average WM capacity[1] spans sequences of only 7 ± 2 non-related chunks (1958: 91). Linking elements together can be a useful strategy for reducing the number of chunks to be retained by WM (Bower and Springston 1970, mentioned in Atkinson 1999). While chunking is possible with digits, it is less probable than with letters (it is not necessary to form a meaningful word to create a chunk of letters, a syllable being sufficient).

WM capacity varies according to the complexity of the items (letters, words, sentences, texts) and the task to be carried out. Management of extended texts can be facilitated by reducing storage demands - for example, words and syntactic structures can be forgotten, to retain only the general meaning of the sentence (Carpenter *et al*. 1994). However, Seleskovitch (1975: 13/15) considers that:

> *ce phénomène général de dissociation du sens et des structures sonores dans la mémoire ne s'applique pas au chiffre, [dont] la signification unique implique une (…) concordance entre leur structure sonore et leur signification.*

If structural complexity increases the burden on WM, *anticipation* can help cope with processing demands. Anticipation is the establishment of the range of all possibilities that might follow a linguistic item, according to language use (Baddeley 1990, 1993). For example, English usage generally expects an article to be followed by a noun or an adjective, while the occurrence of a verb after it is impossible. With numbers, none of the other elements in a sentence anticipates them. Obviously, expressions like “the rate dropped by” or “the

1 Miller originally put forward the “7 ± 2” idea for STM. However, in the present article the reference will be to WM, because in simultaneous interpretation STM is used to carry out cognitive activities (see above).

number of X was" make it clear that a quantitative item of information will follow, but they do not help make inferences about its exact value.

The intrinsic difficulty in remembering numbers is accentuated during simultaneous interpretation (SI), which is a complex cognitive activity. In SI the interpreter listens to the incoming message and translates it, to all intents and purposes, immediately. In practice, there is usually a certain delay (*décalage*) between the source text and the interpreter's output (Setton 1999), necessary to understand a meaningful chunk in the source language, process it, and shift language to produce it.

The role of memory in SI is essential. LTM provides essential lexical, syntactic and semantic information to understand the input and produce the output, while WM is needed to keep the input activated while it is being processed and translated, and to monitor the output

The limitations of WM are even more apparent in SI than in normal text comprehension. Indeed, while mnemonic traces are very short-lived in normal listening conditions (Luria 1976), in interpreting they must arguably be stored for a longer time for continuous checking against the target speech (Kalina 1992). This requires continuous rehearsal, known as *subvocal repetition* (Baddeley 1987). However, the presence of many concurrent tasks in SI is likely to cause interference with this operation: Baddeley (1990: 79) has indeed shown that *articulatory suppression*, i.e. concurrent repetition of items, interferes with subvocal repetition.

Against this background, numerals, especially large ones, could increase demands on the interpreter's WM and disrupt the normal encoding, storing and retrieval of the other elements in the text.

1.3 Gile's Effort Model of Simultaneous Interpretation

The Effort Models are based on the idea that the mind has a limited capacity, and that the difficulties in interpretation stem from time constraints and the need to divide attention between several concurrent operations (Gile 1995: 91). In the Effort Models, SI is seen as requiring a balance in the allocation of processing capacity, according to the requirements of each task performed at a given moment.

The Effort Model of SI is based on three non-automatic interpreting Efforts[2]:

– Listening and Analysis Effort: all the mental operations between perception of discourse by auditory mechanisms and the moment at which the

2 See Gile (1999: 154), and also Setton (1999: 35) for a review.

interpreter either assigns, or decides not to assign, a meaning to the segment s/he has heard;

– Production Effort: all the mental operations between the moment at which the interpreter decides to convey an item of meaning and the moment at which s/he reformulates it in the target speech (TS);
– Working Memory Effort[3]: all the memory operations from the time a speech segment is heard to the time it is reformulated in the TS or disappears from memory.

The management of capacity allocation (Coordination Effort) between the three Efforts causes further energy consumption. At each moment during SI, the individual Efforts deal with different fragments of the original speech (Gile 1995: 99). The sum of the capacity required by all single Efforts at a specific moment represents the total capacity required, which cannot exceed the total capacity available. If one of the three Efforts suddenly needs more resources to deal with a particularly difficult segment, a part of the attentional capacity is diverted from the other two. However, if all the Efforts are operating close to their allotted capacity, this shift of resources can cause errors in one or more tasks.

