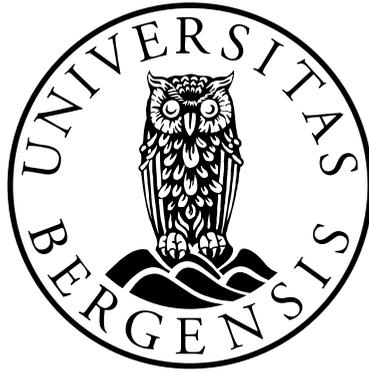


Experience and Expertise in Conference Interpreting

An Investigation of Swedish Conference Interpreters

Elisabet Tiselius



Dissertation for the degree of philosophiae doctor (PhD)
at the University of Bergen

2013

Dissertation date: November 26, 2013

Contents

SCIENTIFIC ENVIRONMENT	11
ACKNOWLEDGEMENTS	12
ABSTRACT.....	15
THE PRESENT THESIS IS BASED ON THE FOLLOWING PAPERS	17
1. INTRODUCTION.....	18
1.1 EXPERTISE AND INTERPRETING RESEARCH	18
1.2 AIM.....	20
2. BACKGROUND	22
2.1 RESEARCH ON THE EFFECT OF EXPERIENCE IN INTERPRETING	22
2.2 INTERPRETING PRACTICE.....	27
2.3 EXPERTISE	29
2.3.1 <i>Different concepts of expertise</i>	29
2.3.2 <i>The concept of deliberate practice in the expertise approach</i>	35
2.3.3 <i>Research on expertise in interpreting</i>	38
2.4 ASSESSMENT	46
2.5 RESEARCH QUESTIONS AND METHODOLOGICAL DEVELOPMENT.....	50
3. DATA AND METHODS	53
3.1 METHODS	53
3.1.1 <i>Investigating expertise</i>	53
3.1.2 <i>Investigating the process</i>	55
3.1.3 <i>Investigating the product</i>	57
3.1.4 <i>Investigating the participants</i>	58
3.2 PARTICIPANTS	58
3.2.1 <i>The cross-sectional data – data set A</i>	59

3.2.2	<i>The long-term data – data set B</i>	61
3.2.3	<i>The raters</i>	62
3.3	INTERPRETING DATA	63
3.3.1	<i>The EU speech</i>	63
3.3.2	<i>The NATO speech</i>	63
3.4	RETROSPECTIVE DATA AND ASSESSMENT FILES	64
3.5	METHODOLOGICAL DISCUSSION	64
3.5.1	<i>Mixed-method design</i>	65
3.5.2	<i>Re-test or not, and other challenges</i>	67
3.5.3	<i>Terminological inconsistencies</i>	68
4.	SUMMARY OF ARTICLES AND GENERAL RESULTS	69
4.1	ARTICLE 1: “REVISITING CARROLL’S SCALES” (DATA SET A)	69
4.1.1	<i>Background</i>	70
4.1.2	<i>Data and method</i>	71
4.1.3	<i>Major findings</i>	71
4.1.4	<i>Discussion</i>	72
4.2	ARTICLE 2: “PROCESS AND PRODUCT IN SIMULTANEOUS INTERPRETING: WHAT THEY TELL US ABOUT EXPERIENCE AND EXPERTISE” (DATA SET A)	73
4.2.1	<i>Background</i>	73
4.2.2	<i>Data and method</i>	74
4.2.3	<i>Major findings</i>	74
4.2.4	<i>Discussion</i>	76
4.3	ARTICLE 3 – “THE DEVELOPMENT OF EXPERTISE – OR NOT: THREE SIMULTANEOUS INTERPRETERS’ DEVELOPMENT OVER TIME” (DATA SET B)	77
4.3.1	<i>Background</i>	77

4.3.2	<i>Material and method</i>	78
4.3.3	<i>Major findings</i>	79
4.3.4	<i>Discussion</i>	80
4.4	ARTICLE 4 – “EXPERTISE WITHOUT DELIBERATE PRACTICE? THE CASE OF SIMULTANEOUS INTERPRETERS” (DATA SET B).....	81
4.4.1	<i>Background</i>	81
4.4.2	<i>Data and method</i>	82
4.4.3	<i>Major findings</i>	83
4.4.4	<i>Discussion</i>	84
4.5	SUMMARY OF THE RESULTS FROM ALL FOUR STUDIES.....	86
4.5.1	<i>Methodological results</i>	86
4.5.2	<i>Research results</i>	86
5.	DISCUSSION	88
5.1	METHODOLOGICAL DISCUSSION	88
5.2	DISCUSSION OF THE RESULTS	90
5.3	THE EXPERTISE THEORY AND SIMULTANEOUS INTERPRETING	91
6.	CONCLUSION	95
	LIST OF REFERENCES	100
	ARTICLE 1	111
	REVISITING CARROLL’S SCALES	111
1.	INTRODUCTION	111
1.1	<i>Purpose and research questions</i>	112
2.	BACKGROUND	113
2.1	<i>Carroll’s scales</i>	113
2.2	<i>Applying grading scales to interpreting</i>	115

3. DATA AND METHOD	117
3.1 ADAPTATION OF THE SCALES	117
3.2 ELICITING MATERIAL.....	120
3.2.1 <i>The speech</i>	120
3.2.2 <i>The interpreters</i>	121
3.2.3 <i>Preparing the transcripts</i>	121
3.3 THE GRADING PROCEDURE	123
3.3.1 <i>The graders</i>	123
3.3.2 <i>Grader training</i>	123
3.3.3 <i>Grading</i>	123
3.4 MEASURING SIGNIFICANT DIFFERENCE AND INTER-RATER RELIABILITY	124
4. RESULTS	125
4.1 INTER-RATER RELIABILITY	125
4.2 <i>INTELLIGIBILITY</i>	125
4.2.1 <i>Intelligibility graded by non-interpreter graders vs. interpreter graders</i>	126
4.3 <i>INFORMATIVENESS</i>	127
4.3.1 <i>Informativeness graded by non-interpreter graders vs. interpreter graders</i>	127
4.4 SPONTANEOUS COMMENTS FROM GRADERS	128
5. DISCUSSION	128
5.1 LIMITATIONS	128
5.2 DISCUSSION OF THE RESULTS	129
6. CONCLUSION	131
REFERENCES	132
ARTICLE 2	136
PROCESS AND PRODUCT IN SIMULTANEOUS INTERPRETING	136

1. <i>Introduction</i>	137
1.1 Aim and Scope	138
2. <i>Background</i>	138
2.1 Expertise Research in Interpreting Studies.....	138
2.2 Studying the Process.....	140
2.3 Evaluating the product.....	140
2.4 The Studies	141
3. <i>First Study: Interpreting Process</i>	142
3.1 Material and Method	142
3.1.1 <i>Input Material</i>	142
3.1.2 <i>The Interpreting Subjects</i>	142
3.1.3 <i>The Retrospection Procedure</i>	144
3.1.4 <i>Transcription and Analysis</i>	144
3.2 Results	145
3.2.1 <i>Processing Problems</i>	145
3.2.2 <i>Instances of Monitoring</i>	148
3.2.3 <i>Strategies</i>	150
4. <i>Second Study: Assessment of the Interpreting Product</i>	152
4.1 Material and Method	152
4.1.1 <i>The Rating Files</i>	153
4.1.2 <i>The Raters</i>	153
4.1.3 <i>The Rating Session</i>	153
4.1.4 <i>The Scales and the Analysis</i>	153
4.2 Results	154
4.2.1 <i>Intelligibility and Informativeness</i>	154

4.2.2 <i>The Effect of Raters</i>	155
5. <i>Discussion</i>	156
5.1 Discussion of Results	157
5.2 Conclusion	158
<i>References</i>	160
ARTICLE 3	169
THE DEVELOPMENT OF EXPERTISE – OR NOT	169
1. <i>Introduction</i>	170
1.1 Aim and scope	170
2. <i>Background</i>	171
3. <i>Material and Method</i>	173
3.1 Participants	173
3.2 Control group	173
3.3 Speeches	174
3.4 Rating	174
3.5 Retrospection	175
4. <i>Results</i>	176
4.1 Assessment of the interpreting	176
4.2 Quantitative data: interpretations and retrospection	177
4.3 Processing problems, monitoring and strategies	178
5. <i>Discussion</i>	181
6. <i>Conclusion</i>	183
<i>References</i>	185
ARTICLE 4	193
EXPERTISE WITHOUT DELIBERATE PRACTICE?	193
1. <i>Introduction</i>	193

2. <i>Background</i>	194
3. <i>Methodology</i>	195
3.1 Participants	195
3.2 Procedure for conducting the in-depth interviews.....	196
3.2.1 <i>Identification of topics and core phenomena</i>	196
3.2.2 <i>The interview setting</i>	197
3.3 Coding and analyses of the interview data.....	198
4. <i>Qualitative Sides of Expertise</i>	198
4.1 Language learning and language knowledge	198
4.2 General knowledge.....	199
4.3 Communicative skills.....	199
4.4 Focus.....	200
4.5 Coping with stress	200
4.6 The interpreting skill	201
5. <i>Deliberate practice, clear goals and openness to feedback</i>	201
6. <i>Discussion and Conclusions</i>	203
<i>References</i>	205

List of tables in summary

Table 1. Age and experience of the cross-sectional interpreters	59
---	----

List of figures in summary

Figure 1. Novice-Expert continuum.	43
Figur 2 Summary of participants, data, analyses and articles.....	69

List of tables in article 1

Table 1 Scale of intelligibility	118
Table 2. Scale of informativeness.....	119

Table 3. Scale of intelligibility on grading sheet.....	120
Table 4. Scale of informativeness on grading sheet	120
Table 5. Age and experience of the interpreters	121
Table 6. Significance in gradings of intelligibility by non-interpreters (n=6).....	125
Table 7. Significance in gradings of intelligibility by interpreters (n=6).....	126
Table 8. Avg. scores of intelligibility graded by non-interpreters (n=6) and interpreters (n=6) .	126
Table 9. Significance for grading of informativeness by non-interpreters (n=6).....	127
Table 10. Significance for grading of informativeness by interpreters (n=6)	127
Table 11. Significance of grading of informativeness graded by non-int (n=6) and int (n=6) ...	127
Table 12. Processing problems. Classification according to Ivanova (1999).	188
Table 13. Monitoring observations (Ivanova 1999).	189
Table 14. Strategies (Ivanova 1999).....	190

List of figures in article 1

Figure 1. Avg scores for intelligibility graded by int (n=6) and non-int (n=6).....	126
Figure 2. Avg scores for informativeness graded by int (n=6) and non-int (n=6)	128

List of tables in article 2

<i>Table 1. Age and experience of the interpreters</i>	143
<i>Table 2. Processing Problems. Definitions of the headings are given in appendix 1</i>	145
<i>Table 3. Processing problems long experience Ivanova vs. the present study.....</i>	147
<i>Table 4. Processing problems Ivanova's novices vs. short and no experience interpreters</i>	147
<i>Table 5. Instances of monitoring</i>	148
<i>Table 6. Instances of monitoring Ivanova's expert vs. the long experience interpreters</i>	150
<i>Table 7. Instances of monitoring Ivanova's novices vs. the short and no experience int.</i>	150
<i>Table 8. Instances of reported strategies.....</i>	151
<i>Table 9. Mean evaluation scores of intelligibility and informativeness.....</i>	154

List of figures in article 2

<i>Figure 1. CA biplot of the association between interpreter experience and processing probl. ..</i>	146
<i>Figure 2. CA biplot of the association between interpreter experience and instances of monit.</i>	149
<i>Figure 3. CA biplot of the association between interpreter experience and strategies.</i>	152
<i>Figure 4. Intelligibility: boxplot of between-group differences.</i>	155
<i>Figure 5. Informativeness: boxplot of between-group differences.</i>	155

List of tables in article 3

<i>Table 1. Intelligibility and informativeness: NATO speech, comparison.....</i>	176
<i>Table 2. Intelligibility and informativeness: EU speech, student rating only.</i>	177
<i>Table 3. Intelligibility and informativeness: cross-sectional and longitudinal data set.</i>	177
<i>Table 4. Length of interpreting, length of retrospection: NATO speech</i>	177

<i>Table 5. Length of interpreting, length of retrospection: EU speech</i>	178
<i>Table 6. Processing problems and categories: NATO speech, longitudinal participants</i>	178
<i>Table 7. Processing problems and categories: EU speech</i>	179
<i>Table 8. Reported monitoring categories: NATO speech</i>	179
<i>Table 9. Reported monitoring categories: EU speech</i>	180
<i>Table 10. Strategies: NATO speech, longitudinal experienced participants</i>	180
<i>Table 11. Strategies: EU speech, longitudinal and cross-sectional interpreters</i>	180
<i>Table 12. Macrostrategies: NATO speech, longitudinal experience participants</i>	181
<i>Table 13. Macrostrategies: EU speech, longitudinal and cross-sectional interpreters</i>	181
<i>Table 14. Processing problems. Classification according to Ivanova (1999)</i>	188
<i>Table 15. Monitoring observations (Ivanova 1999)</i>	189
<i>Table 16. Strategies (Ivanova 1999)</i>	190
<i>Table 17. Macrostrategies (Ivanova 1999)</i>	191

List of figures in article 4

Figure 1. Thematic sketch of topics covered in the in-depth interviews	197
--	-----

Scientific environment

This is a dissertation written at the Department of Foreign Languages at the University of Bergen, the Centre for Research on Bilingualism at Stockholm University and the Institute for Interpretation and Translation Studies at Stockholm University. The dissertation work has been supervised by associate professor Åse Johnsen, professor Birgitta Englund Dimitrova and professor Kenneth Hyltenstam.

Acknowledgements

Writing a dissertation means depending on people for support, encouragement and also critique. There are many people who have done that for me throughout this journey. The first persons to be mentioned, though, should be the ones who were the sine qua non of my work. I'm so grateful that Sarah Williams recorded so many interpreters with such fervency, as she gave me the starting point of my data. Thank you, Sarah, I hope I have given your material justice.

The other very important persons for this project are the interpreters – without their interest and willingness to participate in different research projects, this dissertation would not have gone far. The Swedish interpreting community is a small one, and I have pestered many conference interpreting colleagues with my research. I am forever grateful to you, dear colleagues, that you bore with me all the way from my Master's thesis over ten years ago.

A PhD project does not take off without supervisors to encourage, question and guide the candidate, and put her right from time to time. I have been blessed with the best. You have all been crucial to me, Kenneth for believing in me from the very beginning, Birgitta for inviting me to do research with you, and Åse for helping me keeping the focus. Thank you!

Many other people have also been important for this dissertation: Maria Wingqvist, who sparked the fire when she asked the budding Master's student, "Would you consider writing a PhD?"; Karin S. Lindelöf and Magnus Öhlander, who helped out when I was lost in ethnographic methodology; Lidun Hareide, who has been a wonderful colleague and sounding board on the Norwegian side; Gard Buen Jensen, who agreed to co-author a paper and put me out of my statistical misery; Stig Oppedal, without whose diligent proofreading and editing my prose would only seem like meaningless rambling; and Sandra Halverson, who brought me to Norway. Thank you so much for helping me in this work.

When it comes to my development in interpreting studies, I'm much obliged to the European Society for Translation Studies for generously giving me their summer school grant and making it possible for me to take the CETRA research summer school at KU Leuven. The CETRA course was a wonderful starting point and also gave me the fantastic opportunity to get help and support from Andrew Chesterman, Daniel Gile, Franz Pöchhacker and Miriam Shlesinger. The continued dedication and support that these stars of translation and interpreting studies have shown to a PhD student is truly wonderful. I will never be able to thank you enough, Andrew, Daniel, Franz and Miriam; Miriam, I hope that you continue to keep an eye on us from where you are now.

CETRA was also where I met my wonderful research buddies Šárka Timarová and Veerle Dufrou. Thank you for the great discussions and for keeping a stern “quant” and “qual” eye on the work I've done. Through Šárka, I also finally got to meet Adelina Hild (Ivanova), who came into my research life very early when she kindly sent me the copy of her dissertation that I was so curious of. Adi! What can I say – thank you!

Another great place to hang out has been the Translation Process Research Workshop that Susanne Göpferich initiated and Ricardo Muñoz Martín continued. The workshop has an open-minded atmosphere, and presenting there has helped me develop the argument of my dissertation.

Emilia Iglesias Fernández deserves a particular note in this context. When I was looking for a opportunity to learn more about the work of Angela Collados Aís and the ECIS team, which Emilia belongs to, she organized an Erasmus teacher exchange in Granada and put me up in her home. It was a wonderful week filled with academic inspiration both for the dissertation and beyond. ¡Gracias, Emilia!

And in case you ever come up with the idea to edit another book during your PhD period, Cecilia Alvstad and Adelina Hild are the best friends to do that with.

I have belonged to a greater or lesser extent to three departments: the Department of Foreign Languages at Bergen University, the Centre for Research on Bilingualism at Stockholm University, and the Institute for Interpreting and Translation Studies at Stockholm University. I have been the odd one out (and possibly also the oddball) at all three departments, and thank all of them for including me and helping me in my dissertation work.

In the real world, outside the academic bubble, I'm so fortunate to have Johanna Järvinen, Johanna Broman-Åkesson, Pauline Dandois and Kristina Camitz as friends – thanks for putting up with me as I got increasingly self-centred, confused and busy. A particular thank you to Johanna Broman-Åkesson who also took the time to read all the interviews with me. Merci Pauline pour toujours me sauver. Tack för att ni har sett till att både jag, Carl, barnen, hunden och katten har överlevt, framförallt den här sista perioden.

It would be impossible to write a dissertation without support from my family. Carl, du är ljuset i mitt liv. Följ mig bortåt vägen så skiljs vi längre fram, vi fortsätter gå min väg, och din väg, huvudsaken är att inte var tar sin väg. Eva, Daniel och Anna, ni är bäst! Fortsätt att vara precis som ni är.

Brussels,
June 2013

Abstract

This dissertation investigates the process and product of interpreters with different levels of experience and explores the expertise approach (cf. Ericsson, Charness and Hoffman 2007) as applied to interpreters. The expertise approach claims that highly skilled performers, regardless of their chosen field, use the same type of strategies in order to reach the top levels of their profession. An important feature of the expertise approach is deliberate practice, a specific type of practice that highly skilled performers engage in so as to improve their performance.

The dissertation is based on four different studies featuring two different sets of participants. Two data sets – a *cross-sectional material* with nine participants on three different levels of interpreting experience (none, short and long), and a *long-term material* with three interpreters recorded at two different points in time – were analysed in terms of both processing and product data. The interpreting process was studied by retrospectively analysing and categorizing processing problems, monitoring and strategies, while the interpreting product was analysed by using holistic rating scales for intelligibility and level of information transfer of the interpreting product. In-depth interviews were also conducted with the long-term participants in order to investigate their perception of deliberate practice and their own view of their skill development. An important and integral part of the dissertation, apart from the results, was the development of the holistic rating scales (adapted from Carroll 1966), and the development of an in-depth interview study.

The conclusions of the dissertation are that there are measurable differences of interpreting skill between performers with little or no interpreting experience and performers with long interpreting experience, but this finding could not be supported by the long-term (intra-individual) study. Differences between the groups in the cross-sectional material could also be observed from the process data. Experienced interpreters

encountered fewer processing problems than less experienced interpreters and had more strategies at hand to solve problems. There were also clear differences in terms of instances of monitoring (i.e. controlling the interpreting process and output) between experienced interpreters and other subjects. Monitoring seemed to be a dividing line between experienced and inexperienced interpreters, and experienced interpreters had more processing capacity available to monitor themselves. This was also to a certain extent supported in the in-depth interviews, where the participants reported how they constantly evaluate themselves in terms of improving performance. A key assumption established in the beginning of the project – that experienced interpreters would claim, in the in-depth interviews, that they practise a great deal – was not supported, to our surprise. The interpreters recounted many practice-like activities but stated that they did not actually practise.

