Article

"Think-Aloud-Based Translation Process Research: Some Methodological Considerations"

Sanjun Sun

ÉTUDES ET PROSPECTIVES

Think-Aloud-Based Translation Process Research: Some Methodological Considerations

SANJUN SUN
Kent State University, Kent, USA
ssun2@kent.edu

RÉSUMÉ

Le présent article traite, sous les angles théorique et pratique, de problèmes méthodologiques qui surviennent dans le cadre de recherches sur le processus de traduction se fondant sur les protocoles de verbalisation (PV). Il se fonde sur des questions qui ont été soulevées à partir d'une enquête menée internationalement auprès de 25 chercheurs éminents du domaine. Il est proposé qu'il n'y a pas de données probantes suggérant que les PV changent ou influencent de manière significative le processus de traduction, bien que la validité et la complétude des PV, dans une étude particulière, puissent dépendre de plusieurs variables. Les PV, ainsi que d'autres méthodes comme l'enregistrement de la frappe et l'oculométrie, répondent à des objectifs de recherche différents, et ils peuvent donc être associés dans le cadre d'une étude faisant appel à des méthodes multiples. Plusieurs plans expérimentaux sont alors possibles, ce qui rend possibles des travaux plus approfondis que les études de cas uniques ou les mesures de convergence. Par ailleurs, les méthodes de transcription et d'analyse sont évoquées. Plusieurs idées reçues à ce sujet sont discutées. Ainsi, il est suggéré que les chercheurs ne transcrivent que ce qui est nécessaire, plutôt que de faire une transcription complète, ou ils peuvent même sauter l'étape de transcription ; s'ils choisissent leur matériel expérimental, ils n'ont pas à retenir des passages complets, ils peuvent utiliser des groupes de phrases.

ABSTRACT

Mainly structured around issues revealed in a questionnaire survey among 25 eminent translation process researchers worldwide, this paper deals with methodological issues in think-aloud-based translation process research from two perspectives: theoretical and practical. It argues that there is no strong evidence suggesting that TAP significantly changes or influences the translation process, though TAP's validity and completeness in a specific study might depend more or less on several variables. TAP and such recording methods as keystroke logging and eye tracking serve different specific research purposes, so they can be combined in a multimethod study to answer more complex research questions. Several research designs are available for a multimethod study, and researchers are encouraged to try designs other than one-shot case studies or convergence design. As for the research procedure, this paper touches upon how to transcribe and analyze the protocols. Many stereotypes in this field have been problematized. For example, this paper suggests that researchers transcribe as much as necessary rather than doing a "complete" transcription, or they can even skip the step of transcribing; in choosing test materials, researchers do not have to choose whole passages; they can use a group of sentences.

1. Introduction

Think-aloud-based translation process research emerged in the mid-1980s. In this kind of research, participants are requested to speak out their thoughts while translating a text. To date, there are over 150 journal articles, a few monographs (e.g., Lörscher 1991; Krings 2001; Englund Dimitrova 2005) and doctoral dissertations on this topic (for an annotated bibliography, see Jääskeläinen 2002). However, this field seems to have entered hibernation in recent years. To gauge the current level of interest in Think-aloud Protocols (TAP) research, a survey (see the appendix) was conducted by the present author in early 2009 among 25 eminent translation researchers worldwide who have used TAP in their research.

This survey shows that only 7 responders are currently working on a TAP project. However, when asked whether they think TAP-based translation research is insignificant or uninteresting, 23 of them responded "No." Many of them believe that this method has potential for interesting insights into cognitive processes though with limitations. About the validity and reliability of TAP, 17 of them respond in a positive way and believe that all research methods have their inherent limitations. As for the reason why TAP-based research is on the decline, many researchers attribute it to the emergence of objective recording methods such as computer keystroke logging and eye tracking, which reminds us of Kuhn's "paradigm shift," and some to the time-consuming nature of this kind of research.

There have been two lines of TAP-based translation process research. One is translation process research proper, which aims to identify characteristics of successful translation processes and to understand translation competence acquisition process. The other concerns research methodology, including the validity of the think-aloud method, subject choice (students, bilinguals or professionals), dialogue vs. monologue, between-subjects vs. within-subjects designs, protocol analysis method and others. In her widely read review, Bernardini (2001: 251) mentions that "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)." Jääskeläinen comments in the above-mentioned survey that "[t]he empirical-experimental paradigm is alien to translation scholars."

Mainly structured around issues revealed in this survey, this paper addresses the following methodological issues: 1) TAP's validity and completeness; 2) the emergence of objective recording methods such as keystroke logging and eye-tracking; 3) research design issues in TAP-based studies; and 4) how to transcribe and analyze TAP data.

2. Validity and Completeness of TAP

The causes leading to a research method's decline in popularity in a research field can be many. For instance, one cause can be "the larger tides of intellectual fashion" (Jarausch and Coclanis 2001: 12637). But a major cause in the TAP case is that it has been stigmatized by the concern for verbal reports' validity and completeness. When introducing translation process research, Venuti (2000) in his popular *The Translation Studies Reader* says that:

> Think-aloud protocols are beset by a number of theoretical problems that must be figured into any use made of their data. Verbalization won't register unconscious factors and automatic processes, and it can change a mental activity instead of simply reporting it. Similarly, subjects are sometimes instructed to provide specific kinds of information: description, for instance, without any justification. (Venuti 2000: 339)

These comments by Venuti are problematic, as we shall see. But negative comments made by authoritative figures in a field can be very damaging. In the following paragraphs, we will see what concerns researchers have, whether these concerns are founded or unfounded, and how we can deal with these concerns.

2.1. Validity

One question that constantly surfaces in the literature is whether TAP has an influence on the translation process and alters the cognitive processes (e.g., Hansen 2005). The theory that verbal protocols can be used to elicit data on cognitive processes was proposed by Ericsson and Simon (1980, 1993), and they have provided substantial empirical support for it. Ericsson and Simon hold that "subjects can generate verbalizations, subordinated to task-driven cognitive processes (think aloud), without changing the sequence of their thoughts, and slowing down only moderately due to the additional verbalization" (Ericsson and Simon 1993: xxxii).

2.1.1. Theoretical Discussions about TAP's Reactivity

There have been a lot of theoretical discussions and empirical studies about TAP's reactivity. Many language researchers (e.g., Jourdenais 2001) believe that verbalization of thoughts in language tasks imposes an additional processing load on the participants and is therefore not a pure measure of their thoughts. An oft-quoted critical statement comes from Toury:

> My concern is rather with the possible interference of two modes of translation. Thus what the experiment claims to involve is basically the gradual production of a written translation of a written text. However, the need to verbalize aloud forces the subjects to produce not just mental, but spoken translation before the required written one; and there is a real possibility that spoken and written translation do not involve the exact same strategies. (Toury 1995: 235)

Countering this statement, Tirkkonen-Condit (2006: 683) responds that this is a gross misunderstanding, for it would violate the protocols of thinking aloud if the experimenter asked the translator to translate orally in advance of typing the target text.

2.1.2. Empirical Findings about TAP's Reactivity

Till now, there has been only one empirical study in translation studies testing TAP's reactivity; this study conducted by Jakobsen (2003) showed that thinking aloud delayed translation by about 25%; no significant effects on revision were found; thinking aloud forced translators to process text in smaller segments. The first finding is consistent with Ericsson and Simon's theory. Does the last one indicate that thinking aloud changes translators' cognitive processes? Not necessarily. This finding can actually be deduced from the first finding: as thinking-aloud translators spend more time on each sentence, they surely have more pauses when translating and thus process text in smaller segments. As the translating process includes reading and writing,

empirical studies about TAP's reactivity in these two fields might provide some insights for us.

In the field of writing, Stratman and Hamp-Lyons (1994) compared the results of 12 participants revising a faulty text under thinking-aloud and silent conditions, and found that thinking aloud appeared to stimulate the production of entirely new sentences, inhibit the participants' ability to add words or phrases, and ability to detect and remedy organizational-level errors (e.g., cohesion errors). Ransdell (1995) asked 38 students to write a letter on a computer under three conditions (i.e., thinking aloud condition, silent condition, and silent condition with retrospective replay), and found that thinking aloud slowed the rate of composition, but did not reliably alter the syntactic complexity or quantity of words or clauses written.