Gile has also developed the *Tightrope Hypothesis* (Gile 1999: 157), stating that most of the time interpreters work near saturation level. Therefore, even a limited increase in the attentional requirements could lead to failure. This hypothesis might account for the high frequency of errors and omissions: if interpreters worked well below saturation level, failures should occur only in the presence of very difficult chunks.

Gile mentions numbers several times in his works (e.g., 1984a, 1995, 1999), arguing that they are a problem trigger in SI because of their *low redundancy*. This increases the Listening and WM Effort: it is essential not to miss or forget them, because they cannot be inferred from any other element in the text (Gile 1995: 108).

Numbers are also characterized by *low predictability* (Braun and Clarici 1996). This again increases the Listening and WM Efforts, because no anticipation is possible. The quantity expressed can only be understood the moment it is uttered by the speaker.

In addition, numbers also have a *high informative content* (Alessandrini 1990). Dense speech sections increase processing capacity demands for all Efforts, because the interpreter must process, retain and translate more information per unit of time.

3 Gile originally uses the term STM (see note 2).

According to the Effort Models, numbers are therefore extremely energy-consuming. Each of the three Efforts in SI requires additional resources to deal with them, which is likely to saturate the system or leave insufficient capacity for concurrent Efforts. This could explain the frequency of mistakes with numbers in SI, especially large numbers which are expressed as a string of shorter components.

- Management of numbers in SI

Are there some means, or strategies, by which the interpreter can overcome this difficulty? Obviously, written material (handouts, slides) and the boothmate, who can write down figures and names, are helpful, not only for numbers (Gile 1984b: 84). However, help from these sources is not always available.

One possible solution is reduction of *décalage*. This allows shorter retention in STM, but faster processing and restitution of the input. There is thus more strain on the Listening (Comprehension) and Production Efforts. xxxx

To deal with particularly complex segments, Kalina (1992: 254) suggests that *approximation* could be a useful strategy. She sees it as a means to provide partial information until the interpreter finds a more accurate translation. However, this is probably counterproductive for numbers, because strategies to gain time increase the WM Effort.

Since the main problem with numbers seems to be insufficient WM, a possible strategy could be to write the numbers down while interpreting. Note-taking is widely used in consecutive interpretation, to help LTM during the production stage. Although note-taking can represent a further Effort competing for its share of attentional capacity, it might also prove useful as a means of relieving excessive pressure on WM in SI – particularly for items like numbers and proper names. In addition, the possibility of using a “supra-linguistic code”, the Arabic code, facilitates detachment from the source language. These are obviously theoretical assumptions, which need empirical confirmation.

A few studies, for instance by Crevatin (1991) and Braun and Clarici (1996), have focused on management of numerals in SI, and the usefulness of notes. Results show low overall accuracy for numerals, omissions being the most common mistake. In both these studies, notes proved useful for numbers during “passive” interpretation (i.e. into the A-language). However, detailed and systematic comparison of these experiments is difficult, because the methods used varied considerably. In addition, the languages studied were different (Italian-English in one case, and Italian-German in the other).

2. The experiment

2.1 Aims of the study

This study was set up specifically to investigate the following points:

1) whether numbers cause problems during simultaneous interpretation (and, if possible, why);
2) whether numbers affect the interpretation of nearby numbers and of the surrounding text;
3) which types of numbers cause most problems;
4) whether note-taking can be helpful when interpreting numbers.

Numbers are expected to prove troublesome for subjects in the experiment – especially large numbers, which are likely to require more WM Effort. However, taking notes during SI of numbers might be a good strategy to solve this problem (see above). Whether note-taking is useful for SI of numbers is actually the main focus of the study.

2.2 Materials: interpretation and questionnaire

The experiment was carried out on 15 students of interpretation, aged between 24 and 28 years of age, who had attended interpretation courses for at least 3 years at the Faculty for Interpreters and Translators in Forlì (SSLMIT[4]). All subjects were late bilinguals, Italian being their A-language and English their B-language, and knew at least one other foreign language. Although no test was administered, all subjects were known to be right-handed at least as regards their handwriting. Of the 15 students tested, 12 were female and 3 were male.