The dissertation concludes by calling for more studies on deliberate practice in interpreting, suggesting that the term “interpreter expert” should only be used with caution in scientific studies and that the particular features of expertise and deliberate practice in interpreting should be discussed.

The present thesis is based on the following papers

- I. Tiselius, E. 2009. "Revisiting Carroll's scales". In *Testing and Assessment in Translation and Interpreting Studies*, C. Angelelli and H. E. Jacobson (eds), 95–121. Amsterdam and Philadelphia: John Benjamins.

Reprinted with kind permission by John Benjamins Publishing Company, Amsterdam/Philadelphia. www.benjamins.com.

- II. Tiselius, E. and Jensen, G. B. 2011. "Process and product in simultaneous interpreting: What they tell us about experience and expertise." In *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies*, C. Alvstad, A. Hild and E. Tiselius (eds), 269–300. Amsterdam and Philadelphia: John Benjamins.

Reprinted with kind permission by John Benjamins Publishing Company, Amsterdam/Philadelphia. www.benjamins.com.

- III. Tiselius, E. In preparation. "The development of expertise – or not: Three simultaneous interpreters' development over time."
- IV. Tiselius, E. Submitted. "Expertise without deliberate practice? The case of simultaneous interpreters."

1. Introduction

The conference interpreter community, just as any group, has its stars. There were the Kaminker brothers, George and André, who could translate ninety-minute consecutive speeches without notes (Satow 1979: 511). There was Wolfe Frank, who was considered the best interpreter at the Nuremberg trials (Gaiba 1998). Or Paul Mantoux, the historian turned interpreter during the war, and the only interpreter present at the peace negotiations in Versailles after the First World War (Mantoux 1955). More modern examples include Amanda Galsworthy, who has been the English interpreter for four French presidents (Lefort 2012), or Stalin's interpreter Valentin Berezhkov (Berezhkov 1994). There are also local stars at different institutions or in different booths. Budding interpreters are told stories about older colleagues who are linguistic wizards or multi-sentence jugglers, always getting the message across. Every interpreter has a story of one particular day when the interpreting diploma was still fresh and s/he was assigned to work with one of these stellar colleagues. Stars are created in many different ways, but in order to be an interpreting star an absolute condition is to perform with excellence. In simultaneous interpreting there are few possibilities of cheating: since the cognitive load is extreme and the time is short, smooth-talking is not enough for an expert in interpreting. Interpreting excellence is not achieved overnight – on the contrary, many laborious hours lie behind a seemingly effortless performance, just as for any other profession. Researchers have labelled the type of excellence achieved through many hours of focused training “expertise”.

1.1 Expertise and interpreting research

Expertise theory was introduced to the interpreting research community by Barbara Moser-Mercer, who invited Karl-Anders Ericsson, a leading

proponent of the expertise approach within cognitive psychology, to the Ascona workshops in 1997 (Moser-Mercer 1997). Interpreting research had however focused on the skilled performers long before the expertise theory was introduced. Already early on the skills of experienced professional interpreters were investigated and compared to the performance of less experienced subjects, such as in Gerver's (1971) research on source-language presentation rate and pauses and Goldman-Eisler's (1972) early studies on segmentation and lag.

The expertise approach has turned out to be well-suited for interpreting research, and the study of expertise within interpreting has become a viable and well-established research area in interpreting studies. Moser-Mercer has looked at expertise from the learners' perspective in different studies (e.g. Moser-Mercer 2000; Moser-Mercer, Fraunfelder, Casado & Künzli 2000). Several dissertations have focused on expertise in interpreting, such as Ivanova (1999), who looked at problem-solving strategies; Liu (2001), who investigated working memory; and Vik-Tuovinen (2006), who looked at expertise in a wider perspective by including preparation.

A challenge for expertise research in interpreting is the data collection. Studies tend to be cross-sectional, comparing students or novice interpreters with more experienced interpreters, as in the studies mentioned above. Studies with a more long-term aim often compare how interpreting students develop from the beginning to the end of their programme, as in Moser-Mercer's studies. Longitudinal, or long-term, studies (for a terminological discussion on longitudinal versus long-term, see below in section 2.3.3) are likely to uncover other aspects of expertise than cross-sectional studies can reveal. It is therefore interesting to have both cross-sectional and longitudinal studies of expertise. Very few studies, if any, tend to be truly longitudinal, that is, stretch over many years and follow informants' individual development over time. An obvious reason for this is the long time span needed for such a study

combined with the level of falling off that such a sample would most likely suffer, a feature probably common to all longitudinal studies of expertise.

In this dissertation and in its literary review, different groups of interpreters will be addressed. The groups differ through their interpreting experience. Interpreting experience can be gained from an interpreting program, through professional practice or both. Terms used to describe these different groups are *(highly) experienced interpreters* for interpreters with long or very long professional experience, *little or short experience interpreters* for interpreters with short professional experience or interpreting students, and finally *no experience interpreters or subjects* for subjects participating in different studies and engaging in interpreting activities but with no prior experience of interpreting.

1.2 Aim

The starting point for this dissertation was the investigation of expertise in interpreting. At the beginning of the project I discovered a set of interpreting data that was recorded in the mid-1990s, and it seemed feasible to design a long-term study by making new recordings. With those early recordings in mind, a cross-sectional material was collected in order to complement and mirror the long-term material, and together the two data sets would be used to study the long-term development of expertise. By studying the informants' interpreting process as well as their interpreting product, combining the results from the cross-sectional material and the long-term material, and examining their expertise through in-depth interviews, I aimed to answer the following questions:

- Is there a measurable difference in the interpreting skill from the student level to the highly experienced level?
- Is there a measurable difference in the interpreting skill both when it is measured cross-sectionally (i.e. inter-individually) and long-term (i.e. intra-individually)?

- If there is a measurable difference in the interpreting skill, what does this difference consist of?
- How do experienced interpreters perceive different factors in their long-term competence development?

In order to identify and study the development of the interpreting skill, it had to first be measured, and the measurement of the interpreting skill had to rely on some type of evaluation. Furthermore, the measurement of the product had to be complemented with an investigation of the process.

As the project developed, important methodological issues cropped up. A major part of the dissertation has also consisted of adapting, testing and developing different methods for investigating expertise. Various techniques for retrospection and the categorization of processing problems, monitoring and strategy use were tested and adapted in order to map the process. Scales for assessing product were also tested and developed. Finally, an interview guide for conducting in-depth interviews was developed.

Thus, this project compared the process and product of highly experienced interpreters, novice interpreters and non-interpreting subjects, both cross-sectionally and in the long term.

2. Background

The following section will give an overview of previous research on experienced interpreters from both a cognitive and a quality perspective, before discussing interpreting as a professional practice. The section concludes with an overview of the expertise theory in cognitive psychology, the concept of deliberate practice within the expertise theory, and the definition of an expert in interpreting research.

2.1 Research on the effect of experience in interpreting

Research into conference interpreting studied experienced interpreters long before the expertise theory entered the field. Researchers were interested in the cognitive effort in interpreting and the different skills needed to interpret. Major themes that have been studied within the cognitive framework include processing capacity and different cognitive efforts. Early researchers were, among other things, interested in whether an increased cognitive load affected interpretation and how interpreters handled the allocation of processing capacity.

Precursors in interpreting research found that response time (or lag in interpreting terminology) increases as the task complexity increases (Oléron & Nanpon 1965 in Pöchhacker & Shlesinger 2002: 42). Moreover, subjects with less interpreting experience deliver more fragmented versions than their more experienced peers, and interpreters make use of the speakers' pauses for their interpretations (Barik 1973, 1975). Qualitatively significant differences in the output of highly experienced interpreters as compared with subjects with little or no interpreting experience, have been more difficult to establish (Barik 1975). However, a major difference between highly experienced and subjects without or with limited interpreting experience seems to be that experienced interpreters

segment the input more effectively (Barik 1975: 296). Speech rate, pauses and intonation affect interpreters' output (Gerver 1971), and interpreters produce longer utterances with fewer pauses than the original speaker (*ibid.*). Interpreters' segmentation of input differs from the speakers', and source-language syntax affects both segmentation and ear-voice span, that is, the time from perception of an utterance in the source language until the production of an equivalent in the target language (Goldman-Eisler 1972).

These early findings could not establish conclusive significant differences between subjects with long interpreting experience and subjects with no or limited interpreting experience. But Goldman-Eisler (1972) for one found that cognitive load increased while interpreting compared with other speech production tasks and that highly experienced interpreters seemed to handle that increased cognitive load better than subjects without interpreting experience. She assumed that decoding the input would be the process that required the most attention and that monitoring and encoding are more automatized (Goldman-Eisler 1972: 139). Barik, however, assumed that this difference was due to language direction (Barik 1975: 296). Oléron and Nanpon suggested future studies on how much information can be grasped at one time, or of the interpreters' ability to concentrate on several different things and perform simultaneously, in order to draw conclusions on the process (Oléron & Nanpon in Pöschacker & Shlesinger 2002: 50).

Many studies in early interpreting research compare subjects with limited or no experience to interpreters with professional interpreting experience (cf. Gerver 1971; Barik 1973, 1975; Anderson 1979). Hoffman points out that it seemed natural for researchers in interpreting to examine the performance of professionals and contrast it with the one of trainees (Hoffman 1997: 190). From a cognitive perspective, the clearly perceived differences between subjects with and without interpreting experience could be used to gain information on the cognitive impact or change that develops with the subjects' interpreting skills. As can be seen above,

however, it is not immediately evident how that difference could be measured in order to establish distinctive features between subjects without interpreting experience and interpreters with professional interpreting experience. In fact, some studies, such as Dillinger (1994), could not establish any distinctive difference between interpreters with professional interpreting experience and subjects without interpreting experience.

Sample size has been a methodological issue since the early days of interpreting research. Even today researchers struggle with collecting data large enough for purely experimental designs. The interpreting community, especially the simultaneous interpreting community, is small, and the highly experienced interpreters even fewer. Already Gerver commented that very few interpreters were available for experiments, and furthermore that since

not all of these are willing to take part in experiments, the design and execution of experiments on simultaneous interpretation becomes somewhat of a problem. In effect the choice had twice to be made between an incomplete experimental design or no experiment at all, and in the experiments described [here] the former decision was taken. (Gerver 1971: 26)

Traditional types of analysing tools in early research comprise error analysis (e.g. Gerver 1971; Gile 1985a); ear-voice span measurements (Gerver 1971; Goldman-Eisler 1972; Barik 1973); and assessment of interpreting (Seleskovitch 1975 as cited in Pöchhacker & Shlesinger 2002; Anderson 1979). Assessment is an important part both of interpreting practice and interpreting research and will be dealt with more in detail below in section 2.4.

Early studies of cognitive aspects of interpreting research also propose different models of the interpreting process: Gerver (1976) and Moser-Mercer (Moser 1978) propose general models of simultaneous

interpreting; Gile (1985b) focuses on the different efforts involved in simultaneous interpreting; and Darò and Fabbro (1994) map memory use. Whether focusing on the whole process, the invested effort or the different memory functions in use, all models mirror cognitive constraint of some sort – there is a limit to how much load different processes can handle. The most recent contribution to models of interpreting was proposed by Seeber (2011), whose cognitive load model aims to be a more flexible model and show how cognitive load increases and decreases depending on task.

More recent research into the cognitive aspects of interpreting has also looked at neurolinguistic aspects of interpreting. The plasticity of the brain makes it adapt to new conditions that simultaneous interpreting imposes on students of interpreting. Early neurological research in interpreting was done by Gran and Fabbro (1987). They have been followed by Rinne et al. (2000), who used PET (positron emission tomography) to study simultaneous interpreting, and more recently by Ahrens et al. (2010), who found significant differences in active brain areas between interpreting students' interpreting and their normal speech production. Hervais-Adelman et al. (2011) also found indications of change in the bilingual brain of interpreters. It should be stressed that all four of the aforementioned studies focus on students of interpreting, and their findings also correspond to research in expertise. The brain's single-domain general control network helps us learn new tasks and also plays a key role in controlling working memory in cognitive processing. However, the control network also limits the resources for working memory-dependent tasks and other novel tasks. As processing becomes automatized, the influence of the general control network decreases or disappears. Hill and Schneider (2007), in an overview of different studies on skills acquisition, show how the automatization of acquired skills changes brain plasticity (Hill & Schneider 2007: 675). The acquisition and mastery of new skills change the area activated in the brain, since the brain's plasticity entails that it can

change area and amount of activity as skills are acquired and refined (ibid.).

The function of working memory in interpreting has been studied both from a novice/experienced dichotomy, and also in itself as one of the features involved in simultaneous interpreting processes. Liu (2001), who investigated working memory from an experience/novice perspective through a listening span test, found no significant differences, nor were there any significant differences in working memory span (Liu, Schallert and Carroll 2004); however, experienced interpreters were more accurate in their performance (Liu 2008). Other researchers (e.g. Bajo, Padilla and Padilla 2000) have found that memory span increases with experience. The reason for the contradictory findings on working memory is perhaps partly explained by Timarová (2012), who found that that interpreters' working memory is related to their performance in simultaneous interpreting and that simultaneous interpreting is predominantly related to the central executive functions and not to memory functions. She concluded moreover that there was a link between interpreting experience and some working memory functions.

Research on the effect of experience in interpreting has an underlying assumption of basic translation ability (Englund Dimitrova 2005: 10). An individual who understands two languages also has a basic ability to transfer a message from one of the languages to the other. Englund Dimitrova points out that most research in the field of bilingualism takes for granted a basic translation ability. Englund Dimitrova posits that "basic translation ability is a necessary condition, but no guarantee, for further development of a (professional) competence as a translator, and possibly expertise in translation" (2005: 12). It is fair to assume that the underlying assumptions of studies using subjects with little or no interpreting experience are similar to Englund Dimitrova's postulate. In the present dissertation it can furthermore be noted that for one group of subjects who

did not have any interpreting experience at all, the subjects still produced an interpreting product – it was undoubtedly a difficult task for these subjects, but they did not suffer a complete breakdown. Thus, at least these subjects seemed to possess some type of basic interpreting ability.

As seen above, although some research has shown that interpreters with professional interpreting experience have few omissions, deliver a complete message, segment effectively, handle cognitive load well and have great working memory capacity compared with subjects with little or no interpreting experience, other research has conversely concluded that there is scant difference in accurate delivery, handling of cognitive load or working memory capacity between interpreters with professional interpreting experience and subjects with little or no experience. From earlier research it can be concluded that investigating the effects of experience in interpreting is delicate and difficult. Intuitively, experience ought to improve interpreting performance, but research results are far from conclusive.

2.2 Interpreting practice

Today, interpreting training is common both for public service interpreting and conference interpreting. This section will only focus on conference interpreting training and practice, as the studies in this PhD thesis are restricted to simultaneous conference interpreters. Conference interpreters today are usually trained, although training facilities in this domain are a recent innovation as well. The first interpreting programme was founded in Geneva in 1941. Since then, a pedagogical tradition has evolved through groundbreaking work by Herbert (1952) and Rozan (1979 [1956]), via Seleskovitch and Lederer (1995) to modern classics such as Jones (1998) or Nolan (2005).

Western interpreter training follows more or less the same path. Interpreting pedagogy is firmly rooted in *A Systematic Approach to*

Teaching Interpretation (Seleskovitch and Lederer 1995). Following this and subsequent manuals, interpreting is introduced first through short memory exercises where students are encouraged to let go of the words and look for the meaning of the utterance, a meaning that they should then render in their target language. Memory exercises gradually become longer, and note-taking is introduced. When students master the basics of note-taking, they start to interpret in consecutive mode. Both memory exercises and consecutive interpreting are believed to be a basic preparation for simultaneous interpreting (Gile 2005b). After an extended period of consecutive interpreting (from a semester up to a year), students are typically introduced to the simultaneous mode.

Interpreting students are also taught to practise on their own, outside of teacher-led training. This has been an important characteristic of interpreter training since the early days, although Seleskovitch and Lederer do not provide guidelines for student-led practice but refer to how students should practise and how the teacher should guide that practice (e.g. 1995: 158). Students are expected to practise sub-skills such as language knowledge and general knowledge, often by reading newspapers, watching TV or listening to the radio, but they are also taught to practise interpreting and to record themselves in order to evaluate their performance (Gile 2005b: 135–136).

Interpreter training is guild-like in the sense that active interpreters teach their future colleagues. Furthermore, interpreter training has been developed from a pragmatic rather than from a theoretical perspective. Interpreter training has since the early days been based on active interpreters' perceptions of what needs to be taught in order to succeed as an interpreter. Their views are confirmed as their students graduate and practise successfully. Sawyer (2004) and Iglesias Fernández (2003) have made very comprehensive overviews on interpreter training. Sawyer found that the extensive research and debate on assessment in interpreting has

been poorly reflected in actual interpreter training (2004: 211). He also stresses the need for test validation (2004: 231). It could probably be claimed that interpreter training is more practice-driven than research-driven. This does not necessarily mean that interpreter training is inadequate, but although much has been studied and written on the matter, interpreter training still frequently takes its starting point in tradition, and there are no major empirical studies on interpreting methodology or didactics (cf. Pöchhacker 2004: 183).

As mentioned above, students are taught to practise and assess their interpreting skills at the interpreting programme. Whether interpreters continue to do so in their professional careers has not been mapped in any larger studies, although a study by Leis (2003) suggests that interpreters do assess themselves. A strong norm in conference interpreting is preparation (e.g. Bühler 1986), that is, students are taught to prepare by improving their background knowledge and enhancing their terminology within a certain topic. Professional interpreters are expected to prepare, and professional experience depends not only on hours in the booth but also on practice and preparation.

2.3 Expertise

The following section introduces different approaches to expertise and the concept of deliberate practice. Expertise has been studied from many different perspectives, ranging from theories where talent is the only condition for expertise to those where focused training is deemed more critical for reaching an expert level.

2.3.1 Different concepts of expertise

Already Plato was interested in the expert mentality, contending that humans could be divided into three different types according to their innate aptitude: soldiers, workers and leaders (Ericsson 2009). Over two

millennia later, the British anthropologist and psychologist Francis Galton (Ericsson 2007a: 684) observed that distinguished contributors to society all came from more or less the same background, leading him to assume that talent and excellence were due to an inherited difference in mental capacities. The idea that expertise is developed through training and practice, which is a cornerstone in many current definitions of experts (cf. Ericsson 2007b: 10–12), was a reaction to the prevalent notion that talent was an absolute condition for success in different fields. Therefore, rather than possessing and relying on a unique talent, the aspiring expert must be prepared to spend many hours of focused practice, often from a very early age. A famous contribution to prove this claim is the Polgár couple, who trained their daughters very early on to become elite chess players and thereby demonstrated that pure talent and the supposed male advantage in chess are pure fantasy (Ericsson, Prietula & Cokely 2007). In this view, expertise as opposed to talent means that the expert has spent many years of specific focused training, so-called deliberate practice (Ericsson 2004: see below, section 2.3.2). Certain other experts, for instance athletes and musicians, also start at a very young age.

The theory of expertise, which has become influential in both Interpreting and Translation Studies, has been developed by researchers in cognitive psychology such as Ericsson, Charness, Feltovich and Hoffman (2007). From a cognitive perspective, an *expert* is an individual who has acquired great knowledge in a given field and who can make use of this knowledge to outperform other performers. To continue in Ericsson's words: "*expertise* then refers to the characteristics, skills and knowledge that distinguish experts from novices and less experienced people" (2007b: 3). *Expert performances* are reproducible superior performances of tasks that capture the essence of the respective domains (Ericsson et al. 2007: 3–4). Furthermore, two types of expertise can be singled out, namely routine expertise and adaptive expertise, where routine experts excel in well-known routinized tasks, whereas adaptive experts are able to handle new

tasks and can apply previous knowledge to new situations within their area of expertise (Sonnentag, Niessen & Volmer 2007: 377–378).

