

Think-aloud protocols in translation research

Achievements, limits, future prospects*

Silvia Bernardini

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

Over the last decade, Think-aloud Protocols (TAPs) have been used extensively in process-oriented Translation Studies (TS). The serious questions regarding the experimental validity of this research methodology when applied to translation have nonetheless often remained unspoken. This paper surveys the breakthroughs as well as the limits of the growing body of literature dealing with TAPs in TS, points at the necessity to take issues of experimental, theoretical and environmental validity more seriously, and offers suggestions for improvements. The claim is that the risks involved in the adoption of a lax experimental methodology in TAP studies, often underestimated in the past, may invalidate not only the results obtained in the single projects, but, crucially, the method as a whole.

Keywords: process-oriented TS, concurrent verbalisation, strategies, translation units, affective involvement, automaticity, professionalism, routineness

1. Introduction

Interest in empirical research into the translation process has grown substantially in the last decade, driven by the idea that what goes on in translators' heads while they are translating (versus what scholars had claimed might go on) is crucial to an understanding of translation, and is not derivable solely from an analysis of the final product, the translated text. The latter provides an incomplete and often misleading way into the translation process, hiding both successful strategies and problems.

A number of attempts have been made at accessing the translator's mind, with varying degrees of success. One such attempt has been to ask the translators themselves to reveal their mental processes in real time while carrying out a translation task. Such a method of data collection, known as "thinking aloud", has been imported from the cognitive sciences and applied to translation research, often with little reflection on the consequences inherent in the application of the approach to the new research framework. Theoretical justifications have been imported without questioning their applicability to the new settings, and the validity of the method as a whole has been assumed rather than proved.

We now seem to be witnessing a decrease of interest in Think-aloud Protocols (TAPs), after the enthusiastic efforts of the past ten to fifteen years. A first phase, with identifiable characteristics (anecdotal attitudes, methodological compromises, proliferation of classification schemes) is approaching its conclusion; the birth of a second, more mature testing phase will very much depend on the will of researchers working within this framework to move on to a more rigorous experimental methodology. The aim of this paper is primarily to provide a survey of TAP-based research so far, with its achievements and limits, and then to look ahead, at what a potential second phase might look like, were it ever to see the light.

2. TAPs: A survey

2.1 The theoretical grounding: Think-aloud in psychology and cognitive science

The theoretical framework for TAP experiments is provided mainly by the work of Ericsson and Simon¹ (esp. 1993 (1984)). According to their model ("human cognition is information processing"), information is kept in different memory stores, with varying access and storage capabilities: whereas short-term memory (STM) is characterised by easy access and severely limited storage capacity, long-term memory (LTM) is characterised by more difficult access and larger storage capacity. Only information present in STM, that is, information which is being heeded by the subject (static and conscious "knowledge states" rather than dynamic and unconscious cognitive processes), can be directly accessed and reported. It follows that the cognitive processes these knowledge states are inputs and outputs to, as well as information that is not currently being heeded,

cannot be reported but must be inferred by the analyst on the basis of the verbalisations. A further assumption of this model is that, for verbally encoded information, which can be reported in the same form as the one in which it was heeded, the verbalisation does not interfere with the cognitive process itself, the only effect of thinking-aloud being to slow down the performance. The implications of this model are manifold. Here we shall only consider some of the more closely relevant to our discussion.

First of all, only *concurrent* verbalisation of thoughts can be claimed to exhaustively reflect the mental states of a subject carrying out a relatively long task (let us say, longer than ten seconds). On completion of such “long” tasks, part of the information moves on to LTM, leaving behind retrieval cues only in STM: in such cases, *post hoc* verbalisation has been found to be difficult and often incomplete (Ericsson and Simon 1993 (1984): xvi). Moreover, ruling out the possibility that a subject is interpreting her/his own thought processes or even generating them anew, instead of retrieving them from LTM, can be extremely problematic under these circumstances, thus making results virtually uninterpretable.

Secondly, in order to make sure that the reports actually reflect mental states without distorting them, it is important that subjects do not feel they are taking part in social interaction: albeit obviously a much more natural situation, conversation involves reworking thoughts to make them conform to socially established norms, a process which might sensibly alter the information attended to. The interaction between subject and experimenter (or between subjects) should therefore be avoided or at least reduced to a minimum. A plea for environmental validity is unsustainable in this case: TAPs are either strictly monological or *not* TAPs at all.

Thirdly, practice and experience may affect the amount of processing carried out in STM, so that fewer mental states will be available for verbalisation to subjects experienced in a task. This process, known as “automation”, is explained thus:

Before overlearning has occurred, processes have to be interpreted, with substantial feedback from intermediate processing stages in STM. Overlearning amounts to compiling these processes, so that fewer tests are performed when they are being executed, hence less information is stored at intermediate stages in STM. (Ericsson and Simon 1993 (1984): 127)

Automatic processes are therefore faster and more efficient than processes which are under conscious control. However, they are also less flexible and more difficult to modify at need.

Finally, this model takes into account the effects of personality and personal history over the data collected. The amount of relevant information held in LTM cannot possibly be controlled for, as an experimental situation would require, nor is it possible to control for the amount of knowledge reported on in relation to the performance given. Even though this is a well-known problem in cognitive science research, TAPs are particularly sensitive to it, suggesting that it would be advisable to try and limit the effects of individual differences as much as possible, and to take them into account during the analysis, in order to obtain more reliable and generalisable data.

Summarising:

- Concurrent verbalisation, or thinking aloud, provides data on the mental states heeded by individuals carrying out a task.
- From these states it is then possible to derive information about the relevant mental processes.
- Under the right circumstances (verbally encoded information, no social interaction, no interferences, no instruction to *analyse* thoughts), verbalising is assumed not to interfere with the mental processes and to provide a faithful account of the mental states occurring between them.
- The generalisability and the relevance of the data obtained through TAPs, however, is more difficult to assess, and cannot be guaranteed by the model itself.