The experiment consisted in three simultaneous interpretations from English into Italian, one of a general speech, and two of balanced experimental speeches containing numbers. Note-taking was allowed for the first experimental text only. The English source speeches dealt with recent trends in birth rates and other topics related to motherhood in the United States. All texts were read by an English mother-tongue lecturer at the SSLMIT, and recorded on a cassette recorder with an external microphone. The general text (T0) had the double purpose of ensuring that the subjects were able to produce an acceptable interpretation of a text without numbers, and of providing subjects with relevant information and vocabulary.

The two experimental texts (T1, with note-taking allowed, and T2, without note-taking) were balanced: they contained an introductory part with no

4 Scuola Superiore di Lingue Moderne per Interpreti e Traduttori, University of Bologna.

numbers, lasting about 2 minutes, in which all the specific terms were presented. The second part contained 67 and 60 numbers of different types respectively. “Resting” paragraphs interrupted the flow of numbers now and then, allowing evaluation of whether subjects performed better in these parts without numbers.

Numerals were divided into 5 categories, which were expected to create different problems for subjects. This classification was not used in any other experiment on interpretation of numbers, but was especially created for the present study: A) whole numbers ≥ 1,000 (with 4 or more digits); B) whole numbers < 1,000 (with less than 4 digits); C) decimals; D) ranges; E) dates.

Some numbers were repeated, for instance the date *1998* in both T1 and T2. This could provide information as to whether a number causes the interpreter fewer problems when it is not used for the first time. Other figures formed a *cluster*, that is a passage with a high density of numbers. Here too, the analysis could reveal whether these clusters increase the difficulty of interpretation.

Before starting the experiment, it was made clear that during the interpretation of T1 and T2 a series of numbers would occur, and that note-taking was allowed for T1 only. Subjects were then given a preliminary briefing, which included possible translations for some terms, so that these items could be handled with minimum effort during the trial.

Notes taken during the interpretation of T1 had to be handed in at the end of the session. The only participants who took no notes were subjects no. 1, who is blind, and no. 9. Notes of subject no. 15 are missing. For the interpretation of T2, subjects were not allowed to take notes, but could refer to those taken during the briefing for items other than numbers.

All subjects interpreted the texts in the same sequence, that is T0, T1, and T2, which is their “natural” order. The experiment was carried out at the SSLMIT, using booths and recorders in the interpretation classrooms. All subjects except nos. 1 and 14 interpreted the three texts on different days.

An identical questionnaire was administered to all subjects after the interpretation of T1 and T2, with 10 questions divided into three distinct “blocks” concerning different aspects of the analysis: *Difficulty of text*, *Influence of numbers on performance* and *Importance of notes*.

The purpose of the questionnaire was:

1) to confirm that subjects considered the texts fairly easy except for numbers;
2) to provide a basis for matching subjects’ perception of certain points (their performance in T1 and T2, the influence of numbers, the usefulness of note-taking) against their actual performance. Expectations were that subjects would evaluate both texts as generally easy, and numbers as very difficult; that they would perceive numbers not only as inherently difficult, but also as problem triggers in interpretation of the context; that performance would be

rated worse for T2 than for T1; and that note-taking would be considered important when interpreting numbers.

2.3 Evaluation procedures

The categorization of mistakes is based on that of Braun and Clarici (1996). For the present study, it was decided to simplify the classification and group together similar errors. Six types of mistakes were established:

1) *omissions*: the numeral is left out altogether, or replaced by a generic expression such as *molti*, *pochi* (many, few), etc.;
2) *approximations*: translations respect the right order of magnitude, but are rounded up or down. Usually, the subject is aware that the original number is different, and accompanies his/her interpretation with a lexical element (e.g. 3,941,553 being translated as *circa 3.900.000*, "about 3,900,000");
3) *lexical mistakes*: the order of magnitude of the stimulus is maintained, but the elements composing the numeral are in the wrong order, they have been either misplaced or inverted (e.g. 346 being translated as 436; 1989 as 1998);
4) *syntactic mistakes*: even if it contains the right figures in their correct sequential order, the number is of a wrong order of magnitude, or the nature of the number has been changed, modifying the overall information (e.g. 110,000 becomes 1010; 51.1/1000 becomes 51.1%, 423% becomes "423 units");
5) *phonological mistakes*: the error can be related to a phonemically wrong perception of similar sounding figures in English (e.g. 17, "seventeen", perceived as 70, "seventy");
6) *other mistakes*: this category includes all possible mistakes not belonging to any of the previous types. Causes of such errors are often not apparent, and cannot be readily grouped into categories. These mistakes are therefore kept apart and form a rather miscellaneous group. However, analysis of all mistakes in this category also considers their position in the original text, to identify a possible "echo" effect due to the proximity of another number.

