

# Should Revision Trainees Think Aloud while Revising Somebody Else's Translation? Insights from an Empirical Study with Professionals

Isabelle S. Robert and Louise Brunette

Volume 61, Number 2, August 2016

URI: <https://id.erudit.org/iderudit/1037762ar>  
DOI: <https://doi.org/10.7202/1037762ar>

[See table of contents](#)

## Publisher(s)

Les Presses de l'Université de Montréal

## ISSN

0026-0452 (print)  
1492-1421 (digital)

[Explore this journal](#)

## Cite this article

Robert, I. S. & Brunette, L. (2016). Should Revision Trainees Think Aloud while Revising Somebody Else's Translation? Insights from an Empirical Study with Professionals. *Meta*, 61(2), 320–345. <https://doi.org/10.7202/1037762ar>

## Article abstract

This paper reported on a follow-up study whose aim was fourfold: 1) to determine which variables do seem to influence the amount of verbalization of professional revisers when they verbalize their thoughts while revising somebody else's translation, 2) to determine what kind of revision sub-processes are verbalized, 3) to determine the relation between the type of verbalizations and revision product and process, and 4) to draw conclusions for revision didactics. Results show that variables that could have influenced the verbalization ratio of revisers had no effect on that ratio, except the revision experience. As far as verbalized subprocesses are concerned, it appeared that revisers rarely verbalized a maxim-based diagnosis, but that the more they verbalized such a problem representation, the better they detected, the better they revised, but the longer they worked. Results also show that participants who verbalized a problem representation together with a problemsolving strategy or a solution, detected better, but worked longer. Further research could focus on a particular subcompetence of the revision competence: the ability to explain.

# Should Revision Trainees Think Aloud while Revising Somebody Else's Translation? Insights from an Empirical Study with Professionals

**ISABELLE S. ROBERT**

*Universiteit Antwerpen(TricS), Antwerp, Belgium*  
isabelle.robert@uantwerpen.be

**LOUISE BRUNETTE**

*Université du Québec en Outaouais, Gatineau, Canada*  
louise.brunette.uqo.ca

## RÉSUMÉ

Cet article porte sur les résultats d'une étude poursuivant quatre objectifs : 1) déterminer quelles variables influencent le taux de verbalisation de réviseurs professionnels lorsqu'ils pensent tout haut tout en révisant la traduction d'autrui, 2) déterminer quels types de sous-processus de révision sont verbalisés, 3) déterminer la relation entre le type de sous-processus de révision verbalisés et le produit et le processus de révision et 4) tirer des conclusions pour la didactique de la révision. Les résultats montrent que les variables qui auraient pu influencer le taux de verbalisation des réviseurs n'ont pas d'effet sur ce taux, à l'exception de l'expérience en révision. En ce qui concerne le type de sous-processus de révision verbalisés, il semble que les réviseurs verbalisent rarement un diagnostic basé sur une maxime, mais que plus ils verbalisent ce type de représentation de problème, mieux ils détectent les erreurs, mieux ils révisent, mais plus ils travaillent longtemps. Il s'avère également que les réviseurs qui verbalisent à la fois une représentation de problème et une stratégie de résolution de problème détectent mieux les erreurs, mais travaillent plus longtemps également. De nouvelles recherches pourraient dès lors porter sur une sous-compétence particulière de la compétence de révision, à savoir la capacité à expliquer.

## ABSTRACT

This paper reported on a follow-up study whose aim was fourfold: 1) to determine which variables do seem to influence the amount of verbalization of professional revisers when they verbalize their thoughts while revising somebody else's translation, 2) to determine what kind of revision sub-processes are verbalized, 3) to determine the relation between the type of verbalizations and revision product and process, and 4) to draw conclusions for revision didactics. Results show that variables that could have influenced the verbalization ratio of revisers had no effect on that ratio, except the revision experience. As far as verbalized subprocesses are concerned, it appeared that revisers rarely verbalized a maxim-based diagnosis, but that the more they verbalized such a problem representation, the better they detected, the better they revised, but the longer they worked. Results also show that participants who verbalized a problem representation together with a problem-solving strategy or a solution, detected better, but worked longer. Further research could focus on a particular subcompetence of the revision competence: the ability to explain.

## MOTS-CLÉS/KEYWORDS

révision, processus de révision, didactique de la révision, compétence de révision, protocoles de verbalisation

revision, revision process, revision didactics, revision competence, think aloud protocols

## 1. Introduction

Verbalizing or thinking aloud protocols (TAP) as a data elicitation method has been used in translation process research for more than thirty years and is still used, mainly in combination with other methods, such as keystroke logging, video, screen capture, or eye tracking. In a TAP experiment, participants are asked to verbalize their thoughts or think aloud while performing a task. They are not expected to analyze their performance or justify their actions. The method was borrowed from cognitive psychology (Ericsson and Simon 1993) and it was argued that, when elicited with care and appropriate instructions, think-aloud would not change the course or structure of thought processes, except for a slight slowing down of that process.

A review of the literature in Translation Studies about TAP studies (see section 2.1.) demonstrates that there does not seem to be any research to date on the effect of think-aloud on the revision process of somebody else's translation, which is revision proper, according to the European standard for translation services EN 15038 (European Committee for Standardization 2006).

In her study on the effect of translation revision procedures on the revision product and process, Robert (2012) (see also Robert and Van Waes 2014) used think-aloud protocols as data elicitation method, next to keystroke logging, retrospective interview, and textual revisions for the product part of the study. Although it was not the main objective of her study, she observed that verbalization did not seem to have an impact on the revision process and product, since there was no significant correlation between the verbalization percentage (verbalization duration as compared to the entire revision process duration, see section 3.5.2.) and revision quality, revision duration, and error detection potential (see section 3.5.1.).

Therefore, this paper reports on a follow-up study drawing on Robert's 2012 study data and focusing on the following new research questions:

1. Which variables do influence the amount of verbalization?
2. What kind of revision sub-processes are verbalized?
3. Is there a relation between the type of verbalized sub-processes and revision quality, duration, and error detection potential?
4. Can we draw conclusions for revision didactics?

The following section will be dedicated to a literature review of think-aloud in Translation Studies. It will be followed by a short description of the theoretical framework of the revision process and its sub-processes, in order to frame the second and third research question, and by a section about revision competence, to frame the fourth research question. These more theoretical sections will be followed by the methodology and the results section. In the conclusions, we will assess the relevance of these results for further research in the didactics of revision.

## 2. Theoretical background

### 2.1. *Think-aloud in Translation Studies*

As a research method into the translation process, think-aloud was very popular in the eighties and nineties. In her annotated bibliography of think-aloud studies into translation, Jääskeläinen (2002) discusses more than one hundred studies drawing

mainly on TAP data. However, at about the same time, many researchers were questioning the validity of the method, Jääskeläinen (2000: 80) too, stating that to her, “not enough attention has been paid to testing and refining methods of data analysis in TAP studies on translating [...]” Similarly, Bernardini (2001: 260) argued that it was time to check the validity of all ‘informal’ hypotheses drawing on the use of TAP data by means of more controlled experimental designs and systematic methods of data coding and analysis. Later, Li (2004) reviewed 15 published reports on investigations of translation processes using TAPs and concluded that the research designs left much to be desired. Consequently, this undermined the trustworthiness of many of the findings. Like Bernardini, Li (2004: 310) concluded that the findings from these studies constituted working hypotheses that “needed to be confirmed or refuted by future research relying on more rigor and trustworthiness.”

### *2.1.1. Problems and limitations of TAPs*

In his study on post-editing processes, Krings (2001: 214-233) reviewed the main possibilities and limitations of the method. He discusses the three problems generally associated with it: 1) the problem of consistency or validity, that is, “do consistent (systematic) connections exist between verbalizations and cognitive processes?,” 2) the problem of interference, that is, “does performance of the verbalization task itself change the normal course of cognitive processes?,” and 3) the problem of completeness, that is, “can verbalization contain complete records of cognitive processes?”

The very first problem, that of consistency or validity, is not, according to Krings (2001: 214-233), a typical problem of verbal report data. He argues that although the assumption of a correspondence between verbalizations and processes is in doubt, all other methods using externally observable behavior (such as eye movement) to arrive at conclusions concerning cognitive processes are also based on such assumptions of correspondence and he concludes that “the use of Think Aloud with language-related cognitive processes is the form of verbal reports data use having the least problems with data validity” (Krings 2001: 226).

As far as interference is concerned, Krings (2001: 227) starts with one of the most frequent arguments against TAP, that is, that thinking aloud is an unnatural task that does not occur in everyday life. He argues that it is only partly correct since many people tend to accompany their problem solving processes with a kind of quiet speech. This has also been observed by Robert (2012) and Göpferich and Jääskeläinen (2009).