2.2 TAPs in translation studies

2.2.1 *Achievements*

Viewing translation mainly as a problem-solving process, some TS scholars have put forward the suggestion that it should be possible to study it by means of TAPs, and have set up experiments to test this hypothesis. The varying interests and backgrounds of those involved have resulted in a large variety of approaches, which can only briefly and unsystematically be surveyed here. In this sub-section we focus on achievements, in the following we look at some inadequacies observed and suggest ways of proceeding in the future, if TAPs are to go on having a role in TS.

Most early TAP studies were conducted with foreign language learners or translator trainees. This was mainly due to the availability of subjects and to the pedagogic concerns of the experimenters. However, the hypothesis was also put forward that the verbalisations produced by professionals would be less informative than those produced by non-professionals, due to their more

“automatised” processing style. We shall have more to say about this issue. In the meantime, let us consider one of the major early concerns of researchers working within this paradigm, namely the analysis of “translation strategies”.² We shall then move on to consider more recent foci of attention, such as attention units, automaticity of processing and affective factors.

2.2.1.1 Strategies. The researchers whose work is surveyed in this sub-section have either avoided a terminological discussion of the term strategy (for which alternative definitions abound in linguistics) and used the term in a rather undefined, everyday sense, or endorsed the definition provided by Lörcher (who, in turn, adapts Færch and Kasper’s (1983) definition), according to which a translation strategy is

a potentially conscious procedure for the solution of a problem which an individual is faced with when translating a text segment from one language into another. (Lörcher 1991:76)

Lörcher himself (1986 and 1991) reports on a comparatively large TAP study, in which 48 German learners of English as a foreign language produced 52 translations either into English or into German. They were asked to produce a spoken translation of a written text while thinking aloud and were not allowed to use dictionaries (this was meant to ensure that a larger number of problem-solving processes would be present in the protocols).³ The transcripts of the sessions were then analysed and a number of “translation strategies” were recognised.

According to Lörcher each strategy is formed of a sequence of core elements which can be combined in different ways. A translation process, in turn, is formed of a series of strategies, which can also be combined in different ways. The general conclusions drawn from this study are that:

- TAPs seem to provide reliable and useful data, provided that the analyst interprets them in a systematic and “methodologically controlled” way;
- Despite individual differences and the inherent variability of the translation process, there are regularities that point at the possibility of establishing taxonomies of translation strategies;
- No evidence is found of translation-specific strategies: instead, general text-processing strategies are adapted to the specific task at hand.
- As a side-effect of thinking aloud, the learners’ capacity for problem-solving seems to increase, suggesting that this research methodology might also have important pedagogic applications.

Krings (1986) reports on an experiment in which eight German learners of French as a foreign language translated a text either into or out of the mother tongue. The main focus of attention here is the identification of translation problems and translation strategies on the basis of TAPs. With regards to the former, Krings offers the following list of “problem indicators”:

- The subjects’ explicit statement of problems
- The use of reference books
- The underlining of source-language text passages
- The semantic analysis of source-language text items
- Hesitation phenomena in the search for potential replacements
- Competing potential replacements
- The monitoring of potential replacements
- Specific translation principles
- The modification of written target-language texts
- The assessment of the quality of the chosen translation
- Paralinguistic or non-linguistic features (Krings 1986:267)

With regards to translation strategies that subjects resort to when automatic processing breaks down, Krings suggests that these can be classified as strategies of *comprehension* (inferencing and use of reference works), *equivalent retrieval* (especially interlingual and intralingual associations), *equivalent monitoring* (such as comparing source text (ST) and target text (TT)), *decision-making* (choosing between two competing solutions) and *reduction* (for instance of marked or metaphorical text portions).

A more complex classification of strategies is proposed by Gerloff (1986: 252ff.) who, in her methodologically oriented paper on TAP studies, describes “text-processing strategies” as “any metalinguistic or metacognitive comments made or specific problem-solving behaviours affected, during the decoding and rendering of the translation text”.⁴ The categories she identifies are *problem identification*, *linguistic analysis*, *storage and retrieval*, *general search and selection*, *text inferencing and reasoning*, *text contextualisation*, and *task monitoring*.

In their discussion of the use of lexical search strategies, Mondhal and Jensen (1996) distinguish *production* from *evaluation* strategies. The former are further subdivided into *achievement* strategies and *reduction* strategies (also discussed by Chesterman 1998). Among achievement strategies, which are characterised by an attempt to remain close to the ST, are *spontaneous association* and *reformulation*. Among reduction strategies, which are characterised by their inherently remedial nature, are avoidance and unmarked rendering of

marked items. Finally, evaluation strategies involve, for instance, reflecting on the adequacy and acceptability of translation replacements.

Séguinot (1996) reports on another non-comparative study involving, this time, two professional translators working together at the same task. The underlying assumption in this case is that this everyday setting (the subjects are used to working as a team) would increase the environmental validity of the experiment, without limiting the experimental validity of the results obtained. As a result of this study four types of translation strategies are identified as being typical of “professional” translation, namely *interpersonal* strategies (brainstorming, correction, phatic function), *search* strategies (dictionaries, world knowledge, words), *inferencing* strategies (rereading ST and TT, consulting) and *monitoring* strategies (rereading ST and TT, consulting, comparing units). This translating process is further described as “iterative”, proceeding in a non-linear fashion and operating on the basis of sentence-level “translation units”, which are, however, often interrupted by pauses and hesitations.

