

**The Development of Coherence and Quality of Performance
in Conference Interpreter Training**

by

Kuei-Chuan Peng

Submitted in accordance with the requirements for
the degree of Doctor of Philosophy

The University of Leeds

School of Modern Languages and Cultures

Centre for Translation Studies

June 2006

The candidate confirms that the work submitted is her own and that the appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

I would like to express my gratitude to all those who made the completion of this thesis possible.

First, I am deeply indebted to my principal supervisor, Tony Hartley, for his guidance, inspiration and support all the way through my PhD. Without his encouragement and patience, I would never have achieved the results presented in this thesis. Thanks also to my second supervisor, David Pattinson, for his participation and comments on the drafts.

I owe special thanks to Svetlana Carsten, Director of MA Interpreting and Translation Studies at the University of Leeds, for her trust and support. She encouraged me to participate fully in the training course, allowing me to observe and collect data and to introduce ideas from my research to the students. Other MAITS colleagues were also very generous in sharing with me their insights into the profession of conference interpreting and their training expertise. I want to thank all those MAITS students and professional interpreters who contributed their performances and for their participation and cooperation over the past three years.

I also valued the opportunity to work with academic colleagues from other institutions, particularly with Ian Mason and Isabelle Perez of Heriot-Watt University.

I am grateful for the scholarship awarded to me by the School of Modern Languages and Cultures at Leeds University which allowed me to take my first step into the world of research and also for the funding from the National Centre for Languages (CILT) which supported the joint project with Heriot-Watt.

Finally, I wish to thank all my family and friends at home and abroad for their love and endless support. Many special thanks go to my husband, Martin Thomas, for going through my drafts many times and for making innumerable suggestions which helped to make this thesis more presentable. His patience and support made the final stages of this long and tough journey bearable and ultimately enjoyable.

Abstract

Conference interpreters are made not born, as Herbert (1978) and Mackintosh (1999) point out. The increase in the number of interpreting schools worldwide not only demonstrates the demand for qualified interpreters but also highlights the significance of training for ensuring quality service from conference interpreters. The purpose of training, ultimately, is to make competent interpreters who can deliver quality performance. Research on conference interpreting has tended to focus on issues of quality in professional circumstances (Buhler, 1986; Kahane, 2000; Kurz, 1993; Moser, 1996). Training is also one of the most explored fields in the world of conference interpreting research (Gile, 2000). Yet only limited work has been done to investigate quality standards for pedagogical purposes (Gile, 2001).

To produce quality interpretations, practice alone is not enough. Being reflective is of prime importance. Trainees' awareness of quality is vital for them to become reflective, yet this issue has not been properly addressed in the literature. In addition, in the trainer-centred approach, trainees acquire not only interpreting skills from their trainers, they also inherit the way trainers describe quality. Yet it is often observed that trainers do not share a common meta-language to discuss quality attributes of interpretations. Such confusion is inevitably passed on to the trainees. To address these situations, I gathered quality standards and criteria from professional, training and linguistic fields and devised a feedback tool which spells out those attributes explicitly. This feedback tool is adopted to raise trainees' awareness of quality and ultimately, help them progress in their interpretations.

Talking about quality of interpreting, 'making sense' is generally held to be one of the most important criteria for judging the success of a given interpretation, in both consecutive interpreting (CI) and simultaneous interpreting (SI) (Hatim & Mason, 1997; Kahane, 2000; Kopczynski, 1994; Kurz, 1993; Moser, 1996). For CI in particular, Hatim and Mason (2002: 262) state that the coherence and structure of the rendition are especially important (Peng & Hartley, 2005). Therefore, the significance of coherence should not be overlooked by trainees. Moreover, the development of coherence in their interpretations is a useful measure of their progress.

Building awareness of quality attributes of interpreting, such as coherence, is a process of evolution for trainees, and systematic guidance can facilitate this process (Peng, 2004). In this study, we address the question of how to observe and

investigate the development of coherence in interpreting. I propose that Rhetorical Structure Theory (RST) (1986) is a suitably rich discourse structure framework for exploring how coherence is realised in interpretations. RST has been widely used for describing the hierarchical organisation of natural texts in terms of some 30 functional relations holding between text chunks, thereby characterising the coherence of the whole text. It has also proven to be useful in describing the structure of spoken discourse (Tappe & Schilder, 1998). Its use by Marcu (2000) in automatic text summarisation – which introduces the notions of relevance and salience, and thereby a principled basis for progressively compressing a message – provides further inspiration for the analysis of my data.

My data consist of 66 consecutive interpretations, by eight trainees and three professional interpreters, of three Chinese and three English speeches. Each speech and interpreted discourse is transcribed, segmented into functional units, and mapped into a tree-like RST description. I compare these RST trees using three variables:

- 1) implicit/explicit discourse marking;
- 2) the structure (width and depth) of the tree;
- 3) and the nature of the summary yielded by Marcu's summarisation algorithm

RST also allows me to account for the occurrence of repair/self-correction to explore whether disfluency would impede the coherence of a discourse. The results from the comparison of trainee and professional performances reveal differing approaches to handling the coherence of a discourse. Trainees tend to focus on local cohesion while professionals tend to emphasise the global structure of the discourse. Furthermore, by observing the RST trees of trainee interpretations over time, I witness the development of their capacity for dealing with complex rhetorical structures by using more diverse and more specific connectives. In addition, I observe that a high frequency of self-correction definitely affects coherence, but few repairs do not guarantee good coherence. It is also noted that clear understanding of quality attributes, such as coherence, helps trainees to develop capacities in giving judgements of interpretations (Peng & Hartley, 2005). My evidence suggests this awareness also contributes to the improvement of their own performances.

I believe that RST offers a very useful framework to describe the abstract concept of coherence. It is also worth introducing RST analysis (or at least an RST-aware analysis) to interpreters during their training. Such analysis enables them to capture the structure of coherence better and to give more coherent renditions in their interpreting as a result. This thesis demonstrates that my exploratory approach offers interesting findings and implications for interpreter training, as well as

directions for further research in both the conference interpreting and RST communities.

Table of Contents

Acknowledgements	i
Abstract	ii
Table of Contents	v
List of Figures	ix
List of Tables	xii
List of abbreviations and acronyms	xiii
Chapter 1 Training for Quality Interpreters and Performance	1
1.1. Conference interpreting explained.....	1
1.2. Interpreters are made not born.....	2
1.3. Practice makes perfect.....	5
1.4. Making sense: gold standard for interpreting	8
1.5. Consecutive interpreting.....	10
1.6. Statement of purpose	11
Chapter 2 State of the Art	13
2.1. Consecutive interpreting.....	13
2.1.1. Consecutive interpreting as a profession	13
2.1.2. Consecutive interpreting in training.....	15
2.1.3. Consecutive interpreting and autonomous learning.....	16
2.2. Research on interpreting quality assurance.....	17
2.2.1. Professional standards	17
2.2.1.1. User expectations	20
2.2.2. Educational standards.....	22
2.2.3. Linguistically-informed standards.....	28
2.2.4. An integrated model	30
2.3. Making sense	31
2.3.1. Coherence	34
2.3.2. Cohesion	36
2.4. Rhetorical Structure Theory (RST)	39
2.4.1. Why did I choose RST?.....	39
2.4.2. Origins of RST	40

2.4.3.	RST explained –RST relations and definitions.....	41
2.4.4.	RST analysis.....	45
2.4.5.	Applications of RST	48
2.4.6.	RST for analysing interpretations.....	48
Chapter 3	Trainees’ Awareness of Quality	50
3.1.	Significance of trainees’ awareness	50
3.2.	Aims	52
3.3.	Methodology.....	52
3.3.1.	Subjects.....	52
3.3.2.	Questionnaires and timing.....	53
3.4.	Results and discussion	56
3.4.1.	Questionnaire A (novice vs. advanced trainees).....	56
3.4.2.	Questionnaire B (stage 1 vs. 2; novice vs. advanced trainees).....	58
3.4.3.	Difficult notions: register, coherence and cohesion.....	59
3.5.	Conclusion	60
Chapter 4	Development of the Feedback Grid	62
4.1.	Feedback criteria collection	62
4.2.	Design of the prototype feedback grid (version 1).....	63
4.3.	Methodology.....	67
4.3.1.	Subjects.....	67
4.3.2.	Experiment design	67
4.3.3.	Qualitative analytic framework	71
4.4.	Analysis and results	71
4.4.1.	Input data.....	71
4.4.2.	Analysis of feedback from trainers, professionals and users	71
4.4.2.1.	Categories not used by trainers, professionals and users	71
4.4.2.2.	Categories used by trainers, professionals and users	72
4.4.3.	Criteria used by trainees for self-assessment	75
4.5.	Use of grid by trainees.....	75
4.5.1.	Reports of Usefulness and user-friendliness.....	75
4.5.2.	Actual use by trainees for peer feedback	76
4.6.	Revision of the feedback grid	77
4.7.	Adaptation for consecutive interpreting	80
4.8.	Summary.....	84
Chapter 5	Methodology	85
5.1.	Data collection	86
5.1.1.	Subjects.....	86

5.1.2. Experiments	87
5.1.2.1. Materials	87
5.1.2.2. The recording of interpretations.....	89
5.2. Data annotation.....	90
5.2.1. Sense to texts	91
5.2.2. Text to RST trees.....	91
5.2.2.1. Repair: introduction of a new relation.....	97
5.2.2.2. Weighing RST trees	99
5.2.3. Explicit markers and overtly marked relations	99
5.2.4. Coherence Profile	104
5.3. Feedback grid introduction and administration.....	105
5.4. Trainee Judgement of Peer Performance.....	106
5.4.1. Subjects.....	107
5.4.2. Materials	107
5.4.3. Experiment setup.....	109
Chapter 6 Development of Coherence in Conference Interpreting.....	111
6.1. RST analysis	111
6.1.1. RST trees and tree weight.....	112
6.1.2. Trees and bushes	114
6.1.3. ‘Heavier’ RST trees in Chinese	116
6.1.4. Trainee interpreters start at same level.....	118
6.1.5. ‘Heavier’ RST trees of professional interpretations	119
6.1.6. RST relations vs. RST tree weight.....	120
6.1.7. Quality awareness facilitates performance	122
6.2. Overtly-marked relations with explicit markers	124
6.2.1. Overt ratios in English interpretations	125
6.2.2. Overt ratios in Chinese interpretations	126
6.2.3. Higher overt ratios in trainee interpretations at the beginning	127
6.2.4. Double conjunctions in English.....	127
6.3. Repair	130
6.3.1. More repairs in English than in Chinese interpretations	133
6.3.2. Trainee difference in repair	134
6.3.3. Fewer repairs vs. better coherence	134
6.4. Coherence profile.....	138
6.4.1. Benchmarking and benchmark validation	138
6.4.1.1. Professional interpretation on its own	139
6.4.1.2. Professional interpretation vs. source speech.....	142
6.4.2. Trainees’ coherence profile.....	144

6.4.2.1. Chinese-English interpretations	145
6.4.2.2. English-Chinese interpretations	150
6.5. Trainees' judgements of interpretations	152
6.5.1. Judgement of coherence	152
6.5.2. Judgement results vs. RST tree weight.....	154
6.5.3. Analysis of RST-awareness.....	155
6.6. Summary.....	156
Chapter 7 Conclusion	158
7.1. Synopsis.....	158
7.2. Limitations	160
7.3. Methodology review	161
7.4. Contribution.....	162
7.5. Future work	162
Appendix A Text of source speeches	164
Appendix A.1 Chinese Speech 1	164
Appendix A.2 Chinese Speech 2	166
Appendix A.3 Chinese Speech 3	168
Appendix A.4 English Speech 1	171
Appendix A.5 English Speech 2	172
Appendix A.6 English Speech 3	174
Appendix B RST annotations and RSTTool.....	176
Bibliography	177

List of Figures

Figure 1 Five-stage metacognitive model (Choi, 2004: 181)	17
Figure 2 AIIC quality criteria	18
Figure 3 SCIC quality criteria.....	19
Figure 4 Quality criteria shared by AIIC and SCIC	19
Figure 5 Quality expectations: Buhler (1986) vs. Kurz (1989)	20
Figure 6 EMCI Benchmark with further analysis and clarification	23
Figure 7 ETI assessment criteria.....	24
Figure 8 Examination marking criteria from MAITS.....	25
Figure 9 Evaluation sheet for CI examination (Riccardi, 2002: 124-5)	27
Figure 10 Ideal interpreting by Schjoldager (1996)	28
Figure 11 Sample of RST structure (Peng & Hartley, 2005)	42
Figure 12 Example of ANTITHESIS relation.....	45
Figure 13 Example of JUSTIFY relation	45
Figure 14 Contrasting discourse structure representation (Delin et al., 1994: 62) ..	46
Figure 15 Multilingual discourse structures representations (Delin et al., 1994: 63)	46
Figure 16 English Speech 2	47
Figure 17 Trainee's Chinese interpretation of.....	47
Figure 18 LNT0 performance criteria	56
Figure 19 Segment of LNT0 performance criteria.....	59
Figure 20 First step of feedback grid design.....	63
Figure 21 Experiment design for feedback collection.....	68
Figure 22 Trainees' self-assessment sheet for feedback collection.....	68
Figure 23 Trainers' assessment sheets for feedback collection	69
Figure 24 Professionals' comment sheet for feedback collection	69
Figure 25 Users' assessment sheet for feedback collection.....	70
Figure 26 The structure of Chinese Speech 2.....	89
Figure 27 RSTTool-Text segmenting in English	92
Figure 28 RSTTool-Text segmenting in Chinese	93

Figure 29 RST tree – text annotation	93
Figure 30 Example of RST annotation of English Speech 3	94
Figure 31 RST relations in the RSTTool	94
Figure 32 Export of Chinese RST structure.....	95
Figure 33 Statistics function in the RSTTool.....	96
Figure 34 Tops-English Speech 3	96
Figure 35 Examples of repairs (A).....	98
Figure 36 Example of empty markers	102
Figure 37 Example of recording overtly-marked RST relations.....	103
Figure 38 Coherence profile of English Speech 3.....	105
Figure 39 Chosen interpretations vs. group average.....	108
Figure 40 Trainee’s assessment sheet	109
Figure 41 E>C interpretations: RST tree weight vs. total words	113
Figure 42 C>E interpretations: RST tree weight vs. total words	113
Figure 43 RST bushes and trees.	115
Figure 44 C>E interpretations: RST relations vs. total words	116
Figure 45 Explicit signposting: Professional vs. trainee interpretations.....	116
Figure 46 Prof. C>E & E>C interpretations: RST tree weight vs. total words.....	117
Figure 47 Chinese and English speeches –RST tree weight vs. total words	118
Figure 48 C>E interpretations: total relations vs. RST tree weight.....	121
Figure 49 C>E interpretations: RST tree weight vs. total words – Professional interpretations as benchmark	123
Figure 50 E>C interpretations: RST tree weight vs. total words – Professional interpretations as benchmark	123
Figure 51 Overt ratios of C>E Interpretations	125
Figure 52 Overt ratios of E>C Interpretations	125
Figure 53 RST tree of double-conjunctions of ‘although...but’	129
Figure 54 Examples of repairs (B).....	131
Figure 55 E>C Interpretations: Repair vs. total RST relations	133
Figure 56 C>E Interpretations: Repair vs. total RST relations	133

Figure 57 Example of an error not repaired	135
Figure 58 E>C interpretations: Repair vs. RST tree weight	136
Figure 59 C>E interpretations: Repair vs. RST tree weight	136
Figure 60 Coherence profiles of professional English-Chinese interpretations	139
Figure 61 Coherence profiles of professional Chinese-English interpretations	140
Figure 62 Coherence profile deviations in professional interpretations	141
Figure 63 Coherence profiles of professional interpretations and Speech 2C	142
Figure 64 Coherence profiles of professional interpretations and Speech 3E	143
Figure 65 Coherence profiles of professional vs. trainee interpretations	144
Figure 66 Coherence profiles: Control group C>E interpretations.....	146
Figure 67 Individual coherence profiles of Control group (Speech 2C>E).....	147
Figure 68 Individual coherence profiles of Test group (Speech 2C>E).....	148
Figure 69 Coherence profiles of Test group.....	149
Figure 70 Development of coherence-English interpretations Speech 1 & 2.....	150
Figure 71 Coherence profiles: Control group E>C interpretations.....	151
Figure 72 Development of coherence-Chinese interpretations Speech 1 & 2	152
Figure 73 Trainee judgement on coherence.....	153
Figure 74 Trainee judgements on coherence attributes.....	154
Figure 75 Trainee judgement on coherence vs. RST tree weight	155