In the field of reading comprehension, Olson, Duffy *et al.* (1984: 273) found that "Places where subjects in the [thinking-aloud] task generate more talking… are the same places where independent subjects slow down while reading silently," and concluded that "[t]his supports the claim that the [thinking-aloud] data are related in an important way to what readers are doing during more ordinary types of reading" (Olson, Duffy *et al.* 1984: 273). In the field of L2 reading, Leow and Morgan-Short (2004) empirically found that compared to a silent control group, concurrent verbal reports were not reactive; Bowles and Leow (2005) indicated that compared to a control group, thinking aloud did not significantly affect either comprehension or written production of the targeted form; Wang (2005) replicated Leow and Morgan-Short (2004), and found that thinking aloud is not simply reactive or nonreactive and it is the result of dynamic interactions between several factors (e.g., L2 proficiency level).

Generally speaking, these research findings are favorable for TAP's validity. They also indicate that the question "Does TA have an impact on a translator's thoughts during the translation process and thereby on the translation product?" (Hansen 2005) might be too general to be workable for a specific study. It contains many variables, and the answer will involve too many "It depends." We need to get down to finding potential variables pertinent to the reactivity of TAP. So, what are these potential variables?

2.1.3. Potential Variables Concerning TAP's Reactivity

Russo, Johnson *et al.* (1989) mention that there are at least four potential causes for TAP's reactivity: (1) the additional demand for processing resources, (2) auditory feedback, (3) enhanced learning over repeated trials, and (4) a motivational shift toward greater accuracy. They believe that these causes are "independent and task-specific in that any or all of them may be present depending on the primary task" (Russo, Johnson *et al.* 1989: 764). To us, the first two points are more potent, and will be used as springboards for our discussion in the following paragraphs.

Ericsson and Simon's theory assumes that "only information in focal attention can be verbalized" (Ericsson and Simon 1993: 90). We know that a person's working memory is limited in capacity. If the participant tries to think aloud while performing a task, two attentive processes must interfere with each other if the working memory capacity is stressed. Under such circumstances, the participant faces a choice between thinking aloud and performing the task. (This is why Ericsson and Simon stress that participants should focus on the primary task instead of thinking aloud.) It has been observed that subjects tend to stop talking in situations of high cognitive

load (Jääskeläinen 1999: 101). However, if the verbalization demands are slight and two attentive processes are compatible with the availability of slack cognitive processing resources, there may be no disruption of the primary process (Russo, Johnson *et al.* 1989: 764). The implication for translation process research is that difficulty of the translation task is a potential variable.

Janssen, van Waes *et al.* (1996) made a similar point. They posited that reactivity effects of thinking aloud may be task-specific; knowledge-transforming tasks (which involves development of ideas) might produce more reactive effects than do knowledge-telling tasks (which involves direct retrieval of content from long term memory), for the two entail different degrees of problem solving. The more problem solving, the more difficulty participants have, the more likely it is for them to be affected by having to translate and verbalize at the same time. For translation, tasks like poetry translation involve more knowledge-transforming than do tasks like translating a simple introduction to a company.

Another task (or text) characteristic that might be related to reactivity is the vividness of the source text. Research (e.g., Everding 2009) indicates that readers create vivid mental simulations of the sounds, sights, tastes and movements described in a textual narrative. Ericsson and Simon (1993: 17-18) distinguish three levels of verbalization: when information is originally encoded in verbal form, we will speak of Level 1 verbalization; when the information is originally not encoded in verbal form and has to be translated into that form, we will speak of Level 2 verbalization; when the task instructions ask for verbalization of only selected information, inference or generative processes, we will speak of Level 3 verbalization. Level 1 and Level 2 verbalization refers to thinking aloud. In this case, when participants work on a vivid narrative text, they need to articulate the information that is not originally in verbal form, therefore they engage in Level 2 verbalization, whereas when they work on non-vivid texts (such as argumentative texts), they engage in Level 1 verbalization.

From the above discussion, we can see that task characteristics such as vividness of the source text, its topic, and its difficulty level for the participants are potential variables for reactivity. Difficulty is a relative concept. A text is not evenly difficult, so thinking aloud might not exhibit the same level of reactivity throughout the translation process. Also, a text which is hard for novices might not be difficult for experts. Thus, the task-proficiency level (concerning such factors as L2 proficiency level, direction of translation, work routineness) of participants is an important variable. That is, thinking aloud might have different effects on the translation process of expert translators and novices.

Russo, Johnson *et al.* (1989) propose that "vocalization creates additional aural stimulation that might either facilitate or interfere with performance of the primary task" (Russo, Johnson *et al.* 1989: 764). As vocalization aids memorization, if the task depends on retention of previous messages or results, performance accuracy may increase due to thinking aloud. For translation, if sentences are long or complex, the translator will need to retain some information in working memory to relieve the severely limited short term memory. Also, thinking aloud can lead to more metacognitive activities (e.g., rationalizing their strategies) on the part of translators, thus improving their performance. (In fact, thinking aloud has been used in instruction to improve learners' metacognition (e.g., Block and Pressley 2002).) In addition, if sound-effect is important for the text (e.g., poetry), thinking aloud will help enhance

the translation quality. In a word, thinking aloud might facilitate by giving participants the opportunity to reflect on the primary process. Yet, whether such positive influences are significant enough to distort the basic structure of the translation process is still a question.

Besides the above factors whose effects are subject to experimental research, there are some experimental design-related factors which have a more certain impact on translation cognitive processes. Participants' motivation, availability of warm-up exercises, participant's personality (especially self-consciousness), gender difference between researcher and participant, equipment used (e.g., video cameras) and even surroundings (e.g., a very cold air-conditioned room, availability of drinking water) all influence the participants' amount of verbalizing and performance. Ericsson and Simon's instructions (see Ericsson and Simon 1993: 375) should be taken into account by the researcher in order to control such variables or minimize their negative influences. Russo, Johnson *et al*. (1989) suggest that "there seems to be a natural hierarchy of invalidities: disruption of the primary process is unacceptable, omissions in the verbal report are less serious, and a prolonged [reaction time] is usually inconsequential" (Russo, Johnson *et al*. 1989: 767), and experimenters should try to avoid the more damaging forms.

To sum up, it seems to us that there are four types of factors which might involve reactivity in translation process research: 1) task characteristics such as vividness of the source text, its topic, and its difficulty level; 2) auditory feedback; 3) the task-proficiency level of participants; and 4) experimental design-related factors.

2.2. Completeness

It has been observed by some translation process researchers (e.g., Hansen 2005) that expert translators verbalize less, for most of their cognitive procedure has been automatized and is not available for verbalization. This may not be entirely true. For one thing, several studies (e.g., Gerloff 1988) show that "a higher degree of translational competence does not automatically correspond with a higher degree of translation process automatization" (Krings 2001: 126). The reason is that

> [a]lthough some aspects of the process do grow easier …, other aspects become concomitantly more complex. It is as if greater automaticity at one level (for example, at the level of basic linguistic decoding, e.g. identifying agreement between subject and verb or immediate comprehension of the usual meaning of a word) "frees up" processing capacity which may then be focused on other more complex levels of analysis. (Gerloff 1988: 54)

Instead, some studies show that "novice translators draw more intensively on automatic processes, thus making fewer conscious decisions" (Alves and Gonçalves 2007: 48). For another thing, all translations are domain-specific, and no one is an expert at translating all types of texts of all subjects. In addition, a number of studies (e.g., Jääskeläinen 1999) reveal that routine conditions seem to result in higher levels of automatic processing (which is faster and more efficient than processing under conscious control) by professional translators, whereas non-routine conditions may prompt a less automatic behavior. All these indicate that completeness of verbal reports also involve several variables. A general statement like "expert translators verbalize less or more" is too simplistic.

As mentioned above, Venuti comments that "[v]erbalization won't register unconscious factors and automatic processes" (Venuti 2000: 339). This is true. According to Ericsson and Simon's theory, only those information states that are attended to in short-term memory are verbalized. However, we need to distinguish between characteristics and deficiencies. A man cannot give birth to a baby. Is this a deficiency of men? Likewise, that thinking aloud cannot uncover unconscious thoughts is a characteristic of this method; uncovering unconscious thoughts is not what it is for.