None of the studies described so far attempt to systematically compare strategies across two groups of subjects. However, finding out what it is that distinguishes professional from non-professional (student or lay-person) behaviour has always been a major concern of researchers in process-oriented translation studies. One way of investigating this issue has been to compare the performance of two groups on the same task.

An investigation along these lines is described by Lörscher (1996) who, building on his previous studies (mentioned above), compares the strategies adopted by professional and non-professional translators (foreign-language students). He points out that, although the two groups do not differ qualitatively in their use of translation strategies, they do differ quantitatively, i.e. in the distribution and frequency of the various strategies employed. Furthermore, differences can be detected in the orientation of the approach (towards form in the case of non-professionals, towards sense in the case of professionals), in the size of translation units, in the amount of monitoring of the TT, and lastly in the attention devoted to stylistic and typological adequacy (greater for professionals in all cases).

In the study reported in Séguinot (1991), two similar texts were translated by students of translation at different levels of proficiency (at the beginning and at the end of their courses in specialised translation). French and English mother-tongue speakers translated two advertisements from French into English. The main research focus was once again on the — rather loosely defined — notion of strategies. The author suggests that native speakers of

English (as well as better students; the two categories are unfortunately not distinguished clearly) translating into their mother tongue show more efficient monitoring⁵ and revising strategies, and work more at the textual level, whereas non-native speakers seem to rely more on learned principles and lexical-level processes.

Building on Lörcher's definition, Jääskeläinen (1993) proposes a classification of translation strategies distinguishing between global and local strategies, the former applying to the whole task (considerations about style, readership, etc.), the latter to specific items (i.e. lexical searches). On the basis of this distinction, she is able to claim that global strategies are much more frequently used by professionals and semi-professionals (translator trainees) than by non-professionals in her study. After making a plan, the former appear to follow it systematically through the task, whereas the latter seem to proceed in a more haphazard way. A similar conclusion is reached by Tirkkonen-Condit (1992) who compares the performance of a professional and of a non-professional translator on the same task, and finds that the professional relies more on textual and linguistic knowledge, whereas the non-professional works with smaller units and relies more on extra-textual knowledge. Even though this may be a consequence of the latter's greater familiarity with the subject area as much as a consequence of her lesser familiarity with the translating task, the issue of the interrelation of linguistic/textual and extra-textual knowledge would appear to be a promising topic for further study.

Let me try and chart very briefly the ground we have covered so far. A number of TAP studies, especially early ones, have been concerned with the recognition and classification of translation strategies and with the detection of differences between professional and non-professional strategies. A number of classificatory schemes have been provided, adopting labels like *global/local*, *reduction/achievement*, *monitoring* (and *revising*), *search*, *comprehension*, *equivalent retrieval*, *decision making* and so on. Besides, it has been suggested that the performance of professionals differs from that of non-professionals with regards to both the quantity and the quality (orientation) of the strategies adopted.

Let us now turn to consider three other issues which have been investigated by means of TAPs, namely *translation* (or *attention*) *units* (2.2.1.2), *automaticity* (2.2.1.3) and *affective factors* (2.2.1.4).

2.2.1.2 *Translation units.* Translation (or attention) units are defined as

Those instances in the translation process in which the translator's 'unmarked processing' is interrupted by shifting the focus of attention onto particular task-relevant aspects (Jääskeläinen 1990: 173, cited in Jääskeläinen 1993: 102)

"Unmarked processing" here refers to unproblematic sections of the protocols in which a subject verbalises fluently while reading or writing. Marked processing, then, begins with a problem indicator and ends with a solution to the problem or an indication that the problem is temporarily abandoned.

A "unit of analysis" coding is described by Gerloff (1986), who identifies seven levels of analysis, going from morpheme or syllabic unit to discourse. According to most researchers, the length of translation units is an indication of proficiency, with professional translators working with larger units (sentence and discourse, or group) and moving more comfortably between different unit levels. Clearly, this does not mean that a professional translator never stops midway through a sentence, but only that the sentence is processed as a unit, with more local problems tackled on the way (Séguinot 1996). The suggestion can be put forward, therefore, that attention units are better defined in hierarchical rather than sequential terms, with smaller units being processed within larger units. The search for a term or collocation may be embedded in the processing of a whole sentence, without implying a "word unit" or "phrase unit" analysis.

2.2.1.3 *Automaticity.* Insofar as automaticity of processing is believed to result from experience and proficiency in a task (Ericsson and Simon 1993 (1984); Toury 1988), it is not surprising that researchers have tried to determine whether the performance of professionals is recognizably more automatic than that of non-professionals. In order to do so, they have analysed the amount of marked processing in professionals' and non-professionals' TAPs. The most straightforward hypothesis (that professionals verbalise less than non-professionals) is not endorsed by Jääskeläinen and Tirkkonen-Condit (1991) or by Jääskeläinen (1996 and 1997), who make a distinction between routine and non-routine tasks/situations. In the former, professionals do tend to verbalise less than non-professionals, whereas in the latter the amount of verbalisation is not necessarily smaller. Besides, the nature of the verbalisations tends to differ as well. The explanation offered is that "while some processes become automated, other processes are evoked into consciousness, i.e. the translator becomes sensitised to new kinds of problems" (Jääskeläinen and Tirkkonen-Condit 1991: 105).

This conclusion is supported by the finding that semi-professionals (translator trainees) show more extensive processing than both professionals and non-professionals (Jääskeläinen 1997). This may be because they are aware of the problems involved but have not yet automatised the necessary problem-solving strategies. Equally, professionals are assumed to be better at recognising the need to resort to non-automatic, controlled processes (i.e. problem recognition) than non-professionals. Automatic processes, as we saw above, are typically very efficient but not very flexible, so that there is the danger (pointed out by Wilss 1994: 144) “of problems being forced into a certain structure, because it is believed to offer a solution”. A typical example of this danger would be, for instance, the difficulty experienced by non-professionals in overruling automatic lexical associations (Ivanova 1998: 102), or “false-friends”, a process requiring high control.