In the analysis, all numbers were circled with a different color according to whether they were translated correctly or wrongly and, if so, according to the error concerned. Since one aim of the study was to consider how numbers affect the interpretation of the text as a whole, errors in the surrounding text were identified by differentiating the system of highlighting.

The error analysis conducted on interpretations of T1 and T2 focused mainly on:

– the total amount of wrong numbers;
– which category of numbers (A, B, C, D, E) caused more problems;
– the total amount of different mistakes;

- which was/were the most common mistake/s;
- whether a specific type of number was likely to produce a specific kind of mistake.

To ensure comparability of error rates in T1 and T2, as well as between the various categories of numbers, results were calculated as percentages.

The hypothesis that notes can be useful during SI of numbers was tested statistically, by comparing the results in T1 and T2, note-taking being the distinguishing feature of the two texts. The test used was the Paired Sample T-Test (C.I. = 95%).

Subjects' notes for T1 were examined, to ascertain: (i) what proportion of numbers in the text were noted; (ii) whether subjects who took more notes performed better than the others. The statistical correlation between quantity of notes and quality of performance was examined, matching scores for correctly and incorrectly interpreted numbers in the TS against the frequency of noted numbers.

The analysis of notes also considered which types of numbers were written down. Errors in note-taking were considered for each kind of number.

3. Results

3.1 Performance

In terms of content, all subjects in the sample interpreted the text without numbers (T0) satisfactorily.

Within each of the experimental texts (T1 and T2), fidelity to the source speech was compared in two paragraphs with numbers and two without. To ensure that the amount of information in paragraphs with and without numbers was comparable, they were broken down into items of meaning[5]. Numbers and their immediate context were scored as separate items of meaning, since in many cases subjects omitted the numeral but conveyed the idea of the trend correctly (increase, decrease, etc).

On average, subjects correctly reproduced 81.8% of items in parts without numbers, but only 53.9% of those in parts with numbers, suggesting that interpretation of parts with numbers was more difficult and less accurate. While performances varied (e.g., in fluency, capacity to successfully convey the message, style, etc.), inter-subject differences did not affect the overall trend. In

5 The expression *item of meaning* refers to all items of form judged necessary to understand a sentence or utterance. A sentence of T1, with items of meaning under-lined, is given here as an example: *Babies born to single mothers were 1,293,567 in 1998, that is 32.8% of all births, while in 1997, the percentage was 32.4%.*)

other words, the interpretation was consistently worse in parts with numbers than in those without.

3.1.1 General performance for numbers

The degree of accuracy for numbers varied considerably in both T1 and T2. The error score in T1 ranged from 50 mistakes out of 67 (74.6% of all numerals) to 14 (20.9%). The mean score, 30.2 (45.1%), highlights that on average nearly half the numerals were interpreted wrongly. Conversely, only 8 people translated correctly more than 50% of numbers, and only 3 exceeded 70%. Similar remarks can be made for T2: the error score ranged from 46 mistakes out of 60 (76.7%) to 11 (18.3%). The mean, 29.9 (49.9%) underlines that on average half the numerals were wrongly interpreted, 70% accuracy being reached by only three subjects.

The statistical analysis shows that overall performance is better in T1 than in T2 ($t = 2.882$; $df = 14$; $p = 0.012$). This means that note-taking, the independent variable, probably influenced performance in SI of numbers.

3.1.2 Categories of numbers

Data for both T1 and T2 indicate that numbers are often wrongly interpreted. However, the probability of error varies according to the type of number and its position in the text. Table 1 shows the error score for each category in T1 and T2, and the results of the statistical test carried out to highlight whether the difference in performance is significant (this condition is verified when $p < 0.05$).