Hansen (2005) listed many reasons about the unsuitability of TAP for translation process research, e.g., expert translators who are stammerers cannot think aloud; bilingual translators have trouble verbalizing; thinking aloud during L1 to L2 translation may have an impact on the target text, etc. Most of her criticism addressed at TAP is more related to research design issues than to TAP's suitability. For instance, stammerers cannot think aloud, so why does a researcher have to recruit stammerers as research participants? This is actually a sampling issue. Some researchers even reported that many participants simply could not verbalize. We believe that thinking aloud is an inborn ability everyone possesses. Individuals express their thoughts to themselves during task performance by subvocal speech. In daily life, people sometimes talk aloud to themselves when they are alone; some translators regularly talk aloud to themselves when they translate alone (Kiraly 1995: 93). As noted above, factors like availability of warm-up exercises and participant's personality (especially self-consciousness) can explain why some translators cannot be good thinking-aloud research participants. Again, this is a research design issue. A pilot study should always be conducted in order to recruit suitable participants and set up a workable research design.

Overall, there is "so far at least no strong evidence to suggest that the TA condition significantly changes or influences the performance of these tasks" (Englund Dimitrova 2005: 75). Of course, as suggested by Jääskeläinen (2009), we still need large-scale, systematic studies of the use of TAP as a method to study translation process. In such studies, we need to consider three principles: 1) distinguishing between positive and negative effects, for the purpose of translation process research is to aid translator training; 2) investigating the effects of thinking aloud on the overall process and performance as well as on the specific components of the process (such as inferencing, revision); 3) seeing whether the reactive effects can be controlled or avoided. We optimistically believe that those factors which might lead to validity and completeness issues are controllable.

3. Verbal Reports vs. Objective Recording Methods

3.1. Comparing Verbal Reports and Recording Methods

Recording methods are used to record the overt behavior precisely. Their data often can be transformed into numbers and used in correlational analysis. For example, keystroke logging generates a lot of recorded data consisting of information concerning pausing (where and when pauses occurred, and for how long) and the history of all keyboard actions and cursor movements; it is used to study pause location, pausing in relation to planning and discourse production, and revision behavior (see

Sullivan and Lindgren 2006). One characteristic of keystroke logging is that only the writing process involved in translating is recorded. To track the translator's behavior outside the keystroke logger (e.g., consulting an electronic or online dictionary), screen recording is often used. Such a tool (e.g., CamStudio) can record all screen and audio activity on a computer and create AVI video files. Eye tracking can measure eye movements including the number of fixations, fixation durations, attentional switching, and scanpath similarity (see Duchowski 2007).

While TAP is concurrent verbal report, another form of verbalization is the retrospective verbal report. Retrospective verbal reporting takes place often immediately after the process. It does not interfere with the translation process, and yet it is less reliable compared with TAP as participants may forget what they have done. With the replay function of a keystroke logger or screen recorder, participants can watch their own translation process when they do retrospective verbal report. This measure slightly increases the reliability of retrospective verbal report.

Verbal reports are used to look into thoughts and their sequences. They can only produce verbal data. In contrast, those objective recording methods often cannot help figure out what's really going on in the participants' minds. Hansen (2008) mentions an example in her study: one participant told her during the retrospection, "here I reflected upon… did I really shut our windows at home" (It was a rainy day.) If a participant's mind wanders during the translation process (though this happens rarely), pausing data produced by keystroke logging will be misleading.

3.2. Triangulation and a Multimethod Approach

Since Jakobsen (1999) introduced the concept of triangulation into translation process research, it has come to be regarded as a "best practice" (Shreve and Angelone 2010a: 6). But what does triangulation mean? In its literal sense, triangulation is a technique of physical measurement used in maritime navigation and land surveying to pinpoint a single point with the convergence of measurements taken from two other distinct points (Rothbauer 2008). By analogy, triangulation refers to the use of multiple methods to examine a research problem so that biases can be eliminated and plausible rival explanations can be dismissed (e.g., Campbell and Fiske 1959; Webb, Campbell *et al*. 1966). Denzin (1970/2009) extended this meaning and distinguished four types of triangulation: 1) data triangulation, which involves data collected from different participants or under a variety of conditions; 2) investigator triangulation, which involves multiple researchers in an investigation to gather and interpret data; 3) theoretical triangulation, which consists of using more than one theoretical scheme in interpreting data; and 4) methodological triangulation, which involves the use of more than one method for gathering data. Of the four types, methodological triangulation is the generic one.

In the 1980s and 1990s, the notion of triangulation came under critical review. According to Mathison (1988: 14), discussions of triangulation as a research strategy were based on two assumptions: 1) "the bias inherent in any particular data source, investigator, and particularly method will be cancelled out when used in conjunction with other data sources, investigators, and methods"; 2) "when triangulation is used as a research strategy the result will be a convergence upon the truth" about some phenomenon. Both assumptions are problematic. About the first one, some research-

ers (e.g., Fielding and Fielding 1986) believe that using multiple methods or data sources does not necessarily increase validity, reduce bias or bring objectivity to research, as different methods often measure different aspects of a phenomenon and '[w]hat goes on in one setting is not a simple corrective to what happens elsewhere – each must be understood in its own terms' (Silverman 1985: 21). Triangulation cannot "be meaningfully compared to correlation analysis in statistical studies" (Denzin 2007: 5086). For the second one, a triangulation strategy might produce three kinds of outcomes: convergence, inconsistency, and contradiction among the data. Compared with convergence, the other two outcomes are more likely.

In the recent decade, triangulation has returned to favor, but the focus of its meaning has shifted. For most researchers, it refers to "a multimethod approach to data collection and data analysis" (Rothbauer 2008: 892), and the idea is that the richness and complexity of human behavior can be explored more fully by studying it using both quantitative and qualitative data (Cohen, Manion *et al.* 2000: 112), and convergent, inconsistent or contradictory findings can help the researcher construct explanations of the phenomena. As a result, this metaphor is practically dead; triangulation is now almost synonymous with "multimethod approach" or "mixed methods approach" which involves both quantitative and qualitative data (see Tashakkori and Teddlie 2010).

In the translation process research field, it seems that most recent TAP-based studies (e.g., Alves 2003; Shreve and Angelone 2010b) have adopted a multimethod approach. For example, Englund Dimitrova (2005) combines think-aloud protocols and keystroke logging to study explicitation in translation; Faber and Hjort-Pedersen (2009) use TAP, retropective interview and Translog (a keystroke logger) to investigate the correlation between cognitive processing of legal texts and linguistic explicitation and implicitation in legal translation; Angelone (2010) uses TAP and screen recording to look into the problem-solving behavior of professional and student translators, focusing on the metacognitive phenomenon of uncertainty management. Göpferich (2009) has been using TAP, keystroke logging, screen recording, webcam recording, retrospective interviews and questionnaires in her TransComp project to investigate the development of translation competence over a period of three years.

The more research methods one adopts in one's research, the more complex research questions one might be able to answer. However, the Scope Triangle (time-cost-quality) in project management tells us that there are always trade-offs inherent in any project. Research quality is constrained by time and resources available. Researchers adopting a multimethod approach need more time to collect and analyze data. Research participants might get fatigued or bored if they have to go through a lengthy and complex research procedure. Göpferich (2009) mentions that in her project all research participants preferred think-aloud method to cued retrospection, and one reason was assumed to be that participants felt exhausted after each experiment and did not want to spend more time on the time-consuming immediate retrospection interview part. In addition, if a participant has finished part of the experiment and then decides to drop out, the data he or she has provided will probably be useless for this multimethod project.

Besides these practical considerations, before committing to a multimethod approach, we also need to consider whether these methods are compatible in terms

of reactivity in one study. For instance, if we want to investigate translators' pausing behavior in translating, and choose to use TAP and Translog, this method combination will not work out. The reason is simple: thinking aloud will slow down the translation and (often unproportionately) change the pausing behavior. Screen recording is not intrusive and can be compatible with other methods if running the software does not considerably reduce the computer performance. Keystroke loggers are usually not intrusive. However, such tools have fewer functions compared with Microsoft Word or similar word-processing tools translators work with. For example, spell and grammar checking is an important function for translators. If the keystroke logger used in the experiment does not have this function, participants' translation process might be impacted. Generally speaking, every method has strengths and limitations. But in a study with a specific research question, we need to draw on their strengths and avoid their limitations, and the way is through careful research design.