The second interference aspect is that of the slowing down effect of thinking aloud on the process, which was acknowledged by Ericsson and Simon (1993). In his study, Krings (2001: 227) noted that the participants in the thinking aloud condition worked 30% longer than the participants in the silent condition. By the same token, Jakobsen (2003), who studied the effects of think-aloud on translation speed, revision, and segmentation, observed that thinking aloud delayed translation by about 25%. However, as noted by Krings (2001: 229), the central question of the interference problem is the question of whether there is a qualitative influence of the verbalization task on the primary task. He concludes that based on the current state of research he can assume that that influence is minor, apart from the obvious slow-down effect. The results of later studies do not converge, as explained hereafter.

Jakobsen (2003) observed that thinking aloud did force participants to process text in smaller segments, but that it did not have a significant effect on quality of self-revision. Consequently, he concluded that his findings did not invalidate TAP as a method. On the other side, Hansen (2005) studied the possible impact of thinking aloud on translation product and process and also tried to determine what can be learned from introspective methods like think-aloud and retrospection. She states that “TA must have an impact both on the thought processes, on the translation process and on the translation product” (Hansen 2005: 519). However, Englund Dimitrova concluded, just like Jakobsen (2003), that “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).

Finally, the problem of completeness. According to Krings (2001: 232), it has not yet been sufficiently resolved. He acknowledges that TA data on cognitive processes consciously occurring in short-term memory do not seem to be always complete and that this should be taken into consideration during the analysis of TA data. However, he stresses the fact that think-aloud only provides information about consciously occurring cognitive processes and thus, not about a lot of automatic processes which are also part of any cognitive problem-solving task. Again, Hansen did not seem to be as positive with regard to completeness, stating that “[w]hat is verbalized is a conglomerate of memories, reflections, justifications, explanations, emotions and experiences, and it seems likely that these cannot be separated from each other, even when we use special reminders or retrieval cues” (Hansen 2005: 519).

### *2.1.2. TAPs in recent studies*

Think-aloud as a data elicitation method has continued to trigger the attention of many researchers, in Translations Studies, but also in related disciplines. In 2008, in a major reference book (in German) on translation process research (state of the art, methods, and perspectives), Göpferich (2008: 22) reaffirmed that there was no method, until then, that yielded more information on complex cognitive problem-solving processes than TA. Bowles (2010) in language acquisition research, shed some new light on the problematic. She conducted a meta-analysis of studies with TA. She concluded that the answer to the question of the reactivity of think-aloud (that is, whether it changes the thought process) is not “one-size-fits-all” and that verbal reports can reliably be used as a data collection tool.

In Translation Studies, Jääskeläinen recently announced a study aimed at determining the validity and reliability of verbal report data on translation processes, which will be carried out as a “joint international project consisting of language- and culture-specific sub-projects, with a carefully designed and uniform research design, including strict procedures for the setting of the experiments and instructions to be given to the subjects” (Jääskeläinen 2011: 23). She argues that validity can be an issue with methods in which the situation is somehow manipulated and subjected to experimental control and asks the question whether, in that case, the experimental situation does change the process under study. She also wonders whether thinking aloud does change the process to such an extent that it shows in the product. To her knowledge, there are no systematic studies focusing on the effect of TA on the product. She also points out that the variables potentially affecting the fluency of verbalizing have been overlooked in research and that the actual amount of data elicited is

usually not reported. She concludes that systematic methodological research aimed at identifying the conditions and limitations of verbal reports on translation processes are needed, hence the announced study.

The very same year, in 2011, think-aloud was again the focus of a study dedicated to methodological considerations in translation process research. Sun (2011) reported on a questionnaire survey conducted in 2009 among 25 eminent translation process researchers worldwide. He had observed that think-aloud based translation process research seemed to be on the decline in recent years and wanted to investigate the current level of interest in TAP research. Although results showed that only 7 responders were working on a TAP-project at the time, no fewer than 23 of them still did not consider TAP-based translation research to be insignificant or uninteresting. Many of them still believed that the method has potential for interesting insights into cognitive processes, but also limitations. In his general conclusion, the author states that “there is, to date, no strong evidence suggesting that TAP significantly changes or influences the translation process” (Sun 2011: 946).

Finally, sharing the same interest for methodological issues as Jääskeläinen and Sun, Muñoz Martín (2012: 15) made some suggestions to normalize empirical research in cognitive translatology, pointing to the fact that “our problems [with scientific methods] are by no means exclusive to translatology.”

This review of the literature in Translation Studies about TAP studies demonstrates that there does not seem to be a clear-cut answer to the question of the reactivity of think-aloud in translation process studies and that there is no research to date on the effect of think-aloud on the revision process of somebody else’s translation, hence this follow-up study.

## 2.2. *The revision process*

To our knowledge, there is no model of the translation revision process. However, translation revision is very close to the revision process we find in writing, and of the revision process we find in editing. In the first case, a writer revises his own text, in the second, an editor revises somebody else’s text, and in both cases, the text is not a translation.

Accordingly, writing studies, editing studies, and translation studies appear to have a comparable process as their object of study. The writing process is made of three main subprocesses, that is, planning, translating and reviewing (Hayes, Flower, *et al.* 1987) which can be compared with the three main subprocesses of translation called ‘orientation,’ ‘drafting,’ and ‘revision phase,’ and described by researchers such as Jakobsen (2002) and Norberg (2003), for example. In writing studies, the subprocess of reviewing (also called revision) has received much attention and several revision process models have been developed and discussed (Allal, Chanquoy, *et al.* 2004). However, in translation studies, even if some attention has been paid to revision as a subprocess of translation, that is, self-revision (Breedveld 2002; Breedveld and van den Bergh 2002; Shih 2006), very little attention has been paid to the revision process of somebody else’s translation.

In editing studies, the object of study is the revision of an original text. Definitions of the process of revising a translation and editing a text which is not a translation are consistent:

Revising is that function of professional translators in which they identify features of the draft translation that fall short of what is acceptable and make appropriate corrections and improvements. (Mossop 2007: 109)

[Professional editing is] an activity that consists in comprehending and evaluating a text written by a given author and in making modifications to this text in accordance with the assignment or mandate given by a client. Such modifications may target aspects of information, organization, or form with a view to improving the quality of the text and enhancing its communicational effectiveness. (Bisaillon 2007: 296)

As a consequence, this does not come as a surprise that revision models used in editing studies are based on revision models from writing studies. The follow-up study this paper reports on used Bisaillon's model (2007) of the revision process of a professional editor as a framework for the typology of revision sub-processes. As in Hayes' model (Hayes, Flower, *et al.* 1987), the revision process in Bisaillon's model starts with the reading of the text for comprehension and evaluation. It is followed by the *detection of a problem* whose *representation* is located on a continuum going from ill-defined to well-defined. Accordingly, "representations range from spare representations that contain little information about the problem to richly elaborated diagnoses that offer both conceptual and procedural information about the problem" (Hayes, Flower, *et al.* 1987: 211) Problem representations at the low end of the continuum are labeled 'detections.' As soon as the writer has a relatively well-defined representation of a text problem that is based on a categorization of the problem type, the authors speak of a 'diagnosis.' However, even diagnoses vary in their level of specificity. Consequently, the authors distinguish between 1) intentional diagnosis, 2) rule-based diagnosis and 3) maximbased diagnosis.

Intentional diagnoses are the least well-defined diagnoses. The reviser compares his representation of the text's purpose and/or the author's intentions with what is actually written. The most well-defined diagnoses are rule-based. The reviser has a ready-made problem representation that is based on his knowledge of violations of rules of grammar, punctuation, and spelling. He also has a built-in action to solve the problem. Statements that fall between these two extremes were classified by the authors as maxim-based. "Here, revisers have some established guidelines, but not any clear-cut rules on which to base their diagnoses" (Hayes, Flower, *et al.* 1987: 214).

### 2.3. *The revision competence*

Although translation revision remains marginal in empirical studies, it has been the focus of at least two major handbooks: *Pratique de la révision* (Horguelin and Brunette 1998) and *Editing and Revising for Translators* (Mossop 2001/2007). These handbooks do not provide for a revision process model, but they point to some specific competences of a reviser, such as interpersonal aptitudes and the capacity to justify one's changes. It has to be noted that revision has a long tradition in Canada with authors like Thaon and Horguelin (1980) and again Mossop (1992) who defined the goals of a revision course, such as learning to achieve the mental switch from a 'retranslating' to a 'revising' frame of mind and learning to justify changes, which seems to be specific to the task of a reviser and thus, a specific revision competence.