Ericsson and Smith (1991) pointed out that studies of expertise often only looked at differences between experts and novices or less experienced individuals, instead of investigating expert performance characteristics within a particular domain. They believed that an expert may be socially recognized as an expert without necessarily showing superior performance in absolute terms compared with other performers (Ericsson & Smith 1991). In an earlier article (Tiselius 2010), I argued that being an expert is impossible in a social vacuum and that, in line with Ericsson and Smith's argument, expertise requires both social appreciation and superior performance.

Another concept of expertise, from a more heuristic perspective, is interactional expertise, proposed by Collins and Evans (2007: 2). Interactional expertise is a type of expertise that is co-created between parties in a particular field. Collins and Evans point out that expertise can be defined on several different levels, with the most advanced level being contributory expertise, which they define as the stage when an individual has gained specialist knowledge and can help disseminate and increase such knowledge (Collins & Evans 2007: 2). According to the theory of interactional expertise, an expert is not always the best person to decide how to put his or her expert knowledge into practice, and it is through mutual discussions between experts and lay people that the best solution can be found.

The expertise approach proposed by Ericsson and Smith (1991) focuses on the individual performer. According to Ericsson and Smith, expertise in a field is achieved through a combination of various characteristics (1991: 7, 20–21, 27–28). These characteristics, which are also the ones applied in the research project reported here, consist of at least the following:

(1) *Experts have regular outstanding performances in their field of expertise.* The expert has to show regular proof of expertise; a single top performance is not equal to expertise.

(2) *Experts have access to expert knowledge when needed.* Experts do not necessarily outperform other participants on routine tasks, but excel over novices in difficult situations encountered within their area of expertise.

(3) *Experts have long experience in their field of expertise.* Experts have spent at least ten years or 10 000 hours on task and in practice. It should be stressed that this is the weakest predictor of expertise. In many contexts, in particular in popularized accounts of the theory (e.g. Gladwell 2008), ten years of experience has been put forward as a sole or at least dominant factor to determine or achieve expertise. Clearly, non-expert performers may have spent an equal amount of time on task, without achieving expert levels of performance. However, expertise is hardly possible without extensive experience.

(4) *Experts engage in deliberate practice.* Deliberate practice is a highly focused and regular practice, completed at a time set aside only for practice and solely aimed at improving the given skill. It is also characterized by specific exercises and is often coached.

(5) *Experts have clear goals.* The notion of having clear goals is partly connected with the concept of deliberate practice, as the practice is goal-defined. Furthermore, final goals are usually divided into reachable part-time goals on both the micro and macro levels.

(6) *Experts are open to feedback.* Experts have a positive view of receiving feedback and are good at integrating it, both from superiors and peers.

A challenge for anyone wishing to investigate expertise in domains where there is no external ranking is to define outstanding performance. Ericsson and Smith (1991) give a three-step method for doing this: (1) scientifically analyse the domain, its particular expert skills and performance within the framework of general cognitive theory; (2) identify the task's process and structure and the performers' behaviour; and (3) show how superior performance in that field is built up through the given cognitive processes and how they were acquired. The three-step model is ambitious and far-ranging, and it would entail a very large project to cover these three steps in order to investigate, for instance, simultaneous interpreting. It can be claimed, however, that the growing number of studies on expertise in interpreting helps to build this three-step model. Ericsson (1996) argued that perhaps not all domains are possible for understanding and measuring expertise, as it requires expert performances of objective superiority that can be reproduced. It is indeed a challenge for interpreting research to show that these requirements are achievable.

Ericsson's expert approach has had a great impact on expertise research both in psychology and translation and interpreting studies. There are other proposals and models of how to interpret the notion of expertise from a cognitive perspective. Shanteau's (1992) theory of expert competence aims to reconcile two views existing at that time, namely the cognitive perspective that claimed that experts were cognitively different in every aspect compared to other performers, and research into judgment and decision where experts had made flawed decisions despite their expertise. Shanteau suggests they are both right, but the analysis is incomplete. Instead, in his theory he claims that expertise is built up of five components, namely (1) a sufficient knowledge of the domain, (2) the psychological traits associated with experts, (3) the cognitive skills necessary to make tough decisions, (4) the ability to use appropriate decision strategies, and (5) a task with suitable characteristics. Shanteau may prove useful for this PhD project and for expertise in interpreting, as

measuring expertise in interpreting seems to be a notoriously challenging activity.

If Ericsson's expert characteristics are contrasted with Shanteau's expert components, it is clear that they are neither completely opposite nor completely parallel. Shanteau's second component, "psychological traits associated with experts", could encompass Ericsson's "regular outstanding performance" and "access to expert knowledge when needed". On the other hand, the "access to expert knowledge" would also encompass both Shanteau's "sufficient knowledge of the domain", "cognitive skills necessary to make tough decisions" and "the ability to use appropriate decision strategies". Ericsson also adds "deliberate practice", "clear goals" and "openness to feedback". These three cannot easily be put into any of Shanteau's components. They contribute to for example "psychological traits..." or "cognitive skills...", but they are not an uncontroversial part of them. An important difference between Shanteau's components and Ericsson's characteristics is that Ericsson's characteristics have a developmental part. They encompass the learning perspective by stressing the importance of a subject's deliberate practice and openness to feedback. The sociological part of Ericsson's expertise approach lies in the notion of the subject who engages in deliberate practice and receives feedback and coaching from peers or coaches.

In more recent works, Weiss and Shanteau (2003) have developed an index to empirically assess professional expertise. In their presentation of the index they mention precisely the problem of measuring fields without ranking. They say:

For many tasks at which experts make a living, no measurable outcome exists. How is one to know if the wine taster has judged accurately or if the professor has graded the essays well? Adherents of the expert performance approach would question the merits of studying such domains. Although there is no hint of an objective

external criterion, we believe that some people do these tasks better than others and that people improve their performance. (Weiss & Shanteau 2003: 105)

Unfortunately, when looking closer at their index, it pertains to expert judgments and evaluations and not at the type of expert performance present in interpreting.

Recently, Muñoz Martín (forthcoming) suggested an adaptation of Shanteau's five components to translation expertise. Muñoz Martín suggests five dimensions that consist of (1) knowledge, (2) adaptive psycho-physiological traits, (3) problem-solving skills, (4) regulatory skills, and (5) the self-concept. These five components, although still lacking the concept of deliberate practice as a dimension of its own, are more appealing for studying expertise in interpreting. The five dimensions were put forward very recently and have not yet been empirically tested. For the present thesis the notion of deliberate practice is very much the crux of the matter. So for the purpose of the PhD project reported here, Ericsson's expert approach and characteristics will be followed in the strictest sense possible.

2.3.2 The concept of deliberate practice in the expertise approach

As seen above, one aspect of Ericsson's expert theory is the performers' deliberate practice, a developmental feature of expertise. Ericsson divides the performer's activity into three types (Ericsson, Krampe & Tesch-Römer 1993: 368):

- (1) *work* – an activity is defined as work when it is publicly performed and most often performed for remuneration.
- (2) *play* – an activity is defined as play when it is performed without remuneration, and without a particular goal for the activity, the performer's pleasure during activity is an important part of play.

(3) *deliberate practice* – an activity is defined as deliberate if it is performed at a clearly delimited occasion, with specific exercises (often decided beforehand). The activity is also performed with a clear goal to improve or refine the activity and with an evaluation of the performance.

Experts can be said to be constantly challenging the status quo of their performance, and the expert's deliberate practice is the instrument for that challenge. Highly skilled performers can either stagnate in an automatized mode, or they can excel in expertise by engaging in deliberate practice (Ericsson 2007a: 685). Neither arrested development nor an automatized mode should be confounded with routine expertise (see section 2.3.1). Routine experts have not necessarily stagnated in automatized mode, but nor do they necessarily adapt their expertise to new challenges as adaptive experts do. Deliberate practice is the counteraction to stagnation. The performer who engages in deliberate practice does so over longer periods of time, and the occasions of deliberate practice are focused and well-planned. The performers' practice is also analysed either by the performers themselves or by their peers or coaches according to the set goals or expected levels of achievement (Horn & Masunaga 2007). The planning and evaluation of the practice is thus what contributes to the development of expertise. Horn and Masunaga also define deliberate practice as

focused, programmatic, *carried out over extended periods of time*, guided by conscious performance monitoring, evaluated by analyses of level of expertise reached, identification of errors, and procedures directed at eliminating errors. (2007a: 601, my italics)

Ericsson adds that

the core assumption of deliberate practice is that expert performance is acquired gradually and that effective improvement of performance requires the opportunity to find suitable training tasks that the performer can master sequentially. (2007a: 692)

The individual's deliberate practice is also guided by clear goals and openness to feedback. It is important for experts to be able to break down their activities into reachable part-time goals that can be achieved over shorter periods of time. Performers acquire expertise in their field gradually. Feedback and learning from peers are also important activities in the development of expertise. Experts develop through feedback from coaches, and by observing their peers (Ericsson 2007a: 692).

Although it may be challenging, as has been laid out above, to map and measure superior performance in simultaneous interpreting, the concept of how deliberate practice is executed over longer periods of time seems even more challenging to observe and investigate. It cannot be measured through experiments, but must rather be studied through interviews or journals. Studies that look at subjects' deliberate practice over time in other fields include Sosniak (2007), who used retrospective interviews, and Deakin et al. (2007), who used journal studies. Deliberate practice can also be studied on a micro level, where the use of practice techniques at one particular (often experimental) occasion is studied. The techniques used at this particular session can then be compared between highly skilled performers and less skilled performers (cf. Zimmerman 2007). The fourth article in this thesis is devoted to an in-depth interview study of skilled interpreters' deliberate practice. Prior to the actual interview study, a pilot focus-group study was made (Tiselius 2010). Two un-moderated focus group discussions were carried out over different themes in interpreting. The participants were conference interpreters of the Swedish booth at the European Parliament, both male and female and with a wide age and experience range. The aim of the study was to explore the sociological aspect of expertise, that is, how interpreters viewed their colleagues, work, customers and so forth. The focus group study showed a terminological challenge connected with the concept of deliberate practice. Participants did not intuitively understand the concept of deliberate practice. As a consequence, the analysis of the in-depth interviews in article 4 required

the researchers to interpret participants' responses to different trigger questions in order to study deliberate practice. Deliberate practice is a scientific research concept, a theoretical construct. It is not necessarily the pedagogical or professional term used by professionals themselves. Any interview or questionnaire on deliberate practice will have to address how professionals in a particular field label that particular type of practice and how to make them talk about the different parts of that construct. Participating in an interview could possibly also qualify as deliberate practice, a type of learning through introspection. A full account of the methodology is given in article 4.

As has been shown, deliberate practice is a crucial part of Ericsson's expertise approach. Practice and preparation are also important features of interpreting, as discussed above under section 2.2. When studying expertise in interpreting, it therefore seems inevitable that we must investigate interpreters' possible engagement in deliberate practice as well as their practice habits.

2.3.3 Research on expertise in interpreting

Expertise research in interpreting studies was briefly introduced in section 1 of the introduction. The first part of this section gives an overview of the subjects in earlier studies on interpreting expertise, their profiles, and how are they labelled. The definition of an experienced interpreter or even an interpreting expert differs a lot as shall be seen. The second part of the section will look at results of research that contrasts experienced interpreters with subjects with little or no interpreting experience.

In one of the first articles in interpreting studies on expertise, Hoffman (1997: 192–193) wrote that “both psychological research on expertise and expert system development efforts have actually tended to define expertise rather loosely, or variously [...]. A general challenge to scientific psychology is to generate a definition of expertise that focuses on

cognitive functionality and yet can be used operationally to identify experts”. Moser-Mercer et al. (2000: 108) wrote that “in translation and interpreting it is often acknowledged that the student obtaining his final diploma can call himself an expert with some degree of justification, but that years of experience in the field are still required for him to become a full-fledged professional”. This statement may hold true for the profession, but it is a rather unfortunate wording for interpreting expertise research. In a relative view of expertise, a recent graduate from an interpreting programme will most likely have more expertise than a student, but following Ericsson’s expert characteristics the recent graduate is probably far from being an expert. In his review of interpreting research, Hoffman (1997: 199) divides performers into different categories depending on their former experience of interpreting. He calls subjects without any interpreter training or professional interpreting experience a *naive*, pointing out that “novice” is actually a misnomer of “naive” in many studies, as the term is used for subjects without any previous knowledge of the field in question. Students are called *novices* when they start out in their interpreting programme, *initiates* when they have been initiated to a new skill (e.g. simultaneous interpreting), and *apprentices* when they are in their final stages or are recent graduates. Interpreters with five years of professional experience after graduation are labelled *journeymen*. This categorization loosely follows Dreyfus and Dreyfus’ (1980) proposal of a skill acquisition model comprising the levels of novice, competence, proficiency, expertise and mastery; this system has its roots in the terminology of crafts guilds. Following Hoffman’s categories, the subject with a recent interpreting diploma has reached the level of final-stage apprentice and can most likely not call himself an expert in absolute terms.

In translation studies, Jääskeläinen (2010) asked the pertinent question of whether all professionals are experts. Jääskeläinen points out that early studies on translation process contrasted non-professionals and professionals, and it turned out that many of the professionals did not

produce any outstanding performances. Discussing definitions of expertise, she observes that although expertise research can be divided into two research approaches, *absolute expertise* (study of outstanding performers) and *relative expertise* (comparison between novices and experienced performers), the focus in translation studies has been on relative expertise (2010: 217). She furthermore points out that the professional participants in different studies may not have been completely screened following all expertise criteria (for example, several of the studies comprise participants with less than ten years of experience), and there is no investigation of the type of experience the participants have had.

Going back to interpreting studies and the studies published on expertise, there are several different definitions of experts or professionals in this domain as well. Vik-Tuovinen (2006: 129) points out that sorting the participants in interpreting studies into different experience categories as the ones mentioned above is not a very straightforward task. Reviewing a few studies of expertise in interpreting supports Vik-Tuovinen's observation. In Chincotta and Underwood's (1998: 8) study on bilingual digit span, the professional group (n=12) comprised simultaneous interpreters with at least 100 hours of interpreting practice, while the non-expert control group (n=12) consisted of students of English with no interpreting experience. Ivanova (1999) studied professional interpreters with an average of nine years of experience (n=8) in an expertise study on discourse processing; the novices in her study were interpreting students with three months of experience (n=8). Moser-Mercer et al. (2000: 126) compared the performance of novice and expert interpreters in three different experiments: in the first experiment the professional interpreters (n=5) had between five and ten years of experience, while the students (n=5) were recruited from the first semester of the interpreting programme; in the other two experiments, they labelled their participants professionals interpreters and students without specifying their background

(the number of participants was $n=6+6$ and $n=5+5$ respectively, and five of the participants may possibly be the same as in the first experiment). In a PhD dissertation on working memory (Liu 2001: 32) and a subsequent article (Liu et al. 2004: 24), Liu studied professional interpreters ($n=11$) with at least one year of full-time interpreter training and at least two years of professional experience with at least 40 days of interpreting per year. She had two groups of novices in her study, one were students at the end of their first year of the interpreting programme ($n=11$), the other group was at the end of their second year at the interpreting programme ($n=11$). Vik-Tuovinen (2006: 127) included preparation and debriefing in her study of interpreting at different experience levels. Her expert interpreters ($n=7$) had professional experience of between one to fifteen years, with 10 to 100 days of interpreting. There were two groups of novices in Vik-Tuovinen's material: the first one was recruited at the first term of the interpreting programme, and at the time of the recordings they had 10 hours of interpreting practice ($n=6$); and the second one was recruited at a later stage in the training where the students had well over 180 hours of interpreting practice ($n=8$). Köpke and Nespoulous (2006: 6) studied the differences of working memory between novices and experts. They recruited 21 professional interpreters (12 staff and 9 freelancers) who had between 4 and 35 years of experience. The interpreting students ($n=18$) who participated in their study were recruited in their second and final year of interpreting studies, and they had just started simultaneous interpreting practice. Köpke and Nespoulous also had two control groups, one consisted of bilinguals ($n=20$) and served as the control group for the professional interpreters, the other one consisted of students ($n=20$) and was the control group for the interpreting students.

Though this overview does not claim to cover all studies that have been done with an expert approach or within the expert theory in interpreting, it is quite clear that no consistent profiling exists of either the highly experienced or the less experienced participants. It shows how persuasive

the expert concept is in interpreting studies. In most studies (including my own article 1, because of an unfortunate misunderstanding in the editing process), experienced and professional interpreters are indiscriminately labelled experts regardless of length of experience or performance. In her literature review on experts and interpreting, Liu (2008: 160) points out that since studies on expertise in interpreting use the relative concept of expertise, this entails that when more experienced interpreters are compared with less experienced ones, then any more skilled group can be considered experts and any less skilled group novices. The terminological issues of expertise are the same as Jääskeläinen (2010) emphasized above. Nevertheless, although there may be many indications of expertise, it is impossible to conclude without prior screening that the professional interpreter actually is an expert in the strictest sense of the expertise theory definition. Moreover, the description of the participants or subjects needs to be minute in order to enable literature reviews, study comparisons and replication.

It is clear from the different suggestions of developmental categories by Dreyfuss and Dreyfuss (1980) and Hoffman (1997) described above, as well as from the studies reviewed in this section, that the novice-expertise dichotomy is in fact a continuum. There are other dichotomies at play as well: for example, a student of interpreting acquires interpreting experience in class but no professional experience, while a professional interpreter can have professional experience without having undergone a training programme. The continuum and the different dichotomies in play are described in figure 1. Interpreters who have graduated from an interpreting programme and started to receive remunerated work are professionals, as are experienced interpreters and also expert interpreters. Although novice interpreters can develop into experienced interpreters and also become experts, experience in itself will not make them experts, and far from all professional interpreters can be labelled experts in Ericsson's terms. In this dissertation, the participants of the different studies are

labelled laypersons (meaning they have no experience of interpreting neither from training, nor professionally), interpreters-in-training, short professional experience interpreters and long professional experience interpreters (two groups with fifteen and twenty-five years of experience, respectively). The laypersons have no previous experience of interpreting, but participated in the same interpreting experiments as the other participants. In the case of the no-experience subject, experience refer to interpreting experience, they may have many other areas of experience. These labels are added in the last line of figure 1

	Novices ←-----→ Experts					
	Laypersons	Students of interpr.	Recent graduates	Interpreters w/ exp short	Interpreters w/exp long	Expert Interpreters
Prof. Inter- preting Exp.	No	No	Possible	Yes	Yes	Yes
Inter preting Exp.	No	Yes	Yes	Yes	Yes	Yes
Label in current PhD project	No- experience laypersons (NE)	Subjects- in- training (IT)	-	Short- experience interpreter (SE)	Long- experience interpreters (LE 15 and LE 25)	-

Figure 1. Novice-Expert continuum.