	T1	T2	Paired Sample T-Test		
			T	df	P
A) Large whole numbers	67%	82%	- 2.287	14	0.038
Small whole numbers	37%	49%	- 2.682	14	0.018
C) Decimals	56%	63%	- 2.433	14	0.029
Ranges	47%	57%	- 2.674	14	0.018
Dates	41%	35%	*Difference is not significant*		

Table 1 – Error score for each category of number in T1 and T2

Results suggest particular difficulty in interpreting whole numbers with 4 digits or more (A) and decimals (C), while whole numbers with less than 4 digits (B) and dates (E) are more easily interpreted. Ranges (D) also prove troublesome, especially in T2. However, performance in ranges varies slightly according to the components of the range (e.g. dates, decimals, whole numbers).

The difference in performance between T1 and T2 is significant for all categories except dates. However, category A actually includes the number *1,000*, twice in T1 and once in T2. This is a “borderline” item, in that it comprises four digits but is in fact a simple sequence of a digit (one) and a multiplier (thousand). If it is excluded from the calculation, the error score is the same in both texts (91.7%, for the four “typical” large numbers in both T1 and T2). Despite the statistically significant difference for this category, results thus suggest that fidelity for “normal” large numbers is not affected by note-taking.

3.1.3 Types of mistake

The analysis of the different types of mistakes highlights that the percentage of omissions is the same in both texts, while approximations, lexical mistakes, phonological mistakes and other mistakes are more frequent in T2. Only the proportion of syntactic mistakes is higher in T1.

Table 2 shows the proportion of the various types of mistake in the two texts, and the results of the statistical test.

	Breakdown of mistakes in T1	Breakdown of mistakes in T2	Paired Sample T-Test		
			T	df	P
Omissions	30.1%	30.2%	*Difference is not significant*		
Approximations	6.7%	20%	3.877	14	0.002
Lexical mistakes	4.1%	4.9%	*Difference is not significant*		
Syntactic mistakes	4.6%	2.6%	*Difference is not significant*		
Phonetic mistakes	1.1%	4.8%	3.609	14	0.003
Other mistakes	13.5%	13.5%	*Difference is not significant*		

Table 2 – Breakdown of mistakes in T1 and T2

Omissions are by far the most common mistake. Unlike the other mistakes, omissions mostly involve the surrounding text, that is a reference (e.g. *increase by, drop to*, etc.) or a whole chunk of information. It is also likely that subjects sometimes use it as a strategy to catch up when they are lagging behind.

Approximations are the second most common class of errors (see Table 2). The most significant proportion of approximations is by far that for type-A numbers ($\geq$1,000), in 30% of cases in T1 and almost 60% in T2. This supports the idea that numbers saturate WM, since it shows that the more complex the number, the less able the subjects are to reproduce it correctly. This mistake probably depends in part on a conscious decision to minimize any risk of compounding the error. The interpreter is aware that s/he is missing some specific information, but since s/he cannot store all of it, s/he chooses to simplify the number.

Decimals are the other category for which approximations are relatively frequent (17% of decimals in T1, 20% in T2). This is consistent with the finding that large numbers and decimals are the most troublesome categories (see 4.1). Not only do they tend to be omitted when in clusters, but they are often wrongly interpreted.

Statistical results indicate that approximations are significantly more numerous in T2 (see Table 2). This might suggest that when notes were not allowed, WM could not rely on any help to retain the whole number and the subject was forced to give a rounded version of it.

Lexical mistakes and syntactic mistakes account for low percentages of the error scores. Only type A-numbers show a relatively high incidence of these. This supports the idea that difficult numbers to store cause more problems. The difficulty does not lie in the structure of the numeral, which is linear (from the highest to the lowest order of magnitude), but in limited WM capacity – not all the information needed to accurately reconstruct the number can be stored.

Phonological mistakes are not very significant in the present study. The significant difference between T1 and T2 for phonological mistakes (see Table 2) might depend on the fact that T2 contains more numbers which could be phonetically confused with others (e.g. 14/40, 17/70, etc.).

In the category "other mistakes", about one third of occurrences might be possibly caused by an "echo effect" (see *Evaluation procedures* above), in both T1 and T2.

3.1.4 Context

About 27% of all numbers in T1 and T2 are not in context or have a wrong reference.