Drawing on PACTE's model of translation competence (2003), Künzli (2006) formulates a few suggestions for teaching and learning revision based on a think-



aloud protocol study. He focuses on the acquisition of the strategic competence, the interpersonal competence and the professional and instrumental competence. As far as the interpersonal competence is concerned, Künzli states that, on completion of the module, trainees should be able to “give meaningful feedback” and to “justify the changes made in the draft translation, and to communicate their decisions constructively [...]” (Künzli 2006: 19). This goal is in line with the aptitudes Mossop and Horguelin and Brunette refer to.

In 2009, reporting on a longitudinal study conducted between 2003 and 2008 and aimed at defining the relationship between translation competence and revision competence, Hansen proposed a model of the translation competence with many subcompetences common to translation and revision (‘Empathy,’ ‘Loyalty,’ ‘Ability to abstract,’ ‘Attentiveness,’ ‘Creativity,’ ‘Overview,’ ‘Courage,’ ‘Ability to take decisions,’ ‘Accuracy,’ ‘Ability to use aids, resource persons’). The revision competence model reveals two additional subcompetences that are not included in the translation counterpart, that is ‘Fairness,’ ‘Tolerance’ (changes, improvements, classifications, gradings), and ‘Ability to explain’ (argumentation, justification, clarification of changes). Again, this proposal is in line with Künzli’s suggestions and handbooks on the topic.

Accordingly, revision competence seems to be different from translation competence, but no specific research seems to have been dedicated yet to that aspect of the translation process in the broad sense of the word. However, a translation revision competence model is being developed by Robert, Rемаel and Ureel (2016) at the University of Antwerp.

The importance of revision has nevertheless been acknowledged on the translation market and in translation education. In the translation business, revision has been a compulsory step of the translation process since 2006 for those translation services providers who want to work according to the EN 15038 European standard for translation services (European Committee for Standardization 2006). However, the specific revision competences included in Hansen’s revision model (2009) do not appear in the standard. The document only states that the reviser should have the same translation competences as the translator, plus ‘translating experience in the domain under consideration.’ In translation education, the need to train revisers has been addressed recently by several authors. In 2008, Schjoldager, Wølch Rasmussen and Thomsen reported on a pilot module on *précis*-writing, revision and editing, especially developed for the European Master’s in Translation (EMT). Their findings are based on an exploratory survey of the translation industry internationally and in Denmark and reveal that there is a need for translators to be trained to carry out editing/revision and *précis*-writing. These findings conform to Kruger’s demand to integrate editing courses in university programs (Kruger 2008). In Canada, this need was acknowledged more than 30 years ago by Paul Horguelin, and the first French-English handbook to this end published in 1980 (Thaon and Horguelin 1980).

Similarly, Biel discussed the implications of the EN 15038 European standard of translation services for educational institutions. The author argues that “the standard gains increasingly wider recognition all over Europe, which exerts certain pressure on educational institutions to train translators in line with its requirements” (Biel 2011: 61). She adds “trainers should aim at developing the professional competences specified in the standard and cover all phases of the translation service provision



[...]” (Biel 2011: 70). This obviously includes revision. Finally, although her research concerns the work of revisers and translators at inter-governmental organizations, Lafeber (2012) showed that translators and revisers need analytical, research, technological, interpersonal, and time-management skills, which is in line with the EMT framework.

In conclusion, translation revision is a reality and translation revision competence and related activities such as post-editing, will probably be the focus of more research in a near future.

### 3. Methodology

In the 2012 study on which this follow-up study is based (Robert 2012), sixteen professional revisers were asked to revise four comparable target texts (called text 1, text 2, text 3 and text 4), while thinking aloud. For each text, they had to use one of the four revision procedures to be tested, that is, (a) a single monolingual re-reading (M), (b) a single bilingual re-reading (B), (c) a comparative re-reading followed by a monolingual re-reading (BM), and (d) a monolingual re-reading followed by a comparative re-reading (MB). Consequently, they used all four procedures just one time. The order in which the revisers revised the texts, which is called the ‘task order,’ was determined by the researcher and was counterbalanced, to ensure that neither the same text nor the same procedure was always the first task. For example, for reviser 1, the first task consisted in revising text 1 with procedure M, whereas for reviser 2, the first task consisted in revising text 1 with procedure B. Revisers had to take into account the revision parameters mentioned in the revision brief.

#### 3.1. Participants

The participants (N=16) were all professional revisers with at least one year of experience in revision for the language pair Dutch-French. They were either employees in a translation agency or freelance translators/revisers (between 23 and 41 years old). The majority had either a Master’s degree in translation or a Master’s degree in Germanic languages, with Dutch as one of their working languages and French as their mother tongue. For their participation in the study, the participants received a fee of 40 € (vat not included), based on the usual fee at the time of the experiment. For more details, see table A in the Appendix.

#### 3.2. Materials

The target texts to be revised had to be ‘comparable’ so as to make sure that the variable ‘text’ would be as constant as possible, in order to focus on the impact of the independent variable only (procedure). Therefore, four press releases in Dutch of approximately the same length (500 words) were selected as source texts and translated into French by native Master’s students. Accordingly, the target texts belonged to the same genre and had the same function. In addition, after manipulation by the researcher, they contained approximately the same proportion of authentic ‘errors’ (called ‘items’ in the study), which were considered errors by a panel of translation professionals and/or translation lecturers (N=9). Subsequently, the items were each

classified according to three revision parameters based on Horguelin and Brunette (1998) and Mossop (2001): 1) accuracy, 2) linguistic coding (in the broad sense, that is, including readability), and 3) appropriateness.

### 3.3. *Experimental setting*

The experiments were in agreement with the subjects' schedule, to make sure that the tasks would be considered as any other 'ordinary' planned revision task. No time limit was set, but revisers were asked to try not to work longer than two and a half hours (four tasks). The target texts had to be revised one by one, with a few minutes' interruption.

Almost all experiments were conducted at the subjects' work place, to enhance ecological validity. All instructions previously given by mail were repeated orally just before the experiment. The source texts were available on paper, but the target texts were available only in electronic format in an MS Word file. The subjects were allowed to use all the usual translation and language tools they were familiar with when revising from Dutch into French, but they had to work on the researcher's laptop (to allow Inputlog to record the process, see section 3.4.). In addition, they had access on that laptop to a range of electronic dictionaries and for obvious ecological reasons, they also had access to the Internet.

### 3.4. *Data collection*

Four data collection instruments were used to ensure triangulation: a) revision product analysis based on the changes made in the final version (for the product part of the research), b) think-aloud protocols, c) the keystroke logging software Inputlog (Leijten and Van Waes 2006; 2013), and d) short retrospective interviews (for the process part of the research).

Product data were collected using the 'compare documents function' in MS Word that makes it possible to reveal all changes made to a document. As far as the process data are concerned, a digital recorder was used to record the subjects' verbalizations (think-aloud protocol). The other tool used to collect process data was the keystroke logging software Inputlog<sup>1</sup> that records all keyboard and mouse movements (and related timestamps) without interfering in the process. Finally, short retrospective interviews were conducted to get a detailed profile of each subject. All details are summarized in table A in the appendix.

### 3.5. *Data analysis*

#### 3.5.1. *Revision quality, error detection potential, and revision duration*

As mentioned in the introduction, Robert (2012) observed that verbalization did not seem to have an impact on the revision process and product, since there was no significant correlation between the verbalization percentage or ratio (verbalization duration as compared to process duration) and revision quality, revision duration, and error detection potential.

As far as revision quality is concerned, a score for each reviser was calculated on the basis of the percentage of 'justified changes' for each task. Justified changes are

modifications that correct an error (manipulated ‘item’) in a proper way, on the basis of a consensus among the panel of translation professionals and lecturers (see section 3.2.). In other words, a participant who made 10 justified changes in a text containing 20 items scored 50%. Accordingly, 16 scores for each procedure could be calculated and, consequently, a mean for each series or procedure.

The error-detection potential is a measure of the identification or detection of a problem, or, as explained in section 2.2, a so-called ‘problem representation.’ The error detection potential was calculated for each task on the basis of the sum of ‘justified changes,’ ‘underrevisions,’ and ‘simple detections.’ ‘Underrevisions’ are modifications that attempt to correct an error or item, but that were found unsatisfying by the panel of translation professionals and lecturers. ‘Simple detections’ are items for which no visible change has been made in the final version but for which a detection or identification of a problem was identified by the researcher on the basis of a verbalization (for example, the reviser formulated a doubt) and/or a modification attempt or search action concerning an item, as revealed by the general logging file of Inputlog (for instance, the participant searched for a word in a dictionary). In other words, a participant who made 10 ‘justified changes,’ two ‘underrevisions,’ and three ‘simple detections’ in a text containing 20 items scored 75%. Again, 16 scores for each procedure could be calculated and, consequently, a mean for each series or procedure.