Findings from studies of expertise in interpreting have been thoroughly reported in Liu (2008). She reviews an important part of the cognitive, empirical studies made on interpreters with professional experience, and cognitive, empirical studies where interpreters with professional experience are compared with students of interpreting or subjects without interpreting experience. Liu approaches the studies from the perspective of interpreting skills. She divides the interpreting skill in three parts, as she suggests that there are three obvious processes in interpreting comprehension, translation and production (2008: 161); these three main skills are then divided into sub-skills and cognitive abilities (concurrent articulation, articulatory suppression, working memory and attention

shift). Through findings from different studies (e.g. Goldman-Eisler 1972; Barik 1975; Chernov 1979; Dillinger 1989; Isham 1994; Bajo, Padilla & Padilla 2000; Liu 2001) that have compared interpreters with different levels of interpreting experience and their performances in these cognitive areas, Liu distils some common features of the sub-skills among experienced interpreters that may serve as an indication of expertise in interpreting. Results from the different studies show that experienced interpreters are better at semantic processing than subjects without interpreting experience, and also that experienced interpreters are better at selecting the most important meaning units when circumstances called for that. These results also echo the findings by Vik-Tuovinen (2006), who found that the interpreters with short interpreting experience in her study focused more on the meaning of single words, whereas the interpreter with professional interpreting experience focused on understanding the content. Liu (2008: 164) goes on to show that studies have found that experienced interpreters process longer chunks than subjects with shorter or no interpreting experience, and that experienced interpreters from English into Russian produced fewer syllables (2008: 165). She also mentions that few studies have investigated how experience affects the interpreter's delivery. Vik-Tuovinen (2006: 305) found that the experienced interpreters her material were much more conscientious about delivering the best possible product to their clients. In terms of monitoring output, Liu (2008: 167) points to several studies on delayed auditory feedback that have shown that interpreters with professional interpreting experience are less disturbed by delayed auditory feedback. Working memory studies, including Liu's own (2001), have shown that interpreters with longer professional experience have a larger digit span than their less experienced counterparts. When it comes to attention, Liu (2008: 173) mentions Cowan (2000/2001), who suggested two explanations for attention function in simultaneous interpreting: (1) rapid attention switching between listening and speaking, and (2) well-practised listening and speaking skills. Liu

argues that the studies she reviews support Cowan's two assumptions, for instance, the fact that pauses and pause length increase as cognitive load increases, or the fact that subjects with little or no experience produce more fragmented output. Liu suggests that this is due to the experienced interpreters' ability to have an overall perspective on the interpreting situation, a conclusion that also Vik-Tuovinen puts forward (2006: 308–309). It should be pointed out though, as Liu in fact does (2008: 160), that several of the studies Liu refers to compare experienced and inexperienced interpreters, rather than experts and novices in Ericsson's strictest sense. Liu concludes from the literature review that expert interpreters produce fewer errors and faster responses and use less effort. She goes on to say that there is more to expert interpreting than speed and effort, such as qualitative differences of process and output. She observes that

expert interpreters seem to have developed well-practiced strategies in each of the comprehension, translation, and production processes. [...] These strategies are developed and practiced as a result of the interaction among the comprehension, translation, and production processes that are specific to the needs of simultaneous interpreting. [...] It seems that expert interpreters have developed an ability to efficiently manage their attention so that it can be switched between different processes. (2009: 174)

Liu's conclusions are by no means controversial or questioned here. But from the background of the participants in the material she reviewed, the results are possibly true for experts, but they are first and foremost true for the experienced interpreters who participated in the different studies. These interpreters are most likely both professional and experienced, but can they be called experts in absolute terms? Liu concludes that interpreting studies is only beginning to piece together the evidence to create a more coherent picture of expertise in interpreting (2008: 174). This is quite true, and in order to create this more coherent picture of

expertise we need to clarify definitions, as common definitions are a pragmatic way to determine what is comparable and relevant.

Finally, it should be pointed out that most, if not all, studies comparing different subjects with different types of interpreting experience are cross-sectional. One of the sets of data in this dissertation consists of participants that were recorded at two different points in time: first when attending an interpreting programme in the mid-1990s, and then in the present day for the purpose of this project. As the project developed, different ways of labelling these data was discussed. The term “long-term” was chosen over “longitudinal” for several reasons. First, a typical longitudinal study comprises several points of study or contact over a longer period of time, whereas this study only had two points of study (during the interpreting programme and fifteen years later). Second, as Hansen (in press) points out, if the first experiment is repeated and long periods of time go by between the first data collection and the second time, it can be questioned whether the exact same experiment can be performed, with the same methods and under the same conditions. In analogy to the German difference between *Langzeitstudien* and *Längsschnittstudien*, she decides to use “long-term study” as equivalent to *Langzeitstudien*, and in this text I will follow Hansen’s lead. Similar data sets as in this dissertation project have been used in translation studies by for example Hansen (2008), Azbel-Schmidt (2005) and Göpferich (2013).

2.4 Assessment

Up until this section, the discussion of the theoretical background has dealt with the processes of interpreting and how to study them. This section deals with the product of interpreting. Assessment and evaluation of interpreting assume some type of approach to quality. But the purpose of this PhD project has not been to determine or define quality in interpreting, and looking at their interpreting product is but one possibility for

investigating the difference between groups of subjects. Presumably, interpreters with long professional interpreting experience deliver a product of higher quality than subjects with little or no interpreting experience. In order to find out whether there is a quality difference in the product between these different groups, the product has to be assessed. It should be stressed that the aim of the assessment (and the instrument used) was to test the main skill of interpreters (the interpreting skill) rather than different sub-skills (e.g. language, memory or coordination).

How to assess interpreting is another crucial topic in interpreting studies. The first article of this thesis is devoted to developing an assessment tool, and three of the four articles discuss the evaluation of interpreters. As Angelelli and Jacobson (2009: 3) point out, few assessment instruments or methods in interpreting are based on valid and reliable measures stemming from empirical research. In order to remedy this, they suggest a holistic, rubric-based system that can be tested for validity and reliability (2009: 38–39). They note that there is a tension between theory and practice when it comes to assessment, and that “practitioners believe that expertise in testing is obtained by practical experience” (2009: 45). The trust in practical experience may be a reason for the relative lack of documented testing instruments for interpreting performance. This in turn may also underlie the manifold flora of testing instruments in the research literature. Many researchers develop their own instruments, or rely on traditional, intuitive grading. Kalina (2005) proposed several instruments in order to assure quality by assessment, establishing a model where interpreting should be assessed not only from the output on task, but also from all the different features involved in creating high quality before and after the interpreting assignment (2005: 780). In a more recent contribution (Kalina 2011: 169), she proposes a protocol for assessing students in exams. Both of these proposals are of a componential type. In light of the many questionnaires that have been used with the aim of establishing how quality is perceived by both users of interpreting and interpreters, Moser-

Mercer (2009) stresses the importance of defining the construct of quality “clearly, precisely and unambiguously” (2009: 146), stating that the construct has to be operationalized in order to design valid and reliable measures.

Validity, whether the instrument is actually measuring what it was designed to measure, is crucial in this context. When it comes to validity, Collados-Aís (2011) and her team have made important contributions. In a series of studies they have broken down the components that are typically present in interpreting assessment, for instance, accuracy, accent and speed, and have shown that the assessment of these components are not necessarily valid as other components affect the evaluation too. For instance, the exact same speech received lower scores for accuracy if delivered with a foreign accent, despite raters’ pre-assessment claim that accent was unimportant.

Clifford (e.g., 2001, 2004, 2005) has also contributed to the field of assessment. He points out that assessment has often been linked to a text-semantic similarity or exact reproduction. This stems from the view of the interpreter as a neutral conduit rather than a participant in the event. As explained by Clifford,

the conduit portrays interpreting as an exercise carried out on linguistic forms, one in which even the smallest changes in perspective are not permitted. As noted in the literature, the conduit has at times been called the traditional perception in interpreting [...], its central perspective [...], and even its ideal [...]. (2004: 92)

In an earlier article, however, Clifford (2001: 366) argued that assessment of interpreters should determine whether they have the competencies required for professionals, and he asked the pertinent question of which competencies need to be assessed in professional interpreting. Clifford also refers Berger and Simon’s (1995, cited in Clifford 2001: 373–374) four-step assessment cycle: (1) intention (what is being assessed and why);

(2) measurement (data collection and marking); (3) judgment (when judging, a common system, understood by all assessors, must be used); and (4) decision (a decision is fair and equal if previous steps are followed rigorously). Clifford also cites Berger and Simon's principles of quality assurance in assessment (1995, cited in Clifford 2001: 375): (1) validity (the instrument measures what it was intended to measure); (2) reliability (it gives the same results in tests and re-tests); (3) equity (assessors are aware of possible gaps in performance between different groups); (4) utility (an instrument is practical to use in any given situation, i.e. not too expensive, complicated or bulky); and (5) comparability (the test can be compared although different conditions apply, e.g. different language combinations).

Although not irrelevant for an assessment in a research study, Clifford's assessment cycle may be less applicable, but several of the quality assurance principles should be as important for research as for practical applications. When the instrument used in this PhD project was developed and adapted to the studies, great care was taken to ensure validity, reliability and utility, as can be seen from the presentation in Article 1. In the case of this thesis, an instrument used earlier used by Anderson (1979) – namely Carroll's two scales (1966), one for intelligibility (whether the interpretation in this case is understandable in the target language) and one for informativeness (how much of the information from the source language message that is kept in the target language message) – was chosen for further investigation and development. The adaptation of the scales to this thesis will not be discussed here, as it is done in depth in Article 1 and below in section 3.1.2., suffice to say that they were chosen for their holistic and non-compartmental character. As for Clifford's principle of equity in assessing quality assurance, the conditions for determining equity do not seem quite applicable to this type of research. Clifford's last principle was tested to some extent, as the comparability was checked for assessment from audio files as compared with transcripts.

The comparability of the scales will have to be determined in the future, however, when they are hopefully used in other studies. One thing that springs to mind when re-reading Clifford's 2004 article is that an exact match between interpreting and the original is not desirable, as that would reduce the interpreting into an incomprehensible word-for-word rendition. Hence, full score for informativeness in Carroll's scales may not represent an ideal interpreting.

It is also important to discuss who should rate the interpreting product. Interpreters have knowledge of the interpreting process and of what is required of the interpreting product. They are also trained to assess themselves and their peers from the interpreting programme. On the other hand, they may be familiar with, or even friends with, the interpreter who is being rated, which in turn may bias the rating. Furthermore, interpreters are not the end users of the interpreting product and might not share their perspective on what is important. In a scientific study, it may seem natural that the researcher assesses the interpreting. But the researcher may also be biased, whether by knowing the subjects or by meta-knowledge of the interpreting process. In the case of the present PhD project, both interpreters and laypeople (i.e. non-interpreters) were used as raters.

Assessment is an important part of investigating the interpreters' product. Even though quality is, and has been, a hot topic in interpreting studies since its beginnings, thoroughly researched assessment instruments are still lacking.

2.5 Research questions and methodological development

The research aims of this PhD project were twofold. As described in section 1.2, the project had both methodological goals and research questions. The research was guided by the following questions:

-
- 1) Is there a measurable difference in the interpreting skill from the student level to the highly experienced level?
 - a. It was assumed that there would be a measurable difference in the interpreting skill.
 - 2) Is there a measurable difference in the interpreting skill both when it is measured cross-sectionally (i.e. inter-individually) and long-term (i.e. intra-individually)?
 - a. It was assumed that there would be a measurable difference in the interpreting skill regardless of data.
 - 3) If there is a measurable difference, what does this difference consist of?
 - a. It was assumed that there would be a difference in the assessment.
 - b. It was assumed that there would be a difference in the interpreting process.
 - 4) How do experienced interpreters perceive different factors in their long-term competence development?
 - a. It was assumed that experienced interpreters would claim that they practise a lot.
 - b. It was assumed that experienced interpreters would claim that they constantly strive to improve their interpreting performance.
 - c. It was assumed that experienced interpreters would be able to talk about their goals, on both the micro and macro levels.
 - d. It was assumed that experienced interpreters would claim that they made use of their colleagues for feedback and help.
 - e. It was assumed that experienced interpreters would be able to describe how they work under pressure.

The methodological development comprised the following questions:

- 1) Can holistic scales for measuring intelligibility and informativeness be developed into a valid and reliable measuring instrument for quality in interpreting?
 - a. It was assumed that the scales could be developed and tested so as to form a valid and reliable measuring instrument for quality in interpreting.
- 2) Will holistic scales work equally well as a measuring instrument whether used by laypersons or experienced interpreter raters.

- a. It was assumed that the scales could be used by different raters and still generate valid results.
- 3) How should an in depth-interview be carried out in order to yield results on the concept of deliberate practice?
- a. It was assumed that an interview guide had to be created where participants would be prompted to discuss issues of deliberate practice without being familiar with the research concept of deliberate practice.

3. Data and methods

This section will provide the methodological background for the four articles, discuss the rationale for choosing these particular methods and instruments, and present the participants.

It should be stressed at this point that I am an active conference interpreter myself. I have strived to study my material and conduct my project from an etic perspective, using a scientist-oriented approach. It is, nevertheless, impossible to completely shed my emic knowledge and bias. I hope, however, that I have been sufficiently accurate in my research design and the presentation of both method and results so as not to bias my findings.

Several of the research questions deal with methodological issues, and it has already been pointed out that one of the aims of the PhD project was to develop and test different methodologies. This section will therefore also include a methodological discussion. The instruments used will only be briefly introduced, however, as they are thoroughly discussed in the articles.

3.1 Methods

The first part of this section will deal with different methods for collecting and analysing data.

3.1.1 Investigating expertise

The data in this project have been taken from two groups: a long-term group and a cross-sectional group. The long-term group is unique: the participants in that group were recorded for research purposes for the first time when they attended an interpreting programme in the mid-1990s. Much of the work on this thesis has revolved around which methodological approach would be the most suitable to make use of this

unique material, and the methods have also been developed and refined with this in mind. As will be shown below (3.2.1 and 3.2.2), both the experienced interpreters of the cross-sectional material and the participants in the long-term material, after having gained more than fifteen years of experience, all showed superficial signs of expertise. They had reached the levels of contributory expertise labelled by Collins and Evans (2007: 14). The participants had credentials, experience and a track record, and have acted as examiners, peer-reviewers and so forth. Though they are experts according to Collins and Evans' terminology, they have not been tested according to Ericsson's criteria, the more cognitive side of expertise. Though it was not the main aim of this thesis to determine whether or not these participants are experts in Ericsson's terms, the project's various measurement will presumably provide some insight in that regard.

As pointed out above, the expert performs consistently at a superior level compared with other performers. The investigation of performance is therefore central when studying expertise. In her literature review, Liu (2009) observed that the experienced interpreters in the studies she reviewed showed qualitative differences of both process and output compared with less experienced subjects. Ericsson and Smith (1991: 8) want research within the expertise approach to describe the critical performance under standard conditions. The performance should be analysed, and the components that make it superior should be identified. Since interpreters with long professional interpreting experience and subjects with little or no interpreting experience seem to differ in aspects both product and process, it was deemed important to study both aspects of the performance. Social implications of expertise, although important, have not been investigated in this work, simply for delimitation reasons. A pilot study on sociological aspects of expertise in interpreting was however done as part of a research training course (Tiselius 2010). This study is only referred to from a methodological perspective, however, and not as part of the thesis.

3.1.2 Investigating the process

Englund Dimitrova (2005) convincingly demonstrates the prudence of not drawing too many conclusions on the process simply on the basis of the product. In her study she showed that a hypothesis based on the textual evidence was refuted by process data (2005: 36). It is therefore wise to study both process and product. For investigating the process in interpreting in a non-invasive way, however, methods and instruments of data collection are in sore need. Introspection is a way to make both tacit knowledge (Collins 2010: 4) and invisible processes explicit and accessible. When studying interpreting, concurrent introspection is not available – the interpreter cannot verbalize at the same time as s/he is interpreting. Immediate retrospection is therefore one of the few introspective data collection methods at hand for tapping into the process. Retrospection has a few additional challenges compared with introspection. For example, it needs a cue in order to be appropriate, particularly for longer retrospections (in the case of this thesis, the task was between roughly nine and eleven minutes). It also needs to be immediate: since participants can only be expected to completely remember and verbalize a task of 2–10 seconds (Ericsson & Simon 1993: xvi), the longer time that elapses between task and retrospection, the more of the process is likely to be forgotten. The role of the researcher and the instructions given to participants are also important. The researcher is not a participant in the intro- or retrospective interviews. It is therefore key that the instructions are clear and that the researcher is positioned so as not to invite interaction (preferably obliquely behind). After introductory instructions have been given, the participant is told to keep talking. A potential pitfall is that the participant might start to explain and describe the process rather than just verbalize it. Ericsson also mentions objections raised in psychological research against using the subjects' own verbalizations as scientific data (1993: 1). He argues that information processing models of the cognitive process make it possible to create an

explicit and objective encoding process, which in turn enables the data to be examined objectively (1993: 4). Finally, when analysing the coded protocols, one must keep in mind that subjects might forget, recall something different than the actual process or slip into explaining or describing the process. The use of retrospective protocols and their challenges for both translation and interpreting have been explored in two articles by Englund Dimitrova and Tiselius (Englund Dimitrova & Tiselius 2009; Englund Dimitrova & Tiselius, submitted). These articles are not part of the dissertation but have been essential in understanding and developing retrospection as one of the data collection methods in the thesis. Rather than choosing a single component of the process for further study, I chose retrospection despite its potential limitations, since it was considered to be the least invasive and to possibly provide broader insight into the whole process. Another advantage of choosing retrospection was that it allowed comparison with Ivanova's (1999) study. Ivanova used immediate retrospection with a transcript of the source language speech as cue. She sides with Ericsson and advocates protocol studies "as the most suitable of all currently available methods for the study of skilled and expert performance" (1999: 164). Ivanova also stresses the importance of coding the protocol with an open mind, and that coding for open-ended tasks such as simultaneous interpreting can be done on a more global level with for instance strategy use during the task in mind (1999: 165). There are very few studies within the same theoretical paradigm of expertise in interpreting studies. Hence, it is a strength for this PhD project to be able to compare the results of other studies. It should be stressed however that the data in the present PhD project are only comparable with a small part of Ivanova's data. Retesting will nonetheless give reliability and validity to the methodology. The categories with explanatory notes can be found in article 2, appendix 1 and tables A.1, A.2 and A.3.

3.1.3 Investigating the product

As discussed above under section 1.4, the decision to look at the product and not only the process started a lengthy process of choosing and testing the instrument for the study. At first, I believed that the starting point should be to define how high-quality interpreting should be understood. An example of such a definition could be the European Parliament's description of interpreting, cited in Vuorikoski (2004: 19), where it is stated that interpreting is not a word-for-word translation, but a faithful transmission of the source language message, rendered accurately in the target language. But as Vuorikoski noted in her study (2004: 22), the present dissertation also found that there are too many variables in interpreting, and too many types of interpreting, to identify a static, all-inclusive definition of high-quality interpreting. Or in Pöchhacker's (2004: 153) words, "quality appears not as a self-contained topic but as a complex, overarching theme in which all aspects of the interpreter's product and performance – textuality, source-target correspondence, communicative effect, and role performance – play an integral part". Furthermore, as mentioned above, the works of Collados Aís (2011) point out the difficulties of componential assessment that seem to stem from an idea of absolute quality.

Then I discovered Carroll's scales through the work of Gerver (1971) and Anderson (1979; 1994). These holistic scales were developed for machine translation and used by Anderson and Gerver, but were not used in later studies in interpreting studies. Though the authors of these later studies did not seem to be disappointed with the use of the scales, nor dissuade readers from using them, the research community nonetheless seemed to prefer more componential assessment methods. But since the scales seemed appealing from a holistic perspective, I decided, as described in section 1.4 above, to try them out for this project.

3.1.4 Investigating the participants

The long-term participants (see below) in this study had agreed to participate in in-depth interviews about their interpreting background and their views on interpreting. The interviews in this study were modelled on Kaijser and Öhlander (1999). The interview method chosen was quite different from the structured retrospections described above in section 2.1.2. This is a type of interview similar to the one Koskinen (2008) used in her study of translation in the European Union. The interview is structured inasmuch as the parties involved understand that it is an interview, time is set aside for the interview and the researcher has a clear objective for the interview; moreover, an un-structured interview does not use predetermined questions, but rather a mind map of topics (or something similar) that s/he would like to discuss. The discussions flow without constraint, and it is crucial to allow follow-up questions in all relevant directions. The interviewer also actively participates in the interview, and the interviewer's identity, for instance as an expert in another field (researcher) or in the same field (colleague), is important. A drawback of this method is that the subjects do not necessarily answer the same questions. Article 4 discusses the methods and results of these in-depth interviews.