4. Research Design Issues

Before we get to research designs, we need to know whether TAP-based translation studies are quantitative or qualitative. Why? The present author was engaged in a TAP-based research project in 2005, and that project unfortunately failed. At that time, we collected TAP data before we had a specific research question in mind, for in some qualitative traditions (e.g., Grounded Theory), research questions are allowed or expected to emerge during the data collection and analysis stages. But, do all qualitative beliefs apply to TAP-based studies?

4.1. Qualitative or Experimental/Quantitative?

Translation process researchers have been arguing about research design issues ever since this method was adopted in this field. Many disputes boil down to an essential assumption that TAP-based research is qualitative research so the research should adhere to qualitative trustworthiness safeguards (Li 2004). From this assumption, many researchers argue against the use of non-routine tasks or the laboratory, for it will violate one of the most important rules of qualitative research: "natural situation." As think-aloud protocols are in the verbal form (which is the data form of qualitative research) rather than numerical form (which is the data form of quantitative research), the qualitative opinion is true in this sense. However, Ericsson and Simon say that "[i]n many studies we want to collect verbal reports for cognitive processes that are no different from those occurring in traditional [psychological] experiments" (Ericsson and Simon 1993: 375). This remark implies that TAP is a kind of experimental method. Here arises the conflict. We all know that qualitative research is quite different from experimental research which usually uses quantitative methods and a different set of criteria for rigor. As this question concerns how to conduct and evaluate a TAP study, it begs for clarification.

Qualitative research is modeled on ethnography, which seeks to understand human behavior within its own social setting. Yet, the purpose of TAP is to investigate the sequence of thought processes. Their purposes are not the same. An experiment is "a deliberately planned process to collect data that enable causal inferences and legitimize treatment comparisons" (Morris and Chiu 2001: 5086). Quantitative/

experimental research is meant to test or verify a hypothesis or theory using numerical data and correlation analysis, which is not applicable to TAP. So, in terms of its research purpose, TAP does not align with either qualitative research or quantitative/experimental research.

Qualitative research stresses naturalistic observation and natural setting (the participant's setting) while psychological experiments usually occur in a laboratory in order to better control extraneous and confounding variables and guarantee the study's internal validity. In addition, a lab can be constructed to reproduce a real-world setting, and hence increase the study's external validity. In this aspect, TAP research is similar to psychological experiments; it needs to control confounding variables (such as computer configuration, availability of software tools and dictionaries), and a room with a powerful computer (if possible, using participants' own keyboards and their preferred web browsers) would be a sufficient setting.

Qualitative inquiry typically involves relatively small samples and uses purposeful sampling, whose logic lies in selecting information-rich cases for study in depth, learning a great deal about issues of central importance to the purpose of the inquiry, deriving insight and in-depth understanding rather than empirical generalizations; quantitative/experimental methods usually depend on larger samples and random sampling, whose purpose is generalization from the sample to a larger population and control of selectivity errors (Patton 2002: 230). In TAP research, researchers usually adopt purposeful sampling and use small samples. Thus, it is similar to qualitative research.

Quantitative/experimental research usually adopts a predetermined research design; its research questions, hypotheses and experimental procedure are established at the outset of the study. It often relies on deductive reasoning which starts with theory and tests its applicability. Qualitative research tends to adopt an emergent research design; the research questions, hypotheses and theories emerge during the course of the research and are not specified at the beginning (Denscombe 2007: 250). It traditionally relies on the inductive process, i.e., reasoning from the particular to more general statements then to theory. Most TAP-based translation process studies rely on inductive reasoning, and yet many of them adopt a "between-subjects design," comparing professional translators and trainee translators. Thus, in research design, TAP bears resemblance both to experimental research and to qualitative research.

One of the defining features of experimental research is that it includes a manipulation (also known as stimulus, treatment or independent variable), and at least two conditions that differ only on the particular feature that is manipulated. In contrast, qualitative research stresses naturalistic observation and does not involve any manipulation. In a pure TAP study, researchers simply request the participants to think aloud while performing a task, and there is no manipulation. In reality, TAP method is often embedded in a complex research design which involves manipulation. For instance, in a study, participants are requested to think aloud while translating with access or no access to dictionaries in order to determine whether the access to dictionaries will influence the translator's performance. In such a study, access/no access to dictionaries is the independent variable, the number of translation errors may be the dependent variable. The TAP method is only to help interpret the results; it is subordinate. In this sense, a TAP study is more similar to qualitative research than to experimental research.

Nowadays, many researchers think qualitative and quantitative approaches are not in opposition; instead, they are at the two ends of a continuum. Having discussed the similarities and differences between qualitative, quantitative approaches and TAP studies, I would like to use the following figure to describe the position of TAP research along the continuum from qualitative to quantitative:

FIGURE 1

The relation between TAP and qualitative, quantitative approaches
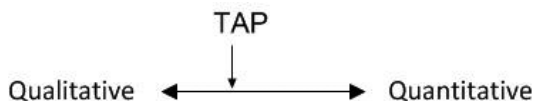
Following this position, trustworthiness safeguards for TAP-based research need to be re-established, and research design issues reconsidered.

4.2. Types of Research Designs

Research designs are frameworks for collecting, analyzing, interpreting data in research studies. They can guide the researchers in choosing research methods and in interpreting results. A typical research design in a TAP-based translation study is like one Göpferich (2009) has adopted in her TransComp project: participants translate texts in Translog on a computer, and are allowed to use the Internet and any other electronic and conventional resources (e.g., print dictionaries); use of electronic resources is registered by a screen-recording tool while use of conventional resources is documented by researchers; immediately after each translation session, the participants need to fill out questionnaires "on how they felt during the translation process, on the problems they encountered, the strategies they employed to solve them and the extent to which they were satisfied with the results"(Göpferich 2009: 28); then, short retrospective interviews are to be conducted with the participants to find out, e.g., whether they are aware of certain problems they may have encountered during the translation process; researchers will evaluate the participants' translations and relate their translation processes to their products in terms of research questions.

4.2.1. Types of General Research Designs

Of course, Göpferich's project is a longitudinal study of the development of translation competence, and its research design is more complex than the above-mentioned procedure. According to de Vaus (2001), there are four main types of research designs in the social sciences in terms of six dimensions (e.g., the number of groups in the design, the methods of allocation of participants to groups, the nature of the intervention): experimental design, longitudinal design, cross-sectional design, and case studies.

In an experiment, participants are usually randomly allocated to experimental or control groups, and only the experimental group is exposed to a treatment (also called the independent variable, such as gender, availability of translation brief, direction of translation) with extraneous variables being controlled. The responses are measured by the investigator. This design enables the investigator to deduce a 'cause and effect' relationship. Till now, few TAP-based translation studies have adopted the experimental design.

In a longitudinal design, a group of participants are measured repeatedly at different times. The investigator compares the pre-treatment measures with the post-treatment measures, and analyzes change over time. The advantage of longitudinal studies is that it keeps participant variables reasonably constant between the conditions, while some of this design's disadvantages are that it may be confounded by extraneous events that occur during the course of the study, and the research results may not generalize over time (Heiman 2001: 571) (e.g., for emergence of new-generation CAT tools, wide adoption of speech recognition software). Göpferich's project adopts this design, and the participants in her study will retranslate the same texts after three years for comparison.

Similar to longitudinal design, cross-sectional design is normally used in developmental research examining how people change as they go through life. In the cross-sectional design, people at varying phases of development are studied at one point in time; this design can be thought of as simulating development over time (Pomerantz, Ruble *et al.* 2004: 408). This design's major advantage is that the study can be conducted rather quickly, while its main disadvantage is that the conditions may differ in terms of other variables (Heiman 2001: 571). Most existing TAP-based translation studies compare trainee translators and professional translators and try to find how translation competence develops.

Case studies are an approach that focuses on one or a few instances of a phenomenon. They can be either longitudinal or cross-sectional, and the objective is "to understand and interpret thoroughly the individual cases in their own special context, and to find information concerning the dynamics and the processes" (Aaltio and Heilmann 2010: 66). A case study is capable of producing hypotheses and providing theoretical breakthroughs. Qualitative case study researchers advocate such strategies as "thick description" and "process tracing" (Blatter 2008). Most existing TAP-based studies are case studies.