The duration of each task was measured in minutes and seconds on the basis of the subjects’ verbalizations.

### 3.5.2. *Verbalization ratio*

The verbalization percentage or ratio was measured drawing on the think-aloud protocols. Each stop of at least one second was registered as a pause, which means that two verbalization sequences were always separated by a pause of at least one second. Accordingly, for each task, the total revision process duration was measured, which consists of the total duration of the verbalization on the one hand, and the total duration of the pauses on the other hand. The verbalization ratio or percentage is therefore the ratio between the verbalization duration and the total process duration. For example, with a total process duration of 30 minutes and a verbalization duration of 15 minutes, a reviser gets a verbalization ratio or percentage of 50%. Verbalization ratio was preferred to the amount of verbalization in minutes and seconds because, obviously, people who worked longer generally verbalized more.

### 3.5.3. *Revision subprocesses*

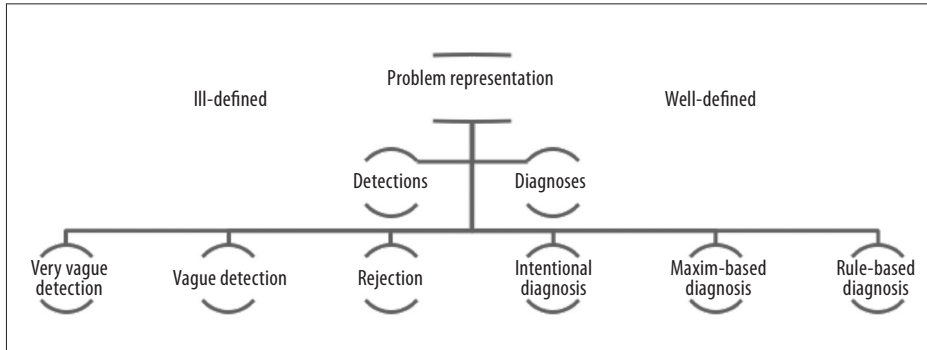
Each verbalization fragment was coded with respect to the subprocesses going on. The coding system was developed drawing on Bisailon’s model of revision (see section 2.2.), with a focus on the type of problem representation, the type of problem-solving strategy, and the immediate solution.

With respect to problem representation, we distinguished between ‘very vague detection’ (for example, interrogative intonation, an interjection expressing doubt, surprise), ‘vague detection’ (for instance, the reviser says he doesn’t like it, that it sounds strange), ‘rejection’ (for example, the reviser says “no,” “it’s not that”), intentional diagnosis (for instance, the reviser says that a part of a sentence is missing), ‘maxim-based diagnosis’ (for example, the reviser says that a sentence is too long, that the formulation is awkward), and ‘rule-based diagnosis’ (for instance, the reviser

says that a capital letter is needed). The least well-defined problem representations (that is, very vague detection, vague detection and rejection) have been put together under the label ‘detection,’ whereas the well-defined ones (that is, intentional, maxim-based and rule-based diagnoses) have been taken together under the label ‘diagnosis.’ This coding is illustrated in figure 1. The coding of problem-solving strategies is detailed in Robert (2012) and further fine-grained in Robert (2014).

FIGURE 1

**Problem representation as structured in the study**



*3.5.4. Richness of the verbalizations*

The ‘richness’ of the verbalizations has also been operationalized. A verbalization fragment can be coded for problem representation, strategy, and solution. However, sometimes, a combination was observed. Consequently, we distinguished between 1) verbalization of a problem representation alone, 2) verbalization of a problem-solving strategy alone, 3) verbalization of an immediate solution alone, 4) verbalization of a problem representation and a problem-solving strategy, and 5) verbalization of a problem representation and an immediate solution.

**4. Results and discussion**

In order to answer the research questions, statistical tests were carried out with SPSS 20.0 for Windows. Parametric tests were conducted when the following assumptions for these tests were met: normally distributed data, homogeneity of variance, interval data, and independence (Field 2009: 133). The Kolmogorov-Smirnov test of normality was used to test the distribution of the scores. The Levene’s test was used to test the hypothesis that the variances in different groups are equal. In repeated-measures design, the participants are measured in more than one condition, which was the case in this study. Since the assumptions related to parametric tests were not always met, the non-parametric Friedman test of analysis of variance had to be used. However, when the test was significant, the parametric variant (one-way ANOVA for repeated measures) was performed as well, because it allows for pairwise comparisons called post hoc tests.

#### 4.1. Variables with a potential effect on verbalization ratio

##### 4.1.1. Subjects' profile

In this section, the potential effect of subjects' profile variables on verbalization ratio will be investigated: gender, age, experience in translation, experience in revision, professional status and status of Dutch in the curriculum.

As far as the gender is concerned, it seems that female revisers verbalized more than male (38% versus 26%), but the difference between the two groups is not significant ( $t=1.182$ ,  $p=.257$ ). However, this test might not be recommended, since both groups were different with regard to the number of participants (2 versus 14). With respect to age, no correlation was found between age and verbalization ratio:  $r=.157$ ,  $p=.281$ . Experience in translation does not seem to play a role either. No significant correlation was found:  $r=.285$ ,  $p=.143$ . In contrast, experience in revision does seem to play a role: a significant positive correlation was found, with  $r=.496$ ,  $p=.025$ , which means that the more experience they have in revision, the more revisers tend to verbalize.

The status of the subjects, that is whether they were employees or freelancers, seemed to have no effect on verbalization ratio. Although freelancers appear to verbalize a little bit more than employees (45.6% versus 33.5%), the difference between the two groups is not significant ( $t=-1.626$ ,  $p=.126$ ). Similarly, the academic degree has no effect on the verbalization ratio either. There is no difference between the 'Masterdegree in translation' group and the 'no Master degree in translation' group ( $t=-.168$ ,  $p=.869$ ), although translators seem to verbalize a little bit more (37% versus 36%) than non-translators.

Finally, the status of Dutch (source language) in the studies of the subjects has no impact either. Subjects who studied Dutch as a B-language during their Bachelor and Master in translation or philology verbalize less than those who studied Dutch in secondary school only or in an additional Master's year (32% versus 43%), but the difference is not significant ( $t=1.789$ ,  $p=.095$ ). All descriptive statistics are shown in table 1.

TABLE 1  
Descriptive statistics related to gender, status, degree, and status of Dutch

	N	Mean	Std. Deviation	Std. Error Mean
<b>Gender</b>				
Female	14	38.08	12.89	3.44
Male	2	26.17	18.06	12.77
<b>Status</b>				
Employee	12	33.58	13.97	4.03
Freelance	4	45.62	7.11	3.56
<b>Degree</b>				
No Master in translation	5	35.72	15.61	6.98
Master in translation	11	36.99	13.25	3.99
<b>Status of Dutch</b>				
B language	9	31.62	14.95	4.98
Other	7	42.99	8.54	3.23

As a conclusion, it can be said that the subjects' profile variables under investigation, that is, gender, age, experience in translation, experience in revision, status, degree and status of Dutch, do not seem to have influenced the verbalization ratio of subjects, with the exception of experience in revision. A significant positive correlation was observed between verbalization ratio and experience in revision, expressed in years.

#### 4.1.2. Task characteristics

In this section, the potential effect of the task characteristics on verbalization ratio will be investigated: revision procedure (M, B, BM or MB), text (1, 2, 3 and 4), task order (1<sup>st</sup>, 2<sup>nd</sup>, 3<sup>rd</sup>, 4<sup>th</sup>, as explained in section 3), environment, and time of the experiment.

The effect of the revision procedure on verbalization ratio was measured through a non-parametric ANOVA test for repeated measures. As can be seen from table 2, the means are very close to each other. The test was not significant:  $\chi^2=.750$ ,  $p=.872$ . The same can be said about the texts: they did not seem to have an effect on the verbalization ratio (means are very close, see table 2) and the test was not significant:  $\chi^2=.825$ ,  $p=.858$ . Similarly, the order of the tasks had no effect either, even if revisers seemed to have verbalized a little bit more at each task, except for the last one. All means are very close to each other (see table 2). The test, again, was not significant:  $\chi^2=2.325$ ,  $p=.534$ .