2.2.1.4 *Affective factors.* Leaving aside cognitive issues for a moment, we shall now turn briefly to consider affective factors in translation. These have been investigated, among others, by Kussmaul (1991); Tirkkonen-Condit (1997); Laukkanen (1996); Tirkkonen-Condit and Laukkanen (1996); and Jääskeläinen (1997). These researchers agree that affective factors such as involvement with the translation task, a relaxed atmosphere and self-confidence correlate positively with what they regard as ‘successful performance’. In routine tasks, where these three elements are likely to be present, subjects are found to produce better translations than in non-routine situations, where they tend to stick more to the ST and avoid reduction strategies as much as possible (Laukkanen 1996: 266). This finding should be carefully evaluated in relation to the validity of the results obtained in experimental conditions, where affective factors are likely to influence the results obtained in unpredictable ways.

A further way into the translation process is offered by the evaluations (of self, task, ST, TT) verbalised by the subjects. According to Tirkkonen-Condit (1997: 83), there is a quantitative as well as qualitative difference between professionals and non-professionals in these regards, due to the fact that “consciousness of the motivations and rationale of one’s own performance seems to grow with translational experience”.

2.2.1.5 *Conclusion.* As can be gleaned from the above discussion, the large amounts of data about the cognitive and affective factors involved in the translation process which could be collected by means of TAPs have no doubt favoured a more empirical approach to the study of translation, and highlighted

the limits of purely speculative models. However, the value of the classifications and observations made so far appears to be still limited. The most obvious limitation of this body of research is the fact that researchers have tended to proceed in a rather anecdotal and unsystematic way in their studies and reports, generally not providing a theoretical justification for the classification schemes they construct and very little information about their methods and findings. The next section focuses on some methodological limits of the studies discussed above, and points at possible ways of overcoming them.

2.2.2 *Limits*

As just mentioned, a major problem with TAP studies has been the lack of an established research paradigm, resulting in a rather loose treatment of methodological issues (research design, data analysis, research report) and in a host of studies setting their own categorisations in a theoretical void. Most of the research reports we have been concerned with so far describe the research design summarily, present findings in an anecdotal fashion, do not provide any statistical analysis of their data (and sometimes not even the data themselves) and leave central theoretical assumptions unexplained. The reader thus finds it difficult to assess the validity of the results obtained. Besides, experimental conditions are often relaxed without as much as a hint that the researcher is aware of possible consequences.

As an example, consider the discussion of “routine vs. non-routine task”, which is a central issue in Laukkanen (1996:257). Here a routine situation is defined as “the kind of task that is familiar to the subject from his/her daily work”, whereas a non-routine situation is defined as “practically any assignment that the subject is not very familiar with”. This definition seems to be inherently ambiguous, as no explanation is given of what familiarity and non-familiarity imply. Are we dealing with content or form? Are we going to label as “routine” a text that deals with a familiar subject, but belongs to a text type the subject has never translated before or rather a familiar text type with unfamiliar subject matter? Or both? Research in applied linguistics, for instance, has alternatively suggested that unfamiliar formal schemata are more likely to result in comprehension problems than unfamiliar content schemata (Oller 1995), and vice versa (Floyd and Carrell 1987). Another aspect to consider is the difficulty of assessing the comparability of texts belonging to different text types. The hypothesis that translators behave in different ways in routine and non-routine situations can only be tested by trying to control all variables apart from familiarity. This is a dismal endeavour, since the two texts might differ

along an almost infinite number of dimensions, the most obviously relevant being difficulty. Though no claim is made that these problems could easily be solved — they are certainly very challenging — this very interesting study would certainly have gained from a more careful treatment of terminological and theoretical issues.

Another problem with most of the studies dealt with here is the excessive reliance on between-subject designs, used to compare the performance of professionals with that of semi-professionals and/or non-professionals. This is a very controversial design, which is nonetheless normally posited without further discussion. Even if we had an uncontroversial way of determining what professionalism involves — the current practice of resorting to measures such as years of experience and official certifications is not fully satisfactory, because it is external — we would still have to take into account individual differences in the ability or disposition to verbalise, interests, involvement with the task, variable effects of the experimental conditions and so on. This preoccupation is shared, among others, by Krings (1987:167) who claims that “individual differences between subjects with regard to their willingness to verbalise might be greater than Ericsson and Simon seem to assume”.

Lastly, it is necessary to mention a general methodological problem with TAP studies in translation research. As a method of data collection in cognitive science, TAPs are recognised as valid only inasmuch as they have been collected under very rigorous experimental conditions. When TAPs are used in translation research, these conditions are often relaxed. Although this is partly due to the justified need to preserve environmental validity, this tendency should be checked, as it may result in the invalidation of the results obtained. Two examples will illustrate the point.

1. According to Ericsson and Simon's (1993 (1984)) theoretical framework, social interaction during the verbalisation should be avoided at all costs, as the need to communicate in a structured way is likely to interfere with the task being carried out in unpredictable and uncontrollable ways. However, a number of studies have investigated *dialogue* TAPs (Séguinot 1996; Kussmaul 1991) and claimed for them the same empirical validity as for *monologue* TAPs.
2. It has been claimed (Færch and Kasper 1987b:15) that “simultaneous introspection ... in terms of concurrent talking or thinking aloud or verbalisation of specific cognitions, presupposes that the modality of language use is not itself oral-productive”. This is because two concurrent tasks of the same kind may interfere with each other in ways still unpredictable at the present

stage of research. However, the influential study conducted by Lörscher (1991) required subjects to think aloud while carrying out a written-to-spoken translation task.⁶

Summarising the points made in this sub-section, there is a lot of work to do in process-oriented translation studies in order to raise the standard of research in this area to the levels required by empirical linguistics research. A crucial problem to tackle at present is the establishment of a more rigorous experimental methodology, which takes into account issues of experimental and theoretical as well as environmental validity, both in the collection and in the analysis of data.