4.2.2. Types of Mixed Methods Designs

In Section 3.2., we talked about triangulation and a mixed methods approach. Mixed methods research involves at least one qualitative strand and one quantitative strand. Since think-aloud protocols is basically a qualitative method while other methods like keystroke logging produce quantitative data, mixed methods designs are pertinent to us.

According to Creswell and Plano Clark (2011), there are six major mixed methods design types: convergent parallel design, explanatory sequential design, exploratory sequential design, embedded design, transformative design, and multiphase design. Of the six, the first four are basic mixed methods designs. To choose an appropriate mixed methods design that fits one's research questions in a study, a researcher needs to consider: 1) whether the quantitative and qualitative strands are mixed before the final interpretation; 2) which strand has the priority (equally important, or, one plays a primary role while the other plays a secondary role); 3) whether the strands will be implemented concurrently, sequentially, or across multiple phases; and 4) how the strands are to be mixed (e.g., merging, embedding) (Creswell and Plano Clark 2011: 105). Each of the six mixed methods designs includes a specific set of research procedures. In the following paragraphs, we will briefly explain the four basic mixed methods design types.

In a convergent parallel design, the researcher collects and analyzes both quantitative and qualitative data separately during one phase of the research at roughly the same time, and then merges the two data sets into an overall interpretation. One data set does not depend on the result of the other. The two strands are equally important for addressing the study's research questions. The purpose of this design is to get complementary data on the same topic. For instance, a translation process researcher can use keystroke logging and TAP in one study, analyze the keystroke logging data quantitatively and TAP qualitatively and then merge the two sets of results.

In an explanatory sequential design, the researcher starts by collecting and analyzing quantitative data, and then collects and analyzes qualitative data. In the two interactive phases, the quantitative phase plays a primary role while the qualitative phase builds on the results of the first phase. The purpose of this design is to use the qualitative results to explain the initial quantitative results (e.g., the relationships, or causes behind the resultant trends). For instance, a translation researcher can collect translation corpus data, identify some features, and then use TAP to explain how these features come into being in the translation process.

In an exploratory sequential design, the researcher starts by collecting and analyzing qualitative data, and then collects and analyzes quantitative data from a larger sample. In the two interactive phases, the qualitative phase plays a primary role while the quantitative phase builds on the results of the first phase in order to test, assess or generalize the initial findings. The purpose of the first, qualitative phase is to develop a measuring instrument, and identify important variables. For instance, a translation process researcher can use TAP to identify some translation process features, and then uses keystroke logging or eye-tracking data to further investigate these features.

In an embedded design, the researcher may add a qualitative strand within a quantitative design (such as an experiment), or add a quantitative strand within a qualitative design (such as a case study). Collecting and analyzing the supporting data can occur before, during, or after the major data collection and analysis procedures. Researchers adopt this design when they have different questions that require different types of data.

4.2.3. Research Design Problems in Existing Studies

There is no best research design. Researchers need to choose one best suited to the particular research question(s) under study. To date, most process-based studies are one-shot case studies, have adopted the convergence design and used multiple methods. As we mentioned, adopting multiple methods in one study is expensive, and it may not be necessary to adopt all the methods available for one specific research question. Meanwhile, we can adopt other mixed methods design types. Of course, this again depends on the research questions.

In the translation process literature, we have several lists of research questions, e.g., Krings (2001: 164-178), Kiraly (1995: 50-51), Shreve and Danks (1997: viii-ix). EXPERTISE (Expert Probing through Empirical Research on Translation Processes) group, which is composed of leading European researchers in the field of translation processes, announced in 2002 that they try to identify the translator's knowledge base and cognitive underpinnings of expert behavior in translation. And their basic

research objectives include translation aptitude, the development of translation expertise, memory structures, monitoring operations in translation, creative mental processes in translation, etc. These research questions give us directions. However, many (if not most) of them are too general for one specific study. They need to be decomposed. Most of us need to focus on one specific issue in one study, and in time, these general questions can be answered by synthesizing specific studies.

A review of translation process research during the past 15 years by Bernardini in 2001 shows that problem indicators and translation strategies together with translation (or attention) units, automaticity of processing and affective factors are researchers' main concerns. Many studies have focused on the same research questions. Although replication is a good practice in empirical research, it is time for us to expand our horizons and turn to other specific research questions (e.g., metaphor translation and explicitation; influence of text types on the choice of one specific translation strategy).

A problem closely related to research design concerns task analysis. Task analysis here refers to having a linguistic and translational analysis of the source text used in the experiment, analyzing its text type/genre/register features, anticipating the range of problems translators could have and strategies/procedures they could use in light of their prior knowledge. Researchers need to take their research questions and hypotheses into account when they do task analysis and select test passages. Krings (2001: 74) provides a list of test passages used by translation process researchers: newspaper articles, tourist brochures, and others. It seems that most researchers did not explain why they chose those passages and how those passages related to their research questions. Lörscher and PACTE are among the few exceptions. When choosing test passages, Lörscher (1991: 89) ensures that "the texts must contain a sufficient number of translation problems to make sure that a sufficient number of (potentially different) problem-solving strategies will be used." PACTE (2009) has been putting their emphasis in data collection and analysis on specific source-text segments that contain Rich Points, which can be linguistic problems, textual problems, extralinguistic problems, problems of intentionality, or problems relating to the translation brief. These Rich Points were determined beforehand in their project, and greatly facilitate their data analysis. Many translation scholars (see Kelly 2005: 117-127 for an overview) have talked about how to select texts in pedagogy in their writings, to which process researchers can refer. Needless to say, choosing test passages casually and without task analysis in a TAP-based translation study will make data analysis difficult or even incur the risk of project failure (especially when researchers select a test passage before they have a specific research question in mind). That said, researchers do not have to select whole passages or focus on so many Rich Points when selecting passages. Sometimes we can use a group of sentences or short paragraphs (e.g., sentences using figures of speech) tied to one translation problem.

From the literature we can see that translation process researchers (e.g., Jääskeläinen 2010) have paid considerable attention to participant selection in terms of their translation competence. However, scant attention has been paid to choosing participants in terms of their suitability for thinking aloud. As mentioned above, although thinking aloud is an inborn ability, some translators (e.g., stammerers, those who are very self-conscious) are not suitable for the experiment. A pilot study is always necessary.

Many researchers complain about the completeness of TAP. According to Ericsson and Simon (1998: 182), "when participants are asked to describe and explain their thinking, their performance is often changed – mostly it is improved." The implication for us is that we can have participants explain their decisions when they think aloud in order to see how much they know and to what degree their performance can be improved. For some research questions, whether cognitive processes have slightly changed due to thinking aloud is not important.

Statistically speaking, to make an experiment more powerful, researchers need to test a large number of participants, maximizing differences between different groups and minimizing differences within each group (e.g., in L1, L2, age, language proficiency, professional background, domain). If the novice group includes first-year and fourth-year college students, and the professional group contains translators who have worked for 2-10 years, comparison between the two groups will not be fruitful as there is too much variability within each group. In addition, as many factors such as anxiety, lack of sleep the previous night, fatigue, and even availability of coffee may impact participants' translation quality, researchers need to test participants on multiple trials. Then, we assume that random differences in these factors on different trials balance out, and we use their average scores (Heiman 2001: 432).

Undoubtedly, TAP-based translation process research is closely related to experimental psychology, and the purpose of psychology is

> […] to discover new psychological variables, to show the relationships among these new variables with already determined variables, and, of course, to discover new relationships among already known variables. (Asher 2001: 1397)

TAP-based translation studies have already uncovered many variables (e.g., translation quality, translation strategies, language proficiency, direction of translation, task routineness), and we still need to discover new variables and relationships among them.

5. Protocol Analysis

The analysis of qualitative data (including TAP data) has a number of common features, which include

> […] simultaneous data collection and analysis, the practice of writing memos during and after data collection, the use of some sort of coding, the use of writing as a tool for analysis, and the development of concepts and connection of one's analysis to the literature in one's field. (van den Hoonaard and van den Hoonaard 2008: 186)

This analysis typically consists of the following steps: transcribing, annotating (or encoding) and analyzing. In the following sections, we will see what problems exist in these aspects and how to proceed.