For most correct numbers, the context and the reference are correctly interpreted. However, there is still a certain proportion of wrongly contextualized correct numbers (5.2% in T1 and 3% in T2), which further reduce accuracy in interpretation of numerals.

A high proportion of wrong numbers is also wrongly contextualized. This result is mainly due to omissions, which are the only category for which there are more mistakes involving the context than those involving the number alone. In T1, there is an approximately 2:1 ratio between omissions involving the whole context and those involving the number alone. In T2, the ratio is 3:1. The high proportion of omissions involving the context might be due to problems in the processing of information near the number, caused by WM shortage – particularly when notes are not used.

If omissions are not calculated, the proportion of wrong numbers in the wrong context is similar to that of correct numbers in the wrong context (from 1% to 2%). This low percentage suggests that usually subjects are able to assess whether their interpretation will be successful. Thus, when they include the context, it is in most cases correct. When they are aware that they did not understand enough to convey the right message, they omit the whole segment. On the other hand, when subjects transpose the overall meaning of a segment with an approximation of the numerical item, the context is usually correct.

In conclusion, these data suggest that numbers are not only a source of difficulty to the interpreter, but also a problem trigger for interpretation of the speech segment in which they occur.

3.1.5 Notes

Notes in T1 cover, on average, 39.4% of numbers. Right notes helping to produce right numbers make up 66% of items noted, that is 25.9% of numbers. Mistakes in the notes, leading to mistakes in production, are common (26.6% of items noted), accounting for 10.5% of the total numbers in the text. Significantly, correct notes failing to prompt the correct number in the target text account for 6.2% of notes and 2.5% of numbers. There is also a minimal incidence of items which are wrongly noted but then correctly interpreted.

When the quality of performance is compared with the quantity of notes, both correct and wrong numbers in the TS show significant correlations (about 0,650 in all cases): (i) with the total amount of numbers noted; (ii) with the amount of correctly noted numbers. The correlation is positive for correct numbers in the TS, negative for incorrect items, giving further support to the view that note-taking improves handling of numbers in SI.

It seems that subjects take more notes for “difficult” numbers. Decimals are the most frequently noted category, followed by large numbers (A) and ranges. However, not even half of the notes for these categories are correct. Small numbers (B) and dates (E), the “easiest” groups, are the categories in which least notes are taken. However, notes are probably also useful for these “easy” numbers in clusters, because they can be rapidly written down to leave room for the next number.

3.2 Perception

The questionnaires for T1 and T2 are identical. Since most of the results were similar in both cases, they were analyzed together.

– BLOCK 1: *Difficulty of the text*

Subjects judged the texts easier than average, while the numbers were considered quite difficult. This strengthens the initial assumption that most errors in the interpretations would be caused by numbers.

Another interesting result regards the difference between two specific variables, “overall speed” and “speed of parts with numbers”. Subjects indicated that the speed of parts with numbers was higher than the overall speed, which suggests that speed was thought to be one of the reasons why numbers were difficult to translate. Actually, the speed was kept constant for the whole text, or even decreased in parts with a high density of numbers, because the reader tried to articulate them as clearly as possible to reduce problems of comprehension. This result stresses the difference between perception and reality. Subjects probably thought they were lagging behind because of a sudden increase in the speed of reading, while the real source of difficulty is more likely to have been the numbers[6].

– BLOCK 2: *Influence of numbers on performance*

Answers to block 2 highlight the general tendency to see numbers as problem triggers. Subjects rated their performance in parts with numbers rather negatively, indicating that both comprehension and production of numbers and surrounding elements were poor. However, they did not seem to know why numbers caused so many problems.

– BLOCK 3: *Importance of notes*

While only 60% of subjects indicated that they usually take notes during SI, as many as 86.7% stated that they take notes for numbers. This clearly underlines the perceived importance of notes when there are numbers in the text.

4. Discussion

4.1 Error analysis

The analysis of general performance highlights that subjects’ accuracy in speech segments with numbers is lower than in those without numbers. This suggests that numbers are indeed problem triggers during SI.

---

6 “Subjective” speed is one of the subjects of Déjean Le Féal’s doctoral dissertation (1978).

The degree of accuracy for numbers is rather low (about 50% on average). Given that the specific structure of numbers is thought to increase the already high processing demands of SI (additional WM capacity to store them), the expectation was that they would also affect interpretation of the context. This was confirmed by the analysis of information items around the number, about 27% of numbers having a wrong reference or lacking the source speech context.