No claim is made here that these problems can be easily solved. Most of them follow almost necessarily from the need to adapt a research methodology developed for relatively simple and straightforward cognitive processes, to the formidable task of studying the translating process. Yet it is the opinion of the present writer that it is too early to give up this model (not least because we do not have many alternative ones to choose from), as TAPs have still not proved all their worth in translation research. In the next and final section we shall discuss a few points which might warrant further reflection, if a second and more mature phase for TAP-based research in translation were to begin.

3. Looking ahead: Issues and suggestions for TAPs in TS⁷

3.1 Issues and aims: Routineness, professionalism and translation into L₂

As discussed in the previous section, a number of studies have been concerned with the effects of routine/non-routine task conditions. One of the conclusions drawn from these studies has been that routine conditions seem to result in higher levels of automatic processing by professional translators, whereas non-routine conditions may prompt a less automatic (and in general less “professional”) behaviour. As we have seen, however, the notion of “routine vs. non-routine condition” is highly controversial. So far, researchers have been contented with positing routineness on the basis of external factors such as experience and similarity among text-types, not a fully satisfactory procedure. A difference between “routine” and “non-routine” conditions can only be determined unambiguously by relying on internal factors (i.e. *if* a translator behaves in different ways in the two conditions, *then* we can suggest that there is a difference). We are faced here with a problem of circularity: we need to

know what routine conditions are in order to investigate how translators behave in such conditions, but the only satisfactory way of determining what is routine is to rely on the behaviour of translators.

A way of bypassing this apparent dead-end is to find a non-gradable routine/non-routine opposition. One such condition, and a promising one, has been found to be “direction of translation”, as opposed to “text-type”. Most translators seem to have a preferred direction of translation,⁸ usually from their second language (L_2) into their first language (L_1), and one would not see why this should not be considered as a routine opposition, involving greater or lesser familiarity. Though I would not go as far as to claim that 100% of translators have zero experience of translating into the L_2 , it is certainly not difficult to find subjects for whom translations into the L_2 constitute less than 10 percent of their professional workload.

A further advantage of this design would be to simulate a professionalism distinction by means of a within-subject design. Professional behaviour is traditionally understood to be definable only on external grounds, by way of reference to the years of experience a translator has or to similar measures. This has resulted in the circularity just mentioned, and worse still, in a host of between-subject designs in which trainees or lay-people and “professionals” were asked to translate the same text. As suggested above (and by Krings 1994), this seemingly straightforward decision has heavy implications for the interpretability of the results obtained, due to the difficulty of controlling subject variables in studies usually involving, for technical reasons, very few subjects (hardly ever more than ten, often as few as two).

The assumptions underlying the suggestions just put forward are far from uncontroversial: the processes involved in translation in the two directions are as yet so little studied that one cannot rule out the possibility that the two might differ on many grounds that have nothing to do with routineness, and, indirectly, with professionalism, the features we are trying to evaluate. However, it seems to be more profitable to arrive at results that cannot be interpreted unambiguously, rather than at results that cannot be interpreted at all: once we have solid evidence that there are differences in the behaviour of translators in the two conditions, further studies can be devised that disentangle the two. For the moment, there seems to be comforting, if limited, evidence that the two factors (routineness/professionalism and translation into the L_1) might be related. To give just one example, translators working out of the mother tongue have been claimed to employ fewer output monitoring strategies — a point made by both Séguinot (1991) and Campbell (1998) — than translators

working into it. Insofar as the same suggestion is put forward by Lörcher (1996) with regards to professional vs. non-professional behaviour, the hypothesis that varying the task conditions as suggested here may simulate both a professional/non-professional distinction, and a “direction of translation” distinction, seems to gain support. Further study is clearly needed to investigate more closely the interrelations between the two variables, which are here merely hypothesised to be related.

Summarising the point made in this sub-section, recommendations for the adoption of within-subject designs stem from considerations about the idiosyncratic nature of the translation process, which cannot be reduced to a series of predictable and formalisable problem-solving steps (as in most of the studies reported by Ericsson and Simon 1993 (1984)), and which is heavily influenced by individual cognitive and affective differences. In order to neutralise the effect of these differences, the performance of the same subject can be investigated across different task conditions, on the assumption, discussed in Section 2.2.2 above, that more signs of professionalism will show in the familiar than in the unfamiliar condition. The learning effect that might result from carrying out the two translation tasks in a sequence is judged to be negligible if compared with the inherent advantages.

3.2 Task administration: Controlling variables and environmental factors

The design and administration of a TAP study is meant to be a very rigorous task: the model itself is unfortunately not robust enough to guarantee the validity of results if the methodological requirements are not met. At the level of experimental design, meeting these requirements would involve, among other things, controlling both subject and task variables as closely as possible, and setting up the least invasive environmental conditions allowed by the experimental need to record the subjects’ verbalisations. In a within-subject design intended to study professionalism, this would involve among many other requirements:

- Having at least four subjects, two for each direction, matched on as many grounds as possible (age, geographical variety of language spoken, professional background, etc.) to ensure that differences observed in the translation processes of individual subjects translating into or out of their mother tongue are not caused by idiosyncrasies, relative difficulty of the two texts, or biases due to specific characteristics of the source/target language.