List of Tables

Table 1 Examples of explicitation in interpreting	39
Table 2 Definition of RST relation: Purpose (Mann and Thompson 1986: 64)	42
Table 3 RST: Nucleus vs. Satellite	42
Table 4 Definition of RST relation: Sequence (1986: 73).....	43
Table 5 RST multinuclear relations.....	43
Table 6 RST subject matter vs. presentational relations	44
Table 7 Questionnaire A for trainee’s awareness of interpretation quality	53
Table 8 Questionnaire B on trainee awareness of quality.....	54
Table 9 Prototype feedback grid for interpreter trainees (version 1).....	66
Table 10 Participation of subjects.....	67
Table 11 Professional participation as interpreters and judges.....	70
Table 12 Revised feedback grid (version 2)	79
Table 13 Inter -textual criteria (version 2)	80
Table 14 Inter-textual criteria for CI (version 3).....	81
Table 15 Feedback grid for CI (version 3).....	83
Table 16 Details of source speeches	88
Table 17 Details of the recording of interpretations	90
Table 18 Data annotation	91
Table 19 Definition of Repair	97
Table 20 English explicit markers from Halliday (1976: 242).....	100
Table 21 Chinese explicit markers from Chan et al. (2000: 13)	101
Table 22 Seven ratios as parameters of coherence profile with rationales.....	104
Table 23 Experiment setup for trainee judgement session	110
Table 24 Examples of dual-conjunctions from trainee English interpretations.....	128
Table 25 Occurrences of double-conjunctions in trainee interpretations	130
Table 26 ‘Repair’ vs. total relations in English-Chinese interpretations	132
Table 27 ‘Repair’ vs. total relations in Chinese-English interpretations	132
Table 28 Scores from trainee judgements	154

List of abbreviations and acronyms

AIIC	Association Internationale des Interprètes de Conférence. It was established in 1953 and now has more than 2000 members in 80 countries worldwide, with full recognition from the European Union, NATO, World Bank, United Nations and most other major international organizations.
A language	The interpreter's native language (or another language strictly equivalent to a native language), into which the interpreter works from all her or his other languages in both modes of interpretation, simultaneous and consecutive. (AIIC classification)
B language	A language other than the interpreter's native language, of which she or he has a perfect command and into which she or he works from one or more of her or his other languages. Some interpreters into a "B" language in only one of the two modes of interpretation. (AIIC classification)
CI	Consecutive interpreting
C > E	Chinese-English interpretations
E > C	English-Chinese interpretations
MAITS	MA Interpreting and Translation Studies at the University of Leeds
RST tree	Tree-like description of text resulting from RST annotation
SCIC	The Joint Interpreting and Conference Service (Service Commun Interprétation-Conférences) was established in April 1981. It provides interpretation in meetings arranged by the European Commission and the other Institutions it serves, and provides a conference organising capacity to Commission services Since October 2003, SCIC became the DG (Directorate General) Interpretation, and also known as DG SCIC. Available: http://europa.eu.int/comm/scic/thescic/mission_en.htm
SI	Simultaneous interpreting

Chapter 1

Training for Quality Interpreters and Performance

1.1. Conference interpreting explained

Viaggio, the former Chief Interpreter in the UN Vienna (1991-2005), points out, ‘interpretation is one of the oldest activities known to man; it has existed ever since two mutually unintelligible languages met’ (2003). Yet as a profession, conference interpreting is relatively young. Conference interpreting only started to attract public attention during the Nuremberg Trials (1945-46), where simultaneous interpreting (SI) was successfully used on a wide scale for the first time. Gaiba (1999) gives a comprehensive account of interpretations at the Nuremberg Trials. To do this, she used both judicial records and interviews with interpreters. She focused on practical arrangements for SI and its effect on the proceedings. Later in 1953, following the formation of the United Nations, where the need of SI was further demonstrated, the first international organization of professional conference interpreting, AIIC (Association Internationale des Interprètes de Conférence), was established. AIIC now has over 2500 members. The adoption of a code of ethics and professional standards in 1957 enabled AIIC successfully to regulate working conditions for interpreters and established a high profile for the profession worldwide. It has also played a significant role in the areas of training and research on important issues of the profession (Pöchhacker, 2004: 29). According to AIIC, ‘a conference interpreter is a professional language and communication expert who, at multilingual meetings, conveys the meaning of a speaker's message orally and in another language to listeners who would not otherwise understand’.¹

There are two major modes of work in conference interpreting: simultaneous and consecutive. In SI, interpreters sit in a sound-proof booth with, usually, a direct view onto the conference room. From there, they ‘listen to a speaker through earphones and simultaneously transmit the message in another language through a microphone to listeners in the room’.² In terms of the process they use to achieve this, interpreters need to listen to the speech, understand it, and translate it into another language - usually their native tongue. Significantly, all the while, they need to

¹ <http://www.aiic.net/en/prof/default.htm>

² http://www.aiic.net/en/prof/how/modes_interpretation.htm

monitor themselves to ensure the quality of the performance while at the same time processing the next part of the speech. It is thus clear that interpreters must exercise great concentration and work under constant pressure to produce accurate and reliable performance, covering a wide range of subjects and dealing with specialised terminology.³

In consecutive interpreting (CI), interpreters listen to a single intervention in its entirety, while taking notes. They then render the meaning of the message in another language. As CI does not need any technical support, such as sound-proof booths and microphones, it was widely used in international conferences. It was taken as ‘the standard medium of debate at the League of Nations, the UN’s ancestor, and continues to be widely used at small, bilingual meetings and ceremonial occasions’ (Setton, 1999: 1).

1.2. Interpreters are made not born

Herbert (1978) and Mackintosh (1999) both argue that professional interpreters are made not born, a stance which highlights the significance of training for interpreters. Over the last decade or so, a significant literature has emerged dedicated to addressing the needs of trainee interpreters, and the processes by which they learn. Given this, it is worth looking in some detail at both the amount of training which is available and the pedagogical paradigms in use.

In the late 20th century, with increasing communication among languages and cultures worldwide, the demand for interpreting services rose correspondingly and led to the establishment of interpreting schools. For example, to support EU enlargement from 2004 and the need for interpreters of additional ten official languages, many postgraduate interpreting programmes in the new member states were established, such as those in Ljubljana, Bratislava and Tallinn⁴. In addition, AIIC’s global survey of Conference Interpreter training programmes in 2004 involved at least 178 interpreting schools.⁵

³ [http://www.unog.ch/80256EE60057CB67/\(httpPages\)/2C87D748E41A2E3880256EF800496BF2?OpenDocument](http://www.unog.ch/80256EE60057CB67/(httpPages)/2C87D748E41A2E3880256EF800496BF2?OpenDocument)

⁴ http://ec.europa.eu/comm/scic/enlargement/genintro_en.htm

⁵ ‘Conference Interpreter Training Programmes: 2005 Survey’. Available: http://www.aiic.net/ViewPage.cfm?page_id=372

The number of postgraduate Chinese-English conference interpreter training programmes has also increased significantly in recent years. In 1997 there were two in the UK (Bath and Newcastle) and two in Taiwan (Fujen Catholic and National Taiwan Normal Universities). In addition to these, three more were established in the UK (Leeds, Heriot-Watt and Salford), and two more in Taiwan also (Chang-Jung Christian University and National Changhua University of Education). In China, the Chinese Interpreter Training Programme supported by SCIC has existed since 1985. By 2004 it had trained more than 300 young diplomats and officials from Chinese ministries and institutions⁶. SCIC has also provided specialist assistance to the University of International Business and Economics in Beijing. With the support from the Chinese Ministries of Education, Foreign Affairs and Commerce, the Sino-EU Interpreter Training Centre, UIBE was established in 2001 to train more professional conference interpreters in order to address the increased need for conference interpreters as a result of China's booming economy. In 2003, China also set up its first 'key' interpreting school at postgraduate level, Graduate Institute of Interpretation and Translation, at Shanghai International Studies University⁷. In short, the rapid increase in number of interpreting schools and programmes around the world over the past decade reflects the increasing need for professional interpreting services.

However, it is important to note that this quantitative increase in training programmes is matched by an increase in awareness of quality in training methods. In the spring of 2006, the AIIC Training Committee published its Best Practice for training conference interpreters⁸.

In training interpreters, as elsewhere in higher education, the aim is to prepare people as reflective practitioners for future professional employment (Aktins et al., 1993). To become a reflective expert, one has to go through a learning cycle including stages of cognition, association and finally autonomy with constant quality practice over time (Anderson, 1995). Training interpreters is no exception.

Firstly, it is vital for trainee interpreters to know what interpreting is and what kinds of skills they must acquire for quality performance to come; secondly, they should relate their awareness of quality in relation to their own actual performances.

⁶ <http://www.delchn.ccc.eu.int/en/Co-operation/Interpreters%20Training.doc>

⁷ <http://giit.shisu.edu.cn/index.htm>

⁸ 'Conference Interpreting Training Programmes: Best Practice'. Available: <http://www.aiic.net/ViewPage.cfm/article27.htm> [Access: 12 May 2006]

Later, with constant practice, either on their own or with their peers during training, they will perform well enough to pass the exams and earn their qualifications. After several years of professional experience, they will finally become experts.

Moser-Mercer (1997) investigates the skill components of professional interpreters. Two later studies explore the many differences between experts and trainees (Ericsson, 2000; Moser-Mercer et al., 2000). However, the stage in which trainees are acquiring an understanding of the reflective features of expertise has not received much attention. The development of awareness of quality performance for trainees is therefore one of the major issues that this thesis aims to address.

Many believe that conference interpreting, a set of specialised skills, should be taught only by experts. For example, the AIIC Training Committee suggests that both CI and SI should be taught by practising conference interpreters, 'since they provide not only knowledge but also know-how essential to acquiring professional expertise'.⁹ Setton (1994) compares interpreting to driving, arguing that just as someone who wants to learn how to drive would go to an experienced driver, trainee interpreters should seek an experienced interpreter as trainer. Thus, in most major interpreting training programmes, practising interpreters are recruited as expert trainers. There is no denying that the expert trainer-centred approach is important for the acquisition of certain skills, especially at a professional level. In such an approach, the trainees have a role model in front of them and a reliable source to consult for practical tips when facing problems.

Moser-Mercer (2000: 340) also uses a driving analogy to illustrate a different point. Like driving, she puts interpreting under the category of procedural knowledge, whose mastery involves complex steps and a lot of practice, as such, she recommends a pedagogical approach which she calls over-learning (Moser-Mercer, 2003). Trainees rely mainly on group practice to achieve over-learning. Thus collaboration among peers becomes vital to the advancement of their interpreting skills and performance.

Andrew Chesterman remarked in a PhD colloquium¹⁰, quality might be discussed from at least two perspectives; 'essential' and 'relational'. In 'essential' terms, the quality of interpretation can be determined by how accurately the

⁹ 'Conference Interpreting Training Programmes: Best Practice'. Available: <http://www.aiic.net/ViewPage.cfm/article27.htm> [Access: 12 May 2006]

¹⁰ PhD Colloquium: 'Research training in Translation Studies: sharing good practice'. May 20 2003, Aston University, Birmingham, UK

information in the source speech is interpreted and delivered. Yet the quality of customer service is also determined by customer satisfaction, which is relational in that, as long as customers do not complain, the service is deemed to be acceptable.

Pearl (1999: 4) states that SI is subject to many contextual variables and different client expectations. It is therefore impossible to achieve total client satisfaction. He suggests that, 'the proper training and education of interpreters is only half, or at best, five eighths of the battle. The other three eighths of the battle have been lost by default' (ibid). In these terms, the function of training of interpreters is to enable them to secure the 'five eighths' of the battle. One area to focus on is achieving this is the provision of clear criteria and guidelines for trainees to follow in order to help each other by giving structured peer feedback.

1.3. Practice makes perfect

Some people might assume that feedback from trainees themselves is not reliable. Others might expect feedback from expert trainers to be more reliable because it is based on solid experience. However, as we will see, expert trainers may lack the ability to share their expertise. Moreover, there are ways in which peer feedback can be helpful. In this section, I will outline some of the problems associated with feedback from expert trainers. I will also explain some benefits of peer feedback, as well as highlighting limitations. Finally I will suggest a way of overcoming these limitations to ensure the peer feedback is given in a structured and informed manner.

Currently, many if not most interpreter training programmes still apply a trainer-centred approach where expert trainers, as the source of expertise and authority, play the major role in judging and assessing trainees' performance. For example, at least 48 schools worldwide which meet the set of criteria by AIIC are practicing trainer-centred teaching, as one of the criteria is that interpretation classes taught by practising conference interpreters.¹¹ Yet the acquisition of interpreting skills by trainees requires not only good professional guidance during classes, but also extensive practice outside these hours (Ericsson et al., 1993; Ericsson, 2000; Moser-Mercer, 2003). In reality, therefore, trainee conference interpreters rely heavily on group practice and feedback from peers – targeting both language

¹¹ 'Conference Interpreting Training Programmes: Best Practice'. Available : <http://www.aiic.net/ViewPage.cfm/article27.htm> [Access: 12 May 2006]

proficiency and communicative competence – to advance their interpreting skills and performance (Hartley et al., 2004).

As Sawyer states, ‘professional judgment alone is an insufficient basis for decision-making’ (2004: 104). As Ficchi suggests, it might be that ‘their teaching lacks a theoretic or systematic basis’ (Ficchi, 1999: 202). There are two ways in particular in which such a lack might affect the effectiveness of expert trainers.

Firstly, expert trainers may find themselves ‘at a loss to account for student performance and to explain students’ difficulties’ (Moser-Mercer, 200:339). They may not have the ‘appropriate meta-language to describe students’ performance’ (ibid). A lack of the necessary meta-language can lead to impressionistic or unhelpfully vague criticism, perhaps claiming, for instance, that an interpretation failed to carry the message, despite the presence of most of the original information. Of course, even when the expert trainer has access to such a meta-language, it is essential that this is effectively shared with the trainees in order to allow them to benefit from comments.

Secondly, expert trainers often evaluate student performances and diagnose problems from the ‘vantage point of their own interpreting practice’ (ibid.; Ficchi, 1999: 202 makes a similar observation). The comparisons which are likely to come from such a vantage point may not be very fair to trainees, as they are not yet fully-fledged professional interpreters and their performance is not ready to be evaluated by professional standards.

Such problems, in part at least, account for the intimidation felt by some trainees when in the presence of the trainer.

Moser-Mercer suggests that the ideal situation would be to have ‘master interpreters’ with both extensive experience and the ability to impart relevant knowledge. Such people are hard to find and, even if found, are likely to be very much in demand.

Perhaps a more practical alternative approach is to encourage students to rely on feedback from their peers. Their peers are by definition in the same boat and are likely to share similar anxieties. Mutual support may overcome problems and frustrations. Peers can share strategies for tackling the common difficulties they face when learning interpreting skills. In short, to maximise the learning experience and result of conference interpreting, I believe in addition to trainers’ guidance, it is important to address the significance of trainee collaboration.

Although it may seem accessible, peer feedback may also lack consistency, organisation and reliability because the fellow trainees do not have a set of agreed criteria and clear guidance to follow. During group practice outside class, trainees normally comment on local problems such as completeness and omission of information, distortion of messages, accuracy of figures and names, to name just a few. As such, unguided peer feedback might fail to address more global concerns. Some can be more critical than others and the comments might not always be properly justified. Under such circumstances, a sense of rivalry might arise within the group. Instead of a supposedly collaborative learning context, group practice is very likely to become a battle field where trainees compete rather than cooperate.