TABLE 2  
Descriptive statistics of verbalization ratio

Verbalization ratio				
	Procedure M	Procedure B	Procedure BM	Procedure MB
Mean (%)	38.11	35.52	38.07	36.20
SD	16.06	14.07	14.06	14.65
Rank	2.56	2.31	2.44	2.69
	Text 1	Text 2	Text 3	Text 4
Mean (%)	38.02	36.56	35.16	38.16
SD	16.28	12.87	14.61	14.96
Rank	2.44	2.38	2.44	2.75
	1st Task	2nd Task	3rd Task	4th Task
Mean (%)	35.69	36.33	38.73	37.16
SD	14.65	13.63	13.00	17.29
Rank	2.31	2.88	2.56	2.25

As far as the environment is concerned, only one subject did not work at his usual work place. He verbalized less than the average (13% versus 37%), but was not the one that verbalized the least. Finally, although subjects that worked in the morning seemed to verbalize more than those who worked in the afternoon (40% versus 33%, see table 3 for descriptive statistics), again, there was no significant difference between the two groups ( $t=1.095$ ,  $p=.292$ ) as far as the timing of the experiment is concerned.

TABLE 3  
Descriptive statistics for timing

	N	Mean	Std. Deviation	Std. Error Mean
<b>Timing</b>				
Morning	8	40.27	10.14	3.58
Afternoon	8	32.92	16.03	5.67

As a conclusion, it can be said that the task characteristics under investigation, that is, revision procedure, text, task order, environment, and time of the experiment, do not seem to have influenced the verbalization ratio of subjects.

#### 4.2. Verbalized revision subprocesses

As explained in section 3.5.3, each verbalization fragment was coded with respect to the subprocesses going on: problem representation, problem-solving strategy, and immediate solution.

As far as problem representation is concerned, we distinguished between six types of representations, that is, 1) 'very vague detection,' 2) 'vague detection,' 3) 'rejection,' 4) 'intentional diagnosis,' 5) 'maxim-based diagnosis,' and 6) 'rule-based diagnosis.' The first three were labeled 'detection' and the last three 'diagnosis,' according to Hayes's model (Hayes, Flower, *et al.* 1987), as shown in figure 1.

Results regarding the strategies alone will not be reported here, since they have been addressed in another publication (Robert 2014). The 'richness' of the verbalizations, however, will also be reported: a verbalization fragment including the verbalization of a problem representation plus the verbalization of a problem-solving strategy, or including the verbalization of a problem representation plus the verbalization of an immediate solution, was considered 'richer' than a fragment including just one of them.

In the following sections, we will focus on the frequency of each type of problem representation (4.2.1), and then on the relation between each type of problem representation and the three variables under study, that is, revision quality, error detection potential, and revision duration (4.2.2). After that, we will investigate the 'richness' of the verbalizations in a descriptive way (4.2.3), and then focus on the relations between that richness and revision quality, error detection potential, and revision duration (4.2.4).

##### 4.2.1. Verbalized problem representation types

The means in table 4 have been calculated drawing on the percentage for each participant, that is, the number of times they verbalized a particular problem representation, divided by the total number of verbalized problem representations, including the absence of a verbalized problem representation.

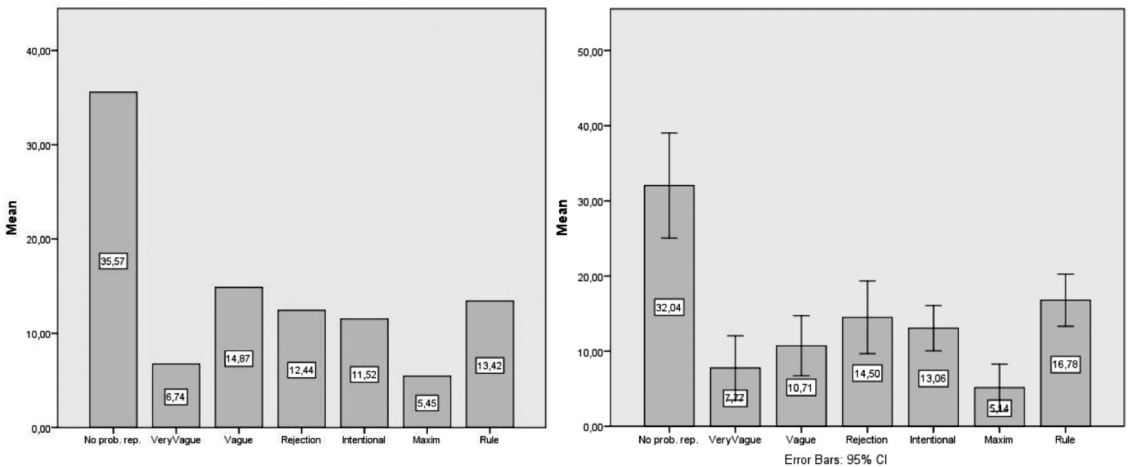


TABLE 4  
**Descriptive statistics for problem representations, per type**

	Error detection		Quality		N
	Mean	SD	Mean	SD	
No prob. representation	35.57	13.64	32.04	13.13	16
Very vague detection	6.74	6.60	7.77	8.04	16
Vague detection	14.87	6.65	10.71	7.50	16
Rejection	12.44	8.08	14.50	9.08	16
Intentional diagnosis	11.52	5.32	13.06	5.64	16
Maxim-based diagnosis	5.45	5.48	5.14	5.89	16
Rule-based diagnosis	13.42	5.12	16.78	6.51	16

As illustrated in figures 2 and 3, the absence of a verbalized problem representation is very frequent. In other words, in these cases, participants did not verbalize a problem representation, but detected and/or revised the problem. When we look at items that have at least been detected (error detection potential, figure 2), the verbalization of vague detection comes second, followed by rule-based diagnoses, rejections, intentional diagnoses, very vague detections, and maxim-based diagnoses. When we look at items that have been detected and revised properly (revision quality, figure 3), the ranking is slightly different: 1) absence of verbalization, 2) rule-based diagnoses, 3) rejections, 4) intentional diagnoses, 5) vague detections, 6) very vague detections, and 7) maxim-based diagnoses.

FIGURES 2 AND 3  
**Mean percentages of problem representations, per type, for error detection and quality**



A test of analysis of variance (ANOVA for repeated measures) was conducted to determine whether the differences between the means were significant, that is, to determine whether there are significant differences in the frequency with which the problem representations were verbalized. Since all series of scores were normally distributed, the parametric variant was chosen, because it allows for post hoc tests, and thus, for pairwise comparisons. The tests were significant, with  $F(6,16)=22.67, p<.001$  (error detection) and  $F(616)=15.25, p<.001$  (quality). Post hoc tests revealed a

significant difference between the absence of a verbalized problem representation and all other types of problem representations, in both the 'quality' and the 'error detection' rankings. All results are shown in table B in the appendix. In the first column, each type of problem representation is included. It is compared to each other type of problem representation mentioned in the second column, for both detection potential and quality. For example, the difference in mean between the vague problem representation and the maximbased problem representation appears to be significant as far as error detection is concerned.

When the results shown in table B in the appendix are combined with the ranking shown in figures 2 and 3, the following observations can be formulated: in the 'error detection' ranking, vague detections, which come second, are significantly less frequent than no problem representation, but as frequent as all other types, except the least frequent one (maxim-based). The same can be said about the third problem representation in the quality ranking, that is, rule-based diagnoses: it is significantly less frequent than no problem representation, but as frequent as all other types, except the least frequent one. Number four (rejection), five (intentional diagnosis), and six (very vague detection) are significantly less frequent than no problem representation, but as frequent as all other types.

In other words, problem representations in the error detection ranking do not seem to differ in frequency, except the maxim-based problem representation which is significantly less frequent than rule-based diagnoses and vague detections. In the quality ranking, approximately the same observations have been made: problem representations do not seem to differ in frequency, except the maxim-based problem representation which is significantly less frequent than rule-based diagnoses, rejections, and intentional diagnoses.

As announced before, problem representations can be grouped under 'detections' on the one hand, and 'diagnoses' on the other hand, as described by Hayes and Flower, *et al.* (1987: 213; see also figure 1). However, when compared, there are no significant differences between no verbalization of a problem representation, the verbalization of a detection, or the verbalization of a diagnosis, for both error detection ( $F(2,16)=0.560, p>.05$ ), and quality ( $F(2,16)=0.179, p>.05$ ). All descriptive statistics are summarized in table 5.

TABLE 5  
Descriptive statistics for problem representation, grouped under 'detections' and 'diagnoses'

	Error detection		Quality		N
	Mean	SD	Mean	SD	
No prob. representation	35.57	13.634	32.04	13.13	16
Detection	34.05	9.88	32.98	9.66	16
Diagnosis	30.39	11.02	34.98	11.72	16

To sum up, it can be said that what revisers did the most, when at least detecting or even correcting an item, is not verbalizing any problem representation at all. When they did, they verbalized all types of problem representations as frequently as one another, except for the maxim-based detection which seems less frequent. When problem representation types are grouped under 'detections' and 'diagnoses,' there is no difference in frequency.