5.1. How to Transcribe

In qualitative research, researchers typically transcribe everything from the tape, then encode and analyze the transcripts. This tradition started since people began to use tape recorders in their research. The reason for transcripts is "to record, to illuminate, to re-present, and to facilitate analysis" (Powers 2005: 2). Transcribing every

word is also the way translation process researchers have adopted. For instance, Krings' transcription rules begin with "The text is taken literally from the tape, as an endless chain of words, meaning without any punctuation and without paragraphs (for exceptions see rules 2 and 3)" (Krings 2001: 208). But is complete transcription absolutely necessary? Technologies are constantly evolving. As mentioned above, in TAP-based translation studies, we usually use computer screen recorders instead of tape recorders or video cameras. Compared with tape and video recordings, screen recordings are much easier to manipulate. Also, screen recordings can also record, illuminate, re-present, and facilitate analysis.

In addition, transcribing has severe shortcomings. First, it is notoriously time-consuming. Transcribing an hour of speech takes over ten hours. In Krings' study, transcribing about 100 hours' VCR recordings took about 1600 working hours (Krings 2001: 213). This alone scares away many potential researchers without any research funding. Second, a transcript is not a full copy of the original event. It can eliminate features of spoken production and lead to missing crucial interpretive resources. For this reason, researchers have to transcribe paralinguistic features (such as sighs, laughter), extralinguistic features (e.g., direction of gaze, gestures, pauses, fillers), as well as verbalizations. And TAP researchers tend to insert timestamps and record pauses in their transcripts. In a way, we can say that these researchers are fighting a losing war. No matter how detailed their transcripts are, they are not 100% accurate representations of the recordings. Transcripts are always selective constructions. Third, a complete transcription always incurs waste though it may bring the researcher a sense of fulfillment. For instance, in Krings' study, nonverbal behavior (including laughs, sighs, groans, whistles and others) was transcribed, but had not appeared in his analysis.

What to transcribe in a study is determined by its research question(s). My belief is that in many TAP-based translation studies, relying on screen and audio recordings is adequate, and researchers do not have to transcribe. Besides saving time, a strong point of this is that researchers can watch the process while listening to the participant's verbalizations, a procedure that can provide more information to the researcher than simply reading transcripts can. A useful tool for watching analytically interesting video clips intensely and helping recognize problem-solving processes is *ATLAS.ti*.

If the researcher feels a need to transcribe, my suggestions are: 1) transcribing selectively following some pre-formulated goal or rules (e.g., only transcribing parts involving a specific strategy you are looking into); 2) using XML in accordance with the *Guidelines for Electronic Text Encoding and Interchange* of the *Text Encoding Initiative* (TEI 2007) (see Göpferich 2010) in transcripts so that they can be analyzed using XML tools (e.g., *oXygen XML*, *Altova XMLSpy*); 3) using standard orthography. If one is not interested in paralinguistic features, extralinguistic features or filled pauses, what is the point of transcribing laughter and sighs? Here is a snippet of a protocol produced by one participant asked to think aloud while mentally multiplying 36 by 24:

> OK, 36 times 24, um, 4 times 6 is 24, 4, carry the 2, 4 times 3 is 12, 14, 144, 0, 2 times 6 is 12, 2, carry the 1, 2 times 3 is 6, 7, 720, 720, 144 plus 720, so it would be 4, 6, 864. 36 times 24, 4, carry the – no wait, 4, carry the 2, 14, 144, 0, 36 times 2 is, 12, 6, 72, 720 plus 144, 4, uh, uh, 6, 8, uh, 864. (Ericsson 2006: 227-228)

Researchers can follow this example.

5.2. How to Encode and Analyze

Codes and coding are integral to qualitative data analysis. Codes are keywords or key concepts. They are used to organize segments of text into key topics defined by researchers and help researchers to find patterns within the data (Maietta 2008: 105). There are two kinds of codes: deductive codes and inductive codes. Deductive codes are formed prior to the analysis process. They may be derived from the researcher's research question(s), research findings in this field, or even other researchers' codebook. Inductive codes emerge from the data.

The most prominent procedure for coding and analyzing qualitative data is Grounded Theory. This approach is strongly committed to emergent themes and inductive coding process, and argues against using pre-established codes in data analysis. However, as Benaquisto (2008: 88) says, "[t]he approach one takes in developing a coding frame depends on a number of factors, including the issue under study, how well the topic is understood, the complexity of the phenomenon, and even the amount of time one has for analysis." So, if a TAP-based translation process researcher is going to look into one aspect already researched by other researchers (e.g., translation strategies), he or she may use deductive codes; otherwise, inductive codes can be used.

In using inductive coding, Grounded Theory provides a specific set of procedures. The coding process has at least two phases: initial coding and focused coding. In initial coding, researchers compare lines of data to define the properties of what is happening, learn how it developed, what it means and then assign brief categories to each line in the data; in focused coding, researchers select these codes as focused codes to sift large batches of data and generate preliminary categories for the emerging theory (Charmaz 2007; see Corbin and Strauss 2008 for details).

Qualitative data analysis tools (e.g., *ATLAS.ti*, *NVivo*) greatly facilitate the coding job. For instance, *ATLAS.ti* provides the following ways of coding: Open Coding, Code In Vivo, Code by list, Quick Coding. Codes can be renamed, combined or divided as the analysis proceeds. Codes are traditionally inserted into transcripts. If there is no transcript, researchers can add codes to video segments directly (e.g., in *ATLAS.ti*). Besides these general-purpose qualitative tools, Alves and Vale (2009) developed a web application for storing, annotating, and querying translation process data.

When coding is finished (at least tentatively), the remaining data analysis will begin. Data analysis can be divided into two categories according to the purpose of the research: inductive and deductive. In inductive data analysis, researchers need to generate meaning from the data; in deductive analysis, researchers try to test or confirm some findings. Miles and Huberman (1994) suggest 13 tactics for the former: 1) noting patterns and themes, 2) seeing plausibility, 3) clustering, 4) making metaphors, 5) counting, 6) making contrasts/comparisons, 7) partitioning variables, 8) subsuming particulars into the general, 9) factor analysis, 10) noting relations between variables, 11) finding intervening variables, 12) building a logical chain of evidence, 13) making conceptual/theoretical coherence. Of these, counting has been a tactic frequently used by TAP translation researchers (e.g., Krings 2001; Lörscher 1991) in their inquiry of the distribution of translation problems, specific processes, or coding units. Numbers can help researchers see rapidly what they have in a large batch of data, verify a hunch or hypothesis and keep themselves analytically honest, protecting against bias (Miles and Huberman 1994: 253). For deductive analysis,

Miles and Huberman (1994: 263) mention a few tactics, including: 1) looking for outliers, extreme cases and negative evidence to verify what a "pattern" is not like, 2) making if-then tests, 3) ruling out spurious relations, 4) replicating a finding, and 5) checking out rival explanations.

Qualitative data analysis tools provide counts of coded instances by code, which can be exported to spreadsheets or SPSS for quantitative analysis. In addition, they can help organize codes and memos, enable the researcher to use diagrams to find relationships between a central category and other major categories.

For the inductive approach, the final product is usually a "model of process and a transactional system" (Smith and Davis 2001: 73). For the deductive approach, researchers can verify a hypothesis, falsify a theory, or reveal more interesting variables.

6. Conclusion

TAP-based translation process research has a relatively short history. Most of the responders in the survey conducted by the present author believe that it has potential for interesting insights into translation-related cognitive processes. Yet many of them have mentioned their doubts and difficulties in using TAP.

This paper clarifies some doubts about TAP's validity and completeness, and argues, based on a literature review, that there is, to date, no strong evidence suggesting that TAP significantly changes or influences the translation process. Meanwhile, it indicates that TAP's validity and completeness might involve several variables. Distinguishing these variables and avoiding those "disturbing" factors through careful research design will further improve this method's validity and verbal report completeness in a study.

Some researchers believe that keystroke logging and eye tracking are to replace TAP. This paper shows that these methods serve different research purposes. To date, TAP is one of the few methods which can help reveal a person's sequence of thoughts in solving a problem. In the recent decade, a multimethod approach is widely practised by translation reseearchers who combine TAP and objective recording methods. Of course, adopting the multimethod approach has advantages and disadvantages.

This paper also points out a misunderstanding that TAP-based translation research is qualitative in nature, and suggests that it is on the continuum between qualitative and quantitative research though closer to the qualitative end. There are several types of research designs which can be adopted in TAP translation research. They all have strengths and weaknesses. Whichever design is to be adopted is up to the research question, and this means that researchers should work backwards, decide which kind of data can be used to prove or falsify the hypotheses, then determine the research design, and finally begin to collect data.