On the other hand, when trainees practice interpreting on their own privately, they may lack knowledge of concepts and standards needed evaluate their own performances. Where they are aware of criteria such as accuracy and coherence, different trainees may understand them differently. In fact, we found this to be the case in evidence presented in Chapter 3 on the development of trainee's awareness of quality and their capacity of describing interpreting performances. It is often observed that they go back to the source speech, or turn to reference materials to figure out issues that they did not manage well in the first run. In the end, they might simply feel bored and give themselves a vague and subjective judgement that 'I messed up' or 'I guess I did all right'. Practices like these, in the end, become a waste of learning opportunities, as trainees cannot reflect on their performances and review their strategies sufficiently thoroughly.

Undoubtedly, interpreting, like many other professions, needs a lot of practice to achieve expertise. As Moser-Mercer has observed, trainees often spend hours every day practising, hoping that they will make good progress (2003). They think that the more they practise, the more they will progress, 'as if they were collecting flight mileage'. Yet when they keep practising without taking a moment to reflect on their performances, they waste their effort and lose the opportunity to identify space for further improvement. In other words, we can plausibly claim that practice does not naturally make perfect, if the quality of practice is not considered.

To make the time and effort invested in practice worthwhile, it is necessary to give trainees a reliable tool as guidance to follow in both their own individual and group practices. To this end, I worked with other colleagues to develop a useable tool, which incorporates criteria from both professional and training institutions in a systematic and user-friendly way, in the hope of enhancing the effectiveness and efficiency of learning results for trainee interpreters. More details of the design process and the use of the tool can be found in Chapter 4.

1.4. Making sense: gold standard for interpreting

When discussing quality standards in translation and interpreting, it is often suggested that identifying a gold standard is problematic. Secara (2005: 39) refers to authoritative studies by Pym and Sager when concluding that ‘there is still no universally accepted evaluation model in the translation world...no generally accepted objective criteria for evaluating the quality of both translations and interpreting performance’. There can always be more than one acceptable translation. This is equally true of interpretation. Like translation, interpreting facilitates communication between speakers of different languages in order to achieve different goals, such as to inform, to instruct, to negotiate, and so on. The dialogic nature of interpreting and the fact that the response is synchronous distinguish it from translation. Expectations of quality interpretations may vary due to contextual factors.

I take Halliday’s definition of text: ‘any passage, spoken or written, of whatever length, that [...] form[s] a unified whole’ (Halliday and Hasan 1976: 1). Accordingly, interpretations, like translations, are viewed as texts. In interpreting, the speech is the source text. Therefore, as Ahn points out, in Hallidayan terms, in order for the interpretation, a target text, to be considered as text, it must be coherent (2005: 699). Ahn also points out, ‘although there has not been very much research on coherence related to interpreting and translation, it is clear that coherence is an important element related to such areas as quality of interpreting and translation, characteristics of TT (target text), analysis of difference between ST (source text) and TT, and the training of interpreters and translators’ (ibid).

Elsewhere, in relation to translation, Reiss and Vermeer (1984) proposed a theory for the concept of translatorial activity. Their rule of intertextual coherence specifies that the target text must be comprehensible and sufficiently coherent for the intended receivers (1984: 119). The limited amount of the literature therefore suggests that coherence as the most important criterion which can be used to judge the quality of translations and interpretations.

This is also true of practice. In conference interpreting, to ensure quality of conference interpreting, both professional organisations and training bodies have published lists of quality standards for their members and trainees to follow. Our search of current literature (see Chapter 2) shows that ‘making sense’, or ‘sense consistency’, is also one of the most frequently proposed attributes to be considered when evaluating the quality of interpretations. Bühler’s (1986) survey was probably the first field study on quality in SI. She recruited 47 AIIC interpreters to take part in her questionnaire-based studies to rate those criteria. She identified 16 linguistic

criteria as specific factors affecting the quality of SI. The results showed that, 'sense consistency with original message' and 'logical cohesion of utterances' were perceived as the most important when evaluating conference interpretation.

In training, coherence is also one of the major quality attributes for trainees to work on. Trainees might be able to preserve most of the information they receive from the speech, but might not be able to deliver the message coherently. The following excerpt from a trainee's Chinese-English interpretation¹² shows the problem of coherence. The use of the pronouns 'it', 'they', 'we' and 'your' is very likely to cause confusion. The lack of signposting before 'Before' (e.g. 'for example') also disconnects the final sentence from its previous statement.

*...In China, we have the same problem as well.
And most of the illegal immigrants are coming from Korea,
coastal areas and also Africa, African countries.
It pose a serious problem
because it has very a bad influence upon societies.
So in China a series of measures have been adopted.
Before they leaving China,
we take some measures to tell the forged document and whether your visa
or passport forged...*

According to the marking criteria for both CI and SI in MA Interpreting and Translation Studies (MAITS) at Leeds University, interpretations must be 'mostly coherent as discourse' to pass the exam.

I have demonstrated that, despite the fact that there can be many acceptable versions of translations and interpretations, the research literature, professional bodies and training institutions all use coherence as a gold standard for acceptability. Since coherence is an important attribute of quality interpretations, it is thus a vital indicator of progress for trainees. Yet to my knowledge, previously there has not been a suitable framework in interpreting studies to describe how coherence is displayed in interpretations and how trainees progress in this aspect.

I propose that Rhetorical Structure Theory (RST) (Mann & Thompson, 1986) is a suitably rich discourse structure framework to describe coherence in interpretations. RST has been widely used for describing the hierarchical organisation of natural texts in terms of some 30 functional relations holding between text chunks,

¹² See Appendix CD: Data Annotation/Ctrl Spch2/CE/C3

thereby characterising the coherence of the whole text. It has also proven to be useful in describing the structure of spoken discourse (Tappe & Schilder, 1998). The set of RST function relations is open, allowing people to add new ones to describe their texts better. I introduced two new relations: Coda to mark the end of speeches, and Repair to account for the common phenomenon of self-correction in interpretations (5.2.2.1). In section 2.4., I will present a more detailed introduction of RST and its particular relevance to this thesis. My results show that RST is a useful framework to use in comparing coherence across different interpretations. RST annotations enable us easily to visualise the improvement of performance over time by trainee interpreters and the difference between professional and trainee interpreters.

1.5. Consecutive interpreting

As a mode of conference interpreting, CI does not always enjoy as much attention as SI in the research literature or in the view of the public. Some state that this mode of interpreting is not as fashionable as SI, yet others believe that it is still widely used and in demand. Gile confirms that CI is disappearing from the Western European market, but still very much alive in Asia and in Eastern Europe, ‘due to its distinct advantages over simultaneous (less costly, less cumbersome in terms of equipment, more flexible over time and space)’ (2001b). For example, in Chinese-speaking markets, CI is much used in business meetings. In the UK, Chinese-English CI is also often requested in high-level dialogues between the two governments.

In training, it is agreed that the training of CI is essential, and many regard it as a prelude to SI. CI is also involved as a prerequisite by international organisations such as SCIC and the United Nations for young interpreters to start their professional career in conference interpreting. In other words, it is a professional threshold that interpreters must cross. In short, the ability to perform CI successfully by giving coherent interpretations demonstrates that the interpreter has the essential ability to engage in interpreting in most work settings.

Moreover, CI also appears to be more suitable than SI in order to observe how coherence is displayed in interpretations. According to Hatim and Mason, specifically in relation to CI, the coherence and structure of the rendition are of great significance: ‘effective CI output shows a clear outline of the way a text is structured’ (2002: 262). They go on to explain that if you want to assess the quality of CI, the structure gives the best indication.

In addition, as Gile points out (2005: 133), CI is a powerful ‘diagnostic tool’. The interpretations given by trainees demonstrate their comprehension of the logic of

the speech as well as their mastery of the target language. This diagnostic aspect of CI explains why it is regarded as a must in most interpreter training programmes.

Finally, in practical terms, it should be noted that the majority of trainee subjects discussed in this thesis were MAITS students with Chinese as their A language and English as B. Since the market demand for CI in Chinese-English is still strong, it is important to address the significance of coherence in their performance in this mode.

1.6. Statement of purpose

In summary, considering all the discussion above, I identified four major goals of this thesis.

1. To explore the basis for judgement about quality for conference interpreting (Chapter 2)
2. To abstract and organise systematically the performance criteria for conference interpreter training (Chapter 4)
3. To establish a framework to capture coherence of conference interpreting in such a way that we can make comparative and qualitative judgements about interpretations by professional and trainees (Chapter 6)
4. To investigate the development of awareness of these criteria in trainees (Chapter 3) and its impact on their judgement of their peers and on their own performances (Chapter 6)

First of all, it is vital to know how quality is currently viewed in different areas of conference interpreting by collecting and reviewing standards judging interpreting quality in professional practice. Secondly, to raise trainees' awareness of those quality issues and facilitate the development of their interpreting skills especially in terms of reflections and evaluation, it is important to provide them with explicit criteria as guidance. These are based on my exploration of the literature and consultation with professional interpreters. RST appears to provide a suitable framework to capture the development of coherence as the vital attribute of quality. It thus enables me to make qualitative and comparative judgements about the interpretations by professional and trainee interpreters. Finally, it is important to investigate the impact of raised awareness of quality criteria on trainees' judgement

and their performances. These experiments carried out in order to achieve and monitor this are described in detail in Chapter 5.

Chapter 2

State of the Art

I set out here to outline the state of the art in conference interpreting and issues which are directly relevant to the problem statement presented in Chapter 1. To begin with, since I focus on CI, I will address the status and importance of this mode of interpreting in both training and professional fields. I will also explore the elements which contribute to successful CI (2.1). In addition, I review current criteria that are used in both professional and training organisations for judging the quality of interpreting (2.2). I will briefly explain how my review of criteria was used to inform the design of an evaluation tool, which is the focus of section 2.2.4. With evidence and research on interpreting quality (Bühler 1986 and Kurz 1993), I will consider what constitutes ‘making sense’ (Seleskovitch, 1978b), one of the most widely used criteria in conference interpreting (2.3). As crucial factors which contribute linguistically to the formation of meaning, cohesion and coherence will be discussed in sections 2.3.1 and 2.3.2. I will then introduce Rhetorical Structure Theory (RST), a framework which I used to observe and explore the differences between trainee and professional interpretations (2.4).

2.1. Consecutive interpreting

Within conference interpreting research, as well as in the view of the public, CI does not always enjoy as much attention as SI. Specifically, research on the quality of interpreting has mainly focused on SI, rather than on CI (Kalina, 2002: 122). Nevertheless, CI is in constant demand in some parts of the professional world and also plays a crucial role in conference interpreter training.

2.1.1. Consecutive interpreting as a profession

In professional practice, CI crosses both fields of conference and community interpreting, and the need for this mode of service varies across different markets. In conference settings, some consider that CI has been largely replaced by SI, and is seen as a second best being used when SI is not feasible (e.g. due to technical failure).

Kalina points out that the use of CI is changing professionally. She observes that increasingly, presentations in the mass media tend to be interpreted consecutively. In addition, in some very formal events, in order to be able to monitor the immediate effect of their speeches, speakers prefer CI done phrase by phrase, rather than as ‘extended formal consecutives (with turns of 15 to 20 minutes or more)’ (2002: 174).

Elsewhere it is reported that traditional CI is still widely used and very much in demand. In SCIC, for instance, CI still ‘remains relevant for certain kinds of meetings (e.g. highly technical meetings, working lunches, small groups, field trips)’¹³. From my own experience, I have observed that, in Chinese-speaking markets, CI is much used in various settings, such as in press conferences, where there is a need for local journalists to be able to interact with the invited speaker. In the UK, Chinese-English CI is often requested in high-level dialogues, technical discussions and policy negotiations between the two governments.

Different markets also have different expectations of CI, in terms of language directions, for instance. As Gile points out, some ‘purists’ in the West, ‘demand that interpreters only work into their A-languages in SI. Yet interpreters of certain languages such as Chinese and Arabic in the UN need to work both in and out from their mother tongues. Likewise, interpreters of some new member states in the European Union are expected to do the same. In the marketplace, a large proportion of the interpreters work both ways’ (2005: 141).

In East Asian countries, such as China, Taiwan, Hong Kong, and Japan, the same expectation is widespread in the market. Moreover, with particular regard to Taiwan, Setton notes that the market expects interpreters to work both ways, ‘be fully “bi-active”, i.e. able to work accurately, and produce acceptable grammar, vocabulary, register, etc. in two languages, both in CI and SI’ (1994a). Setton further explains that, ‘the virtual absence of Western language native speakers as potential interpreters means that the present European model of task distribution will not be generally viable in the foreseeable future’ (Setton, 1994b: 58).

In the private market in the West, interpreters of Spanish, German and East European languages are often required to work both ways, too. Therefore, working both ways is a requirement for interpreters of certain language combinations, such as Chinese-English in this study. In order to support this professional reality, trainees are trained to work into both Chinese and English. As we shall see, this practice has yielded interesting results regarding textual coherence in the interpretations.

We shall see in the following section, CI is often taken as a prerequisite for young interpreters to start their professional career in conference interpreting in most of the international organisations, such as in SCIC, UN, etc. Indeed, CI is taken into consideration by most training programmes as a threshold which must be crossed in order to be recognised as a professional.

¹³ ‘What is consecutive interpreting?’ Available:
http://europa.eu.int/comm/scic/interpreting/tech_cons

2.1.2. Consecutive interpreting in training

In conference interpreting training, there is consensus that CI is essential. As Setton points out, it is important to teach students ‘hand-to-hand-combat’ for equipment breakdowns and other emergencies (1994b: 63). Other pedagogical motivations for teaching CI are given in the following paragraph. Moreover, trainee interpreters need to be trained to work in both ways for CI and they are expected to pass the degree examination in both directions in most of the training programmes, too. Gile supports the view that working both ways should not be problematic for CI (2005: 141).

As was noted previously, some regard CI as a prelude to SI. For example, Seleskovitch of the Paris School firmly believes that students should have complete mastery of CI before receiving training on SI, as the skills of CI could be transferred to SI if fully mastered (Seleskovitch 1978a). The practice in ETI (École de Traduction et d'Interprétation) in Geneva is that students need to pass the CI test in order to receive training in SI. Therefore, if a student fails the CI exam, he or she is judged to be ‘unsuitable’ for the profession of conference interpreting.

Others, such as Gile, however, consider CI as ‘the “highest” form of interpreting, above SI, essentially because it requires the comprehension phase to be completed before the formulation phase’ (2005:132). Gile goes on to explain that ‘the short time lag between perception and production in SI allows production from verbal traces, whereas in CI, production is done from traces of the content.’ Therefore, SI is partly possible at word-identification level, without deeper comprehension, whereas this does not work in CI (ibid). Setton also points to the significance of CI in developing analytical and information handling skills (1994b: 33). Of course, SI and CI differ in other respects. For example, in CI one thinks back in order to organise one’s rendition, while in SI one is constantly anticipating the upcoming discourse.

Interpreter training programmes, such as the one in Trieste University in Italy, acknowledge that CI involves a distinct set of skills which should be acquired separately. In such programmes, CI is taught in parallel with SI training.

Gile (2005: 133) believes that the significance of CI is that ‘it fosters analysis and reformulation’. It is a powerful ‘diagnostic tool’ which highlights strengths and weaknesses, especially in terms of comprehension and target language ability.

During the initial stage of CI training, many training programmes, including MAITS at Leeds University, engage trainees in memory exercises. Trainees need to work with their memory without note-taking. It is a good opportunity to demonstrate to trainees how their memory works. Gile suggests that ‘if they listen carefully and

understand the logic of the speech, its content will be stored in their memory even without a conscious effort to memorise it' (2005). To this end, trainee interpreters are constantly reminded of the significance of the logic of the speech and the links between messages.

Logical links remain important when trainees do take notes in CI. As Rozan points out, 'an idea can be distorted completely if its relation to the previous idea is not clearly indicated. When taking notes, then we should never miss out the links' (Rozan, 2003).

2.1.3. Consecutive interpreting and autonomous learning

In addition to the trainer-centred approaches adopted in most training schools, Ficchi (1999: 205) believes that it is important to introduce autonomous learning for CI. Autonomous learning has been defined as 'the ability of take responsibility for one's learning' (Victori & Lockhart, 1995: 223). However, taking such responsibility can have several benefits which will be discussed in this section.