4.2.2. *Relation between verbalized problem representation types and error detection potential, revision quality, and revision duration*

To determine the potential relation between the type of verbalizations and error detection potential, revision quality, and revision duration, Pearson correlation tests were carried out. All results are summarized in table 6.

TABLE 6

**Correlation tests for error detection, quality, and duration, and problem representations (N=16)**

	No prob. rep.	Very vague	Vague	Rejection	Intentional	Maxim-based	Rule-based
	r	r	r	r	r	r	r
	p (1-tailed)	p (1-tailed)	p (1-tailed)	p (1-tailed)	p (1-tailed)	p (1-tailed)	p (1-tailed)
Error	-.298	-.403	.455*	-.126	.183	.508*	.188
detection	.131	.061	.038	.321	.249	.022	.243
Quality	-.216	-.466*	.612**	-.184	.219	.507*	-.086
	.211	.034	.006	.247	.208	.023	.375
Duration	-.357	-.633**	.182	.171	.405	.494*	.310
	.088	.004	.250	.263	.060	.026	.121

\* Correlation is significant at the .05 level (1-tailed).

\*\* Correlation is significant at the .01 level.

With regard to items that have been at least been detected (error detection potential), a significant correlation has been observed with vague detections and with maxim-based diagnoses. As far as quality is concerned, the same correlations have been observed, plus a negative correlation with very vague detections. Finally, as to revision duration, a positive correlation has been observed with maxim-based diagnoses, and a negative correlation with very vague detections.

As described above, problem representations can be grouped under 'detections' on the one hand, and 'diagnoses' on the other hand. As shown in table 7, there is a positive correlation between the verbalization of diagnoses and the error detection potential, but again a negative correlation with the revision duration.

TABLE 7

**Correlation tests for error detection, quality, and duration, and problem representations, grouped under 'detections' and 'diagnoses' (N=16)**

	No prob. rep.	Detections	Diagnoses
	r	r	r
	p (1-tailed)	p (1-tailed)	p (1-tailed)
Error	-.298	-.066	.428*
detection	.131	.404	.049
Quality	-.216	-.086	.312
	.211	.376	.120
Duration	-.357	-.161	.586**
	.088	.276	.009

\* Correlation is significant at the .05 level (1-tailed).

\*\* Correlation is significant at the .01 level.

As a conclusion, it can be said that the more participants verbalized a vague detection or a maximbased diagnosis, the better they detected. The same can be said about quality: the more participants verbalized a vague detection or a maxim-based diagnosis, the better they revised. Inversely, the more very vague detections they verbalized, the worse they revised. As far as duration is concerned, it appears to be significantly associated with the verbalization of maxim-based diagnoses, which means that the verbalizations of these diagnoses take time. Inversely, duration is negatively associated with the verbalization of very vague detections, which is rather logical, since these detections are generally just an interrogative intonation or an interjection. When problem representations are grouped under ‘detections’ on the one hand, and ‘diagnoses’ on the other hand, it seems that the more revisers verbalize a diagnosis, the better they detect. However, it takes time: the verbalization of diagnoses is also associated with duration. As far as quality is concerned, no correlation has been found.

#### 4.2.3. ‘Richness’ of verbalizations

As explained in the methodology sections, a verbalization fragment can be coded for problem representation, strategy, and immediate solution. However, sometimes, a combination was observed. Consequently, we distinguished between 1) verbalization of a problem representation alone, 2) verbalization of a problem-solving strategy alone, 3) verbalization of an immediate solution alone, 4) verbalization of a problem representation and a problemsolving strategy, and 5) verbalization of a problem representation and an immediate solution. The means in table 8 have been calculated drawing on the percentage for each participant, that is, the number of times they verbalized one of the five verbalization types described before, divided by the sum for all types.

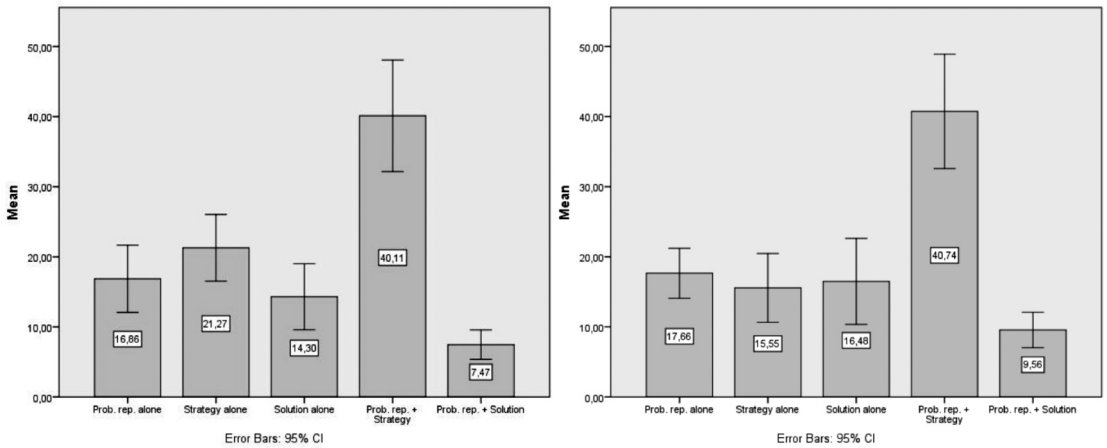
TABLE 8  
Descriptive statistics for verbalizations types, per type

	Error detection		Quality		N
	Mean	SD	Mean	SD	
Problem repres. alone	16.86	8.99	17.66	6.69	16
Strategy alone	21.27	8.93	15.55	9.24	16
Solution alone	14.30	8.84	16.48	11.51	16
Problem repres. + Strategy	40.11	14.93	40.74	15.31	16
Problem repres. + Solution	7.47	3.93	9.56	4.75	16

As illustrated in figures 4 and 5, the verbalization of a combination of a problem representation and of a strategy is the most frequent, in both the error detection and quality rankings. The combination of a problem representation and a solution is the least frequent, in both rankings as well. The second, third and, fourth places vary: in the error detection ranking, the verbalization of a strategy alone comes second, before the verbalization of a problem representation alone, and the verbalization of a solution alone. In the quality ranking, the verbalization of a problem representation alone comes second, before the verbalization of a solution alone, and the verbalization of a strategy alone. However, the differences are very small.

FIGURE 4 AND 5

Mean percentages of verbalization type, per type, for error detection (left) and quality (right)



A test of analysis of variance (ANOVA for repeated measures) was conducted to determine whether the differences between the means were significant, that is, to determine whether there are significant differences in the frequency of each verbalization type. Since all series of scores were normally distributed, the parametric variant was chosen because it allows for post hoc tests, and thus, for pairwise comparisons. The tests were significant, with  $F(5,16)=23.31, p<.001$  and  $F(5,16)=17.76, p<.001$ . Post hoc tests (table C in the appendix) revealed a significant difference, in both rankings, between the verbalization of a combination of a problem representation and a strategy, and all other types of verbalizations.

These different types of verbalizations can be grouped under ‘single verbalizations’ (all verbalizations types when alone) and ‘combined verbalizations’ (both combinations). All descriptive statistics are summarized in table 9. However, the means are rather close. The parametric t-test did not reveal any difference:  $t=.660, p=.519$  and  $t=-.085, p=.933$ .

TABLE 9  
Descriptive statistics for verbalization types, grouped under ‘single’ and ‘combined’ types

	Error detection		Quality		N
	Mean	SD	Mean	SD	
Single verbalizations	52,42	14,70	49,69	14,38	16
Combined verbalizations	47,58	14,70	50,31	14,38	16

To sum up, the combination of a problem representation and a strategy is significantly more frequent than all other types. Another salient result is the very low frequency of the verbalization of a combination of a problem representation and an immediate solution. All other types are approximately as frequent as one another.

4.2.4. *Relation between the richness of verbalizations and error detection potential, revision quality, and revision duration*

To determine the potential relation between the type of verbalizations and error detection potential, revision quality, and revision duration, correlation tests were carried out. Results summarized in table 10 show that there does not seem to be a relationship between the type of verbalizations and error detection and quality. However, the verbalization of a solution alone is negatively associated with duration, and the verbalization of the combination of a problem representation and a strategy is positively related to duration.

TABLE 10  
Correlation tests for verbalization types, per type (N=16)

	Problem repres. alone	Strategy alone	Solution alone	Problem repres. + Strategy	Problem repres. + Solution
	r	r	r	r	r
	p (1-tailed)	p (1-tailed)	p (1-tailed)	p (1-tailed)	p (1-tailed)
Error	-.273	-.285	-.173	.420	.064
detection	.154	.143	.261	.053	.407
Quality	-.304	.087	-.316	.340	-.073
	.127	.374	.117	.099	.393
Duration	-.393	-.038	-.512*	.650**	-.333
	.066	.445	.021	.003	.104

\* Correlation is significant at the .05 level (1-tailed).