3.2 Participants

All the participants in the different studies of this project had Swedish as their mother tongue. All groups of interpreter-subjects with professional interpreting experience (n=9) had English as a passive working language (C-language in AIIC terminology, that is, a language the interpreter has full understanding of and works from but not into). The subjects in interpreter training (n=3) had English as one of their passive working languages in the training programme. The no-experience subjects (n=3) had English as a strong foreign language; they were not screened,

however, and merely self-rated their proficiency. The professional interpreter participants (n=9) constitute an undeniably small group, but it should be stressed that the entire population of English-Swedish conference interpreters is also very small. AIIC lists 34 interpreters with that combination; even if that figure is boldly doubled (given that not all conference interpreters are members of AIIC), it still makes a population of N=68. Using that estimate it would mean that the participants make up approximately 13 % of the entire population; in that perspective, the group may be small, but would still be representative. Furthermore, the fact that the data contain both a cross-sectional and a long-term group make both inter-individual and intra-individual comparisons possible.

3.2.1 The cross-sectional data – data set A

The cross-sectional data include nine participants, divided into three groups according to the interpreting experience of the participant (see table 1, also reproduced in Article 2). None of the participants received any economical remuneration.

Table 1. Age and experience of the cross-sectional interpreters

Group	Age span	Years at university	Int. training diploma	Years of int. experience
No experience	20–29	4	No	0
Short experience	30–49	4	Yes	2
Long experience	50–60	4	Yes	25+

The no-experience group (from now on the NE group, cf. figure 1 above) consists of students recruited at the Institute for Interpretation and Translation Studies at Stockholm University. The NE group comprises three females. They were recruited from the first semester introductory course to translation and interpreting: students there were sent an e-mail asking them to participate in a study; several students volunteered, and those with Swedish as their mother tongue and who claimed to have a high proficiency of English were chosen. Although they had no prior experience of interpreting, the NE group had been in an interpreting booth

during their introductory course and were thus familiar with the concept of interpreting. The NE group did the interpreting task in the training booths at the Institute for Interpretation and Translation Studies at Stockholm University.

The short interpreting experience group (from now on the SE group, cf. figure 1 above) consists of interpreters who had graduated from an interpreting programme two years previous to the data collection event. The SE group comprises one male and two females. They were all accredited to the European institutions, but were not yet AIIC members.¹ AIIC membership is subject to application and screening, and it is fair to assume that the subjects were not AIIC members because of their fairly limited professional experience (they would be eligible for AIIC membership first after 150 working days). Since the SE group consisted of recent graduates, they had not had any experience with teaching or assessing interpreting, though they did have experience with self and peer assessment during the interpreting programme. Their recruitment to the study was based on the number of years of experience after graduation, but they also constituted a convenience sample in the sense that they were directly approached through e-mail and asked if they were willing to participate in a study. The SE group performed their task at the researcher's workplace and not in a booth.

Finally, the long interpreting experience group (from now on the LE 25 group, cf. figure 1 above) consists of highly experienced interpreters. It comprises one male and two females. They had all the superficial signs of expertise. The LE 25 interpreters all had a diploma from an interpreting programme. They had been working actively for at least twenty-five years, with an average of at least 100 days per year. They all had both teaching

¹ AIIC membership is gained with at least 150 days of working experience and the signature of three AIIC members guaranteeing the the applicant's quality and work ethics; see <http://aiic.net/node/2395/joining-aiic/lang/1> (accessed 12 April 2013).

and assessment experience. They were accredited freelance interpreters at the European institutions and were AIIC members. All interpreters of the LE 25 group were furthermore considered “good interpreters” in the interpreting community. As was also shown under 2.1.1, the LE 25 group could be defined as having contributory expertise (Collins 2007: 14). It was a convenience sample, in that I asked fellow interpreters with the right profile (i.e. accreditation to EU institutions, AIIC membership, experience as interpreting teachers and examiners, long professional interpreting experience, and, finally, a reputation for being “good interpreters”) whether they would like to participate in a study. The LE 25 group participated in the experiment in the booth at their workplace outside of working hours.

3.2.2 The long-term data – data set B

The participants in the long-term group were recruited the first time when studying at an interpreting programme in the mid-1990s. Williams (1995) was designing a major study on processes in simultaneous interpreting that aimed to study factors such as anomalous stress, prosody and pitch in interpreting. For the purpose of the project, she recruited both interpreting students and professional interpreters. The students recruited in her study had several different combinations of working languages. They were recorded interpreting in both simultaneous and consecutive mode and when talking freely (both in their mother tongue and their foreign languages). For this dissertation, the tapes from Williams’ project were generously made available from the Centre for Research on Bilingualism in Stockholm. The tapes were studied and four possible subjects were identified. The subjects were identified on the following criteria: (a) having Swedish as their mother tongue; (b) having English as a C-working language; (c) remaining active interpreters; and, (d) most importantly, willingness to participate in a new study. Moreover, the interpreting on the

tapes needed to be from the same source speech in order to make comparisons between the subjects possible.

Of the four possible participants, three were available at the time of the new recordings. They were all staff interpreters at the European institutions and had been so for some fifteen odd years. They had experience both as interpreter trainers and as assessors, and they were all considered “good interpreters” by their peers. One was an AIIC member. At the time of the new recordings, this group had also reached levels of contributory expertise (Collins & Evans 2007). All three were female, and they kindly agreed to participate in new recordings and in-depth interviews. As described in figure 1 above, the subjects in data set B will be referred to as subjects in training (the IT group) when their student data are referred to and as long experience interpreters (the LE 15 group) when their professional data are referred to.

3.2.3 The raters

Though the raters are not the main participants per se, they are nevertheless important to the studies. The raters were interpreters and non-interpreters who rated the quality of the interpretations using the holistic scales. They all had Swedish as their mother tongue. There were several different groups of raters, two for data set A and two for data set B. For data set A (cross-sectional), the raters were (a) university students without previous specific knowledge of interpreting (n=6), and (b) interpreters with professional interpreting experience and experience with examination and peer assessment (n=6). They are described in more detail in article 1 of this dissertation. For data set B (long-term), the raters consisted of (a) university students without previous specific knowledge of interpreting (n=12); (b) interpreters with professional interpreting experience and experience from examination and peer assessment (n=12); and (c) another set of university students without previous specific knowledge of interpreting (n=9). Groups (a) and (b) rated the NATO speech (see section

3.3.2 below) and group (c) rated the EU speech (see section 3.3.1 below). The raters for data set B, and the rationale for the different raters of the different speeches, are described in more detail in article 3 below.

3.3 Interpreting Data

This section presents the two speeches that were used to elicit interpreting data from the subjects. All subjects interpreted the EU speech, and the long-term group also interpreted the NATO speech twice with fifteen years in between.

3.3.1 The EU speech

All subjects (n=12) interpreted a speech from the European Parliament, given originally by Commissioner Byrne in 2001. It was a fairly general speech, but very fast (141 words per minute on average) and also pronounced with a heavy Irish accent. For this project, the speech was transcribed and tweaked to add some additional difficulties (names and figures). The speech was then re-recorded by a native English speaker with Received Pronunciation. In its re-recorded version, it was also controlled for speed (119 words per minute on average). The EU speech can be found in article 2, appendix 2.

3.3.2 The NATO speech

The three long-term participants also interpreted a NATO speech that was used as teaching material during their training. The speech (and the interpreting from the IT group) was chosen for the following reasons: (a) the point in time in the training programme (the chosen speech was delivered when the students had been practising in the simultaneous mode for a couple of months, and the speech would then supposedly also present some difficulties for the LE 15 group); (b) it was not an exam-level speech for the students (as exam-level speeches have their own genre); (c) relative

difficulty (speed, terminology, themes; as said above, it was important that the speech should present some challenges for the LE 15 group and yet not be too difficult for the IT group); and (d) sound quality of both speech and interpretations (these were old tapes, some of them recorded on small recording devices, and it was important to hear clearly, both for interpreting and transcription). The NATO speech can be found in article 3, appendix 1.

3.4 Retrospective data and assessment files

As described above in section 3.1.1, all participants, except the IT group, performed retrospection immediately after the interpreting task. The retrospection was cued with a transcription of the source speech and then recorded and transcribed. The transcripts served as protocols in the categorization process following Ivanova's categories (1999). The categories are found in article 2, appendix 1.

The interpretations in both the cross-sectional and the long-term data were transcribed and transformed into assessment files. The transformation consisted of dividing the interpretations into smaller units according to idea and then mixing them randomly. This is described in detail in article 1, and an example of the assessment files can be found in article 1, appendix 2.

3.5 Methodological discussion

This section will reflect on some of the methodological issues and challenges that have not been touched upon earlier in this account, such as the use of mixed-method design and the choice to pursue the investigation despite a very small material in the long-term study.

3.5.1 Mixed-method design

Just as in other disciplines, researchers have highlighted the benefits of triangulation in interpreting studies too (see for instance Gile 2005a or Hild 2007). In trigonometry and geometry, triangulation means finding the unknown third point by using two known points; in social studies, triangulation is the use of at least three (but preferably more) different studies, theoretical perspectives, investigators and data sets to examine a certain topic (see for instance Denzin 2006 or Scott and Marshall 2009). Presumably, researchers would obtain more robust results by using a variety of means, such as different researchers or data sets, to investigate a certain concept or construct in interpreting. Not many studies use triangulation in interpreting studies, however, and the few researchers who do in fact triangulate use mostly the within-method, that is, they triangulate with different varieties of similar methods (Denzin 2006: 472).

Quantitative method designs dominate in conference interpreting research. And many research objects, such as cognitive load, working memory or the effects of interpreting under pressure, are easily and more appropriately researched quantitatively. But there are other topics, such as the perception of role or identity, that are not so straightforwardly refuted or supported by a yes/no hypothesis, traditionally used in studies with a quantitative approach. Diriker (2004) and Monacelli (2009) are among the few who use qualitative methods to investigate conference interpreting. Monacelli investigated voice (not the physical voice in this case, but the voice as a representation of the interpreter's persona) in simultaneous interpreting from a constructivist epistemology in order to study the speaker's discourse and the interpreter's rendering of that discourse. Diriker used critical discourse analysis and semi-structured interviews in her study, which aimed at distinguishing between different discourses on simultaneous interpreting and how the interpreters put that discourse into practice as they interpret.

Though both quantitative and qualitative research methods are in use in Interpreting Studies, mixed-method designs do not seem to be widely employed. Such designs can for example be used to enrich the understanding of quantitative results by providing certain insights into the subjects in question: for example, a questionnaire with multiple choice or Likert Scale answers can be illustrated by quotes from the open-ended questions at the end of the questionnaire (Patton 2002: 5). In short, mixed methods help the researcher to approach an object of study from different angles. When studying a particular area, such as expertise in interpreting (as will be described below), mixed methods may be used to enlighten parts of the issues raised by the expertise approach that cannot be reached through an experimental design

The reason in this particular case for combining qualitative and quantitative methods lay in the struggle to find the best way to analyse an existing and unique material, namely that of the late Sarah Williams (1995). Her article from 1995 gives an overview of her intentions with the material. This material consisted of recordings of interpreting students in the 1990s, as described above in section 2.2.2. By deciding to use that material, it was also necessary to find out what made sense to the material, to paraphrase Quinn Patton (2002: 72). It also meant that instead of starting to work within a specific theory or with a clear hypothesis or research question in mind, the project began with a material, and research questions and hypotheses needed to be adapted accordingly. In addition, William's material was recorded for other aims than this project. When the interpreters who were recorded as students had been identified and agreed to participate in new recordings, the challenge was to design a study that would yield interesting results from this unique data. Since the starting point was the possibility of obtaining a material that could pave the way for a long-term study, the theory chosen was the expertise approach. Furthermore, a mixed-method design was developed in order to analyse both interviews and experimentally yielded material. In order to obtain a

broader perspective, the long-term material was supplemented (as described above) with recordings of the same subjects fifteen years later and with the cross-sectional material that featured nine subjects. It is hoped that the mixed-method design in this thesis, although presented in different articles, will contribute to a fuller picture of the development of experience and maybe also expertise.

3.5.2 Re-test or not, and other challenges

As discussed above in section 3.1.2, an advantage of using the same methodology as part of Ivanova's (1999) study was to be able to relate to her results. It would also provide stronger support for the results here. There are two challenges to this claim though. First, none of the studies reported here is an exact replication, since Ivanova's two groups and the three plus two groups in this thesis are not identical; and although the source language is English in both cases, the mother tongue differs (Bulgarian in Ivanova's case and Swedish in this case). Second, Ivanova's novices are interpreting students in their second and final year at the interpreting programme. If the different groups are labelled according to Hoffman's terminology (cf. section 2.3.3 p. 31), the novices in Ivanova's study are apprentices, whereas the novices in this dissertation can be divided in three groups: novice bilinguals (NE group), apprentices (IT group) and journeymen (SE group). According to a strict scientific definition, where replication is to reproduce an experiment with the exact same conditions and with the aim to obtain the same results (Scott & Marshall 2009: 646), this is not a replication. It will however qualify as a re-test, or a re-study (Scott & Marshall 2009: 647).

The differences between the different groups pointed out above could also be of importance when the groups are compared using the Carroll scales.

An important drawback was the fact that one retrospective interview from the long-term material was lost because of a technical mishap. Since the

long-term material was scarce to begin with, the loss of that file was clearly a blow. I decided to continue analysing the material and carry on with the study, as the material was truly unique, and deserved every effort to be shared and analysed.

Despite the challenges and drawbacks of the material, it is my hope that the results will still be considered solid and interesting.

3.5.3 Terminological inconsistencies

Finally, it should be mentioned that because this is an accumulative thesis with some articles published before the project's end, the terminology will vary slightly in the various articles. In particular, the subjects evaluating the interpreting will be variously labelled *graders* and *raters*. The terms *subject* and *participant* will be used interchangeably throughout the dissertation to refer to the members of the various groups. In social sciences, "participant" is often preferred over "subject" as researchers stress the active involvement of the individuals participating in a study. For the in-depth interviews, "participant" was thus more natural.

4. Summary of articles and general results

This section summarizes the results of the different parts of the study reported in four different articles. The material was collected as one whole unit that was studied and analysed from different perspectives, each reported in a different article. Article 1 is methodological, article 2 reports the results from the cross-sectional data and article 3 reports the results from the long-term study; finally, article 4 is a report of the in-depth interviews of the long-term interpreters. The mind map in figure 2 illustrates the different data sets, analysis methods and articles. The focus of this summary is mainly on results and somewhat on methodology, but as issues have already been discussed in section 3, the methodological sections have been kept at a minimum.

	Eliciting material	Primary subjects	Raters	Retro-spection	Rating
Article 1	EU-speech	Cross-sectional NE, SE, LE25	Interpreters Non-interpreters	No	Yes
Article 2	EU-speech	Cross-sectional NE, SE, LE25	Interpreters Non-interpreters	Yes	Yes
Article 3	EU-speech Nato-speech	Long-term IT, LE15 Cross-sectional NE, SE, LE25	Interpreters Non-interpreters	Yes NE, SE, LE 15, LE 25	Yes
Article 4	Interview mind map	LE15	No	No	No

Figure 2. Summary of participants, data, analyses and articles.

4.1 Article 1: “Revisiting Carroll’s scales” (data set A)

Article 1 was the first article published in this PhD project. The article used data set A, that is, material from the cross-sectional group of interpreters.

4.1.1 Background

The study tests the scales instrument used to assess the interpretations in the PhD project. Focusing mainly on methodology, the study provides an overview of Carroll's scales, the notion of quality in interpreting and the rationale for using Likert-type, holistic scales to rate interpreting. Since the assessment of interpreters' end product was central to the whole project, Carroll's scales were chosen to be adapted and evaluated further since they appeared to be valid and easy to use.

The purpose of the study was first to investigate whether holistic scales give valid results when assessing simultaneous conference interpreting products, and second whether there was any difference between laypeople and professional interpreters when rating with holistic scales.

Carroll's research areas were assessment tools for language testing (Stansfield & Reed 2004) and machine translation. He developed two scales for evaluating machine-translated texts (Carroll 1966). Carroll favoured an integrative testing design. He established the need for two evaluative scales based on the two constructs of *intelligibility* and *informativeness*, as he claimed that a translation could be perfectly intelligible but lack fidelity to the original, while another text could be completely unintelligible and yet be completely faithful to the original (Carroll 1966: 57).

Gerver (1971) and Anderson (1979) used Carroll's scales to assess interpreting. Gerver did not provide any critical analysis of the application of the scales, but Anderson questioned whether the scales as instruments were sufficiently fine-tuned for measuring the output of interpreting.

Carroll's scales can account for central aspects of the interpreted event, but not for its entirety as a communicative event.

4.1.2 Data and method

In the study, the scales were adapted to interpreting before being tested. They were adapted to spoken language and Swedish, and highly similar steps were merged so as to reduce the number of steps from nine to six. The adapted scales and are found in tables 1–4 in article 1.

The nine interpretations of the cross-sectional material were used (see section 2.3.1 above) as eliciting material and transformed into assessment files. The three groups were labelled “long-experience (LE) interpreters”, “short-experience (SE) interpreters” and “no-experience (NE) interpreters”, respectively. Their interpretations were carefully transcribed and then transformed into a written text, adding punctuation according to intonation. The text version of each rendition was then divided into 18 interpreting units according to Lederer’s (1978: 330) units of meaning or translation units (Gile 2009: 101). Two criteria were taken into consideration: intonation and idea. The assessment files were built up from the divided interpretations. Each rating file comprised excerpts from all interpreters, randomly mixed.

The raters were six university students (non-interpreter raters), and six professional interpreters (interpreter raters) who had both trained and evaluated interpreters. The raters were instructed at the beginning of each rating session. After rating, the protocols were checked for significant difference and inter-rater reliability.

4.1.3 Major findings

The inter-rater reliability was acceptable for all groups and raters, although slightly higher for interpreters. For intelligibility, as rated by both non-interpreter and interpreter raters, a t-test showed significant differences for all groups except between the SE and LE 25 groups. Informativeness, as rated by both non-interpreter and interpreter raters, showed significant

difference between all groups. Rating scores and *p*-values can be found in tables 6–11 of article 1.

4.1.4 Discussion

The two questions posed in the study (i.e. whether holistic scales give valid results when assessing simultaneous conference interpreting products, and whether there was any difference between laypeople and professional interpreters when rating with holistic scales) were both answered in the affirmative. Since three groups of interpreting subjects with clearly different profiles were tested, it could be expected that both their intelligibility and informativeness ratings would significantly differ. Furthermore, in all but one case both rater groups found significant differences between the interpretations; the case where there was no significant difference was intelligibility (defined as understandable, spoken Swedish) between the SE and LE 25 interpreters' product. This near-equal intelligibility is perhaps not surprising: since the SE interpreters had graduated from an interpreting programme, passed a freelance test for the EU institutions and worked for two years, they had thereby most likely gotten experience in delivering understandable, spoken Swedish.

There are some limitations to this study of the scales' applicability, though. The sample was small, both in terms of the raters and the interpretations rated, and it was limited to English-Swedish, which may in a wider perspective be a potential limitation (though not for this dissertation project). Despite these limitations, the study supported the choice of holistic scales as a grading instrument for the rest of the PhD project, something that also opens up for testing them in a broader application. The first step of such a broader application could be to test them in a live interpreting context, for example an entrance test or an exam.

4.2 Article 2: “Process and product in simultaneous interpreting: What they tell us about experience and expertise” (data set A)

Article 2 was co-authored with Gard B. Jensen. I was responsible for collecting the data and choosing the instrument for analysis, and was the main writer of sections 1, 2, 3.1 (except second part of 3.1.4), 4.1 and 5. Jensen was responsible for choosing the statistical analysis instrument and for the statistical analysis, and he was the main writer of sections 3.1.4 (second part), 3.2 and 4.2. Both authors participated in the development and editing of the whole article.