As was noted above, Ficchi (1999) has advocated autonomous learning in CI training. He attempts to introduce autonomous trainees to improve their CI skills, such as listening, note-taking and message delivery through practice. However he fails to provide any concrete criteria for trainees to evaluate their performances. The strategies suggested have more to do with improving language skills.

Similarly, Choi suggests it is vital to equip trainees with a 'metacognitive approach' to practice self-evaluating for their CI performance (2004). He assumes that 'if students were taught to self-evaluate, they could build up confidence by realizing that they have the potential to perform better tomorrow than today' (2004: 171). He adopts the Interpretive Theory (Seleskovitch and Lederer, 1986) as the basis for trainees to practice self-evaluation.

According to the Interpretive Theory, there are three phases of the interpreting process: comprehension, deverbalization, and reformulation. Firstly, interpreters need to understand the intended message of the source speech. To achieve this, interpreters need to use both linguistic and extralinguistic knowledge. In the phase of deverbalization, interpreters need to separate the message itself from the linguistic packaging. After that, interpreters need to reformulate the message into the target language.

In Choi's study, trainees were asked to identify their problems from the three phases. Choi proposes a five-stage metacognitive model as follows:

- 1) *Stage 1: self-evaluation by student (S) / feedback by teacher (T)*
- 2) *Stage 2: problem-finding (S, T) / student profiling (T)*
- 3) *Stage 3: prioritization (S, T)*
- 4) *Stage 4: practice (S)*
- 5) *Stage 5: reevaluating (S)/ monitor (T)*

Figure 1 Five-stage metacognitive model (Choi, 2004: 181)

He claims that 'the metacognitive evaluation process can serve as a useful guideline for students to strive towards self-development and for teachers to provide more meaningful feedback' (Choi, 2004: 183). The reported disadvantage of this model is that it might not be effective for 'less-proactive' students, who would be more dependent on teachers' feedback (ibid). Yet I suspect that the lack of a set of agreed criteria for the evaluation of performance might be another source of weakness. Without criteria and a common metalanguage, trainees who share similar problems might describe them differently (see detailed discussion on trainees' awareness of quality in Chapter 3). In turn, the inability to identify common problems might lead to missed opportunities for sharing strategies for handling them.

2.2. Research on interpreting quality assurance

Despite the fact that there are criteria for judging interpreting performances proposed by professional bodies, it is often observed that many professionals and trainers still rely on their intuitive and subjective judgements to decide whether a candidate interpreter is competent enough for the profession (Campbell & Hale, 2003). The ultimate aim of eliciting criteria for trainee conference interpreters to follow is to help them form better understandings of those criteria and develop better awareness of the quality of their performances. It is hoped that by doing so they will acquire better interpreting skills and produce quality performances. Thus it is important to review how quality assurance is practised and judged in both professional and training contexts.

2.2.1. Professional standards

In the professional world, the quality of conference interpreting is determined by both subjective and objective factors. Objective factors such as working conditions, environments and issues such as professional ethics are regulated by

professional associations (AIIC, 1990; 1991). Both AIIC and SCIC set their own criteria for the quality of interpreters' performance. To be admitted by AIIC as a member, an applicant needs to seek sponsorship from existing members who have worked with and evaluated him or her in real-life settings for quality performance.¹⁴ The Admissions Committee of AIIC uses Bühler's criteria (1986) to judge the performance of its applicants. These are listed below.

- 1) *Sense consistency with original message*
- 2) *Logical cohesion of utterance*
- 3) *Correct grammatical usage*
- 4) *Completeness of interpretation*
- 5) *Fluency of delivery*
- 6) *Native accent*
- 7) *Pleasant voice*

Figure 2 AIIC quality criteria

Some of these criteria are not sufficiently clear. Criterion number 7) *pleasant voice*, for instance, summarises a number of constituent factors, such as pitch, (tone), intensity (loudness), and timbre (quality), which remain unspecified and which are largely judged subjectively. Although 6) *Native accent* is specified in this set of criteria, research shows that this is not a major concern for users (Bühler, 1986). There is no denying that there are many advantages to having interpreters work into their mother tongues, yet in reality interpreters are often asked to work into another of their working languages. In such cases, the standard of native accent is not realistic. Nor is it vital. Criterion 1) *sense consistency with original message*, again lacks definition and is therefore difficult to attain and evaluate. Often the intended message of a discourse is not readily apparent from the surface. Even a very capable and experienced interpreter may convey only part of the many underlying messages in the interest of getting the main point across. In 5) *fluency of delivery*, both terms are vague. If someone can speak nonsense non-stop at a stable pace, does it make them fluent in delivery? All in all, the criteria suggested by AIIC Admissions Committee lack clear definition and proper organisation.

In addition to the ongoing Chinese Interpreter Training Programme described earlier, SCIC used to run in-house training, where novice interpreters worked with senior interpreters while being under their supervision. The interpretations given by the novices were evaluated against a set of quality standards. SCIC's in-house scheme terminated in 1997. According to consultation with some accredited SCIC

¹⁴ 'Applying to AIIC'. Available: <http://www.aiic.net/ViewPage.cfm/article118.htm>

interpreters, who were trained under that scheme, criteria similar to those used by AIIC were used, though they were described differently. In judging SI, the following criteria were applied:

- 1) *Rigour and consistency*
- 2) *Faithfulness to original (substance and style)*
- 3) *Quality of communication with audience*
- 4) *Calm, regular delivery*
- 5) *Avoid literal/word for word translation*
- 6) *Correct, spontaneous use of TL*

Figure 3 SCIC quality criteria

In SCIC's standards, 1) *Rigour and consistency* appears to be an attribute of the interpreting performance alone. As such, it is different from the *sense consistency with original message* suggested in AIIC's criteria. When discussing sense consistency with the original message, SCIC uses 2) *faithfulness of substance and style*. In addition, SCIC's 4) *delivery* specifies pace and regularity, rather than involving fluency.

SCIC also sets out its criteria to judge the quality of CI. These are similar to those for SI. In addition to the first three criteria used for SI, CI will also be judged by 'clarity and elegance of expression', as well as 'speed and fluency'.

Having compared and contrasted the standards adopted by both professional organizations, it is apparent that they share common ground when judging their candidates. For instance, both AIIC and SCIC judge performance into the interpreter's active language (their mother tongue) by the following criteria:

- 1) *Content (accuracy/faithfulness)*
- 2) *Structure (consistency/cohesion)*
- 3) *Delivery (fluency)*
- 4) *Language expression (vocabulary, grammar and style)*

Figure 4 Quality criteria shared by AIIC and SCIC

Clearly, giving a successful interpreting performance requires more than linguistic competence alone. According to AIIC's 'Advice to Students Wishing to Become Conference Interpreters'¹⁵, excellent language skills and a broad knowledge base are prerequisites for anyone intending to train as a conference interpreter. Besides linguistic performance, both organisations acknowledge the significance of knowledge (general and subject matter), communication skills (public speaking,

¹⁵ 'Advice to Students Wishing to Become Conference Interpreters'. Available: <http://www.aiic.net/ViewPage.cfm/article25.htm> [Access: 12 May 2006]

communication with audience, tact and diplomacy) and personality (concentration, persistence and pressure-resistance) of the candidate when recruiting new members.

2.2.1.1. User expectations

Of course, within the professional world of conference interpreting, subjective factors affect perceptions of the quality of conference interpreting. These include expectations, backgrounds and roles of participants (delegates, audience, organisers and the interpreters) in a communicative situation.

Moreover, interpreting should be viewed ‘within a conceptual spectrum from international (conference) to intra-social (community) spheres of interaction’ (Pöchhacker, 2001), expectations of quality are likely to vary along this spectrum.

Where there are people, there are variables. Such subjective factors have been investigated by a series of scholars (Bühler, 1986; Kahane, 2000; Kurz, 1993; Moser, 1996) who attempted to explore perceptions of conference interpreting quality from various perspectives.

In 1986, Bühler conducted the first field study into user expectations. Bühler identified fifteen linguistic and extralinguistic factors which affect the quality of SI and asked AIIC interpreters to score them for importance. Bühler concluded that ‘these criteria reflect the requirements of the user as well as fellow interpreters in a (hopefully) well-balanced mixture’ (Bühler 1986 in Kurz, 2001: 398). As Kahane points out, Bühler’s set of criteria ‘have the virtue of the being the first and in addition have been used in subsequent studies thus enabling a degree of comparability’ (ibid, 2000).

In 1989, Kurz used eight of Bühler’s criteria with 47 delegates in a medical conference. Kahane (ibid) compared the two studies and compiled the following table which provides the percentage of participants for each sample who considered each of the eight criteria to be important.

	Bühler 1986 Interpreters %	Kurz 1989 Users %
Sense consistency with original message	96	81
Logical cohesion of utterance	83	72
Correct grammatical usage	49	45
Completeness of interpretation	47	36
Fluency of delivery	49	28
Correct grammatical usage	48	11
Native accent	23	11
Pleasant voice	28	17

Figure 5 Quality expectations: Buhler (1986) vs. Kurz (1989)

Both studies confirm the view that *Sense consistency with original message* and *Logical cohesion* are the two criteria which are valued most highly by both interpreters and users. *Accent* and *voice*, by comparison, do not seem as vital. It also seems that interpreters value all the criteria more than users do, apparently they are more concerned about the quality of their work than are their users. The criteria were later used again by Kurz in 1993 to test another two distinct user groups, an international conference on quality control (n=29) and during a Council of Europe meeting (n=48). Kurz's studies (1989 and 1993) on user expectations showed that among all the groups, *Sense consistency*, *Logical cohesion* and *Correct terminology* were regarded as more important than other criteria (Kurz, 2001: 398).

Moser (1996) conducted another large scale survey, commissioned by AIIC, to investigate quality based on the judgements, needs and expectations of users of conference interpreting. This brought together evaluations produced by a wide range of users. He discussed the degree to which respondents' individual characteristics may have influenced their responses to or expectations of, interpreter performance. From the above studies, it seems reasonable to conclude that, as in other service industries, criteria for judging quality vary according to the perceptions, expectations and attitudes of the end users.

Other researchers have argued that users are not good judges of quality, because they were not in a position to perform the task (Collados Ais, 2002; Shlesinger et al., 1997). Shlesinger et al. (1997) reasoned that since listeners (users) could not understand the source message, they lack the most crucial means of assessing quality. Consequently, smooth delivery may create a false impression of high quality, when much of the message may in fact be distorted or even missing. Interestingly, AIIC actually also encourages its members to pay attention to voice and delivery, because 'less able, less accurate colleagues have been preferred because of a pleasant voice and reassuring delivery' and conversely, without special attention to voice and delivery, professional performance might be mistakenly under-rated¹⁶.

In brief, although quality criteria in the professional practice are not readily applicable for trainee interpreters, they should be integrated into the training criteria. Preparing trainee interpreters for the various demands arising from different real-world situations that they will face is undoubtedly one of the ultimate goals of the training.

¹⁶ 'Practical Guide for Professional Conference Interpreters'. Available: <http://www.aiic.net/ViewPage.cfm/article21.htm> [Access: 04 June 2003]

2.2.2. Educational standards

Criteria are used to judge both prospective students on application to a training programme and also the performance of successful applicants throughout training. The following discussion will focus on the latter use, but it is worthwhile considering the former, in particular the use of aptitude tests as an aid to selection, briefly first. Many studies also emphasise the use of aptitude tests in recruiting students for training programmes, in the hope of selecting the right candidates and maximising training results (Harris, 1992; Lotriet, 2002; Moser-Mercer et al., 2000; Russo & Salvador, 2004; Setton, 1994). Yet, according to some expert trainers, who also manage training programmes, some candidates strong at the beginning might not shine as expected at the end; while other, apparently weaker students, with suitable learning strategies and positive learning attitudes, will progress steadily over time and eventually stand out¹⁷. In other words, with the benefit of explicit guidance and effective learning approaches, including guidance for giving peer feedback and for peer collaboration, trainees who appear less able at initial screening might stand out in the end.

It may be that the diffuseness of the AIIC and SCIC guidelines discussed above reflects the difficulty of accounting for the widely varying situations in which professional conference interpreters work.

However, both organisations play vital roles in training conference interpreters. AIIC has developed a ten-point statement of quality criteria for training programmes¹⁸ and provides trainer training seminars to pass on good practice. SCIC, in addition to its pedagogical assistance to conference interpreter training programmes¹⁹, was also instrumental in drawing up the assessment criteria used by the European Masters in Conference Interpreting (EMCI). This programme was launched as a pilot project by the joint effort of SCIC and the European Parliament in 1997 to address the need for qualified conference interpreters of less common language-combinations which arose as a result of EU enlargement. The EMCI group initially comprised eight university-level institutions which drew up a core curriculum

¹⁷ Internal report on 2003 AIIC Trainer Training Seminar: Simultaneous Interpreting Training and its Applications by B. Moser-Mercer in Porto, Portugal.

¹⁸ 'Conference Interpreting Training Programmes: Best Practice'. Available: <http://www.aiic.net/ViewPage.cfm/article27.htm> [Access: 12 May 2006]

¹⁹ 'Pedagogical Assistance from DG SCIC'. Available: http://ec.europa.eu/comm/scic/interpreter/ass_ped_en.htm [Access: 12 May 2006]

for interpreter training at postgraduate level. In the core curriculum, it specifies the assessment criteria for both CI and SI:²⁰

For CI:

At the end of the programme students will be capable of giving a fluent and effective consecutive interpretation of a speech lasting at least 10 minutes into the target language, accurately reproducing the content of the original and using appropriate terminology and register.

For SI:

At the end of the programme students will be able to provide a fluent and effective simultaneous interpretation of speeches of at least 20 minutes into the target language, accurately reproducing the content of the original and using appropriate terminology and register.

Clearly the criteria for both modes of interpreting are very similar. Indeed, the only difference is the time of the input speech. Personal experience of years as an interpreter trainer suggests that the criteria listed above are too vague for trainee interpreters to follow. I have therefore analysed the components of the EMCI criteria used in final exams. Details are given in Figure 6.

EMCI Final Exam Benchmark ²¹		My attempt of further analysis and clarification	
<u>Content</u>	accuracy/fidelity	source text vs. target text	→ observable in output
	coherence/logical links	target text as a whole	
	cultural comprehension, general knowledge		→ inferable from output (cognitive resources and processes)
	linguistic comprehension		→ observable in output (accuracy and fidelity)
<u>Form</u>	concision, clarity grammar and usage appropriate vocabulary style, register		→ linguistic attributes (phonetic, grammatical, lexical, semantic) observable in output
	delivery		→ fluency or presentation skills?
<u>Skills</u>	communication		→ function of the output, judged by the end users
	analysis, reasoning, problem-solving		→ inferable from output, yet not observable

Figure 6 EMCI Benchmark with further analysis and clarification

²⁰ <http://www.emcinterpreting.net/whatis.htm> [Access: 12 May 2003]

²¹ <http://www.emcinterpreting.net/curriculum.htm> [Access: 12 May 2003]

This set of criteria appears to be more structured than the criteria for CI and SI given in the above discussion, yet each criterion listed for the assessment is not clearly defined and can easily cause confusion. For instance, in terms of form, both style and register are listed, but the difference is not explained. Moreover, criteria are not ranked or weighted in an overall assessment of quality. How can one judge whether the performance has achieved the function of communication? Are the same criteria applicable for both SI and CI? It is evident that work is required to restructure the benchmarks systematically so that they will be explicit enough for trainees to follow. First, I will review the criteria in use at other training institutes.

Among the conference interpreting training schools worldwide, ESIT (Ecole Supérieure d'Interprètes et de Traducteurs) in Paris is the most prestigious and has a unified, structured and detailed training doctrine. The training methodology in SCIC is essentially based on this Paris school. ESIT was established in 1957 and its clear pedagogical instructions are laid down in the text book (Seleskovitch & Lederer, 2002). Quality standards are elaborated in description, which are comprehensive, but too lengthy to be readily adopted as a set of criteria.

In the final exam, for example, trainees' interpretations are judged on three aspects: 1) knowledge of languages; 2) Interpreting skills; and 3) Some specific mistakes (ibid: 307). None of these aspects is elaborated in a formal manner. These criteria leave much room for subjective judgement to form, and also trainees might find them too vague to follow. Moreover, important skills are not explicated. For example, based on Seleskovitch's famous 'théorie du sens' (1986), 'deverbalization' is an important interpreting skill, which trainees need to demonstrate in their exam performance.