However, performance varies according to the type of number. Numbers with four or more digits (type A) show the highest error score (about 91%)[7], followed by decimals (59.5%), ranges (51.7%), small whole numbers (43.3%) and dates (38.1%)[8]. The most common mistakes are omissions, which involve about 30% of all numbers in both texts. This is consistent with the results of other studies on numbers in SI (Crevatin 1991; Braun and Clarici 1996).

Clusters increase the difficulty of numbers[9], even for items like small numbers which do not usually trigger mistakes. Subjects seem to rely more on note-taking when they have to deal with a dense passage, since nearly all the most frequently noted numbers in T1 are in clusters. This suggests that when the difficulty increases, the need for notes is greater. Results therefore support the hypothesis that notes may be a useful strategy in relieving the burden that numbers place on the WM.

Performance for numbers does not seem to improve when they occur more than once in the speech. Subjects sometimes even repeat the same numbers in their notes. Repeated numbers are thus treated like all the other numbers, probably because they are discarded from WM as soon as they are translated and have to be assimilated again when they recur.

Results of the error analysis are in line with the Effort Model (Gile 1995). Numbers are likely to require considerable effort to be assimilated and stored, and often exceed the attentional capacity threshold, causing errors during SI.

4.2 Comparative analysis

The overall performance is worse in T2, suggesting an effect of the independent variable – i.e. note-taking. The relative frequency of approximations for longer numbers (large whole numbers, ranges and decimals), for which notes could be expected to prove most useful, is consistent with this observation. However,

---

7 This percentage refers to "normal" large numbers, without "borderline" cases (i.e. *1,000*).

8 Percentages represent the error score for each category as a mean between performance in T1 and T2.

9 See also Pearl (1999: 20).

results for type-A numbers (> 1,000) actually show that notes are not very useful with this category (in most cases they are wrong), probably because large whole numbers are often too dense even to be successfully noted. In other words, any advantage in terms of reduced demand on WM is offset by the extra processing capacity required for note-taking. On the contrary, decimals and ranges contain too many components to be remembered correctly all the time, but few enough to be quickly written down; notes are often correct for these two categories.

An interesting remark concerns subjects who usually take notes. In both texts, but especially in T2, they have a higher mean error score than subjects who do not usually take notes. An explanation could be that the exclusion of note-taking has a more marked effect on subjects who usually rely on them. These subjects' lower accuracy in T2 might also be due to their conviction that, since they cannot take notes, they will surely perform badly. On the contrary, those who are used to interpreting without notes will find themselves more at ease in the conditions created for T2.

Another difference concerns the degree of accuracy for subjects who said that they were not affected by numbers during interpretation. These subjects have a lower error score than others in both T1 and T2, suggesting that they are reasonably able to evaluate their performance. Despite this, their performance in parts with numbers is actually worse than in parts without.

In conclusion, the study suggests that the error score for numbers increases when note-taking is not allowed. However, this seems to be true only of decimals, ranges and (to a lesser extent) small whole numbers (see Table 2). The error score increases most in terms of approximations, probably because these categories of number overload WM during interpretation of T2 and are therefore only partially reproduced.

4.3 Final remarks

When the present experiment was set up, it was attempted to limit possible flaws and create a situation as close as possible to real life, despite the need to run the study in a classroom and record shorter interpretations than would be common in actual practice.

Admittedly, the texts were constructed with the specific aim of studying numbers, the speeches were recorded, and the environment was artificial (the classroom). However, this was in part due to the logistic impossibility of having all subjects interpret a single "live" speech at the same time. Lack of sufficient booths meant that this would have been possible only by repeating the speech several times, with the risk of changing such features as speed of delivery, intonation and pronunciation of numbers.

The sample was kept small to eliminate as far as possible any variables which could have influenced the outcome of the experiment. It is composed of students, which means that results are not representative of professional interpreters. However, only English first language students with at least three years of practice were chosen, to ensure that all had a certain minimum level of experience.

The analysis considered the variable "note-taking" only in passive interpretation (i.e. into the A-language). It would therefore be interesting to study direction of interpretation as a variable in future research, or to repeat this study with active interpretation (i.e. into the B-language). Moreover, languages other than Italian and English could give different results, in relation to features such as word length (Ellis 1992) or syntax. This underlines the interest of carrying out similar experiments with other languages, to compare the results obtained, and focus more on the effect of word length in SI of numbers.