4.2.1 Background

The results from the analysis of the cross-sectional material are presented in this article, with the expertise approach as the theoretical background. The study of the cross-sectional material was divided into two parts, one studying the process through retrospection and the other studying the product through assessment. The overall aim of the article was to answer the following question: Can performance differences be established between three groups of interpreters with different levels of experience (NE, SE and LE 25)? The first part investigated whether the three groups differed in their interpreting process, as manifested through reported processing problems, instances of monitoring and strategies (see appendix 1 of article 2). The second part examined whether the three groups differed in their interpreting product, as assessed by raters using Carroll scales; this part is an enhanced analysis of the material in article 1. Ericsson and Smith (1991: 15) relate quality to the investigation of expertise by stating that “although judges can reliably assess the superior quality of the product, it is difficult to analyse such products in order to identify the measurable aspects capturing the superior quality of the product.” It cannot be stressed enough the importance of combining the assessment of quality with the

investigation of the cognitive processes underlying the production of that quality.

4.2.2 Data and method

Both parts of article 2 analysed the cross-sectional material in data set A (described above in section 3.2.1). The first part is based on retrospective protocols coded according to Ivanova (1999), and the occurrences of the different categories were further analysed by using correspondence analysis (CA). Correspondence analysis is a type of multivariate statistical analysis where one variable (experience in this case) has a more or less explanatory value, and different responses can be studied from this variable (in this case how much e.g. omission is related to experience). The interpreters in this experiment interpreted an EU speech, and immediately afterwards they performed a retrospection from a transcription of the source speech as cue. The interpreting and the retrospection were then transcribed for the analysis.

In the second part of article 2, the ratings made in article 1 were analysed further in order to support the results more solidly, as article 1 had focused more on methodology than actual results. This time the ratings were run through two Friedman rank sum tests as well as a Nemenyi-Damico-Wolfe-Dunn post-hoc test for pairwise comparisons.² A final test was performed to establish whether ratings differences between the two rater groups (non-interpreters and interpreters) could be a distorting factor.

4.2.3 Major findings

The analysis in part one showed, first, that the main difference in processing problems lay between NE interpreters on the one hand and SE and LE 25 interpreters on the other. When the different target language

² The R code for this test was provided by T. Galili from <http://www.r-statistics.com/2010/02/post-hoc-analysis-for-friedmans-test-r-code/>.

processing problems were studied more in detail, it was clear that the NE group struggled more with comprehension and simultaneity issues, while the SE and LE 25 groups were typified by problems such as perception or finding an equivalent. The LE 25 group was positively associated with difficulties stemming from source language input rate and syntactic processing, while problems finding linguistic equivalents were positively associated with the SE group. The processing problems encountered by the subjects in this study were compared with Ivanova's (1999) two groups using Spearman's rank test, and no significant difference was found between the groups in this experiment and Ivanova's.

Second, in the case of monitoring, a difference was found between the NE and SE groups on the one hand and the LE 25 group on the other. The LE 25 group was positively associated with controlling the accuracy of the translation before utterance. There is a difference between the NE and SE groups in time management issues and internal comments on the speaker, both of which are positively associated with SE interpreters. A qualitative analysis of the monitoring instances of translation showed that despite the similarities in raw data for LE 25 and NE interpreters, the LE 25 interpreters reflect on better ways to interpret a certain utterance, whereas the NE interpreters try to find a general coherence in the output. When compared with Ivanova's groups (1999), the raw figures suggested differences between the experienced interpreters, but the Spearman rank test did not confirm those differences. The novice groups showed no significant difference in the Spearman rank test.

Third, in the case of strategies, the greatest difference lay between the NE and LE 25 groups. Considering the task was simultaneous interpreting, deletion was unsurprisingly the most common strategy for all groups. Overgeneralization was strongly associated with LE 25 interpreters, while creative interpreting was strongly associated with NE interpreters. The strategy figures could not be compared with Ivanova's groups, as the analysis in the two studies differed here: Ivanova counted strategies related

to processing problems only, whereas this study counted all strategies. A separate count in order to compare the two studies could of course have been done, but since participants in this study often did not connect their reported strategy to a particular processing problem, such a comparison seemed deficient.

In the second part of article 2, the statistical analysis showed that the processes differed significantly for all three levels of interpreting experience. Moreover, the second round of statistical testing confirmed the results from article 1, there were significant differences for all the groups except for intelligibility between SE and LE 25 interpreters. Finally, it was also found that there were no significant differences between the two groups of raters.

4.2.4 Discussion

The data in article 2 are too small to draw any major conclusions on the differences between the NE, SE and LE 25 groups. Significantly, however, the results support Ivanova's (1999) findings where the groups and data collection are comparable. The results also support other researchers' results: For instance, Vik-Tuovinen's (2006) conclusion that beginners focus on source text and linguistic expression is supported by the NE subjects, who struggle with problems of lexical access (a typical source text and linguistic problem). For the LE 25 group, the monitoring category translation was prevalent, a fact that may support Vik-Tuovinen's (2006) findings that experienced interpreters focus more on situational factors than less experienced interpreters. The results also agree with Liu's (2001) result that experienced interpreters monitor output better. The findings of this study show that LE 25 interpreters make more use of monitoring strategies, in particular to check the appropriateness of the utterance and reflect on the speech or the speaker. Certain processing problems, instances of monitoring and strategies may indeed tell us something about expertise, and they may indicate which components of the performance

should be studied in detail to find the superiority Ericsson and Smith (1991) encourage us to look for. The present study suggests that these components might be the ability to monitor and the skill to deliver interpretations with very little information loss compared with the original.

It was significant for this dissertation that the three groups of interpreters showed measurable differences, and that two reliable instruments and a consistent control group were available for investigating the long-term subjects in articles 3 and 4.

4.3 Article 3 – “The development of expertise – or not: Three simultaneous interpreters’ development over time” (data set B)

4.3.1 Background

Article 3 reports the result of the investigation of the quantitative process and product data from the three long-term interpreters. The aim of this study is to investigate whether and how interpreting performance improves over time. The starting hypothesis is that there will be improvements over time in the interpreters’ performance, and that their performance the second time around will correspond to that of experienced peers.

As discussed above in section 3, it is important to study both process and product when studying an interpreter’s development. The results in article 2 showed that there were many differences in terms of both process and product between the NE interpreters and SE and LE 25 interpreters. Following Englund Dimitrova (2010), the process is defined as the cognitive activity of producing a target speech in one language from a source speech in another language. The product is defined as the target speech.

The study compared long-term interpretations from both a process perspective, using retrospection, and a product perspective, using holistic scales.

4.3.2 Material and method

The participants in this study are the three interpreters featured in the long-term data (data set B), and are described in depth above in section 3.2.2. They are variously labelled the IT group (when recorded at the interpreting programme) and the LE 15 group (when recorded fifteen years later). In addition, the cross-sectional participants, described in section 3.2.1, were used as a control group.

The LE 15 interpreters interpreted two speeches: the EU speech (see section 3.3.1) and the NATO speech (see section 3.3.2). The NATO speech was interpreted on two occasions, the first time during the interpreting programme and the second time fifteen years later. After this latter interpreting, the subjects carried out retrospections of their interpreting. The retrospection was cued with a transcript of the original speech, with normalized orthography and punctuation. Unfortunately, one of the retrospection files was lost due to a technical mishap. The five remaining retrospections were analysed by the author together with a research colleague and coded for processing problems, instances of monitoring and strategy use. The interpretations of the EU speech were divided into smaller units and randomly mixed into six rating files. The interpretations of the NATO speech by the LE 15 group were mixed into smaller units together with the interpretations of the NATO speech by the IT group, and assembled randomly into six rating files with examples from all three interpreters both as an IT (student) and as a LE 15 (professional and experienced). The NATO files were assessed by both interpreter raters and non-interpreter raters, whereas the EU files were only assessed by non-interpreter raters. The raters used the holistic scales tested and adapted in article 1.

4.3.3 Major findings

The main reason for the NATO speech being assessed by both interpreter raters and non-interpreter raters was that the inter-rater variability was very high for the interpreter raters, so it was decided that the assessment had to be redone. Surprisingly enough, the non-interpreter raters had a very low inter-rater variability, but the means of the ratings were more or less the same for the two groups.

The surprising finding of this study was that there were only small differences in the ratings between the interpretations produced by the IT and LE 15 groups. In some cases the scores were even worse for the LE 15 interpretations than for the IT interpretations. The results did not therefore support the assumption that these interpreters had continued to develop interpreting experience that would make them perform better.

When it comes to the EU speech and the comparison with the cross-sectional data, the scores for the LE 15 interpretations were comparable with the SE interpretations and hence worse than the LE 25 interpretations. Furthermore, the scores of the LE 15 NATO and EU speech interpretations are remarkably similar; as the long-term subjects thus received similar scores for their three interpretations (IT NATO, LE 15 NATO and LE 15 EU), they did not evince any long-term development.

In terms of the analysis of the process, it is hard to draw any firm conclusions, as one of the retrospections was lost. But the little processing data that remained confirmed the findings of article 2 and Ivanova (1999), namely that experienced interpreters encounter fewer processing problems and have more strategies at hand to solve the ones they do encounter. No correspondence analysis was done in this analysis, as data from only two interpreters seemed too meagre.

4.3.4 Discussion

The results of the study are challenging. Do they suggest that the three experienced interpreters are not experts according to Ericsson and Smith's (1991) characteristics? Do they indicate that the performance level reached at the end of the interpreting programme was some type of final stage? Or were perhaps the instruments used to measure performance too blunt?

The quantitative results from the cross-sectional material (discussed above in section 4.2.4) go against the assumption that the performance did not improve from the end of the interpreting programme. In fact, the results in article 2 strongly suggest improvement, at least inter-individually. The LE 15 interpreters have many superficial signs of expertise that support the assumption that they are experts, as explained above in section 3.2.2 (they work at EU institutions, are members of AIIC, are labelled "good interpreters" by their colleagues, and have long experience). The results from the study that tested the scales (4.1.1) and the cross-sectional study (4.2.2) also undermine the idea that the instruments are too blunt, as the instruments showed a clear difference between the groups in the former studies. Another fact that could be in play here is that the same individuals were tested on different occasions; perhaps intra- and inter-individual variation differ. Another factor that may change the results is interpreting style. In her study of translation expertise, Azbel-Schmidt (2005) found that style seemed to be established early on in an interpreters's career. An interpreting style that favours message compression and thereby word deletion is not necessarily a low-quality interpreting style, but may result in lower scores in an assessment where differences on the word level may affect the overall judgment. It is thus possible that the interpreters developed their style early on, and that this style had a negative impact on the ratings.

Alternatively, the results may be due to a flaw in the design. The EU speeches were only interpreted by the experienced interpreters, perhaps

prompting the raters to assess them more strictly. The raters presumably contrast the different assessment units against each other, and if an assessment file conversely contains interpretations from both inexperienced and experienced subjects, then the latter will “look better”.

Whatever the reasons for the results of this study, they are interesting and unexpected and put the whole project in a different light.

4.4 Article 4 – “Expertise without deliberate practice? The case of simultaneous interpreters” (data set B)

The study reported in article 4 had a different methodological approach than articles 1–3, and carried out in-depth interviews with the LE 15 interpreters of data set B.

4.4.1 Background

Article 4 reports on the qualitative study of the PhD project. The method of in-depth interviews was chosen to cover aspects of deliberate practice, as has been described above in section 2.3.2. To the best of my knowledge, this is the only in-depth interview study on deliberate practice in interpreting, and it may well be the only study so far on interpreters’ engagement in deliberate practice in interpreting practice.

A subject’s engagement in deliberate practice could conceivably be observed in an experiment, but then only for short tasks and not for the extended periods of time that typically characterize deliberate practice. Moreover, the individual’s ability to set clear goals and be open to feedback are important features of expertise. The performer must be able to specify intentions, results or outcomes. Research in goal-setting has shown that practitioners perform better when they specify detailed goals or break a goal down into different sub-objectives (Zimmerman 2007).

A study of deliberate practice will need to discuss the interpreting skill and its different sub-skills, such as skills in linguistics, concentration, analysis and listening, speaking and reading. There is no exhaustive list of skills needed for interpreting, and none that is both empirically tested and generally agreed upon (see e.g. Jones 1998, New Jersey Courts 2007 and Corsellis 2008). Research by McNamara et al. (2011) and Napier and Bontempo (2011) in sign-language interpreting have singled out conscientiousness, emotional stability, self-esteem and openness to experience as factors that predict interpreting success. The openness to experience would be something like intellectual curiosity and thus not far from one of the components of deliberate practice. The aim of the study in article 4 was to explore whether the three LE 15 participants engaged in deliberate practice, set clear goals and received feedback, and whether and how they applied such experiences in their professional life.

4.4.2 Data and method

The participants were the three LE 15 interpreters in the long-term material, described above in section 3.2.2. It should be stressed that these participants were more than happy to participate in new recordings and in in-depth interviews; indeed, their very willingness may reveal something about their (unconscious) view of deliberate practice and learning through introspection.

As the study required a different methodology than the other studies, some time had to be spent on developing this method. In-depth interviews as a tool are described both in section 3.1.4 above and in section 2.2 of article 4. A mind map was also developed to serve as the basis of the interviews. In the above mentioned pilot study on sociological factors in expertise (Tiselius 2010), the participants clearly did not understand the concept of deliberate practice and considered it unimportant or unclear; however, the participants did in fact mention examples that the research leader classified as deliberate practice, goal-setting or openness to feedback. A list of

topics, presented in section 2.2.1 and figure 1 of article 2, was therefore developed that could be used as triggers for the core topics. The interviews were then conducted around these triggers. The interviews, lasting from an hour to ninety minutes, were carried out at the participants' workplace immediately after the interpreting task reported in article 3. The interviews were recorded, transcribed and then analysed, with ATLAS.ti software used to examine core concepts. The protocols were coded following how the interpreters described learning and practising interpreting skills and sub-skills. The interviews were also reread together with a research colleague from another domain with experience in narrative analysis in order to look for topics that perhaps had been overlooked in the initial analysis.

4.4.3 Major findings

The three interpreters who participated in this study were all focused language learners, although two of them did not focus on language learning until a relatively late stage (late teens, early twenties). None of them was born or grew up bilingually, but they were all highly dedicated once they started focusing on language learning. All three also talked about how they constantly broadened their general knowledge by listening to the news and reading newspapers, books and so forth, in addition to the usual meeting preparations. Moreover, they all stressed the importance of teamwork and of listening to one another, both to help out and to learn. Another issue that stands out is their general ability to focus, with the interpreters talking about their skill to concentrate and to be present in the situation. When it comes to interpreting skill, if by that we mean the ability to transfer a message from one language to the other, none of these interpreters deliberately practised that particular skill. They also considered the interpreting skill to be more or less innate.

When it comes to the notion of deliberate practice, there are a multitude of examples in the interviews of the three interpreters practising several sub-

skills (though not the main skill). However, none of the examples seems to be “deliberate” in the sense defined by Ericsson et al. (1993), that is, that time is set aside with defined exercises and clear goals to refine the main skill. But the three interpreters talk about how they endeavour to improve their interpreting and how they absolutely do not want to work on autopilot, which could suggest that they are intuitively counteracting stagnation in an automatized mode (Ericsson 2007: 685). They also feel a sense of elation when performing well, which could be seen as a type of monitoring. Mood (cf. article 2) is a type of monitoring, where the interpreter reacts in positive or negative terms to his or her own interpreting. Many instances of mood in article 2 pertained to how satisfied the participant was with a certain solution or interpreting. In view of the results of the in-depth interview, there seems to be a connection between monitoring and the type of self-evaluation the participants talk about. Another point is that the goals they spoke of were task goals (i.e. goals for what they want to achieve when performing) and not training goals (i.e. goals for improving a certain skill). The feedback these interpreters talk about is not direct feedback from colleagues, but rather from evaluating themselves.

4.4.4 Discussion

The findings are interesting from several perspectives. Ericsson et al. (2007: 685) note the importance of social networks and support for deliberate practice and the development of expertise. The three LE 15 interpreters’ responses suggest a dearth of support and encouragement in their environment, where there are no competitions, rankings, coaches or performance-based salary increases. In order to develop, it is probably necessary for interpreters in such an environment (one that is typical for interpreters) to use their own strategies for improving their interpreting skill as they themselves understand this skill. The question then is whether such strategies – that is, such practising of sub-skills – can qualify as

deliberate practice. Or perhaps if the time for deliberate practice of the main skill, in the case of highly experienced simultaneous interpreters is merged with the time for work. If none of these explanations is valid, is it possible to be an expert without engaging in deliberate practice, or is it the very concept of “no expertise without deliberate practice” that is inapplicable to interpreting studies? The performers’ deliberate practice is a basic tenet of the expertise theory (Ericsson 2007b), which stipulates that practice should be deliberate and isolated from work (Ericsson, Krampe & Tesch-Römer 1993). One reason why the three participants of this study do not seem to engage in deliberate practice may quite simply be that they are not experts. Yet these interpreters seem very much engaged in practice and strive to improve themselves. But their practising is not done in isolation from work, and much of it seems more intuitive than deliberate. Assuming that the three interpreters are indeed experts, this indicates that expertise in interpreting is possible without deliberate practice.

For the PhD project as a whole, the findings in this study provide an interesting perspective on the three long-term interpreters. Articles 1–3 studied expertise from a quantitative perspective and did not take deliberate practice into account, but focused solely on performance and process during the task. The in-depth interview study provides a fuller picture of possible interpreting expertise and whether highly experienced interpreters engage in deliberate practice exercises.

On a pedagogical note, an interesting implication may be to introduce the notion of practice and skill development during an interpreter’s whole career, as well as the type of continued education that Bontempo and Napier also suggest for sign-language interpreting (2007: 295–296).

4.5 Summary of the results from all four studies

4.5.1 Methodological results

Holistic, Likert-type scales based on Carroll's scales (1966) provided reliable and valid results when tested on interpreters with different levels of experience. The results of the participants' retrospective recollection of the task and the researcher's subsequent protocol analysis were also corroborated by Ivanova's (1999) results. Finally, a mind map was developed as an interview guide for in-depth interviews on deliberate practice.

4.5.2 Research results

The results on the interpreting process showed that experienced interpreters (LE 15 and LE 25) encounter fewer processing problems and have more strategies at hand when they encounter problems than interpreters with short (SE) or no (NE) professional experience. This was true both for interpreters in the cross-sectional material (data set A) and in the long-term material (data set B). Furthermore, experience is decisive when it comes to monitoring: in the cross-sectional material (data set A), the most experienced interpreters (LE 25) were more associated with monitoring than the other two groups (SE and NE).

The results regarding the assessment of the interpretations are clear when it comes to the cross-sectional material (data set A). There is a statistically significant difference concerning product between the NE group and the SE and LE 25 groups. There is also a statistically significant difference in the transferred information (i.e. the content) between SE and LE 25 interpreters. However, the results from the long-term material (data set B) showed no difference between the product of the IT and LE 15 groups (that is, the same interpreters recorded fifteen years apart).

In-depth interviews with the three LE 15 interpreters showed that they were extremely goal-focused from early on in life and engaged in many practice-like activities (including for several of their sub-skills). They did not however give any indication of engaging in deliberate practice as defined by Ericsson et al. 1993. They did talk about constantly striving to produce better interpretations and also about the positive physical experience of performing well.

5. Discussion

This section discusses both the methodology and results of the dissertation.

5.1 Methodological discussion

First, it should be said that the combination of quantitative and qualitative methods has been fruitful and makes to a certain extent up for the low number of participants. The surprising findings in the various studies may be due to the research instruments used, but as discussed in section 4.3.4, there is also strong support for the functionality of the methodological design and thereby the validity of the results.