\*\* Correlation is significant at the .01 level.

When the different types of verbalizations are grouped under 'single' and 'combined' verbalizations, results (table 11) are slightly different, and shed more light on the potential effect of particular types of verbalizations on revision quality, error detection potential, and revision duration. There is no significant correlation between any type of verbalization and revision quality. However, there is a significant positive correlation between error detection and combined verbalizations, and a significant negative correlation between error detection and single verbalizations.

TABLE 11  
Correlation tests for verbalization types, grouped under 'single' and 'combined' types

	Single verbalizations	Combined verbalizations
	r	r
	p (1-tailed)	p (1-tailed)
Error	-.444*	.444*
detection	.043	.043
Quality	-.338	.338
	.100	.100
Duration	-.571**	.571**
	.010	.010

\* Correlation is significant at the .05 level (1-tailed).

\*\* Correlation is significant at the .01 level.

In sum, it can be said that the more solutions alone were verbalized, the quicker participants worked, and the more they verbalized a combination of a problem representation and a strategy, the longer they worked. Besides, the more combined verbalizations participants uttered, the better they detected, whereas it is the opposite with single verbalizations. Conversely, combined verbalizations take time, contrary to single verbalizations.

## 5. Conclusions

This paper reported on a follow-up study whose aim was fourfold: 1) to determine which variables do influence the amount of verbalization of professional revisers when they verbalize their thoughts while revising somebody else's translation, and 2) to determine what kind of revision sub-processes are verbalized, 3) to determine the relation between the type of verbalizations and revision quality, duration, and error detection potential, and 4) to draw conclusions for revision didactics.

Results show that variables that could have influenced the verbalization ratio of revisers had no effect on that ratio, except the revision experience. A correlation was found between experience measured in years, and the proportion of verbalizations, which means that the more revisers have experience with revision, the more they seem to be willing to verbalize. This does not seem to be relevant for the didactics of revision, since the discipline is generally aimed at trainees with little or no experience in revision.

However, results about the type of verbalized revision sub-processes are interesting: it seems that participants did not verbalize a problem representation for about a third of the problems they encountered. When they did, they verbalized detections approximately as often as diagnoses. When detections are split up into very vague detections, vague detections and rejections, and when diagnoses are split up into intentional diagnoses, maxim-based diagnoses, and rule based-diagnoses, only maxim-based diagnoses appear to have been verbalized less than others, the differences between all other types of problem representations being generally not significant. However, maxim-based diagnoses are one of the very few problem representations to be associated with revision quality and error detection, but also with duration. In other words, the more revisers verbalize such a problem representation, the better they detect, the better they revise, but the longer they work.

Vague detections are also associated with quality and error detection. From a didactic point of view, vague detections are not a useful learning tool. What is interesting as regards revisers' training is that a significant correlation has been observed between diagnoses (as one category, regrouping intentional, maxim-based, and rule-based diagnoses) and error detection, but also with duration. In other words, the more revisers verbalized a diagnosis, the better they detected. Even if no correlation was observed with quality, that result should be taken into account, since error detection is the very first step in revision. Moreover, results also show that participants who verbalized a problem representation together with a problemsolving strategy or a solution, detected better, but worked longer.

Consequently, the two main effects of TAP on the revision process seem to be that revisers tend to detect better when they verbalize a diagnosis and/or when they verbalize a problem representation together with a problem-solving strategy.



Therefore, we think that further research could focus on a particular subcompetence of the revision competence: the ability to explain. In a follow-up study, we intend to test the hypothesis that ‘justify aloud’ has a positive effect on the error detection potential of revision trainees.

#### NOTE

1. Inputlog, [www.inputlog.net](http://www.inputlog.net).

#### REFERENCES

- ALLAL, Linda, CHANQUOY, Lucile and LARGY, Pierre (2004): *Revision: Cognitive and Instructional Processes*. Dordrecht: Kluwer Academic.
- BERNARDINI, Silvia (2001): Think-aloud protocols in translation research: Achievements, limits, future prospects. *Target*. 13(2):241-263.
- BIEL, Lucja (2011): Training translators or translation service providers? EN 15038: 2006 standard of translation services and its training implications. *JoSTrans*. 16:61-76.
- BISAILLON, Jocelyne (2007): Professional Editing Strategies Used by Six Editors. *Written Communication*. 24(4):295-322.
- BOWLES, Melissa A. (2010): *The Think-Aloud Controversy in Second Language Research*. New York: Routledge.
- BREEDVELD, Hella (2002): Translation processes in time. *Target*. 14(2):221-240.
- BREEDVELD, Hella and VAN DEN BERGH, Huub (2002): Revisie in vertaling: wanneer en wat. *Linguistica Antverpiensia New Series*. 1:327-345.
- ERICSSON, Karl Anders and SIMON, Herbert Alexander (1993): *Protocol Analysis: Verbal Reports as Data, A Bradford book*. Cambridge Massachusetts Institute of Technology.
- EUROPEAN COMMITTEE FOR STANDARDIZATION (2006): *European Standard EN 15 038. Translation services – Service requirements*. Brussels: European Committee for Standardization.
- FIELD, Andy (2009): *Discovering Statistics Using SPSS*. London: SAGE Publications.
- GÖPFERICH, Susanne (2008): *Translationsprozessforschung: Stand, Methoden, Perspektiven, Translations-Wissenschaft*. Tübingen: Narr.
- GÖPFERICH, Susanne and JÄÄSKELÄINEN, Riitta (2009): Process research into the development of translation competence: Where are we, and where do we need to go? *Across Languages and Cultures*. 10(2):169-191.
- HANSEN, Gyde (2005): Experience and Emotion in Empirical Translation Research with Think-Aloud and Retrospection. *Meta*. 50(2):511-521.
- HANSEN, Gyde (2009): The speck in your brother’s eye – the beam in your own. Quality management in translation and revision. In: Gide HANSEN, Andrew CHESTERMAN and Heidrun GERZYMISCH-ARBOGAST, eds. *Efforts and Models in Interpreting and Translation Research: A Tribute to Daniel Gile*. Amsterdam: John Benjamins, 255-280.
- HAYES, John R, FLOWER, Linda, SCHRIEVER, Karen A., et al. (1987): Cognitive processes in revision. In: Sheldon ROSENBERG, ed. *Advances in Applied Psycholinguistics*. Cambridge: Cambridge University Press, 176-240.
- HORGUELIN, Paul A. and BRUNETTE, Louise (1998): *Pratique de la révision*. Brossard, QC: Linguattech.
- JÄÄSKELÄINEN, Riitta (2000): Focus on Methodology in Think-aloud Studies on Translating. In: Sonia TIRKKONEN-CONDIT and Riitta JÄÄSKELÄINEN, eds. *Tapping and Mapping the Processes of Translation and Interpreting*. Amsterdam: John Benjamins, 71-82.
- JÄÄSKELÄINEN, Riitta (2002): Think-aloud protocol studies into translation: An annotated bibliography. *Target*. 14(1):107-136.
- JÄÄSKELÄINEN, Riitta (2011): Back to Basics: Designing a Study to Determine the Validity and Reliability of Verbal Report Data on Translation Processes. In: Sharon O’BRIEN, ed. *Cognitive Explorations of Translation*. London: Continuum, 15-29.