Another well-known interpreter school, the School of Translation and Interpretation (ETI) in Geneva has been influential since its establishment in 1941. Figure 4 shows the criteria²² which appear on a grid developed by ETI that both teachers and students can use to assess SI in class and in practice.

- *Content (accuracy, faithfulness, completeness, terminology, makes sense)*
- *Language (appropriate, natural, correct, lexis varied, register firm)*
- *Voice (pitch, timbre, accent, lively intonation)*
- *Delivery (smooth, regular, articulation)*
- *Mic. use (distance, noise in microphone)*

Figure 7 ETI assessment criteria

²² Source: provided by Isabelle Perez at Heriot-Watt University

ETI's list appears to be more structured and more explicit than that of EMCI, in not only itemising the content of each category, but also covering technical issues like *microphone use* which are easy to follow and observe for trainees when they work together. Yet scope for confusion remains. To what extent are the components independent? Can a rendition be faithful and complete but not accurate? Also what is the difference between being appropriate and using a 'firm' register?

For MAITS at Leeds, we have in-house marking criteria (Figure 8) which are used by both internal and external examiners. They are also included in students' hand-book²³.

To achieve 70% or higher (first class performance), a student's interpretation should:

- *show a very high degree of reliability in relaying meaning*
- *be entirely coherent as discourse*
- *show command of appropriate TL expression*
- *achieve a standard of presentation which demonstrates mastery of the skills involved in keeping pace and addressing an audience*

To achieve 50% (the pass mark), a student's interpretation should:

- *relay meaning without systematic distortion and without major unwarranted omissions or additions*
- *be mostly coherent as discourse*
- *achieve a standard of TL expression which does not impede communication to a significant extent*
- *achieve a standard of presentation which shows some evidence of ability to keep pace and address an audience*

Figure 8 Examination marking criteria from MAITS

The above assessment criteria remind examiners of certain aspects such as reliability of meaning (source text vs. target text), coherence (target text), target text expression (target text) and the effect of communication (target text) when giving judgements. However, what lies under the headings remains implicit. For trainees, the criteria are supposed to make clear the different standards that they will reach by giving different performances, but the difference between the best performance (70%) and the level close to failure (50%) is not clearly defined from the trainee's point of view.

Riccardi (2002) introduces the macrocriteria and the microcriteria to evaluate interpretations in the University of Trieste, Italy (SSLMIT). She adopts the four

²³ MAITS Handbook 2005-2006. Appendix M. Marking Criteria. University of Leeds.

quality objectives proposed by Maurizio Viezzi (1996) as the macrocriteria, which are equivalence, accuracy, appropriateness and usability. She suggests that these criteria enable people to evaluate interpretations from different perspectives. 'Equivalence and accuracy examine the relationship between ST and TT, whereas appropriateness and usability examine the relationship between TT and the audience within a specific communicative event' (Riccardi, 2002: 119). This set of criteria is intended to cover as many aspects of communication as possible, such as the 'communicative intention of the speaker, the characteristics and needs of the target listener and...the restrictions imposed by the situation and by the text-typology' (ibid: 118). This might serve as a sound framework for trainee interpreters to reflect on their performance in the long run, yet in the training context, they are not so concerned by such contextual factors.

The microcriteria, against which the trainees' CI and SI performance is evaluated by the teachers in SSLMIT, might be more helpful in training. These criteria mainly focus on interpreting errors and are largely drawn from the author's personal experience as both interpreter and interpreter trainer (ibid: 121). The microcriteria are presented in the form of two evaluation sheets: one for CI and the other for SI. Both modes of interpreting share fifteen linguistic criteria and interpreting elements. Two additional criteria, 'eye contact' and 'posture', are for specifically for CI, and 'incomplete sentences' is only used for SI. The evaluation sheet is designed to cover the most frequent types of deviation and indicates areas for improvement by trainees. Figure 9 gives the Evaluation sheet for CI examination.

Evaluation sheet for CI examination

consecutive from L1 into L2

consecutive from L2 into L1

phonological deviations

none some many

prosody deviations

none some many

production deviations

none some many

pauses (> 3 sec.)

none some many

eye contact

none occasional frequent

hand control and/or gesticulation and/or posture

none satisfactory good

lexical deviations (common)

words none some many

collocations none some many

lexical deviations (technical)

words none some many

collocations none some many

morphosyntactic deviations

none some many

logical/semantic deviations

none some many

acceptable serious unacceptable

omissions

none/some few many

useful (coherence) negligible sense alteration

additions

none/some few many

useful (coherence) negligible intrusive

reformulation

good satisfactory poor

calques

none some many

register

maintained modified altered

technique

good satisfactory poor

successful solutions

none/some few many

overall performance

good satisfactory poor

Figure 9 Evaluation sheet for CI examination (Riccardi, 2002: 124-5)

The major disappointment of the SSLMIT evaluation sheet is the lack of any hierarchical organisation for the criteria. Indeed they seem to be presented in an arbitrary order. For example, it is not clear why ‘production deviation’ and ‘pauses’ are next to each other, or why ‘lexical deviations’ and ‘hand control’ are neighbouring criteria. In addition, the criteria presented in the evaluation sheet are described at once vaguely and technically, which might cause trainees or even trainers

difficulty in understanding and using them. For example, ‘production deviation’ can imply semantic errors; and the meaning of ‘morphosyntactic deviations’ and ‘calques’ might not be readily apparent.

In addition to sets of criteria used in institutions, certain individuals have also made efforts in this regard. Schjoldager (1996) produced a feedback sheet for trainees to judge their own SI performances, as well as those of their peers, and for trainers to assess trainees’ performance in class. She endeavoured to make her criteria as explicit as possible and suggested that in ‘ideal interpreting’:

1. *The listener can understand what the interpreter says and can bear to listen to him/her.*
2. *The interpreter's language is adequate* → linguistically and extralinguistically?
3. *The interpreter's rendition is coherent and plausible.*
4. *The interpreter is a loyal communicator of the speaker's message.*

Figure 10 Ideal interpreting by Schjoldager (1996)

In Schjoldager’s feedback sheet, both strengths and weaknesses of performance are included to help trainees become confident and more skilled at giving and receiving criticism. Moreover, explications of assessment criteria are provided to help students better understand criteria couched in technical terms. For instance the explication of coherence reads, ‘if an interpreter’s performance lacks coherence, the listener loses interest in the message’ (ibid). This explication is over-simplified, as listeners might lose interest for other reasons. Incoherence might cause more confusion than the mere loss of interest.

In sum, I have identified weaknesses in all of the sets of criteria reviewed in this section. However, this material provides a solid foundation, as well as the motivation, for the development of a new feedback tool for trainees to use themselves. This corresponds to my second research objective.

2.2.3. Linguistically-informed standards

In addition to criteria applied in training settings, researchers in linguistic disciplines such as pragmatics, text analysis and discourse analysis have offered perspectives and approaches for judging the quality of conference interpreting.

To begin with, the pragmatic approach to defining quality covers a broad range of issues to consider when discussing the quality of interpreting. Kopczynski (1994) believes that in addition to equivalence (semantic level), congruence (semantic and grammatical levels), and correspondence (closest translation overriding differences between two languages), pragmatic issues such as the goal of communication (the

special effect produced by a speaker's speech) in a situated speech (i.e. the expectations/backgrounds of hearers) should not be ignored. The quality of interpreting in this field is explicitly related to communicative goal, which is contextually determined.

Hatim and Mason (2002) suggest text linguistics as a framework for discussing the product of interpreting. In defining textuality, this approach posits the three vital 'basic domains of textuality': texture, structure, and context (ibid: 255). They suggest that the three domains can be applied to three basic modes of interpreting; SI, CI and liaison interpreting respectively.

SI relies heavily on texture, and there are 'various devices used in establishing continuity of sense and thus making a sequence of sentences operational (i.e. both cohesive and coherent).' (ibid: 255). Hatim and Mason suggest that the use of devices such as 'anaphoric and cataphoric reference, substitution, ellipsis, conjunction, lexical cohesion' help to hang texts together (ibid: 259). In CI, awareness of structure helps the interpreter to 'perceive specific compositional plans', to outline a text, and further assists us to 'flesh out the details' (ibid: 255).

Liaison interpreting is also known as 'ad hoc' or 'public service' interpreting in the UK, 'dialogue' and 'three-cornered' interpreting in Austria, and is also often called 'community interpreting'. It is 'performed in two language directions by the same person' and 'widely used when the interpreter must be present in order to bridge the communication gap' (Gentile et al, 2001: 17). In liaison interpreting, contextual factors 'determine the way in which a given sequence of sentences serves a specific rhetorical purpose' (Hatim & Mason, 2002: 255). These contextual clues tend to assume greater importance as long-term guides, since the input of text is not always complete in terms of texture and structure. Also, there are no reliable clues to the way the two-way interaction will develop and conclude.

Works on discourse analysis are also worth mentioning. Clifford (2001) suggests that the conventional lexico-semantic assessment of interpreters' performance has its limitations. When describing the full set of interpreting competencies, both content and context should be considered. In his *Performance-based Assessment Rubric for Interpreting*, competencies needed for interpreting are categorized into three major parts: deixis, modality and speech acts, and the performances are classified into three levels: basic, intermediate and advanced. In terms of deixis, interpreters should be able to interpret a passage with many voices by shifting roles effectively. In terms of modality, interpreters need to use prosodic effects, such as intonation, to show relationships between units of information. Last

but not least, interpreters need to organise their utterances into expected argumentative structures to meet the requirement of speech acts. Thus, if interpreters achieve the ‘advanced’ level for all of the three aspects, they should be able clearly to differentiate all actors in the discourse, use prosody effectively to indicate addition of information, and use argumentative structures which are expected by the target language culture.

In the pragmatic, discourse and text analysis approaches, I observe that interpretations (target texts) are discussed independently as well as in relation to end-users. In other words, the correspondence between source text and target text seems to be of less concern. The research completed by Shlesinger et al (1997), however, fills in this gap. They again stress that in order to define the concept of quality, two major issues need to be considered: identity of the evaluator (i.e. quality for whom) and the methods of assessment. When evaluating the target text, they also argue that we should consider the performance in three dimensions: ‘intertextually, intratextually and instrumentally’ (ibid: 128). To evaluate the target text ‘intertextually’, we should compare and contrast the source text and the target text for both agreement and deviation. To evaluate the target text ‘intratextually’, one should evaluate the target text as a whole according to its acoustic, linguistic and logical features. To evaluate the target text ‘instrumentally’, we need to confirm the usefulness and comprehensibility of the target text. For instance, going too fast will mean that the interpretations is poor in instrumental terms, no matter how good it might be intertextually and intratextually. All three aspects are very relevant to my research design, and informed the development of my integrated model of criteria.

2.2.4. An integrated model

My review of the relevant literature on quality assurance in conference interpreting has revealed that, in the professional world, SI has been discussed more. In training settings, and other studies of quality of interpreting performance, both SI and CI modes are explored. Moreover, while most criteria used to judge performance are product-oriented, in order to design criteria which are useful as guidelines for trainees, I need to transfer them into a process-oriented approach. Gile (2001) points out that, interpreting quality assessment for trainees is essentially different from professional assessment, due to its guiding function. Training should guide trainees to progress step by step to acquire interpreting skills and overcome psychological barriers. Initially, in the early stages of training, trainees should be assessed according to their progress. Later, it is necessary to move toward product-oriented assessment by judging the quality of the performance given. This will maximise the value of guidance from trainers and help trainees to optimise their performances.

The review of current criteria in training institutions or for training purposes shows that coherence appears to be one of the most prominent criteria in operation. As we have seen, different terms are used in different institutions: ‘deverbalization’ at ESIT, ‘make sense’ at ETI Geneva, ‘coherent’ in the MAITS programme at Leeds, ‘semantic/logic deviation’, ‘omission’ and ‘addition’ at SSLMIT, and ‘coherent and plausible’ by Schjoldager.

Given its ubiquity in sets of quality criteria for evaluating interpretations, the development of coherence is taken as the principal quality feature that this thesis aims to describe and investigate further (Research statement 3). Secondly, I aim to locate coherence and other quality criteria discussed so far within a hierarchical structure. This structure and the relationships between the various criteria will be discussed in Chapter 4. The process of consultation and validation of the hierarchy will also be presented.

The feedback tool I propose for conference interpreting trainees will be based on a hierarchy of criteria that trainees can follow easily. The review of current criteria used to judge the performance of conference interpreting has revealed some scope for improvement. In addition to the technical language used when discussing criteria, the lack of organisation and clarification of each quality attributes contributes to the confusion of trainee interpreters. Thus, in Chapter 4, I propose an integrated model of feedback criteria. The integrated model categorises as comprehensive a set of quality criteria as possible within a hierarchical structure. It aims to provide trainee interpreters with a tool for formative, rather than a summative, evaluations. It does this by facilitating the exchange of useful feedback about interpreting performances. I will also explain how the proposed feedback grid was created, validated, and piloted with trainee interpreters. Finally, the process of refining and revising the feedback grid will be introduced.

2.3. Making sense

As I argued in Section 1.4, ‘making sense’ appears to the gold standard for interpreting in general. ‘Making sense’ is important both at the point where the interpreter receives the speech and at the point where the audience receives the target text produced by the interpreter. Pöchhacker describes the major steps of the interpreter’s work as: ‘understanding (“making sense of”) what had been expressed in a source language, and expressing the ideas grasped, i.e. the “message”, in another language so that they would “make sense” to the target audience’ (2004: 56).

In comprehending texts, psycholinguists claim that after reading a long text, people can mostly remember the gist but leave out details. The gist being remembered is called 'macrostructure' (Van Dijk and Kintsch, 1983, cited in Greene and Coulson, 1995: 44). This 'macrostructure' corresponds with the 'sense' that Seleskovitch proposed in her interpretive theory. Seleskovitch (1977: 28) states that interpreting is 'a conversion from source language to sense,' and once one has grasped the 'intermediate link', the sense can be reproduced in any language, regardless of the words used in the source speech. This intermediate link is the deverbalization mentioned above.

It is important to note, however, that language users often have different prior knowledge and expectations of the topic under discussion and therefore end up with different interpretations of a text. As Green and Coulson explain, 'they are continually making inferences in an effort to make sense of what they hear and read, based on their general knowledge and expectations' (1995: 45). In trainee interpretations, it is often observed that trainees produce interpretations which 'make sense' internally, but that this sense is totally different from that of the source speech. This might be largely attributed to the result of a lack of shared knowledge and expectations between them and the speaker. In order to minimise this phenomenon, when selecting speeches for trainees to interpret for subsequent analysis, I deliberately chose topics which would not require special background knowledge (see Section 5.1.2.1).

Having said this, it is essential to acknowledge that texts do not cohere by chance. Relations link the various parts of a text together to achieve overall coherence. Shlesinger observes that a text hangs together by a 'network of relations which establish links between its various parts; these links, or cohesive ties, enable the reader or hearer to process the text in a coherent way' (Shlesinger, 1995: 193). Thus links in the source text undoubtedly provide interpreters with clues to construct the interpretation.

Dam (1998) has demonstrated out that CI is not exclusively meaning-based, or totally 'deverbalised' in Seleskovitch's terminology. Dam collected five professional consecutive interpretations from Spanish into Danish. She then segmented the interpretations into manageable units in order to compare the degree of lexical similarity (form-based interpreting) and lexical dissimilarity (meaning-based interpreting) between the speech and the interpretations. Her results show that only 11% of the target text segments were constructed without any traces of the linguistic form of the source speech. Moreover, her data shows more lexical similarity than lexical dissimilarity (1998: 64). She concludes that, 'form-based and meaning-based

interpreting appears to go hand in hand' and that 'target text production in consecutive interpreting is based on a constant alternation between verbal and non-verbal source text features' (1998: 65).

From this it would seem that some traces of the linguistic form of the source speech, such as discourse markers and cohesive links, are used as clues by interpreters to comprehend the speech on the one hand. On the other hand, those linguistic clues might be re-used as signposts to facilitate users' comprehension of their interpretation. Successful CI, in particular, as Hatim and Mason rightly suggest, should show 'a clear outline of the way a text is structured' (2002: 262).