One of the factors that may skew the results is the scales chosen to measure the quality of the interpreting product. However, they were tested and retested and ought to be considered reliable, and they provided the expected outcome when they were used in the cross-sectional material that supports their validity.

Another factor that may affect the results is the elicitation tool. It is a fact that experts do not excel in routine tasks (see section 2.3), and there is a possibility that the interpreting task in the study was seen as a (too simple) routine task by the LE 15 participants; however, this was not observed in the LE 25 participants. Moreover, the elicitation tool (the speech) was to a certain extent adapted. Fairly short speeches were used for experimental reasons, and one of the criteria for choosing the speech was its generality, so as not to make it impossible for the laypersons participating in the experiment. The first speech was nevertheless tweaked in order to add difficulties such as figures, names and difficult reasoning. For the long-term group (LE 15), the second speech was the same one they had interpreted at the interpreting programme for comparative reasons; this speech presented difficulties to the LE 15 interpreters since it was dated

and dealt with matters outside their area of expertise. Except for the IT group in the long-term material, the speeches were not interpreted in a routine setting. LE 15 and LE 25 interpreters interpreted from an interpreting booth, but they did not have any ordinary listeners and the elicitation speech was recorded; they were also surrounded by recording equipment. The SE and NE groups performed their interpreting either at university or at the researcher's workplace. The setting for the LE 15 and LE 25 interpreters does not indicate a routine task, which in this case would be in a booth with a live speaker, live audience and at least one colleague. In addition to the unusual conditions, the research leader was also present next to the participants listening to their performance. The experimental setting of the data collection event did not affect the performance of the LE 25 group, however, and the SE group performed at the same level as the LE 15 group. The elicitation situation was moreover a stressful event where experts could potentially excel because of their access to expert knowledge.

Furthermore, the result may be skewed since the interpreters' voice was not part of the assessment in the ratings, which were done from transcripts with normalized spelling and syntax (based on intonation). Collados Aís et al. (2011) point out that voice quality is so important that it may actually overshadow other key factors. The aim of the rating in this study was to assess the interpreter's ability to reproduce in the target language an understandable message that contained the information in the source language message, and it was assumed that the possibility of a rater identifying the rated subject through the voice would affect the evaluation. However, as important as voice may seem when grading interpreting, an experiment that compared grading from sound files and grading from transcripts showed that there was no significant difference between the two (Tiselius 2010). Finally, the same conditions were true for the cross-sectional material, and in that material the differences between the groups were clear.

Based on the arguments above I assume that the instruments were valid and reliable and that the data were representative. The following section (5.2) will discuss the results.

5.2 Discussion of the results

The quantitative results of the long-term material were the most surprising ones, as they went against the assumption that experience would enhance the interpreting performance and yield high assessment scores. As mentioned above, there are many superficial indications that the three LE 15 interpreters are highly skilled. However, the assessment results indicate they are not experts as defined by Ericsson and Smith (1991). It cannot be excluded that their expert knowledge is found in other areas than the one tested in these studies. In the in-depth interviews they talk about adding languages, improving their general knowledge and working on delivery, so perhaps that is where their expert performance can be found.

Hervais-Adelman et al. (2011) showed that changes take place in an interpreter's brain during training; given the results of the long-term study in this dissertation, we may wonder whether such neurological and cognitive changes take place during training and then remain fairly stable. As mentioned above in section 2.1, Hill and Schneider (2007: 675) say that as processing becomes automatized, the influence of the general control network (necessary while learning a task) either decreases or disappears entirely. If we assume that the automatization is completed during the training programme, and if automatized processes vs. non-automatized processes is what influence the results of the grading, then the difference between the subjects may be too small on the intra-individual level to be measurable, as the processes in this scenario were acquired and automatized during the interpreting programme and then perhaps unchanged over the years.

The experienced interpreters in the cross-sectional material (LE 25), who had more than twenty-five years of experience, received considerably higher scores than the experienced interpreters in the long-term material (LE 15), who had fifteen years of experience at their second recording. This may be due to experiment design, as discussed above, but it may also indicate that it takes longer than the supposed ten years, or even fifteen years, to gain expert knowledge in interpreting. In-depth interviews with the LE 25 group could have shed light on those issues, but practical constraints entailed that such interviews were unfortunately only carried out with the LE 15 group.

Finally, as is also discussed in article 4, there are few incentives for professional interpreters who have reached the highest level of their field to continue practising their skills. Having been accredited to international institutions, they are subject to constant quality monitoring so as not to perform below a certain minimum, but there are no mechanisms aimed at improving the main skill. Professional development consists of improving general knowledge or language skills (clearly important), but there are no rankings or pay raises for the best interpreters. The incentive for improvement lies instead in personal well-being and in the satisfaction of a job well done (as assessed by themselves). This is not necessarily a bad incentive, but the importance of the environment should not be underestimated.

5.3 The expertise theory and simultaneous interpreting

The findings are not conclusive concerning expertise in interpreting. In the cross-sectional material the results were clear and conclusive. There is a clear dividing line between the NE group on the one hand and the SE and LE 25 groups on the other, and there are also measurable differences both for process and product between the SE interpreters and the LE 25

interpreters. With these results we can assume that training and experience clearly matter in interpreting, and that extensive experience makes a difference.

The results from the cross-sectional material are however not confirmed by the interpreters in the long-term material (the LE 15 group), who did not interpret a given speech measurably better despite fifteen years of active experience. As years in the profession is a weak factor of expertise, the other parts of the superficial side of expertise was also taken into account, and all three of them had credentials that would put them in the expert category. But the LE 15 group outperformed neither the early recordings of themselves nor their experienced colleagues (the LE 25 group) in the cross-sectional material. Their performance was stable when improvement was expected. Either it must be assumed that these interpreters were not experts in absolute terms as defined by Ericsson and Smith (1991), or further investigations are needed to obtain more information.

The in-depth interview also shows that it is still an open question whether interpreters engage in deliberate practice as defined by Ericsson et al. 1993. Although the interpreters described examples of what would be labelled practice, none actually said that they practise and two of them even explicitly said that they never practise. This is supported by Vik-Tuovinen's (2006: 308) finding that experienced interpreters are less occupied with preparation than their less experienced counterparts. None of the three interpreters in the present study talks about or provides any examples of working to improve their main skill. So although they practise, it is hard to label it deliberate practice in the sense assumed in the expertise theory. On top of that, they all more or less think there is an X factor or an innate talent in interpreting. If interpreters thus believe that their main skill is innate, there may be less reason to continue practising this skill. As seen above, students are taught to practise and assess their

interpreting skill, and individuals who prepare for an accreditation test for interpreters presumably also practise this skill. Once a student has graduated or a novice interpreter has passed the accreditation test, however, there are few if any courses or initiatives to refine the interpreting skill. Although interpreters learn to improve their language skill, learn new languages, enhance their general knowledge and possibly also practise their consecutive interpreting skill as they prepare to add that new language to their combination, there seem to be few incentives to improve the interpreting skill itself once the student has graduated or passed an accreditation test. There are no particular merits for improving the interpreting skill, such as higher salaries, prestigious prizes nor an improved ranking. Freelance interpreters can presumably get more assignments and thereby more money if their interpreting skill improves. But many other factors are involved when freelancers are assigned jobs, such as language combination, availability, geographical proximity and not least personal connections. Hunt (2007: 35) points out that since expertise requires both motivation and support, society greatly influences where expertise is produced: in areas where remunerations are high and excellence in a field is remunerated even higher, experts are likely to prosper. In the interpreting world, conference interpreters are paid the highest while staff conference interpreters at various institutions also receive a comparatively high monthly salary. According to Hunt's theory, interpreting experts could thus be expected to be found as staff interpreters at international institutions.

For the expertise theory, the subjects' engagement in deliberate practice is an absolute condition. Interpreting studies have only started to discover what characterizes an expert performance, and the findings here suggest that research must be carried out on deliberate practice in interpreting as well. These findings indicate either that interpreters practise their skills in a naive manner (that is, without a conscious understanding of the deliberate dimension, yet with features of deliberate practice), or that the

three interpreters do not engage in deliberate practice, something that might also apply to conference interpreters in general. More studies must examine deliberate practice in interpreting before a definitive conclusion can be drawn, but there may be a need to redefine either the concept of deliberate practice in interpreting in particular or the criteria of the expertise theory in general.

Moreover, it is clear from these results that the expertise label needs to be used cautiously. Liu's overview of expertise in interpreting (2009) shows that much is known about interpreters with a certain amount of experience and what they do or not. But not much is known about what *experts* do, since the definition of an expert interpreter remains unclear. This is not to say that expertise in interpreting would have to take the Weiss and Shanteau (2003) definition that no measurable outcome exists concerning the expertise of interpreters, but the identification of expertise in interpreting, following the expertise theory (Ericsson, Charness & Hoffman 2007), requires more methodologically minute studies, before we can sketch the traits necessary for expertise in interpreting. In fact, it may be difficult to assess an outstanding performance because the group of highly experienced interpreters is small and homogeneous, and presumably they could be all excellent or all mediocre. These and other question marks must be addressed if interpreting studies is to adhere to the strictest definition of the expertise approach as described above.

6. Conclusion

The aim of this dissertation was twofold, as it contained both a methodological and a research side. In regard to methodology, it aimed to test holistic scales for assessing interpreting and develop an interview guide for in-depth interviews on deliberate practice. In regard to research, it aimed to establish a measurable difference in the interpreting skill (concerning both process and product) among interpreters with different levels of experience, and to explore what this difference consisted of. The data consisted of a cross-sectional material (n=9) and a long-term material (n=3). The following text repeats the questions in section 2.5 and provides answers to them:

- 1) Is there a measurable difference in the interpreting skill from the student level to the highly experienced level?
 - a. The assumption that there would be a measurable difference in the interpreting skill among performers with little or no experience and performers with long experience was supported for one of the data sets, the cross-sectional data (A).
- 2) Is there a measurable difference in the interpreting skill both when it is measured cross-sectionally (i.e. inter-individually) and long-term (i.e. intra-individually)?
 - a. The assumption that there would be a measurable difference in the interpreting skill regardless of data was not supported, as there was no measurable difference between the IT interpreters and the LE 15 interpreters in the long-term data (B).
- 3) If there is a measurable difference, what does this difference consist of?
 - a. The assumption that there would be a difference in rating between participants with little experience and participants with long experience was supported by the cross-sectional material (A), but not for the long-term material (B).
 - b. The assumption that there would be a difference in the interpreting process between participants with little experience and participants with long experience was

supported by the cross-sectional material (A) and could not be tested for the long-term material (B).

- 4) How do experienced interpreters perceive different factors in their long-term competence development?
 - a. The assumption that experienced interpreters would claim that they practise a lot was not supported. However, the participants they talked frequently about other practice-like activities.
 - b. The assumption that experienced interpreters would claim that they constantly strive to improve themselves was supported.
 - c. The assumption that experienced interpreters would be able to talk about their goals, on both the micro and macro levels, was not directly supported (although they often talked about how they had achieved different goals and generally seemed goal-oriented).
 - d. The assumption that experienced interpreters would claim that they made use of their colleagues for feedback and help was partly supported (interpreters help their colleagues in the booth and also listen for inspiration, but coaching outside the booth was never mentioned).
 - e. The assumption that experienced interpreters would be able to describe how they solve issues under pressure was supported (all participants talked about their ability to focus and perform under difficult conditions).

For the methodological development part, the following questions were answered:

- 1) Can holistic scales for measuring intelligibility and informativeness be developed into a valid and reliable measuring instrument for quality in interpreting?
 - a. Yes, the assumption that the scales could be developed and tested so as to form a valid and reliable measuring instrument for quality in interpreting was supported.
- 2) Will holistic scales work equally well as a measuring instrument whether used by laypersons or experienced raters?
 - a. Yes, the assumption that the scales could be used by different raters and still generate valid results was supported.

-
- 3) How should an in depth-interview be carried out in order to yield results on the concept of deliberate practice?
 - a. An interview guide was created and participants were prompted to discuss issues of deliberate practice through different trigger topics.

Interpreting performance was compared, in regard to both process and product, in many combinations between subjects without interpreting experience and subjects with different levels of interpreting experience, both cross-sectionally and long-term. As reported above, there was a measurable difference between the groups in the cross-sectional material, but not in the long-term material. Another conclusion, supported by other studies, is that experienced interpreters, when interpreting, have more strategies at hand and encounter fewer processing problems than less experienced interpreters or laypersons to interpreting non-interpreters. The results from the project supported the findings in Liu's literature review (2009), where she notes that experienced interpreters seem to have developed well-practised strategies in the comprehension, translation and production processes that are specific to the needs of simultaneous interpreting. And finally, these experienced interpreters have developed an ability to efficiently manage their attention so that it can be switched between different processes (Liu 2009: 174). Vik-Tuovinen's (2006) results are also supported by the results in this project. Her finding that more experienced interpreters were conscientious about their delivery was confirmed in the in-depth interview. From the interviews it could also be concluded that the experienced interpreters allocate much time for practice, although they don't consciously label it as such. Furthermore, they were also highly goal-oriented both in life in general and when interpreting.

From a methodological point of view, the holistic scales that were adapted from Carroll (1966) and used in articles 1–3 are well-tested by now and await further testing by for instance examiners. The retrospective method used for exploring the interpreter's process also produced valid and

reliable results, although, as shown by Englund Dimitrova and Tiselius (submitted), the use of retrospective protocols needs to be done with great exactitude. Retrospection is nevertheless warmly recommended as a method for investigating interpreting processes.

This is a large study, albeit with few participants. In order to really evaluate and map outstanding performance, further studies and more extensive data per interpreter are needed, with each participant ideally being recorded on several different occasions.

It is encouraging that the quantitative results from the cross-sectional study supported other results from other researchers (Ivanova 1999 and Vik-Tuovinen 2006). Hopefully, those results can in turn be supported by other researchers too.

Further longitudinal or long-term research would also be welcome in the field, as little is known within interpreting studies about intra-individual development beyond interpreting programmes. In order to understand the development of not only expertise in interpreting but also of interpreters after graduation, the field needs to see more longitudinal or long-term studies. Surely there must be many audio cassettes or mp3 files from various interpreter training programmes lying around waiting to be followed up.

Comparison with expertise in other fields that also lack rankings and reward systems would also be welcome, together with discussions on how to identify a practitioner's main skill. The main skills of nurses, researchers or blacksmiths could presumably also be the object of discussion.

Finally, I repeat my call for more studies on deliberate practice. In order to study all the domains of expertise in interpreting, deliberate practice needs to be part of the tradition. In this thesis, interpreters' deliberate practice has been studied through in-depth interviews, but it could also be studied

through for instance diary studies or on a micro-level with a quantitative design.

List of references

- Ahrens, B., Kalderon, E., Krick, M. C. and Reith, W. 2010. "fMRI for exploring simultaneous interpreting." In *Why Translation Studies Matters* [Benjamins Translation Library. EST Subseries: Vol. 6.], D. Gile, G. Hansen and N. K. Pokorn (eds), 237–248. Amsterdam and Philadelphia: John Benjamins.
- Anderson, L. 1979. "Simultaneous interpretation: Contextual and translation aspects." (Master's thesis, Concordia University, Canada.)
- . 1994. "Simultaneous interpretation: Contextual and translation aspects." In *Bridging the Gap. Empirical Research in Simultaneous Interpretation* [Benjamins Translation Library: Vol. 3.], S. Lambert and B. Moser-Mercer (eds), 101–120. Amsterdam and Philadelphia: John Benjamins.
- Angelelli, C. 2004. *Revisiting the Interpreter's Role: A Study of Conference, Court, and Medical Interpreters in Canada, Mexico, and the United States* [Benjamins Translation Library: Vol. 55]. Amsterdam and Philadelphia: John Benjamins.
- . 2009. "Using a rubric to assess translation ability: Defining the construct." In *Testing and Assessment in Translation and Interpreting Studies* [ATA Monograph Series], C. V. Angelelli and H. E. Jacobson (eds), 13–47. Amsterdam and Philadelphia: John Benjamins.
- Angelelli, C. V. 2007. "Assessing medical interpreters: The language and interpreting testing project." *The Translator* 13 (1): 63–82.
- Angelelli, C. V. and Jacobson, H. E. (eds). 2009. *Testing and Assessment in Translation and Interpreting Studies* [ATA Monograph Series]. Amsterdam and Philadelphia: John Benjamins.
- Azbel-Schmidt, M. 2005. *How Do You Do It Anyway?: A Longitudinal Study of Three Translation Students Translating from Russian into Swedish*. Stockholm: Acta Universitatis Stockholmiensis.
- Bajo, M. T., Padilla, F. and Padilla, P. 2000. "Comprehension processes in simultaneous interpreting." In *Translation in Context: Selected Papers from the EST Congress, Granada, 1998*, A. Chesterman, N. Galardo San Salvador and Y. Gambier (eds), 127–142. Amsterdam Philadelphia: John Benjamins.
- Barik, H. C. 1973. "Simultaneous interpretation: Temporal and quantitative data." *Language and Speech* 16: 237–270. Retrieved from <http://las.sagepub.com/content/16/3/237>.
- . 1975. "Simultaneous interpretation: Qualitative and linguistic data." *Language and Speech* 18 (3); 272–297.

-
- Berezhkov, V. M. 1994. *At Stalin's Side: His Interpreter's Memoirs from the October Revolution to the Fall of the Dictator's Empire*. Secaucus, NJ: Carol.
- Bontempo, K. and Napier, J. 2007. "Mind the gap! A skills analysis of sign language interpreters." *The Sign Language Translator and Interpreter* 1 (2): 275–299.
- Bühler, H. 1986. "Linguistic (semantic) and extra-linguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters." *Multilingua* 5 (4): 231–235.
- Carroll, J. B. 1966. "An experiment in evaluating the quality of translations." *Mechanical Translation and Computational Linguistics*, 9 (3–4): 55–66.
- Chincotta, D. U. G. 1998. "Simultaneous interpreters and the effect of concurrent articulation in immediate memory: A bilingual digit span study." *Interpreting* 3 (1): 1–20.
- Chernov, G. V. 1979. "Semantic aspects of psycholinguistic research in simultaneous interpretation." *Language and Speech* 22: 277–295.
- Clifford, A. 2001. "Discourse theory and performance-based assessment: Two tools for professional interpreting." *Meta* 46 (2): 365–378. Retrieved from <http://www.erudit.org/revue/meta/2001/v46/n2/002345ar.pdf>.
- . 2004. "Is fidelity ethical? The social role of the healthcare interpreter." *TTR: Traduction, Terminologie, Rédaction*, 17 (2): 89–114. Retrieved from <http://www.erudit.org.ezp.sub.su.se/revue/ttr/2004/v17/n2/013273ar.pdf>.
- . 2005. "Putting the exam to the test: Psychometric validation and interpreter certification." *Interpreting* 7 (1): 97–131.
- Collados Aís, Á., Iglesias Fernández, E. P. M. E. M. and Stévaux, E. 2011. *Qualitätsparameter beim Simultandolmetschen: Interdisziplinäre Perspektiven*. Tübingen: Narr Verlag.
- Collins, H. M. 2010. *Tacit and Explicit Knowledge*. Chicago and London: University of Chicago Press.
- Collins, H. and Evans, R. 2007. *Rethinking Expertise*. London and Chicago: University of Chicago Press.
- Corsellis, A. 2008. *Public Service Interpreting: The First Steps [Palgrave Textbooks in Translating and Interpreting]*. Basingstoke: Palgrave Macmillan.
- Daro, V. and Fabbro, F. 1994. "Verbal memory during simultaneous interpretation: Effects of phonological interference." *Applied Linguistics* 15 (4): 365–381.