For solving practical difficulties in carrying out TAP studies, this paper provides suggestions in how to transcribe, encode and analyze protocols. Since researchers tend to use computer screen recorders in their research, transcription, which can be very time-consuming, is no longer an indispensable step.

This paper has dealt with methodological issues in TAP-based translation process research from two perspectives: theoretical and practical, and has problematized many stereotypes in this field. I hope rules and suggestions mentioned in it will be a useful guideline for the investigation of cognitive aspects of translation.

REFERENCES

Aaltio, Iiris and Heilmann, Pia (2010): Case Study as a Methodological Approach. *In*: Albert J. Mills, Gabrielle Durepos and Elden Wiebe, eds. *Encyclopedia of case study research*. Los Angeles: SAGE Publications, 66-76.

Alves, Fabio, ed. (2003): *Triangulating Translation: Perspectives in Process Oriented Research*. Amsterdam: John Benjamins.

Alves, Fabio and Gonçalves, José Luiz (2007): Modelling translator's competence: Relevance and expertise under scrutiny. *In*: Yves Gambier, Miriam Shlesinger and Radegundis Stolze, eds. *Doubts and directions in translation studies*. Amsterdam: John Benjamins, 41-55.

Alves, Fabio and Vale, Daniel Couto (2009): Probing the unit of translation in time: Aspects of the design and development of a web application for storing, annotating, and querying translation process data. *Across Languages and Cultures*. 10(2):251-273.

Angelone, Erik (2010): Uncertainty, uncertainty management, and metacognitive problem solving in the translation task. *In*: Gregory M. Shreve and Erik Angelone, eds. *Translation and cognition*. Amsterdam: John Benjamins, 17-40.

Asher, John William (2001): Research Methodology. *In*: W. Edward Craighead and Charles B. Nemeroff, eds. *The Corsini encyclopedia of psychology and behavioral science*. 3rd ed. New York: Wiley, 1396-1402.

Benaquisto, Lucia (2008): Coding frame. *In*: Lisa M. Given, ed. *The Sage encyclopedia of qualitative research methods*. Los Angeles: SAGE Publications, 88-89.

Bernardini, Silvia (2001): Think-aloud protocols in translation research: achievements, limits, future prospects. *Target*. 13(2):241-263.

Blatter, Joachim K. (2008): Case study. *In*: Lisa M. Given, ed. *Sage encyclopedia of qualitative research methods*. Los Angeles: SAGE Publications, 68-71.

Block, Cathy Collins and Pressley, Michael (2002): *Comprehension instruction: research-based best practices*. New York: Guilford Press.

Bowles, Melissa A. and Leow, Ronald P. (2005): Reactivity and type of verbal report in SLA research methodology: Expanding the scope of investigation. *Studies in Second Language Acquisition*. 27(3):415-440.

Campbell, Donald T. and Fiske, Donald W. (1959): Convergent and Discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 56(2):81-105.

Charmaz, Kathy (2007): Grounded Theory. *In*: George Ritzer, ed. *The Blackwell encyclopedia of sociology*. Malden: Blackwell Pub.

Cohen, Louis, Manion, Lawrence and Morrison, Keith (2000): *Research methods in education*. 5th ed. London: RoutledgeFalmer.

Corbin, Juliet M. and Strauss, Anselm L. (2008): *Basics of qualitative research: techniques and procedures for developing grounded theory*. 3rd ed. Los Angeles: SAGE Publications.

Creswell, John W. and Plano Clark, Vicki L. (2011): *Designing and conducting mixed methods research*. 2nd ed. Thousand Oaks: SAGE Publications.

De Vaus, David (2001): *Research design in social research*. London: Thousand Oaks: SAGE Publications.

Denscombe, Martyn (2007): *The good research guide: for small-scale social research projects*. 3rd ed. Maidenhead: Open University Press.

Denzin, Norman K (1970/2009): *The research act: a theoretical introduction to sociological methods*. Chicago: Aldine Pub. Co.

Denzin, Norman K. (2007): Triangulation. *In*: George Ritzer, ed. *The Blackwell Encyclopedia of Sociology*. Malden: Blackwell Publishing, 5083-5088.

Duchowski, Andrew T. (2007): *Eye tracking methodology: theory and practice*. 2nd ed. London: Springer.

Englund Dimitrova, Birgitta (2005): *Expertise and Explicitation in the Translation Process*. Amsterdam: John Benjamins.

Ericsson, K. Anders (2006): Protocol Analysis and Expert Thought: Concurrent Verbalizations of Thinking during Experts' Performance on Representative Tasks. *In*: K. Anders Ericsson, Neil Charness, Paul J. Feltovich*, et al.*, eds. *The Cambridge handbook of expertise and expert performance*. Cambridge: Cambridge University Press.

Ericsson, K. Anders and Simon, Herbert Alexander (1980): Verbal reports as data. *Psychological Review*. 87(3):215-251.

Ericsson, K. Anders and Simon, Herbert Alexander (1993): *Protocol analysis: verbal reports as data* (Rev. ed.). Cambridge: MIT Press.

Ericsson, K. Anders and Simon, Herbert Alexander (1998): How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*. 5(3):178-186.

Everding, Gerry (2009): Readers build vivid mental simulations of narrative situations, brain scans suggest. *Record*. Visited on 12 August 2010, <http://record.wustl.edu/tips/page/normal/13383.html>.

EXPERTISE (Last updated: 2002). *Expression of interest*. Visited on 24 October 2007, <http://www.hf.uio.no/ilos/forskning/forskningsprosjekter/expertise/expression_of_interest/full.pdf>.

Faber, Dorrit and Hjort-Pedersen, Mette (2009): Manifestations of inference processes in legal translation. *In*: Arnt Lykke Jakobsen, Inger M. Mees and Susanne Göpferich, eds. *Behind the mind: Methods, models and results in translation process research*. Frederiksberg: Samfundslitteratur Press, 107-124.

Fielding, Nigel and Fielding, Jane L. (1986): *Linking data*. Beverly Hills: SAGE Publications.

Gerloff, Pamela Ann (1988): *From French to English: A Look at the Translation Process in Students, Bilinguals, and Professional Translators*. Doctoral dissertation, unpublished. Boston: Harvard University.

Göpferich, Susanne (2009): Towards a model of translation competence and its acquisition: the longitudinal study TransComp. *In*: Arnt Lykke Jakobsen, Inger M. Mees and Susanne Göpferich, eds. *Behind the mind: Methods, models and results in translation process research*. Copenhagen: Samfundslitteratur, 11-37.

Göpferich, Susanne (2010): Data Documentation and Data Accessibility in Translation Process Research. *The Translator*. 16(1):93-124.

Hansen, Gyde (2005): Experience and Emotion in Empirical Translation Research with Think-Aloud and Retrospection. *Meta*. 50(2):511-521.

Hansen, Gyde (2008): The Dialogue in Translation Process Research. *In*: TRANSLATORS ASSOCIATION OF CHINA, ed. *Translation and Cultural Diversity*. (XVIII FIT World Congress, Shanghai, 2-7 August 2008). Beijing: Foreign Languages Press, 386-397.

Heiman, Gary W. (2001): *Understanding research methods and statistics: an integrated introduction for psychology*. 2nd ed. Boston: Houghton Mifflin.

Jääskeläinen, Riitta (1999): *Tapping the Process: An Explorative Study of the Cognitive and Affective Factors Involved in Translating*. Joensuu: University of Joensuu Publications in the Humanities.

Jääskeläinen, Riitta (2002): Think-aloud protocol studies into translation: an annotated bibliography. *Target*. 14(1):107-136.

Jääskeläinen, Riitta (2009): Think-aloud protocols. *In*: Mona Baker and Gabriela Saldanha, eds. *Routledge encyclopedia of translation studies*. 2nd ed. London: Routledge, 290-293.

Jääskeläinen, Riitta (2010): Are all professionals experts? Definitions of expertise and reinterpretation of research evidence in process studies. *In*: Gregory M. Shreve and Erik Angelone, eds. *Translation and cognition*. Amsterdam: John Benjamins, 213-227.