Of course, interpretation should 'make sense' to the listeners. If an interpretation can not be comprehended by its audience, it fails to facilitate communication between two languages. Many studies suggest that one way to decide the quality of a text is to see how easy it is for readers or listeners to comprehend the intended message (Beaugrande & Dressler, 1981; Scott and Souza, 1990; Shlesinger, 1995).

Beaugrande & Dressler (1981) suggest that a text, whether oral or printed, should serve as a communicative discourse, communicating between the intentions of the speaker and the needs of the listeners. In other words, if the text is not comprehensible for the listeners, it does not fulfil its communicative function. With regard to translation and interpreting, Shlesinger similarly emphasises that, 'successful translation, after all, will depend on whether target text recipients can achieve second-degree interpretation with minimal extra processing effort' (1995: 209).

Coherence plays an important role in making text comprehensible. Comprehending interpreted texts is no exception: the more structured the text is, the easier it will be for the listener to follow it. Beaugrande and Dressler (1981) propose cohesion and coherence as the two most significant of seven standards by which the communicative value of a text can be measured. Similarly, Scott and Souza explain that, 'the more structured the input is, the easier it will be for the reader to derive its underlying message' (1990: 53). Sanders and Noordman also believe that people need coherence to understand a text (2000: 37). Just as successful comprehension is necessary for a coherent representation of the input text, it is reasonable to claim that a coherent underlying representation of the evolving output text is a condition of successful production.

It also follows that a text with low perceived coherence will be difficult to comprehend, since it will lack the cues enabling the construction of a coherent

representation by the listener. For example, Ficchi observes that the discourse of poor CI performances by trainee interpreters can be confusing and imprecise, 'lacking coherence and cohesion', with sentences not linked but juxtaposed (1999: 202).

2.3.1. Coherence

In addition to the importance of coherence to text comprehension addressed above, I will review the treatment of coherence in the literature from other perspectives in this section.

Some regard coherence as an internal mental phenomenon in both text production and comprehension (Gernsbacher & Givón 1995; Sanford and Moxey 1995), while others see coherence as the result of the interaction between texts and text users (de Beaugrande & Dressler, 1981).

Gernsbacher and Givón see coherence as 'a property of what emerges during speech production and comprehension - the mentally represented text, and in particular the mental processes that partake in constructing that mental representation' (1995: vii). Sanford and Moxey suggest that coherence is the 'result of the interpretation and integration of interpreted text elements by the listener (reader) in relation to the intentions of a speaker (writer)' (1995: 181), and psychologically, people tend to produce a 'coherent mental representation' of the text when they try to comprehend it (ibid: 183).

On the other hand, de Beaugrande and Dressler claim that coherence concerns how the concepts and the relations underlying a text are 'mutually accessible and relevant' (1981:3-7). They also address that to understand a text, people make inferences based on their knowledge and expectations (ibid: 4).

However, without language features, these cognitive representations would not be communicable. Lexical, semantic and syntactic features are therefore needed, so that the interaction between knowledge can take place.

Gernsbacher & Givón suggest that, to construct coherence, lexical knowledge and 'grammatical processing cues' are vital in achieving both 'local and global coherence links' (1995: viii). Hobbs (1979) suggests that coherence markers such as anaphora are normally considered as clues to coherence (Hobbs 1979 cited in Sanford & Moxey, 1995: 163). Sanders and Noordman also suggest that coherence relations can be made explicit by the use of linguistic markers (2000: 38).

In other words, coherence, no matter whether it comes from cognitive representation of a text or the interaction between text and text users, relies on linguistic features to display in a text.

Research also shows that ‘cognitive relationship such as contrast, equivalence, cause and consequence, and temporal sequence, which present and organise information in a logical manner’ are vital for writers to construct paragraphs and readers to understand text structures (Higgins et al., 1999: 347).

Logical presentation of such relationships is of equal importance in interpretations, in order that the target audience can understand the text structure of the interpretation.

Reinhart believes that a coherent (‘ideal’) text needs to be ‘connected’, that the clauses of a text should be formally connected, ‘in which adjacent pair is either referentially linked, or linked by a semantic connector’. Also it needs to be logically consistent with the previous sentence, and sentences need to be ‘relevant’ to both the discourse topic and to the context of the utterance (Reinhart, 1980 cited in Sanford & Moxey, 1995: 162).

Reinhart’s study appears to support something often observed in interpreter training sessions, in which trainees are often reminded to give ‘connected’ and ‘logical’ interpretations and to use ‘linking words’, so that their interpretations make sense.

In ‘The Basic Principles of Consecutive Interpreting’, it is clearly stated that special attention needs to be given those relationships in order to analyse the ‘links’ (Jones, 1998).

‘Ideas may be linked by logical consequences, logical causes, put together without cause-effect relations, and may also be expressed by a series of opposing concepts. In the first and second case, the interpreter will devote special attention to the connectors used - e.g. therefore, so, consequently, as a result, due to, owing to, as, since, because - whereas in the case of sequential ideas the interpreter should not abuse the word and, thereby avoiding the risk of stylistically impoverishing the translation’. (Jones, 1998)

All in all, the description of coherence in interpreting by Blum-Kulka is found to be suitable, and will serve as the definition used in the following discussion:

‘...coherence can be viewed as a covert potential meaning relationship among parts of a text, made overt by the reader or listener through the process of interpretation’ (Blum-Kulka, 2000: 299).

2.3.2. Cohesion

Cohesion, like coherence, relates to the comparability of texts in so far as it is used to realise relations between elements of meaning in language. Halliday and Hasan (1976) define cohesion as is ‘relations of meaning that exist within the text, and that define it as a text’ (ibid: 4). It is also ‘the set of semantic resources that exists for linking a sentence with what has gone before’ (ibid: 10). Cohesion relates sentences to each other. When the sequence of sentences in a text is disturbed, the meaning of the text is very likely to be changed or destroyed (ibid: 28). Moreover, as Cawsey points out, without cohesion, a text will become ‘unnatural’ and ‘inefficient’. In addition, ‘incorrect discourse structures (and intentions) may be inferred’ (1990: 76).

Shlesinger regards cohesion as ‘the network of relations which allow us to interpret a text by providing links between its various elements’ (1995: 193). Blum-Kulka defines cohesion as ‘an overt relationship holding between parts of the text, expressed by language specific markers’ (2000: 299). Thus cohesion provides continuity. She goes on to explain that, the choice of cohesive markers can affect the texture, style and even meaning of a text (ibid: 302).

Indeed, the use of cohesive markers, variously known as ‘connectives’ (Crystal, 1991), ‘cue phrases’ (Knott & Dale, 1992) and ‘linguistic markers’ (Sanders & Noordman 2000), is significant in many ways.

First of all, they link different parts of texts together. Crystal defines the function of connectives as to connect linguistic units at any level (1991: 74).

Secondly, cohesive markers also serve as signposts for people to work towards a coherent mental representation of a text (Gernsbacher & Givón, 1995; Graesser et al., 1997; Noordman & Vonk, 1997; Sanders & Noordman, 2000). Empirical evidence demonstrates that ‘cue phrases’ help readers to construct a mental representation of a text (Knott & Dale, 1992: 13). Caron believes that connectives provide ‘procedural instructions’ when people construct semantic representations (1997: 70).

Thirdly, the use of cohesive markers helps to make the coherence relations between text segments explicit (Sanders & Noordman, 2000: 45). The coherence

relations can be seen as the building blocks of a coherent text. When the coherence relations of a text are made explicit with linguistic markers, the text becomes easier for people to comprehend. Haberlandt (1982) demonstrated that when a sentence is led by a linguistic marker which makes the relation with the preceding text explicit, the reading time of that sentence is speeded up (Haberlandt, 1982 cited in Knott & Sanders, 1998: 138). Research results by Millis and Just (1994) also suggest that the presence of connectives influences the representation immediately after reading, that some linguistic markers 'give rise to...faster and more accurate reactions and a probe task, to faster and more accurate responses to comprehension questions' (Sanders & Noordman, 2000: 42).

Clearly, with cohesive devices such as conjunctions, anaphoric and cataphoric references as signposts, readers can follow the relationships between units of discourse more easily (Higgins et al., 1999). Conjunctions, for instance, connect text parts which take place in succession (Halliday & Hasan 1976: 227), and can signal readers or listeners to relate the segments which precede and follow the conjunction (Shlesinger, 1995: 203).

It follows that, in interpreting, the presence of cohesive markers also helps to decide whether the interpretation can be efficiently understood by the audience. If cohesive devices which are present in the from the source speech are missing from the corresponding interpretation, it is likely that, while elements of meaning are presented, the overall meaning is distorted and hard for the audience to understand (Shlesinger, 1995: 212). However, the presence of coherence markers also makes a significant contribution by help helping interpreters to grasp the gist of the speech.

It has been shown that after reading an explicitly structured text, readers can easily reproduce the structure of the original text (Meyer et al., 1980 in Sanders & Noordman, 2000: 41). However, some researchers found that there are no such effects when participants 'just listened to passages' (Rickards et al., 1997 in Sanders & Noordman, 2000: 41). They found, however, if participants are forced to process information more deeply in a task in which they had to take notes while listening, explicit marking does help them to recall the speech later (ibid).

Of course consecutive interpreters follow just such a process in their work. They need to listen to a speech attentively not only for messages, but also for the structure of the speech. They then process information either with or without notes. Finally they attempt to reproduce the speech in the target language coherently.

In both translation and interpreting, 'explicitation' is a common textual feature of target texts. 'Regardless of the languages concerned, the interpreters tend to

render implicit forms more explicitly’ (Shlesinger, 1995: 210). Below are two interpretations of a part of a speech (Speech 3), which is used for data collection in this thesis. The speech is delivered in Chinese and interpreted into English by trainee and professional interpreters. In both interpretations, there are more explicit markers than in the source speech. In Example 1, a trainee’s interpretation, almost every chunk of information is opened by an explicit discourse marker. In the professional interpretation (Example 2), the use of ‘on the other hand’ appears to emphasise the contrast between two different viewpoints.

<p><u>Speech</u></p>	<p>不過有些人覺得這是沒有問題 這樣做不會有問題 是因為他們是所謂的既得利益者 以一些歐洲汽車 . 製造廠來講 , 他們同意在十年內 , 使新車的效能增加 25% 這些公司常常非常沾沾自喜地認為 他們對生態環保都盡了心力 我們應該要加入他們的行列嗎 我非常難以苟同 ! 他們這些作為 , 為的不過就是要增加自己車子的銷售量 (Speech 3)</p>
	<p>Literal translation: <i>But some people feel that this is not a problem there is no problem by doing so this is because that they are the so-called the benefited. Take some European car...manufacturers for example They agree that in ten years, they will increase car efficiency by 25%. These companies often congratulate themselves They have made contribution to environmental protection. Should we join them? I find it hard to agree with them. What they want to do is to increase their sales of cars.</i></p>
<p>Example 1</p>	<p><u>But</u> some people think there's no problem <u>because</u> they will get some profit from this. <u>For example</u> the vehicle makers, they would like to reduce 25% emissions, <u>and</u> they feel rather proud of this. <u>But</u> I can't agree with this, <u>because</u> they do so <u>since</u> they think they can increase their sales by doing this. (C1_3ce)</p>
<p>Example 2</p>	<p><u>But on the other hand</u> there are also those people <u>who</u> claim that it's not a big issue to reach such a target at all. <u>For instance</u> those car manufacturers in Europe. They claim that there won't be any problem at all to bring about such reduction in the CO2 emission. <u>However</u> they are talking about such reduction from a purely parochial interest from a purely parochial perspective (<u>repetition</u>) <u>i.e.</u> by making such a claim</p>

	it will help them to increase the car sales every year. (P2_3ce)
--	--

Table 1 Examples of explicitation in interpreting

As Blum-Kulka explains, successful translation requires complex text and discourse processing. Specially, ‘the process of interpreting performed by the translator on the source text might lead to a TL (target language) text which is more redundant than the SL (source language) text’, and more cohesive explicitness is observed in TL texts regardless of the difference between the two linguistic and textual systems involved (2000: 300). One way to achieve such explicitation is to mark the coherence relation with discourse markers and connectives. Indeed Niska (1999) observes that interpreters make ‘extensive’ use of cohesive devices to enhance coherence. As Ballester and Jimenez simply put it, conjunctions “convey relationships between ideas” (1992: 241). It is therefore not surprising that they are often used by interpreters to make explicit relations in their interpretations. Proper use of conjunctions, therefore, becomes one important step for trainee interpreters can take to realise the relationships between ideas of a speech. Without linguistic markers, one can still establish coherence relations, but it takes more time because the relation will have to come from the basis of the content of the clauses without being facilitated by the markers (Sanders & Noordman, 2000: 54).

To summarise, I have demonstrated the two-fold significance of ‘making sense’ for interpreters. On the one hand, they make use of the coherence relations of the speech to ‘make sense’ of the messages intended by the speakers. On the other hand, to facilitate listeners’ comprehension of their interpretation, interpreters mark the relations and outline the structure of the speech with linguistic markers as cohesive devices. In CI, in which the structure of discourse matters the most, however, there appears to be a lack of substantial research on coherence and cohesion. Apart from looking into the shifts of cohesive devices in translation by comparing the source text (speech) and the target text (interpretation), to my knowledge, no one has proposed a framework to represent textual coherence, which allows comparison of coherence in different valid interpretations.

2.4. Rhetorical Structure Theory (RST)

2.4.1. Why did I choose RST?

I explored fields other than interpreting and translation studies for a suitable framework with which to represent the coherence of interpretations and compare the coherence of different valid interpretations of a single speech.

According to Halliday and Hasan's definition, an interpretation qualifies as a text (1976) (see Section 1.4). In 'Cohesion in English' (1976), they provide a comprehensive description of cohesion in the language system. Yet it would be difficult if not impossible to use the description as a framework for comparing the coherence features of different interpretations. Shlesinger used the theory in a small-scale study of interpretations of a single speech by 13 trainees. However, it does not suit my needs. Halliday and Hasan's theory is descriptive and operates at local levels of cohesion rather than of the level of global coherence of the text. It lacks guidelines for implementation. Moreover, as it does not seek to represent text structure, it does not facilitate comparisons of the type I wish to make.

I also looked at theories proposed by text linguists, such as that of de Beaugrande and Dressler (1981). They suggest that a text should satisfy seven standards to be functional in a communicative situation. The seven standards are: cohesion, coherence, intentionality, acceptability, informativity, situationality and intertextuality. This framework describes the success of a text based on the result of the interaction between the speaker and the listener. However, this is too broad for the current study which seeks to maximise training benefits by focusing on key issues. Contextual factors are not central. Moreover, I have demonstrated the significance of coherence and cohesion in relation to making sense, these two factors appear as two of the seven standards in this framework.

On the other hand, Rhetorical Structure Theory (RST) allows me to describe how different parts of a text relate to each other in terms of function, and how they contribute to the overall coherence of the text. In addition, it enables me to represent the coherence relations between the parts of a text in a hierarchical structure which facilitates comparison. Mann and Taboada explain that 'RST is intended to describe texts, rather than the processes of creating or reading and understanding them. It posits a range of possibilities of structure – various sorts of building blocks which can be observed to occur in texts' (Mann & Taboada, 2005).

2.4.2. Origins of RST

RST was originally developed by William C. Mann at the Information Sciences Institute of the University of Southern California (CSC/ISI) and Sandra Thompson at the University of California at Santa Barbara during the mid-to-late 1980s. According to Bateman and Delin (2005), Mann was interested in looking for a textual coherence model explicit enough to drive automatic text generation and Thompson had been researching coherence and textual signals of discourse relations for years. Their

cooperation gave birth to RST, ‘A Theory of Text Organization’, as it was called to begin with (Mann & Thompson, 1986).

According to Mann & Thompson, RST is a framework which describes the relations between text parts in functional terms. RST analysis results in a hierarchical representation of the text. Significantly, texts of different sizes and types, including ill-formed speeches, can be analysed using RST.