In conclusion, no valid comprehensive strategy has been identified to deal with these extremely difficult linguistic items called numbers. The only strategy that seems to work is the tried and tested method of having the boothmate write down names and figures. If the interpreter is alone, s/he can use omissions and approximations when numbers are too large or dense for notes to be useful. Note-taking is not enough to definitively solve the problem of numbers, arguably among the main features which make SI a "finite and fallible function" (Pearl 1999: 3).

References

Alessandrini M.S. (1990): "Translating numbers in consecutive interpretation: an experimental study", *The Interpreters' Newsletter* 3, pp. 77-80.

Atkinson R. (1999): "Memoria", in *Introduzione alla Psicologia*. Ed. by R. Atkinson *et al.*, Piccin, Padova, pp. 265-305.

Baddeley A. (1987): *Working memory*, New York, University Press.

Baddeley A. (1990): *Human memory: Theory and Practice*, Hove, Lawrence Erlbaum Associates.

Baddeley A. (1993): *La memoria. Come funziona e come usarla*, Bari, Laterza.

Braun S. and Clarici A. (1996): "Inaccuracy for numerals in simultaneous interpretation: neurolinguistic and neuropsychological perspectives", *The Interpreters' Newsletter* 7, pp. 85-102.

Carpenter P.A., Miyake A. and Just M.A. (1994): "Working memory constraints in comprehension: evidences from individual differences, aphasia, and aging", in *Handbook of Psycholinguistics*. Ed. by M.A. Gernsbacher, NewYork, Academic Press, pp. 1075-1122.

Craik F.I.M. and Lockhart R.S. (1972): “Levels of processing: a framework for memory research”, *Journal of Verbal Learning and Verbal Behavior* 11, pp. 671-679.

Crevatin, A. (1991): “La traduzione dei numeri in interpretazione simultanea: un contributo sperimentale”, unpublished dissertation, SSLMIT, Trieste.

Ellis N. (1992): “Linguistic relativity revisited: the bilingual word-length effect in working memory during counting, remembering numbers, and mental calculation”, in *Cognitive Processing in Bilinguals*. Ed. by R.J. Harris, Amsterdam, Elsevier Science Publishers, pp. 137-155.

Gile D. (1984a): “Des difficultés de la transmission informationnelle en interprétation simultanée”, *Babel* 28, pp. 18-25.

Gile D. (1984b): “Les noms propres en interprétation simultanée”, *Multilingua* 3/2, pp. 79-85.

Gile D. (1995): *Regards sur la recherche en interprétation de conférence*, Lille, Presses Universitaires de Lille.

Gile D. (1999): “Testing the Effort Models’ tightrope hypothesis in simultaneous interpreting: a contribution”, *Hermes* 23, pp. 153-172.

Harrington M. (1992): “Working memory capacity as a constraint on L2 development”, in *Cognitive Processing in Bilinguals*. Ed. by R.J. Harris, Amsterdam, Elsevier Science Publishers, pp. 123-135.

Hurford J.R. (1987): *Language and Number: The Emergence of a Cognitive System*, Oxford, Basil Blackwell.

Kalina S. (1992): “Discourse processing and interpreting strategies – an approach to the teaching of interpreting”, in *Teaching Translation and Interpreting. Training, Talent and Experience*. Ed. by C. Dollerup and A. Lindegaard, Amsterdam-Philadelphia, John Benjamins, pp. 251-257.

Luria A.R. (1976) *The neuropsychology of memory*, Washington, V.H. Winston & Sons.

Miller G.A. (1956): “The magical number seven, plus or minus two: some limits in our capacity for processing information”, *The Psychological Review* 63, pp. 81-97.

Pearl S. (1999): “The other three eighths and the four ‘F’s”, *The Interpreters’ Newsletter* 9, pp. 3-28.

Seleskovitch D. (1975): *Langage, langues et mémoire*, Paris, Minard Lettres Modernes.

Setton R. (1999): *Simultaneous interpretation. A cognitive-pragmatic analysis*. Amsterdam-Philadelphia, John Benjamins.