- Administering four tasks, two warming-up tasks and two experimental tasks, chosen and ordered so as to perform the required function with as little a learning effect as possible.
- Avoiding almost any form of social interaction between subject and experimenter (apart from reminders to verbalise), and renouncing the wealth of information provided by video-recording so as to check the well-known tendency of subjects to monitor their verbal performance more carefully in this condition. Rather, alternative, much less invasive techniques could be used, such as eye-movement tracking and sound recording. It is also possible to write a simple macro to instruct the computer to record every single keyboard stroke and mouse-click performed by the subject. These can then be replayed in the same order and with the same timing, allowing the researcher to observe corrections, hesitations, movements backwards and forwards through the text, and so on.⁹

3.3 Transcription, coding, analysis

TAP studies are very labour-intensive, requiring experimenters, once they have designed the experiment and carried it out, to transcribe and code the transcripts appropriately before they can proceed with the analysis. This intermediate phase is time-consuming and does not appear at first to be particularly rewarding. For this reason, there seems to be a tendency for researchers to transcribe quickly, and then proceed swiftly to a coding of the most obvious features relevant to their hypotheses. This is an understandable but unfortunate practice. The obvious methodological problem here is that such an unsystematic collection of process indicators risks introducing a strong, albeit involuntary, bias. The human eye is not always good at recognising patterns, and is even less so when there might be a chance for such patterns to invalidate a cherished hypothesis. Because of the extreme wealth of information available, a TAP can end up supporting virtually any claim, if a selective, unconstrained coding procedure is applied to it.

Whilst not claiming that the human bias can be, or indeed should be fully removed from any experimental situation, I would like to suggest that it might be possible to limit its negative effects by stating, prior to coding and even prior to transcribing:

1. What are the features of the translation process one is trying to infer from observation of a TAP;

2. What indicators are likely to signal such features;
3. What values should be recorded for each of these indicators.

To give an example, pauses have been suggested in the literature to signal that cognitive processing is going on, for which reason the subject is temporarily unable to verbalise. Let us hypothesise therefore that we are interested in recording pauses. We should subsequently decide what features of pauses are likely to be relevant. *Length* is probably one such feature. *Frequency* is another good candidate, together with *reason for pausing*. With reference to the latter, one should also decide what level of specificity is required by the experiment. Summarising:

- List of process indicators
 - Pauses
 - The subject is thinking
 - The subject is typing
 - The subject is looking up references
 - Source context references
 - Dictionary references
 - Monolingual dictionary
 - Bilingual dictionary
- ...

Other entries within the process indicators list might include *intonation profiles*, *paralinguistic features*, *emphasis* and so on. Suggestions can be derived from the available literature, where one can find numerous inspiring hypotheses. Table 1 reports some of the features discussed in this paper, together with the relevant process indicators identified in the literature.

Clearly, these suggestions are not meant to guarantee perfect retrieval of process indicators, hardly a feasible endeavour, but rather to increase the chances of seeing patterns that criss-cross the protocol, and being consistent with each other lend strong support to one conclusion rather than another. Transcription, coding and analysis of protocols thus attempt to account as systematically as possible for a number of clearly stated priorities, by sticking to a thought-out project that is formulated early on in the experiment and remains constant throughout the whole process. The methodological value of such a procedure over the more or less conscious picking out of substantiating examples is obvious. But there are two further advantages. First of all, a more systematic retrieval of process indicators can provide the raw data for factorial

Table 1. Summary of the features and process indicators discussed in the literature

Strategies	Comprehension (inferencing and dictionary use) Equivalent retrieval (collocation, association)
Krings 1986	Equivalent monitoring (choosing equivalents)
Gerloff 1986	Reduction (simplification, unmarking, avoidance) Linguistic analysis (syntactic, grammatical, lexical) Inferencing and reasoning (world knowledge, experience) Contextualization (restates information, uses context) Editing (correction, congruity assessment, changes to TT etc.)
Evaluations	Of ST Of equivalents
Laukkanen 1996	General evaluation/comments Of translation performance Of reference material
Involvement markers	1st person references References to mental processes (<i>I think</i>)
Östman 1986, cited in Jääskeläinen 1997	Vagueness (hedges, empty pronouns) Monitoring information flow (<i>anyway, right, then</i>) Emphatic particles (<i>really</i>) Precision (examples, analogies) Evaluation of ST, TT, reference material, self
Length of units of analysis	Morpheme or syllable Word Phrase
Gerloff 1987	Clause Sentence Discourse Group
Translation maxims	Does the informant read the ST? Does s/he note down potential problems (mentally or on paper)?
Mondahl/Jensen 1996 Königs/Kaufmann 1996	Is evidence of macrocontextualization present (text awareness)? Is audience considered? Is verbatim translation the goal? Is the procedure linear or circular? Is the briefing influential?

analyses along the lines described by Biber (1988) in relation to text-type classification. With regard to our potential experiment looking at routineness and translation into/out of the mother tongue, it might be possible to find

correlations among parameters, which could then be grouped together and interpreted functionally. We would thus be able to develop an internal measure of routineness and, potentially, gain insights into the nature of “translation into L₂” competence (Campbell 1998). Furthermore, a clear statement of features, process indicators and values would substantially increase the value of individual pieces of research, providing interested readers with the information needed to assess and reuse the experimental design and results obtained, and possibly even duplicate the experiment with different subjects, language pairs, tasks, etc.