- Deakin, J. M., Côté, J. and Harvey, A. S. 2007. "Time, budgets, diaries, and analyses of concurrent practice activities." In *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), 303–318. Cambridge: Cambridge University Press.
- Denzin, N. K. 2006. *Sociological Methods: A Sourcebook* (5th ed.). New Brunswick, NJ: Aldine Transaction.
- Dillinger, M. 1994. "Comprehension during interpreting: What do interpreters know that bilinguals don't?" In *Bridging the Gap. Empirical Research in Simultaneous Interpretation* [Benjamins Translation Library: Vol. 3], S. Lambert and B. Moser-Mercer (eds), 101–120. Amsterdam and Philadelphia: John Benjamins.
- . 1989. "Component processes of simultaneous interpreting." Doctoral dissertation, McGill University, Montreal.
- Diriker, E. 2004. *De-/Re-Contextualising Conference Interpreting: Interpreters in the Ivory Tower?* [Benjamins Translation Library: Vol. 53]. Amsterdam and Philadelphia: John Benjamins. Retrieved from <http://www.benjamins.nl/#catalog/books/btl.53/main>.
- Dreyfus, S. E. and Dreyfus, H. L. 1980. "A five-stage model of the mental activities involved in directed skill acquisition." University of California, Berkeley. Unpublished report supported by the Air Force Office of Scientific Research, USAF (contract F49620-79-C0063). <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA084551> (retrieved May 21, 2013).
- Englund Dimitrova, B. 1991. *När två samtalar genom en tredje: Interaktion och icke-verbal kommunikation i medicinska möten med tolk* [Rapporter om tvåspråkighet: Vol. 7]. Stockholm: Centrum för tvåspråkighetsforskning vid Stockholms universitet.
- . 2005a. *Expertise and Explicitation in the Translation Process*. Amsterdam and Philadelphia: John Benjamins.
- . 2005b. "Combining product and process analysis: Explicitation as a case in point." *Bulletin suisse de linguistique appliqué* 81; 25–39.
- . 2010. "Translation process." In *Handbook of Translation Studies*, Y. Gambier and L. van Doorslaer (eds), 406–411. Amsterdam and Philadelphia: John Benjamins.
- Englund Dimitrova, B. and Tiselius, E. 2009. "Exploring retrospection as a research method for studying the translation process and the interpreting process." In *Methodology, Technology and Innovation in Translation Process Research* [Copenhagen Studies in Language: Vol. 38], I. M. Mees, F. Alves and S. Göpferich (eds), 109–134. Copenhagen: Samfundslitteratur.

-
- . (in press). “Retrospection in interpreting and translation: Explaining the process.”
- Ericsson, K. A. 1996. *The Road to Excellence: The Acquisition of Expert Performance in the Arts and Sciences, Sports, and Games*. Mahwah, NJ: Lawrence Erlbaum Associates.
- . 2004. “Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains.” *Academic Medicine: Journal of the Association of American Medical Colleges* 79 (10 Suppl.): 70–81.
- . 2007a. “The influence of experience and deliberate practice on the development of superior expert performance”. In *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), 683–703. Cambridge: Cambridge University Press.
- . 2007b. “An introduction to *Cambridge Handbook of Expertise and Expert Performance*: Its development, organization and content”. In *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), 3–19. Cambridge: Cambridge University Press.
- . (ed.). 2009. *Development of Professional Expertise: Toward Measurement of Expert Performance and Design of Optimal Learning Environments*. New York: Cambridge University Press.
- Ericsson, K. A., Charness, N. F. P. J. and Hoffman, R. R. (eds). 2007. *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.). Cambridge: Cambridge University Press.
- Ericsson, K. A., Krampe, R. T. and Tesch-Römer, C. 1993. “The role of deliberate practice in the acquisition of expert performance.” *Psychological Review* 100 (3): 363–406. doi:10.1037/0033-295X.100.3.363.
- Ericsson, K. A., Prietula, M. J. and Cokely, E. T. 2007. “The making of an expert.” *Harvard Business Review*. Retrieved from http://141.14.165.6/users/cokely/Ericsson_Preitula_&_Cokely_2007_HBR.pdf.
- Ericsson, K. A. and Simon, H. A. 1993. *Protocol Analysis: Verbal Reports as Data* (rev. ed.). Cambridge, MA: MIT Press.
- Ericsson, K. A. and Smith, J. 1991. *Toward a General Theory of Expertise: Prospects and Limits*. Cambridge: Cambridge University Press.
- Gaiba, F. 1998. *The Origins of Simultaneous Interpretation: The Nuremberg Trial* [Perspectives on Translation]. Ottawa: University of Ottawa Press.

- Gerver, D. 1971. "Aspects of simultaneous interpretation and human information processing." (Doctoral dissertation, Oxford University, UK.)
- . 1976. "Empirical studies of simultaneous interpretation: A review and a model." In *Translation: Applications and Research*, R. W. Brislin (ed.), 165–202. New York: Gardner Press.
- Gile, D. 1985a. "La sensibilité aux écarts de langue et la sélection d'informateurs dans l'analyse d'erreurs: une expérience." *The Incorporated Linguist* 24 (1): 29–32.
- . 1985b. "Le modèle d'efforts et l'équilibre d'interprétation en interprétation simultanée." *Meta* 30 (1): 44–48. Retrieved from <http://id.erudit.org/iderudit/002893ar>.
- . 2003. "Quality assessment in conference interpreting: Methodological issues." In *La evaluación de la calidad en interpretación: Investigación* [Colección interlingua: Vol. 37], A. Collado Aís, M. M. Fernández Sánchez and D. Gile (eds), 109–123. Granada: Comares.
- . 2005a. "Empirical research into the role of knowledge in interpreting: Methodological aspects." In *Knowledge Systems and Translation*, H. V. Dam, J. Engberg and H. Gerzymisch-Arbogast (eds), 149–171. Berlin and New York: Mouton de Gruyter.
- . 2005b. "Teaching conference interpreting." In *Training for the New Millennium: Pedagogies for Translation and Interpreting* [Benjamins Translation Library, EST subseries: Vol. 60.], M. Tennent (ed.), 127–151. Amsterdam and Philadelphia: John Benjamins.
- . 2009. *Basic Concepts and Models for Interpreter and Translator Training* (rev. ed.) [Benjamins Translation Library: Vol. 8]. Amsterdam and Philadelphia: John Benjamins.
- Gladwell, M. 2008. *Outliers: The Story of Success*. New York: Little, Brown and Company.
- Goldman-Eisler, F. 1972. "Segmentation of input in simultaneous translation." *Journal of Psycholinguistic Research* 1 (2): 127–140.
- Gran, L. and Fabbró, F. 1987. "Cerebral lateralization in simultaneous interpreting." In *Across the Language Gap. Proceedings of the 28th Annual Conference of the American Translators Association, Albuquerque, New Mexico, October 8–11, 1987*, K. Kummer (ed.), 323–335. Medford, NJ: Learned Information.
- Göpferich, S. 2013. "Translation competence: Explaining development and stagnation from a dynamic systems perspective." *Target* 25 (1): 61–76.

-
- Hansen, G. 2008. "Kommunikative Kompetenz in Übersetzungsprozessen: Eine Langzeitstudie." *Redit* (1): 48–56.
- . (in press). "Many tracks lead to the goal: A long-term study on individual translation styles." In *Tracks and Treks in Translation Studies. Proceedings of the 6th EST Conference*, M. Bartlomiejczyk, R. Meylaerts, S. Vandepitte and C. Way (eds), 49-62 Amsterdam and Philadelphia: John Benjamins.
- Herbert, J. 1952. *Manuel de l'interprete: Comment on devient interprete de conferences* (2nd ed.). Geneva: Librairie de l'Universite Georg.
- Hervais-Adelman, A. G., Moser-Mercer, B. and Golestani, N. 2011. "Executive control of language in the bilingual brain: integrating the evidence from neuroimaging to neuropsychology." *Frontiers in Psychology* 2. Retrieved from <http://www.frontiersin.org/Cognition/10.3389/fpsyg.2011.00234/full>.
- Hild, A. 2007. "Establishing rigour in a between-methods investigation of expertise." In *Doubts and Directions in Translation Studies. Selected Contributions from the EST Congress, Lisbon 2004* [Benjamins Translation Library: Vol. 72], Y. Gambier, M. Shlesinger and R. Stolze (eds), 99–12. Amsterdam and Philadelphia: John Benjamins.
- Hill, N. M. and Schneider, W. 2007. "Brain changes in the development of expertise: Neuroanatomical and neurophysiological evidence about skill-based adaptations." In *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), 653–682. Cambridge: Cambridge University Press.
- Hoffman, R. R. 1997. "The cognitive psychology of expertise and the domain of interpreting." *Interpreting* 2 (1–2): 189–230.
- Horn, J. and Masunaga, H. 2007. "A merging theory of expertise and intelligence." In *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), 587–611. Cambridge: Cambridge University Press.
- Hunt, E. 2007. "Expertise, talent, and social encouragement." In *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), 31–38. Cambridge: Cambridge University Press.
- Iglesias Fernández, E. 2003. "La interpretación en la universidad española. Estudio empírico de la situación académica y didáctica." (Doctoral dissertation, Universidad de Granada, Spain).
- Isham, W. P. 1994. "Memory for sentence form after simultaneous interpretation: Evidence both for and against deverbilization". In *Bridging the Gap. Empirical Research in Simultaneous Interpretation*

- [Benjamins Translation Library: Vol. 3], S. Lambert and B. Moser-Mercer (eds), 191–211. Amsterdam and Philadelphia: John Benjamins.
- Ivanova, A. 1999. “Discourse processing during simultaneous interpreting: An expertise approach.” (Doctoral dissertation, University of Cambridge, UK.)
- Jones, R. 1998. *Conference Interpreting Explained* [Translation Theories Explained: Vol. 5]. Manchester: St. Jerome.
- Jääskeläinen, R. 2010. “Are all professionals experts? Definitions in expertise and reinterpretation of research evidence in process studies.” In *Translation and Cognition* [ATA Monograph Series: Vol. 15], G. M. Shreve and E. Angelone (eds), 213–227. Amsterdam and Philadelphia: John Benjamins.
- Kaijser, L. and Öhlander, M. (eds). 1999. *Etnologiskt fältarbete*. Lund: Studentlitteratur.
- Kalina, S. 2005. “Quality assurance for interpreting processes.” *Meta* 50 (2): 768–784. Retrieved from <http://www.erudit.org.ezp.sub.su.se/revue/meta/2005/v50/n2/011017ar.pdf>.
- Kalina, S. 2011. “Maß für Maß: Eine vergleichende Profilanalyse von Diskursen beim Dolmetschen.” *trans-kom* 4 (2): 161–175. Retrieved from http://www.trans-kom.eu/bd04nr02/trans-kom_04_02_03_Kalina_Profilanalyse.20111205.pdf.
- Koskinen, K. 2008. *Translating Institutions: An Ethnographic Study of EU Translation*. Manchester and Kinderhook: St. Jerome.
- Kurz, I. 1993. “Conference interpretation: Expectations of different user groups.” *The Interpreters’ Newsletter* 5: 13–21. Retrieved from <http://www.openstarts.units.it/dspace/handle/10077/4908>.
- Köpke, B. and Nespoulous, J.-L. 2006. Working memory performance in expert and novice interpreters. *Interpreting*, 8 (1), 1–23.
- Lederer, M. 1978. “Simultaneous interpretation: Units of meaning and other features.” In *Language Communication and Interpretation*, D. Gerver and H. W. Sinaiko, 323–333. New York and London: Plenum Press.
- Lefort, I. 2012. “Je me dois de n’être qu’une voix, la voix de l’autre: Entretien avec Amanda Galsworthy, interprète free lance pour l’Elysée et présidente d’Alto-International.” *La Tribune*. Retrieved from <http://www.latribune.fr/accueil/a-la-une.html>.
- Leis, Ü. 2003. “Tõlkide elukestev õpe Eestis – olukorra ja vajaduste võrdlev analüüs.” (Master’s thesis, University of Tartu, Estonia.)
- Liu, M. 2001. “Expertise in simultaneous interpreting: A working memory analysis.” (Doctoral dissertation, University of Texas at Austin, USA.)

-
- . 2008. “How do experts interpret? Implications from research in Interpreting Studies and cognitive science.” In *Efforts and Models in Interpreting and Translation Research. A Tribute to Daniel Gile* [Benjamins Translation Library: Vol. 80], G. Hansen, A. Chesterman and H. Gerzymisch-Arbogast (eds), 159–178. Amsterdam and Philadelphia: John Benjamins.
- Liu, M., Schallert, D. L. and Carroll, P. J. 2004. “Working memory and expertise in simultaneous interpreting.” *Interpreting* 6 (1): 19–42.
- Macnamara, B. N., Moore, A. B., Kegl, J. A. and Conway, A. R. 2011. “Domain-general cognitive abilities and simultaneous interpreting skill.” *Interpreting* 13 (1): 121–142.
- Mantoux, P. 1955. *Les délibérations du Conseil des Quatre (24 Mars–28 Juin 1919): notes de l’officier interprète. Vol. 2: Depuis la remise à la délégation allemande des conditions de paix jusqu’à la signature du Traité de Versailles*. Paris: CNRS.
- Monacelli, C. 2009. *Self-Preservation in Simultaneous Interpreting: Surviving the Role* [Benjamins Translation Library: Vol. 84]. Amsterdam and Philadelphia: John Benjamins. Retrieved from <http://www.benjamins.nl/#catalog/books/btl.84/main>.
- Moser, B. 1978. “A hypothetical model and its practical application.” In *Language Communication and Interpretation*, D. Gerver and H. W. Sinaiko (eds), 353–368. New York and London: Plenum Press.
- Moser, P. 1995. *Survey on Expectations of Users of Conference Interpretation*. Retrieved from http://aiic.net/community/attachments/ViewAttachment.cfm/a525p736-918.pdf?&filename=a525p736-918.pdf&page_id=736.
- Moser-Mercer, B. 1997. “Methodological issues in interpreting research: An introduction to the Ascona workshops.” *Interpreting* 2 (1–2): 1–11.
- . 2000. “The rocky road to expertise in interpreting: Eliciting knowledge from learners.” In *Translationswissenschaft*, M. Kadric, K. Kaindl and F. Pöchhacker (eds), 339–352. Tübingen: Stauffenburg Verlag.
- . 2009. “Construct-ing quality”. In *Efforts and Models in Interpreting and Translation Research. A Tribute to Daniel Gile* [Benjamins Translation Library: Vol. 80], G. Hansen, A. Chesterman and H. Gerzymisch-Arbogast (eds), 143–156. Amsterdam and Philadelphia: John Benjamins.
- Moser-Mercer, B., Frauenfelder, U. H., Casado, B. and Künzli, A. 2000. “Searching to define expertise in interpreting.” In *Language Processing and Simultaneous Interpreting. Interdisciplinary Perspectives* [Benjamins Translation Library: Vol. 40], B. Englund Dimitrova and K.

- Hyltenstam (eds), 107–131. Amsterdam and Philadelphia: John Benjamins.
- Muñoz-Martín, R. (in press). “Situating translation expertise: A review with a sketch of a model.”
- Napier, J. and Bontempo, K. 2011. “Evaluating emotional stability as a predictor of interpreter competence and aptitude for interpreting.” *Interpreting* 13 (1): 85–105.
- New Jersey Courts. 2007. *Knowledge, Skills, and Abilities (KSAs) for the Profession of Court Interpretation*. Retrieved from <http://www.judiciary.state.nj.us/interpreters/ksa.pdf>.
- Nolan, J. 2005. *Interpretation: Techniques and Exercises* [Professional Interpreting in the Real World: Vol. 2]. Clevedon: Multilingual Matters.
- Patton, M. Q. 2002. *Qualitative Research and Evaluation Methods* (3rd ed.). Thousand Oaks and London: Sage.
- Pöchhacker, F. 2004. *Introducing Interpreting Studies*. London: Routledge.
- Pöchhacker, F. and Shlesinger, M. (eds). 2002. *The Interpreting Studies Reader*. London and New York: Routledge.
- Rinne, J., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., Teräs, M., Sipilä, H. and Sunnari M. 2000. “The translating brain: cerebral activation patterns during simultaneous interpreting.” *Neuroscience Letters* 294: 85–88.
- Rozan, J. F. 1979 [1956]. *La prise de notes en interprétation consécutive* [Publications de l'École d'interprètes, Genève]. Geneva: Librairie de l'Université Georg.
- Satow, E. M. 1979. *Satow's Guide to Diplomatic Practice: Edited by Lord Gore-Booth* (5th ed.). London and New York: Longman.
- Sawyer, D. B. 2004. *Fundamental Aspects of Interpreter Education: Curriculum and Assessment* [Benjamins Translation Library: Vol. 47]. Amsterdam: John Benjamins.
- Scott, J. and Marshall, G. 2009. *A Dictionary of Sociology* (3rd ed.). Oxford and New York: Oxford University Press.
- Seeber, K. G. 2011. “Cognitive load in simultaneous interpreting: Existing theories – new models.” *Interpreting* 13 (2): 176–204. doi:10.1075/intp.13.2.02see.
- Seleskovitch, D. 2002. “Language and memory: A study of note-taking in consecutive interpreting.” In *The Interpreting Studies Reader*, F. Pöchhacker and M. Shlesinger (eds), 121–129. London and New York: Routledge.

-
- Seleskovitch, D., Lederer, M. and Harmer, J. 1995. *A Systematic Approach to Teaching Interpretation*. Silver Spring, MD: Registry of Interpreters for the Deaf.
- Shanteau, J. 1992. "Competence in experts: The role of task characteristics." *Organizational Behavior and Human Decision Processes* 53: 252–266.
- Sonntag, S., Niessen, C. and Volmer, J. 2007. "Expertise in software design." In *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), 373–387. Cambridge: Cambridge University Press.
- Sosniak, L. A. 2007. "Retrospective interviews in the study of expertise and expert performance." In K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), 287–301. Cambridge: Cambridge University Press.
- Stansfield, C. W. and Reed, D. J. 2004. "The story behind the Modern Language Aptitude Test: An interview with John B. Carroll (1916–2003)." *Language Assessment Quarterly* 1 (1): 43–56.
- Timarová, Š. 2012. "Working memory in conference simultaneous interpreting." (Doctoral dissertation, University of Leuven/Charles University, Belgium/The Czech Republic).
- Tiselius, E. 2010. "A sociological perspective on expertise in conference interpreting: A case study on Swedish conference interpreters." In *Translation Effects: Selected Papers of the CETRA Research Seminar in Translation Studies 2009*, O. Azadibougar (ed.), Leuven: KU Leuven.
- . (2010). "Methods for assessing simultaneous interpreting: Is there a difference using sound files or transcripts when assessing simultaneous interpreting?" Conference paper delivered at Research Methodologies in Translation and Interpretation Studies, Montreal, May 28–30, 2010.
- Vik-Tuovinen, G.-V. 2006. *Tolkning på olika nivåer av professionalitet*. Vasa: Universitas Wasaensis.
- Vuorikoski, A.-R. 2004. "A voice of its citizens or a modern Tower of Babel? The quality of interpreting as a function of political rhetoric in the European Parliament." (Doctoral dissertation, University of Tampere, Finland.)
- Weiss, D. J. and Shanteau, J. 2003. "Empirical assessment of expertise." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 45 (1): 104–116. doi:10.1518/hfes.45.1.104.27233.
- Williams, S. 1995. "Observations on anomalous stress in interpreting." *The Translator* 1 (1): 47–64.

Zimmerman, B. J. 2007. "Development and adaptation of expertise: The role of self-regulatory processes and beliefs." In *The Cambridge Handbook of Expertise and Expert Performance* (5th ed.), K. A. Ericsson, N. F. P. J. Charness and R. R. Hoffman (eds), 705–722. Cambridge: Cambridge University Press.