- JAKOBSEN, Arnt Lykke (2002): Translation drafting by professional translators and by translation students. In: Gyde HANSEN, ed. *Empirical Translation Studies. Process and Product*. Frederiksberg: Samfundslitteratur, 191-204.
- JAKOBSEN, Arnt Lykke (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.
- KRINGS, Hans Peter (2001): Repairing texts: empirical investigations of machine translation: postediting processes. Kent, Ohio: Kent State University Press.
- KRUGER, Haidee (2008): Training Editors in Universities: Considerations, Challenges and Strategies. In: John KEARNS, ed. *Translator and Interpreter Training: Issues, Methods and Debates*. New York: Continuum, 39-65.
- KÜNZLI, Alexander (2006): Teaching and learning translation revision: Some suggestions based on evidence from a think-aloud protocol study. In: Mike GARANT, ed. *Current Trends in Translation Teaching and Learning*. Helsinki: Helsinki University, 9-24.
- LAPEBER, Anne (2012): Translation Skills and Knowledge – Preliminary Findings of a Survey of Translators and Revisers Working at Inter-governmental Organizations. *Meta*. 57(1):108-131.
- LEIJTEN, Mariëlle and VAN WAES, Luuk (2006): Inputlog: New Perspectives on the Logging of On-Line Writing. In: Kirk P. H. SULLIVAN and Eva LINDGREN, eds. *Computer Key-Stroke Logging and Writing: Methods and Applications*. Oxford: Elsevier, 73-94.
- LEIJTEN, Mariëlle and VAN WAES, Luuk (2013): Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication*. 30(3):358-392.
- LI, Defeng (2004): Trustworthiness of think-aloud protocols in the study of translation processes. *International Journal of Applied Linguistics*. 14(3): 301-313.
- MOSSOP, Brian (1992): Goals of a revision course. In: Cay DOLLERUP and Anne LODDEGAARD, eds. *Teaching translation and interpreting: training, talent and experience: papers from the first Language International Conference Elsinore, Denmark, 31 May - 2 June 1991*. Amsterdam: John Benjamins, 81-90.
- MOSSOP, Brian (2001/2007): *Revising and Editing for Translators*. Manchester: St. Jerome.
- MUÑOZ MARTÍN, Ricardo (2012): Standardizing translation process research methods and reports. In: Isabel GARCÍA-IZQUIERDO and Esther MONZÓ, eds. *Iberian Studies on Translation and Interpreting*. Oxford: Lang, 11-22.
- NORBERG, Ulf (2003): *Übersetzen mit doppeltem Skopos. Eine empirische Prozess- und Produkstudie, Studia Germanica Upsaliensa*. Uppsala: Uppsala University.
- PACTE (2003): Building a translation competence model. In: Fabio ALVES, ed. *Triangulating translation: Perspectives in process oriented research*. Amsterdam: John Benjamins, 43-66.
- ROBERT, Isabelle S. (2008): *Translation revision procedures: An explorative study*. In: Pieter Boulogne, ed. *Translation and Its Others. Selected Papers of the CETRA Research Seminar in Translation Studies 2007*. Visited on 1<sup>st</sup> August 2016, <<http://www.arts.kuleuven.be/cetra/papers/files/robert.pdf>>.
- ROBERT, Isabelle S. (2012): *La révision en traduction: les procédures de révision et leur impact sur le produit et le processus de révision*. Antwerp: University of Antwerp.
- ROBERT, Isabelle S. (2014): Investigating the problem-solving strategies of revisers through triangulation. An exploratory study. *Translation and Interpreting Studies*. 9(1):88-108.
- ROBERT, Isabelle S., REMAEL, Aline and UREEL, Jim J. J. (2016): Towards a model of translation revision competence. *The Interpreter and Translator Trainer*. 10(2). DOI: 10.1080/1750399X.2016.1198183.
- ROBERT, Isabelle S. and VAN WAES, Luuk (2014): Selecting a translation revision procedure: do common sense and statistics agree? *Perspectives*. 22(3):304-320.
- SCHJOLDAGER, Anne, WØLCH RASMUSSEN, Kirsten and THOMSEN, Christa (2008): Préciswriting, Revision and Editing: Piloting the European Master in Translation. *Meta*. 53(4):798-813.

- SHIH, Claire Yiyi (2006): Revision from translators' point of view. An interview study. *Target*. 18(2): 295-312.
- SUN, Sanjun (2011): Think-Aloud-Based Translation Process Research: Some Methodological CONSIDERATIONS. *META*. 56(4):928-951.
- THAON, Brenda M. and HORGUELIN, Paul A. (1980): *A Practical Guide to Bilingual Revision*. Montreal: Linguattech.

## APPENDICES

TABLE A

**Overview of the subjects' profile**

Reviser	Gender	Age	Experience in translation (left) and in revision (right) in years		Status	Degree	Status of Dutch (source language)
1	female	33	0	8	employee	Master in philology	B
2	female	32	7	6	freelance	Master in philology + 1-year Master in translation	B
3	male	37	3	3	employee	Master in philology + 1-year Master in translation	B
4	female	39	13	13	employee	Master in translation	B
5	female	37	13	13	freelance	Master in translation	C
6	female	31	8	6	employee	Master in translation	Secondary school
7	female	33	9	5	employee	Master in translation	B
8	female	29	7	5	employee	Master in translation	Secondary school
9	female	35	12	9	employee	Master in translation	C
10	female	23	0	1	employee	Master in philology	B
11	female	32	7	7	employee	Master in translation	B
12	female	33	10	10	employee	Master in translation	Secondary school
13	female	38	15	15	employee	Master in translation	B
14	male	24	2	2	freelance	Master in translation	C
15	female	41	14	10	freelance	Other Master	Secondary school
16	female	32	8	8	employee	Master in translation	B

Note: C for Dutch means that Dutch was not part of the main degree, but of an additional year in Dutch within the same Translation department. B means that Dutch was one of the languages learned during the Master degree, in Translation or philology. N.B.: The Master's in translation and in philology were preceded by a bachelor in the same discipline and the same languages.

TABLE B  
**Pairwise comparison (df=6)**

Pairwise comparison		Mean difference	Std. error	Sig.
<b>Error detection</b>				
No prob. repres.	Very vague	28.822**	3.429	.000
	Vague	20.698**	4.595	.009
	Rejection	23.127**	4.657	.004
	Intentional	24.050**	4.328	.001
	Maxim	30.112**	4.394	.000
	Rule	22.149**	3.902	.001
Very vague	Vague	-8.123	2.590	.133
	Rejection	-5.695	2.842	.748
	Intentional	-4.772	2.434	.776
	Maxim	1.290	2.485	1.000
	Rule	-6.673	2.625	.381
Vague	Rejection	2.428	2.793	1.000
	Intentional	3.351	1.866	.870
	Maxim	9.414**	1.723	.001
	Rule	1.451	1.932	1.000
Rejection	Intentional	.923	2.197	1.000
	Maxim	6.985	2.444	.224
	Rule	-.978	2.576	1.000
Intentional	Maxim	6.062	1.679	.053
	Rule	-1.901	1.750	.999
Maxim	Rule	-7.963**	1.540	.002
<b>Quality</b>				
No prob. repres.	Very vague	24.264**	3.091	.000
	Vague	21.326**	4.339	.004
	Rejection	17.541	4.967	.062
	Intentional	18.983**	4.210	.009
	Maxim	26.898**	4.426	.000
	Rule	15.255*	3.918	.030
Very vague	Vague	-2.939	3.515	1.000
	Rejection	-6.723	3.241	.700
	Intentional	-5.282	2.740	.797
	Maxim	2.634	3.058	1.000
	Rule	-9.009	3.046	.186
Vague	Rejection	-3.784	3.070	.997
	Intentional	-2.343	2.072	.999
	Maxim	5.573	2.085	.308
	Rule	-6.070	2.380	.375
Rejection	Intentional	1.441	2.577	1.000
	Maxim	9.357*	2.393	.029
	Rule	-2.286	3.063	1.000
Intentional	Maxim	7.916**	1.632	.004
	Rule	-3.727	2.321	.945
Maxim	Rule	-11.643**	1.949	.001

\* Correlation is significant at the .05 level (1-tailed).

\*\* Correlation is significant at the .01 level.

TABLE C  
**Pairwise comparison (df=4) 165 mots**

Pairwise comparison		Mean difference	Std. error	Sig.
<b>Error detection</b>				
Prob. repres. alone	Strat. alone	-4.411	3.747	.949
	Sol. alone	2.563	3.019	.995
	Prob. repres. + Strat	-23.251**	5.026	.003
	Prob. repres. + Sol.	9.393*	2.584	.024
Strat. alone	Sol. alone	6.974	2.849	.241
	Prob. repres. + Strat	-18.840*	5.133	.022
	Prob. repres. + Sol.	13.804**	2.496	.001
Sol. alone	Prob. repres. + Strat	-25.814**	5.619	.004
	Prob. repres. + Sol.	6.830	2.334	.099
Prob. repres. + Strat	Prob. repres. + Sol.	32.644**	4.035	.000
<b>Quality</b>				
Prob. repres. alone	Strat. alone	2,100	3,288	1,000
	Sol. alone	1,172	3,007	1,000
	Prob. repres. + Strat	-23,089**	4,746	,002
	Prob. repres. + Sol.	8,093*	2,017	,011
Strat. alone	Sol. alone	-,928	4,058	1,000
	Prob. repres. + Strat	-25,190**	4,956	,001
	Prob. repres. + Sol.	5,993	2,703	,352
Sol. alone	Prob. repres. + Strat	-24,261*	6,281	,015
	Prob. repres. + Sol.	6,921	2,991	,302
Prob. repres. + Strat	Prob. repres. + Sol.	31,183**	4,382	,000

\* Correlation is significant at the .05 level (1-tailed).

\*\* Correlation is significant at the .01 level.