A final, related suggestion regarding reusability relates to the transcription and coding standards adopted. To the best of my knowledge, there is no standard way of compiling TAPs, with the consequence that comparisons and exchange of transcripts are seriously hindered. A very welcome improvement in this area would be the development of coding procedures adopting standardised mark-up languages developed for the encoding and exchange of electronic texts, such as XML. Apart from the obvious advantage of favouring access and comparability, these suggestions are aimed at making transcripts browsable with software programmes developed for corpus studies, which would substantially ease the analysis. Finally, the transcripts would also become available to form the core of a think-aloud corpus for the computational analysis of translation processes, undoubtedly a very promising resource.

4. Conclusion

The aim of this paper has been to discuss some methodological issues relating to the use of Think-aloud Protocols in process-oriented translation studies. This research methodology has been shown to provide a very promising framework for the investigation of certain cognitive aspects of translation, a field of study that could so far only be tackled speculatively. In the last few years substantial effort has been put into this area of research, resulting in a large amount of valuable insights about the cognitive and affective factors involved in translation. At this early stage of research, the data have been mainly used in a rather informal way, as a source of suggestions and examples about the behaviour of translators: their strategies, affective involvement, units of analysis, evaluations, translation maxims and so on. The ultimate goal of this work has obviously been to shed light on the characteristics of successful translation processes in terms of their underlying constituents. For this reason, the main focus of attention of researchers has been the comparison between producers of “good”

and “bad” translation, on the assumption that the quality of the products might correlate with some features of the processes.

There is nothing inherently wrong with this approach. However, now that experience with empirical translation studies has started to pile up, and a substantial number of “informal” hypotheses have been made, it would seem to be time for researchers in the field to start questioning the methodological assumptions of their work more systematically. It is time, in other words, to check the validity of these informal hypotheses by means of more controlled experimental designs and systematic methods of data coding and analysis, accurately reported upon to favour the interpretability and reusability of the data and results provided. Suggestions have been made in this paper with reference to all these aspects of TAP research.

Apart from the obvious necessity to adopt a scientifically sounder methodology of data collection, analysis and report, the way ahead in process-oriented translation studies would appear to involve the development of a relatively uncontroversial classification of process indicators. Such a classification could limit the proliferation of terminological distinctions in the literature, and provide researchers with an instrument for the systematic analysis and description of TAPs. Presently, these seem to be necessary steps if the discipline is to proceed beyond the somewhat rudimentary stages this paper has been concerned with.

Notes

* In the next issue of *Target*, an annotated bibliography of think-aloud protocols studies into translating compiled by Riita Jääskeläinen will be published. (The Editors)

1. A discussion of think-aloud as a method of data collection exceeds the scope of this paper: the interested reader is referred to the extensive discussion and reference list provided by Ericsson and Simon (1993 (1984)).
2. The choice to classify studies according to their main object of analysis constitutes an attempt to provide as systematic a survey as possible of a body of research which is unfortunately still rather anecdotal in nature.
3. The environmental validity of this decision seems dubious, the extra processes triggered by the absence of reference tools being an obvious effect of the experimental condition, of arguably little descriptive value. Moreover, an important set of strategies (those involving the use of reference literature) remain unaccounted for.
4. It should be pointed out that Gerloff is adopting, somewhat misleadingly, a very broad notion of “strategy” in which problem indicators (cf. Krings, above), cognitive factors and

affective factors are conflated. So, for instance, “laughing” and “addressing the experimenter” are listed under the heading “text-processing strategies”.

5. Similarly, Toury (1986) suggests that expert translators (if operating in cultures which stigmatise transfer) show a greater incidence of strategies involving the monitoring of positive and negative discourse transfer than less expert translators. The assimilation and operation of norms as an internal monitoring device would seem to be one of the research areas where the TAP methodology might prove particularly useful.

6. This procedure also begs the question of the relationship between the processes involved in oral vs. written translation. In this study it is assumed that the difference between the two is not relevant (but see Toury (1991) for a different viewpoint).

7. The following section is based on a pilot study carried out at the Advanced School of Modern Languages for Interpreters and Translators of the University of Bologna at Forlì. My thanks to those colleagues who kindly accepted to take part.

8. If not stated otherwise, claims relating to the characteristics of the translation market refer to the Italian situation only.

9. I did not use any commercially available version of this simple piece of software, though there may very well be some. For the pilot study I conducted I opted for a relatively simple *Windows* macro that my computing officer offered to write.

References

- Biber, Douglas. 1988. *Variation across speech and writing*. New York: Cambridge University Press.
- Campbell, Stuart. 1998. *Translation into the second language*. London and New York: Longman.
- Chesterman, Andrew. 1998. “Communication strategies, learning strategies and translation strategies”. *Malmkjær* 1998: 135–143.
- Ericsson, K. Anders and Herbert A. Simon. 1993 (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Færch, Claus and Gabriele Kasper. 1983. “On identifying communication strategies in interlanguage production”. Claus Færch and Gabriele Kasper, eds. *Strategies in interlanguage communication*. London: Longman, 1983. 210–238.
- Færch, Claus and Gabriele Kasper, eds. 1987a. *Introspection in second language research*. Clevedon: Multilingual Matters.
- Færch, Claus and Gabriele Kasper. 1987b. “From product to process: Introspective methods in second language research”. Færch and Kasper 1987a: 5–23.
- Floyd, P. and Patricia L. Carrell. 1987. “Effects on ESL reading of teaching cultural content schemata”. *Language learning* 37:1. 89–108.
- Gerloff, Pamela. 1986. “Second language learners’ reports on the interpretive process: Talk-aloud protocols of translation”. *House and Blum-Kulka* 1986: 243–262.