Jakobsen, Arnt L. (1999): Logging target text production with Translog. *In*: Gyde Hansen, ed. *Probing the process in translation: methods and results*. Copenhagen: Samfundslitteratur, 9-20.

Jakobsen, Arnt L. (2003): Effects of think aloud on translation speed, revision, and segmentation. *In*: Fabio Alves, ed. *Triangulating translation: perspectives in process oriented research*. Amsterdam: John Benjamins, 69-95.

Janssen, Daniël , Van Waes, Luuk and Van Den Bergh, Huub (1996): Effects of thinking aloud on writing processes. *In*: C. Michael Levy and Sarah Ransdell, eds. *The Science of Writing: Theories, Methods, Individual Differences and Applications*. Mahwah: Lawrence Erlbaum Associates, 223-250.

Jarausch, Konrad H. and Coclanis, Peter A. (2001): Quantification in History. *In*: Neil J. Smelser and Paul B. Baltes, eds. *International encyclopedia of the social & behavioral sciences*. Amsterdam; New York: Elsevier, 12634-12638.

Jourdenais, Renée (2001): Cognition, instruction, and protocol analysis. *In*: Peter Robinson, ed. *Cognition and second language instruction*. Cambridge: Cambridge University Press, 354-375.

Kelly, Dorothy (2005): *A handbook for translator trainers: a guide to reflective practice*. Manchester: St. Jerome.

Kiraly, Donald C. (1995): *Pathways to translation: from process to pedagogy*. Kent: Kent State University Press.

Krings, Hans P. (2001): *Repairing texts: empirical investigations of machine translation postediting processes* (Translated by Geoffrey Koby, Gregory Shreve, Katjz Mischerikow, *et al.*). Kent: Kent State University Press.

Leow, Ronald P. and Morgan-Short, Kara (2004): To think aloud or not to think aloud: The issue of reactivity in SLA research methodology. *Studies in Second Language Acquisition*. 26(1):35-57.

Li, Defeng (2004): Trustworthiness of think-aloud protocols in the study of translation processes. *International Journal of Applied Linguistics*. 14(3):301-313.

Lörscher, Wolfgang (1991): *Translation Performance, Translation Process, and Translation Strategies: A Psycholinguistic Investigation*. Tübingen: Gunter Narr.

Maietta, Raymond C. (2008): Computer-assisted data analysis. *In*: Lisa M. Given, ed. *The Sage encyclopedia of qualitative research methods*. Los Angeles: SAGE Publications, 103-108.

Mathison, Sandra (1988): Why Triangulate? *Educational Researcher*. 17(2):13-17.

Miles, Matthew B. and Huberman, Alan Michael (1994): *Qualitative data analysis: an expanded sourcebook*. 2nd ed. Thousand Oaks: SAGE Publications.

Morris, Carl N. and Chiu, Calvin (2001): Experimental Design: Large-scale Social Experimentation. *In*: Neil J. Smelser and Paul B. Baltes, eds. *International encyclopedia of the social & behavioral sciences*. Amsterdam: Elsevier, 5086-5090.

Olson, Gary M., Duffy, Susan A. and Mack, Robert L. (1984): Thinking out loud as a method of studying real time comprehension processes. *In*: David E. Kieras and Marcel Adam Just, eds. *New methods in reading comprehension research*. Hillsdale: Erlbaum, 253-286.

PACTE (2009): Results of the validation of the PACTE translation competence model: Acceptability and decision making. *Across Languages and Cultures*. 10(2):207-230.

Patton, Michael Quinn (2002): *Qualitative research and evaluation methods*. 3rd ed. Thousand Oaks: SAGE Publications.

Pomerantz, Eva, Ruble, Diane and Bolger, Niall (2004): Supplementing the snapshots with video footage: taking a developmental approach to understanding social psychological phenomena. *In*: Carol Sansone, Carolyn C. Morf and A. T. Panter, eds. *The Sage handbook of methods in social psychology*. Thousand Oaks: SAGE Publications, 405-426.

Powers, Willow Roberts (2005): *Transcription techniques for the spoken word*. Lanham: AltaMira Press.

Ransdell, Sarah (1995): Generating Thinking-Aloud Protocols: Impact on the Narrative Writing of College Students. *The American Journal of Psychology*. 108(1):89-98.

Rothbauer, Paulette M. (2008): Triangulation. *In*: Lisa M. Given, ed. *Sage encyclopedia of qualitative research methods*. Los Angeles: SAGE Publications, 892-894.

Russo, J. Edward, Johnson, Eric J. and Stephens, Debra L. (1989): The validity of verbal protocols. *Memory & Cognition*. 17(6):759-769.

Shreve, Gregory M. and Angelone, Erik (2010a): Translation and cognition: Recent developments. *In*: Gregory M. Shreve and Erik Angelone, eds. *Translation and cognition*. Amsterdam: John Benjamins, 1-13.

Shreve, Gregory M. and Angelone, Erik, eds. (2010b): *Translation and cognition*. Amsterdam: John Benjamins.

Shreve, Gregory M. and Danks, Joseph H. (1997): Preface. *In*: Joseph H. Danks, Gregory M. Shreve, Stephen B. Fountain*, et al.*, eds. *Cognitive processes in translation and interpreting*. Thousand Oaks: SAGE Publications, xii-ix.

Silverman, David (1985): *Qualitative methodology and sociology: describing the social world*. Brookfield: Gower Publishing.

Smith, Randolph A. and Davis, Stephen F. (2001): *The psychologist as detective: an introduction to conducting research in psychology*. Upper Saddle River: Prentice Hall.

Stratman, James F. and Hamp-Lyons, Liz (1994): Reactivity in concurrent think-aloud protocols. *In*: Peter Smagorinsky, ed. *Speaking about writing: reflections on research methodology*. Thousand Oaks: SAGE Publications, 89-112.

Sullivan, Kirk P. H. and Lindgren, Eva, eds. (2006): *Computer keystroke logging and writing: methods and applications*. Amsterdam: Elsevier.

Tashakkori, Abbas and Teddlie, Charles, eds. (2010): *Sage handbook of mixed methods in social & behavioral research*. 2nd ed. Thousand Oaks: SAGE Publications.

TEI (Last updated: 7 July 2010). *Guidelines for electronic text encoding and interchange*. Visited on 1 September 2010, <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>.

Tirkkonen-Condit, Sonja (2006): Think-Aloud Protocols. *In*: Keith Brown, ed. *The encyclopedia of language and linguistics*. 2nd ed. Boston: Elsevier, 678-686.

Toury, Gideon (1995): *Descriptive translation studies and beyond*. Amsterdam: John Benjamins.

van den Hoonaard, Deborah K. and van den Hoonaard, Will C. (2008): Data analysis. *In*: Lisa M. Given, ed. *The Sage encyclopedia of qualitative research methods*. Los Angeles: SAGE Publications, 186-188.

Venuti, Lawrence, ed. (2000): *The translation studies reader*. London/New York: Routledge.

Wang, Jiawen (2005): *Another Look at Reactivity in L2 Think-aloud Protocols: A Replication Study*. M. A. Thesis, unpublished. East Lansing: Michigan State University.

Webb, Eugene J., Campbell, Donald T., Schwartz, Richard D.*, et al.* (1966): *Unobtrusive measures: nonreactive research in the social sciences*. Chicago: Rand McNally.

APPENDIX

A survey on TAP-based translation process research

Please choose/highlight the items applicable to you, and elaborate if possible.
1. I am currently working on a TAP-based research project.
Yes
No
If yes, could you describe your topic?

2. I am unsure about the implications or significance of such research.
True
False
Please elaborate:

3. I no longer find research questions in this field interesting.
True
False
Please elaborate:

4. I am not convinced about the use (validity, reliability, etc.) of think-aloud methods; I do not think this method is sufficiently rigorous.
True
False
Please elaborate:

5. There are no clear methods for transcription.
True
False
Please elaborate:

6. Have you developed any annotation scheme and used it in transcription?
Yes
No
Please elaborate:

7. There are no clear methods for data analysis.
True
False
Please elaborate:

8. Have you used any specific method or software tool to analyze the transcript/protocols?
Yes
No
Please elaborate:

9. TAP-based research is too labor-intensive and time-consuming.
Yes
No
Please elaborate:

10. Please describe other difficulties you have encountered (e.g., research design, difficulty in finding participants) that might explain why publications on TAP-based research are on the decline recently.