Driven by research in natural language generation, RST became one of the most popular theories for describing the connectedness of written non-dialogue discourse (Hovy, 1998, Scott and de Souza, 1990, Moore and Swartout, 1991, Cawsey, 1991, McCoy and Cheng, 1991, Horacek, 1992, Hovy, 1993, Moore and Paris, 1993, Vander Linden and Martin, 1995 cited in Marcu, 2000: 19). More recently, RST has also proved useful when describing the structure of spoken discourse (Tappe & Schilder, 1998).

2.4.3. RST explained –RST relations and definitions

According to Mann and Thompson’s definition, RST relations ‘hold between two non-overlapping text spans, here called the nucleus and the satellite’ (1986: 4). Bateman and Delin further explain that ‘the importance of nuclear element is defined in terms of its contribution to the rhetorical goals of the text as a whole’ (2005: 3). Within RST, these rhetorical goals are to correspond with the intentions of the speaker (Bateman and Delin, 2005: 3).

A nuclear element cannot be removed from a text without damaging its coherence, whereas satellites can often be removed without compromising overall coherence (i.e. the text would still be perceived as attempting to fulfil the same broad communicative function) (ibid).

In other words, if the satellites are deleted from a text, it tends still to make sense, while a text from which the Nuclei have been deleted does not. Generally there are two kinds of rhetorical relations: ‘asymmetric relations, where one of the related rhetorical units is singled out as the rhetorical head, or nucleus, and symmetric relations, also termed multinuclear, where all of the related units are of equal status’ (Bateman and Delin, 2005: 2).

Figure 11 shows that the result of RST analysis can be represented as a tree-like structure of relations. The first two spans, ‘She picked up the phone’ and ‘She dialled the number’, form a multi-nuclear relation, while the third span ‘in order to call the airline’ is the Satellite.

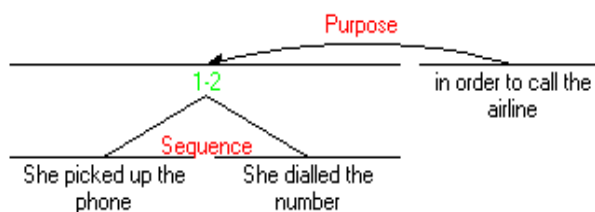


Figure 11 Sample of RST structure (Peng & Hartley, 2005)

Adjacent spans are linked by rhetorical relations. In the example above (Figure 11), the rhetorical relation holding between the first two spans is ‘Sequence’ and the relation holding between that unit and the Satellite is ‘Purpose’.

Mann and Thompson provide formal definitions for all the rhetorical relations. Each definition of an asymmetric relation covers four fields: 1) constraints on the Nucleus (N); 2) constraints on the Satellite (S); 3) constraints on the combination of Nucleus and Satellite (N+S) and 4) the Effect on the readers (R)

For instance, the relation ‘Purpose’, an asymmetric relation present in the example in Figure 11, is defined as follows:

Relation name: Purpose	
1)	Constraints on the N: presents an activity
2)	Constraints on the S: presents a situation that is unrealized
3)	Constraints on the N +S combination: S presents a situation to be realized through the activity in N
4)	The Effect: R recognizes that the activity in N is initiated in order to realize S

Table 2 Definition of RST relation: Purpose (Mann and Thompson 1986: 64)

Mann and Taboada’s website on RST (2005) provides a table which describes the relationships, between the Nucleus and the Satellite in most RST relations. To illustrate this relationship, I have reproduced several entries for RST relations (Antithesis, Background, Justify and Purpose) in Table 3.

Relation Name	Nucleus	Satellite
Antithesis	ideas favoured by the author	ideas disfavoured by the author
Background	text whose understanding is being facilitated	text for facilitating understanding
Justify	Text	information supporting the writer’s right to express the text
Purpose	an intended situation	The intent behind the situation

Table 3 RST: Nucleus vs. Satellite

If a relation does not have a single span of text which is central to the author's purposes, it is said to be Multinuclear (e.g. Sequence in the example in Figure 11). Multinuclear relations are also formally defined. As there is no Satellite in multinuclear relations, the definition does not cover constraints on the Satellite. The definition of Sequence given by Mann and Thompson is reproduced below in Table 4.

<p>Relation name: Sequence</p> <p>1) Constraints on the N: multi-nuclear</p> <p>2) Constraints on the combination of nuclei: A succession relationship between the situations is presented in the nuclei</p> <p>3) The Effect: R recognizes the succession relationships among the nuclei</p>

Table 4 Definition of RST relation: Sequence (1986: 73)

As with the asymmetric relations, the relationship between spans in multinuclear relations is described in a table on Mann and Taboada's website (2005). I have reproduced entries for the multinuclear relations used in my annotation in Table 5.

Relation Name	Span	Other Span
Sequence	an item	a next item
Contrast	one alternate	the other alternate
Joint	(unconstrained)	(unconstrained)
List	an item	a next item

Table 5 RST multinuclear relations

Mann and Thompson also suggest a taxonomy of the RST relations by labelling them as Subject Matter and Presentational relations (1986: 17). The effect on the reader distinguishes between the two groups of relations. Subject matter relations are 'those whose intended effect is that the reader *recognizes* the relation in question'. Presentational relations, on the other hand, are 'those whose intended effect is to *increase some inclination* in the reader, such as the desire to act or degree of the positive regard for, belief in, or acceptance of the nucleus' (ibid: 18). Table 6 presents the classification by Mann and Thompson (1986: 18).

Subject-matter relations	Presentational relations
Elaboration	Motivation (increases desire)
Circumstance	Antithesis (increases positive regard)
Solutionhood	Concession (increase positive regard)
Volitional cause	Background (increases ability)
Volitional Result	Enablement (increases ability)
Non-Volitional Cause	Evidence (increase belief)
Non-Volitional Result	Justify (increases acceptance)
Purpose	
Condition	
Otherwise	
Interpretation	
Evaluation	
Restatement	
Summary	
Sequence	
Contrast	

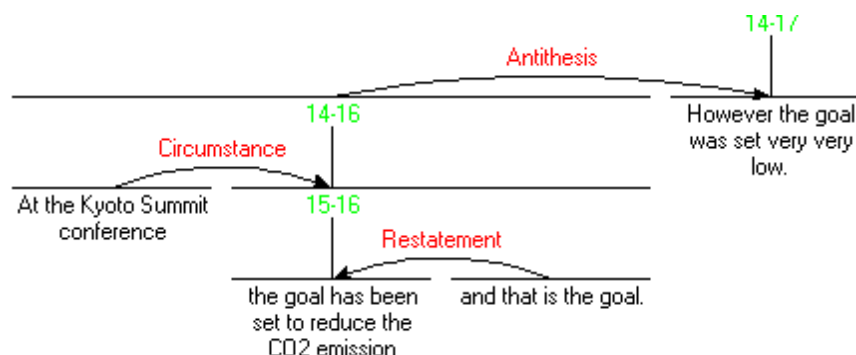
Table 6 RST subject matter vs. presentational relations

Mann and Thompson carried out ‘a detailed examination of the kinds of rhetorical relationships and corresponding rhetorical structures needed to carry out text analysis of texts of any kind’ (Bateman and Delin, 2005: 2). They collected about 25 relations, now known as ‘classical RST’. These relations are reported to be able to cover most of the relations in English texts (Hovy, 1990: 19). I found that the same relations also supported Chinese text.

Although Mann and Thompson explicitly stated this list of relations is open-ended, ‘it has in fact proved very stable over the years’ (Bateman & Delin, 2005: 2). I adopted this set of classical RST relations to data annotating, and added two of my own: Coda and Repair. A Coda is often used to mark the end of a conference speech, such as ‘thank you for your attention’. Repair is often observed in spoken texts, where speakers give up on a sentence halfway through and restart it straight afterwards. This is also true in interpretations and is observable in my data. Detailed discussion of this issue and a definition for this RST relation is presented in Section 5.2.2.1.

I will now present two examples of RST relations (Antithesis and Justify) used in interpretations collected as data to be used in this thesis. This illustrates the suitability of RST as a framework for representing the textual coherence of interpretations.

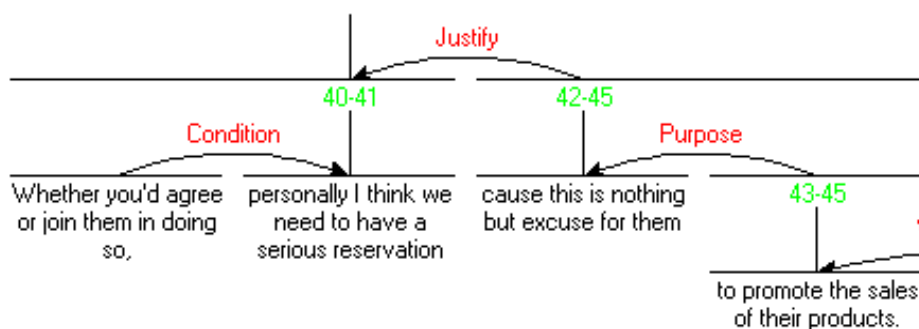
In Figure 12, the text span from segment 14 to 16 holds the idea disfavoured by the speaker and the text span containing segment 17, on the other hand, is favoured by the speaker. Therefore, the Nucleus of this text fragment is segment 17, and the span from segment 14 to 16 is the Satellite. We also see that the Nucleus is marked by ‘However’, an explicit cue which signals the ‘Antithesis’ relation.



[Appendix B: Data Annotation/Prof/CE/P1_3ce]

Figure 12 Example of ANTITHESIS relation

In Figure 13, the span containing segments 40-41 is an opinion presented by the speaker (Nucleus), and the span containing segments 42-45 supports the speaker’s opinion (Satellite). Again, it is clear that the ‘cause’ (because), another cohesive device, in segment 42 explicitly marks the relation of ‘Justify’.



[Appendix B: Data Annotation/Prof/CE/P1_3ce]

Figure 13 Example of JUSTIFY relation

2.4.4. RST analysis

According to Mann and Thompson (1986: 19), ‘an RST analysis always constitutes a plausible account of what the writer wanted to achieve with each part of the text. An RST analysis is thus a functional account of the text as a whole’. They also observe that ‘virtually every text has an RST analysis, as most texts are ‘hierarchically structured and functionally organised’ (ibid: 20). They recognise that it

is normal and predictable that a text has more than one RST analysis, and indeed, a single analyst may sometimes produces more than one analysis of the same text.

Delin et al. (1994) demonstrate that in a monolingual context, the same user instructions expressed differently (see Example 1 and 2 in Figure 14) can result in two different RST analyses (Figures reproduced from Delin et al., 1994: 61).

- (1) Pull down and remove the white plastic tray that holds the video cable and unpack the cable. (Apple) (2) Pull down and remove to unpack the video cable. (Apple)

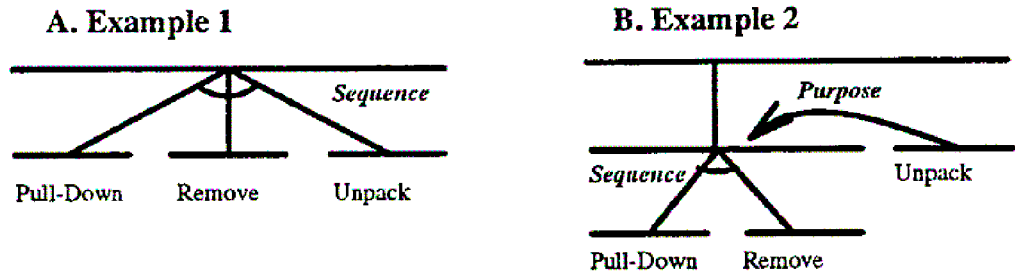


Figure 14 Contrasting discourse structure representation (Delin et al., 1994: 62)

They further demonstrate that such phenomena are very common in multilingual environments. They use an example from a trilingual instruction manual, in English, French and German, to show how the discourse structure varies (see examples 3, 4 and 5 representing in Figure 15) (Delin et al., 1994: 63).

- (3) The stepping load can be altered by loosening the locking lever and changing the position of the cylinder foot. (Liftmaster)
- (4) Pour modifier la charge d'appui, desserrer les leviers puis déplacer le pied des vérins. (Liftmaster)
To modify the load stepping loosen the levers then change the foot of the cylinder foot
- (5) Nach Lockern der Klemmhebel kann durch Verschieben des Zylinderfußes die Tretbelastung verändert werden. (Liftmaster)
After loosening of the levers can by pushing of the cylinder foot the load changed be.

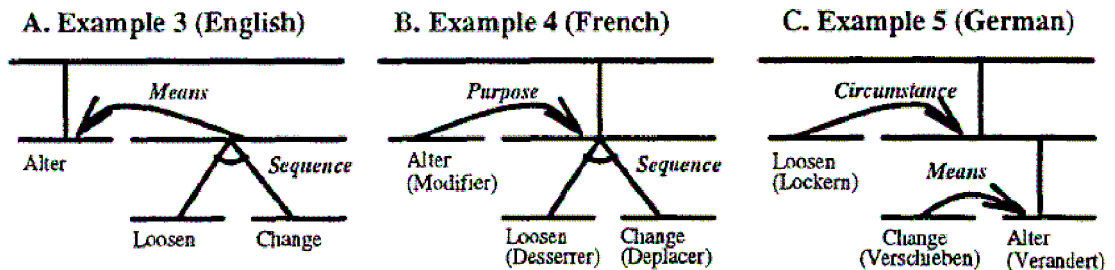


Figure 15 Multilingual discourse structures representations (Delin et al., 1994: 63)

Similarly, when analysing the data of interpretations in this thesis, it is clear that the RST representation of interpretations vary from their corresponding source

speeches. The examples below offer a comparison of a source speech in English (Figure 16) and its corresponding Chinese interpretation by a trainee interpreter (Figure 17).

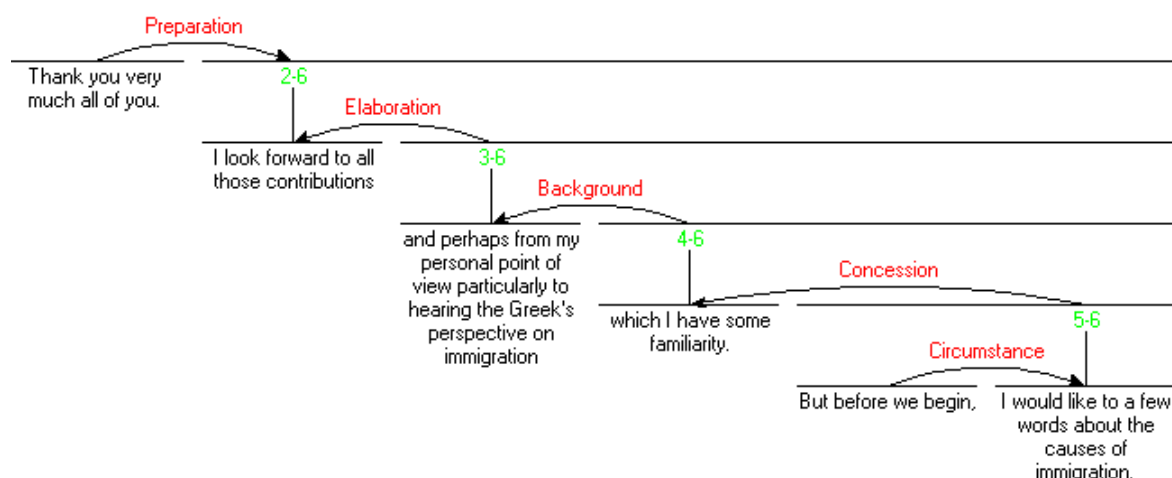
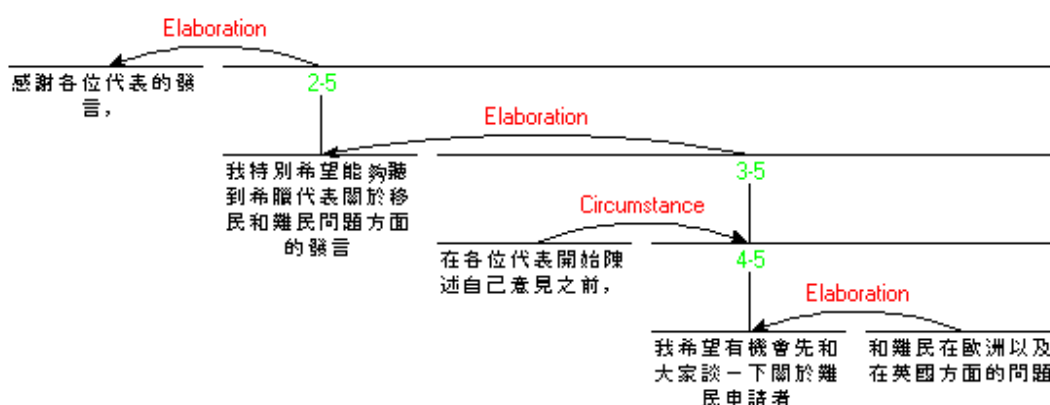


Figure 16 English Speech 2



Literal translation

1. Thank you for your interventions
2. I especially hope to hear something from the Greek delegate on issues about immigrants and refugees)
3. Before you start giving your opinion
4. I hope I have the opportunity to discuss with you about asylum seekers
5. and problems about refugees in Europe and in the UK

[Appendix B: Data Annotation/Ctrl Spch2/EC/C1]

Figure 17 Trainee's Chinese interpretation of

Mann and Thompson believe that ambiguity with regard to text structure is normal in RST. Marcu also points out that, 'discourse is ambiguous the same way sentences are: usually, more than one discourse structure is produced for any given text' (2000: 137). When an RST analyst finds ambiguity, 'it is a recognition that any of several incompatible analyses are plausible, and that the text does not provide a sufficient basis to disallow any of them' (Mann and Thompson, 1986: 28). Simultaneous analyses happen often in RST analysis. According to Mann and

Thompson, ‘sometimes there is a pair of spans in a text for which the analyst recognizes that more than one relation definition holds’ (ibid). They call this ‘overlap’. They go on to explain that ‘the difference between ambiguity and simultaneous analyses is in the compatibility of the alternate analyses’ (ibid).