- House, Juliane and Shoshana Blum-Kulka, eds. 1986. *Interlingual and intercultural communication: Discourse and cognition in translation and second language acquisition studies*. Tübingen: Gunter Narr.
- Ivanova, Adelina. 1998. "Educating the 'language élite': Teaching translation for translator training". *Malmkjær* 1998:91–109.
- Jääskeläinen, Riitta. 1990. *Features of successful translation processes: A think-aloud protocol study*. Joensuu: University of Joensuu. [Unpublished licentiate thesis.]
- Jääskeläinen, Riitta. 1993. "Investigating translation strategies". Sonja Tirkkonen-Condit and John Laffling, eds. *Recent trends in empirical translation research*. Joensuu: University of Joensuu Faculty of Arts, 1993. 99–120.
- Jääskeläinen, Riitta. 1997. *Tapping the process: An explorative study of the cognitive and affective factors involved in translating*. Joensuu: University of Joensuu Faculty of Arts. [Doctoral dissertation.]
- Jääskeläinen, Riitta and Sonja Tirkkonen-Condit. 1991. "Automatised processes in professional vs. non-professional translation: A think-aloud protocol study". *Tirkkonen-Condit* 1991:89–109.
- Königs, Frank G. and Reinhard Kaufmann. 1996. "Processus mentaux étudiés chez des sujets allemands apprenant le français lorsqu'ils sont en train de traduire". *Meta* 41:1. 7–25.
- Krings, Hans P. 1986. "Translation problems and translation strategies of advanced German learners of French (L2)". House and Blum-Kulka 1986:263–275.
- Krings, Hans P. 1987. "The use of introspective data in translation". *Færch and Kasper* 1987:159–175.
- Kussmaul, Paul. 1991. "Creativity in the translation process: Empirical approaches". Kitty M. van Leuven-Zwart and Ton Naaijken, eds. *Translation Studies, the state of the art: Proceedings of the first James S Holmes symposium on Translation Studies*. Amsterdam: Rodopi, 1991. 91–101.
- Laukkanen, Johanna. 1996. "Affective and attitudinal factors in translation processes". *Target* 8:2. 257–274.
- Lörscher, Wolfgang. 1986. "Linguistic aspects of translation processes: Towards an analysis of translation performance". House and Blum-Kulka 1986:277–292.
- Lörscher, Wolfgang. 1991. *Translation performance, translation process, and translation strategies: A psycholinguistic investigation*. Tübingen: Gunter Narr.
- Lörscher, Wolfgang. 1996. "A psycholinguistic analysis of translation processes". *Meta* 41:1. 26–32.
- Malmkjær, Kirsten, ed. 1998. *Translation and language teaching: Language teaching and translation*. Manchester: St. Jerome.
- Mondahl, Margrethe and Knud Anker Jensen. 1996. "Lexical search strategies in translation". *Meta* 41:1. 97–112.
- Oller Jr., John W. 1995. "Adding abstract to formal and content schemata: Results of recent work in Peircean semiotics". *Applied linguistics* 16:3. 273–305.
- Séguinot, Candace. 1991. "A study of student translation strategies". *Tirkkonen-Condit* 1991:79–88.
- Séguinot, Candace. 1996. "Some thoughts about think-aloud protocols". *Target* 8:1. 75–95.
- Tirkkonen-Condit, Sonja, ed. 1991. *Empirical research in translation and intercultural studies: Selected papers of the TRANSIF seminar, Savonlinna 1988*. Tübingen: Gunter Narr.

- Tirkkonen-Condit, Sonja. 1992. "The interaction of world knowledge and linguistic knowledge in the processes of translation: A think-aloud protocol study". Marcel Thelen and Barbara Lewandowska-Tomaszczyk, eds. *Translation and meaning* Part 2. Maastricht: Rijkshogeschool Maastricht, 1992. 433–440.
- Tirkkonen-Condit, Sonja. 1997. "Who verbalises what: A linguistic analysis of TAP texts". *Target* 9:1. 69–84.
- Tirkkonen-Condit, Sonja and Johanna Laukkanen. 1996. "Evaluations: A key towards understanding the affective dimension of translational decisions". *Meta* 41:1. 45–59.
- Toury, Gideon. 1986. "Monitoring discourse transfer: A test-case for a developmental model of translation". House and Blum-Kulka 1986: 79–94.
- Toury, Gideon. 1988. "The coupled pair of 'solution + problem' in Translation Studies". Patrick Nigel Chaffey, Antin Fougner Rydning and Solveig Schult Ulriksen, eds. *Translation theory in Scandinavia: Proceedings from the Scandinavian symposium on translation theory (SSOTT) III, Oslo, 11–13 August 1988*. Oslo: University of Oslo, 1988. 1–23.
- Toury, Gideon. 1991. "Experimentation in Translation Studies: Achievements, prospects and some pitfalls". Tirkkonen-Condit 1991: 45–66.
- Wilss, Wolfram. 1994. "A framework for decision-making in translation". *Target* 6:2. 131–150.

Résumé

Au cours de la dernière décennie, les méthodes qui commandent aux sujets d'exprimer leurs pensée à voix haute ont largement investi les études du processus de la traduction. Néanmoins, d'importantes questions relatives à la validité expérimentale de ces méthodes, en tant qu'appliquées à la traduction, sont demeurées sans réponse. Le présent article passe en revue aussi bien les percées réalisées par un nombre croissant de travaux inspirés de ces méthodes que les limites auxquelles ils font face; il souligne la nécessité de traiter avec plus de sérieux les questions de validité expérimentale, théorique et d'environnement, et suggère des moyens d'amélioration. Le nochalance avec laquelle la méthodologie expérimentale a été adoptée, notamment dans le passé, risque d'invalider non seulement les résultats obtenus par des projets particuliers, mais également la méthode en sa totalité.

Author's address

Silvia Bernardini
 Corso della Repubblica 136
 47100 FORLÍ
 Italy
 e-mail: silvia@sslmit.unibo.it