As we saw in the previous section (see Figure 12, Figure 13 and Figure 16), explicit discourse markers in the text often serve as cues to assist in identifying RST relations. In addition to the empirical evidence in my own data, there is support for this in studies by Hovy (1990) and Scott & Souza (1990).

As we saw previously, interpretations tend to be heavily marked by cohesive devices such as conjunctions, it would be interesting to investigate this phenomenon further.

2.4.5. Applications of RST

In addition to the applications in natural language generation that I have already mentioned, over the last decade or so, RST has recently been used in new and varied ways.

‘(RST) has also been taken further by text linguists, who have applied it to a wider range of texts than the original starting point of Mann and Thompson. The extended RST list has been also validated across several languages; contrastive RST analyses have been performed, for example, for Dutch (Abelen et al., 1993), Chinese (Marcu et al., 2000), French, Portuguese and German (German, Standard) (Delin et al., 1996), Spanish (Taboada, 2004) and a host of other languages. RST therefore continues to be an active area of research into text organization’ (Bateman and Delin, 2005: 2).

In his work on automatic text summarisation, Marcu (2000) introduces the notions of relevance and salience, and thereby provides a principled basis for progressively compressing a message. As we shall see, this development provides further inspiration for the analysis of my data.

2.4.6. RST for analysing interpretations

As we saw earlier, RST has also proved to be useful when working with spoken text (Tappe & Schilder, 1998). However, to my knowledge, it has previously not been used when working with conference interpretations.

Interpreting involves constructing relationships between the component parts of the message in order to capture their contribution to the discourse and their function in a given situation. Given the time constraints interpreters face during interpreting, they have to prioritise incoming information, which requires evaluating the relative salience of the different message parts. As Scott and Souza (1990: 48) rightly point out, ‘RST can be used to represent both the message and the text plan and that it provides a means for capturing the notions of relevance and coherence within the representation of messages’. In short, RST appears to be a very suitable framework for us to compare interpretations by professional and trainee interpreters in terms of textual coherence.

Chapter 3

Trainees' Awareness of Quality

Trainees' awareness of quality interpreting and the development of a shared meta-language for describing interpreting performance are vital to this thesis, especially important to the design of a feedback tool of performance criteria (Statement 2). By learning how trainee interpreters perceive quality attributes, it is hoped to identify the basis for raising their quality awareness. Ultimately, these issues are vital for better interaction between teaching and learning, and therefore for the goal of training: improved performance.

3.1. Significance of trainees' awareness

In this section, I will discuss three scenarios in which trainees' awareness of quality come into play: trainer-trainee interaction; attributes of autonomous learning and collaborative learning. I will also highlight the significance of reflection in professional development.

As Kiraly states (2000), teaching and learning is not a one-way transmission process; it is a 'mutually beneficial process of sharing perspectives'. However, interpreter training has long been trainer-centred. Professional interpreters, in the role of trainer, serve as the main source of authority and expertise, passing on knowledge and skills to novices. Their judgement of trainees' performance, however, 'is an insufficient basis for decision-making' (Sawyer, 2004: 104). It 'fluctuates widely and therefore should not be relied upon exclusively to ensure equity and fairness in testing' (ibid: 103).

In addition, when comments are too 'technical', using terms such as register and coherence, novices struggle to follow them and may not be able to benefit fully.

With better awareness of quality issues concerning interpretations, I expected that trainee interpreters would not only develop better understanding of the comments and suggestions from the trainers but also be able to engage in better discussion with the trainers and among themselves. In other words, the interaction between training and learning is facilitated.

In addition, if autonomous learning is to be promoted in interpreter training, trainee awareness is essential. Being reflective is considered a milestone skill for

professional practice (Schon, 1987 cited in Reiman, 1999: 598). When people engage in reflective practices, 'they demonstrate a capacity (or disposition) to analyze the process of what they are doing, and to reconstruct their professional and personal knowledge schemes' (Reiman, 1999: 598). On the other hand, it has been shown that for adult learners, trainee interpreters in this case, learning new and complex skills, such as conference interpreting, without reflection will make no impact on learners' cognitive structure (Conrad & Hedin, 1981; Sprinthall & Scott, 1989 cited in Reiman, 1999: 602). Moreover, it is suggested that 'reflection is not automatic' (Reiman, 1999: 598). It needs to be guided (Reiman 1988 cited in Reiman, 1999: 602). An awareness of quality issues is essential as a basis for reflection on the part of trainee interpreters.

Last but not least, awareness of quality issues plays an important role in facilitating collaborative learning among interpreter trainees. It is a common practice for trainee interpreters to spend much time practising with their peers outside class and very often they give feedback to each others' performances. Gile observes that, 'the automation of cognitive skills and stamina build-up require much practice, more than can be given in class... That is why students in conference-interpreter training programs are required to set up informal groups of two to four or five people and practice on a daily basis' (Gile, 2005: 135). In addition, 'students often enjoy working in a group and they value learning from and with other people' (Jacques 2000 cited in Elliott & Higgins, 2005: 40). Research also shows that when a new learning experience 'involves "helping others and taking the perspective of others"', it becomes a very powerful and complex activity that can promote learning and development across a variety of professions, as well as a variety of interpersonal and intrapersonal domains' (Sprinthall & Scott, 1989; Sprinthall & Scott, 1989; Oja & Sprinthall, 1978; Peace, 1992; Reiman & Thies-Sprinthall, 1993; Watson, 1995 cited in Reiman, 1999: 600). In addition to the acquisition of interpreting skills, such interpersonal and intrapersonal development is also important for trainee interpreters to have and can be acquired by practising peer feedback during their training.

It is often observed, however, that trainees did not benefit much from such practice. In no small part this may be due to a lack of consensus regarding quality criteria, and a lack of consistency in understanding and using and describing any such criteria. For instance, a comment such as 'you didn't sound very smooth' can point to problems of pace, hesitation due to poor understanding of the speech, poor language structure due to grammatical errors, or due to many other issues.

Finally, by looking into the issues of trainee awareness, I hope to gain a better understanding of the development of trainee awareness for novice interpreters and of

the role training plays in the formation of consensus and over time a shared meta-language among the novices.²⁴ Of course, the ultimate test of the significance of trainee awareness is to see how it affects performance.

3.2. Aims

Having understood the significance of trainees' awareness of quality and of their cognitive development in this regard, I aim to answer three important questions at this stage. For one thing, it is essential to uncover what quality interpreting is in the eye of trainee interpreters. What they perceive might be very different from what professionals perceive. Irrespective of the possible gap between the professionals and the trainees, I also need to consider whether trainees agree or disagree with each other. Most important of all, I aim to find out whether training helps to clarify confusion and enhance mutual understanding between trainers and trainee interpreters and among the novices themselves and how this affects performance.

3.3. Methodology

3.3.1. Subjects

This step of the investigation involved two groups of trainees at two different levels of training: novice and advanced.

I recruited 22 novice trainee interpreters (novices) with various language combinations from the MA in Interpreting and Translation Studies (MAITS) at Leeds University. At the time of participation in this survey, most had no professional interpreting experience, while a few had previously received some interpreting training from university modules.

I involved 15 advanced trainees from Eastern Europe who hoped to work for the European Parliament. With the support of the Parliament, they came to Leeds for a four-week English enhancement programme in the summer 2003. All of these subjects were trained interpreters with work experience.

²⁴ This chapter was presented at the Fourth Conference on Quality in Translation and Interpreting – Academic & Professional Perspectives (Peng, 2004).

3.3.2. Questionnaires and timing

Novice trainees answered my questionnaires at three different stages. Firstly, prior to any formal training in the postgraduate programme in Leeds (week 0), they were given the first questionnaire (A) (Table 7) on their perception of good/bad interpreting performances.

Immediately after completing the first questionnaire with minimal instructions, they were then invited to complete a semi-structured questionnaire (B) on nine quality criteria commonly referred to in both training and professional circumstances (Table 8).

What makes for a good/bad interpreting performance? Please list all the criteria that you find important and describe their characteristics against which a good/bad interpreting performance is judged.		
Criteria	Characteristics	Good/Bad interpreting performance

Table 7 Questionnaire A for trainee's awareness of interpretation quality

What is your understanding of these terms in the table when you receive/use them as the feedback to interpreting performance?	
Terms	Your Understanding
Accuracy	
Cohesion & Coherence	
Communication	
Completeness	
Delivery & Fluency	
Register	
Terminology	
Voice	
Booth manner	

Table 8 Questionnaire B on trainee awareness of quality

After completing these two questionnaires, they had the first four weeks of their MAITS training. During this period, trainees were trained on public speaking skills and were engaged in memory exercises. In training for public speaking, trainees were asked to prepare some content for presentations but to speak without a script in either of their working languages, i.e. their native language or English (as their B language). Others listened attentively, to comprehend the speeches without taking notes, and then reproduced the speeches in their own words, either in English or in their own mother tongues. By doing so, trainees were trained to comprehend then memorise the logic of a speech, rather than reciting the speech word for word.

In short, it was a period of time when trainees learn the basics of how to give speeches which make sense to their audience, and also how to make sense of speeches given by their peers, and then reproduce their in their own words, while still making sense to their audience.

After the four weeks, in week 5 the group of novice trainees were given questionnaire (B) again, yet this time it included only eight attributes instead of nine. I removed 'booth manner' because I found, in the previous round, that this item was confusing and they had in any case not yet practised much SI in the booth.

In week 10 when the novice trainees finished their first term, had learned note-taking for CI, and had their first few classes on SI, I organised a workshop. This time I only involved the Chinese-English group novice trainees. An in-depth discussion on interpreting quality criteria was held, with a list of performance criteria proposed by LNTO²⁵ as a prompt (Figure 18).

The advanced trainees filled in both questionnaires (A and B) during their stay in Leeds, but without the benefit of participating in this workshop.

²⁵ The Languages National Training Organisation (LNTO) came into being in 1998.

As part of the network of National Training Organisations, its aim was to set standards for the development and use of language skills in and for the work place and to promote a greater national capability in languages for business and employment purposes. Now LNTO was merged with the Centre for Information on Language Teaching and Research (CILT) in 2003 and formed The National Centre for Languages. Available: <http://www.cilt.org.uk/about.htm> [Access: 12 May 2006]

Element Int 5.2.1 Interpret one-way specialist assignments

When you interpret one-way, you must show that:

1. you interpret the meaning of a sustained presentation:
 - precisely and fluently in the target language
 - maintaining a consistently satisfactory performance throughout the assignment
2. you reflect the source language user's:
 - register, tone and speed of production
 - attitude, irony, sarcasm and innuendo
 - non-verbal communication
 - social and cultural norms
 - role and relationship with the audience
3. you accurately interpret:
 - factual information, concepts and opinions
 - standard language and any regional or national dialects
 - complex language, specialist terminology and jargon
4. you paraphrase the meaning of complex terminology and phrases, if there is no direct equivalent in the target language
5. your conduct is consistent with the professional code of conduct
6. you support effective communication throughout the assignment and take action if communication breaks down

Figure 18 LNT0 performance criteria²⁶

3.4. Results and discussion

The results of the two sets of questionnaires from the two groups of trainees not only addressed part of my research goals stated in Chapter 1 but also provided abundant data for further discussion.

3.4.1. Questionnaire A (novice vs. advanced trainees)

There was a big difference between the novice and advanced trainees in the way they answered the first questionnaire. For the advanced trainees, responses followed the structure given in the instructions, with further elaboration on the proposed

²⁶ Available: <http://www.languagento.org.uk/quality/tistandards.htm>. [Access: 20 March, 2003]

criteria. For the novices, the instructions were largely neglected and most subjects responded by giving arbitrary points they could recall on the spot.

The questionnaire revealed several interesting findings with respect to the novice interpreters. First of all, it was apparent that many subjects confused good interpreting performance with good interpreters. Many of them elaborated on what interpreters should have (e.g., knowledge, language competence, memory, communication skills, note-taking skills, good memory, quick reaction, etc) or how they should be (professional and confident), instead of how interpreting performance should be judged as good or bad (as originally requested by the questionnaire). In other words, apart from such features of interpreters, observable features of an interpreting performance (e.g., clarity, fluency, specific language features such as intonation, pronunciation, accent, voice, pace, etc) were raised unsystematically or merely implied within the description of ideal interpreters.

In addition, it emerged that there was huge divergence in understanding the proposed attributes among the novice subjects. 'Knowledge', the most commonly raised attribute, was understood by the novices as 'general knowledge', 'cultural/societal knowledge', 'current affairs', 'knowledge on specific fields', 'wide range of knowledge', etc. Work ethics, morality, preparation for assignments, relevant education, posture, body language (eye-contact, in particular) and even lifelong learning were all offered as synonyms for professionalism.

With regard to the more 'observable' features, a similar level of confusion arose. For instance, 'Clarity' elicited definitions such as 'clear and logical', 'clear and concise', 'clear diction and delivery', 'correct grammar', 'clear voice and language', and even 'a clear mind'. In terms of 'Accuracy', it included 'accurate information', 'accurate language' (grammar & pronunciation), 'accurate translation' and 'accurate messages'. To describe 'Language', they used terms such as 'source/target language', 'active/passive language', 'A/B languages', 'translation into/from', etc, to distinguish between the two languages involved in the interpreting process. They made no connection with aspects of language, such as vocabulary, grammar, idioms, fluency, and pronunciation. Intonation and accent, likewise, are features of language, too, yet they were listed as individual criteria, not as subordinate attributes of 'Language'.

All in all, despite the fact that the novices were aware of a few essential features of interpreting performance (languages, clarity, accuracy, etc), they obviously lacked an efficient and systematic way to describe interpreting performance and a consensus on what constituted 'quality' in interpreting. Without a systematic way to describe

interpreting performance, their descriptions of either good interpreters tended to be loose and disorganized, with limited, yet diverse, terms. Without a consensus on interpreting quality, the general picture of good interpreting performance for the subjects was rather sketchy.

Interestingly, the novices appeared to be more interested in and capable of describing interpreters rather than interpreting performance. This is perhaps because, before joining the training programme, they had already thought about this when considering whether they were likely to make good interpreters in the future. Discriminating between good and poor interpreting performances, however, was beyond their experience and understanding. In short, the results of the first questionnaire verified that the advanced trainees were more capable of giving a systematic description of quality interpreting, while novice trainees confused 'good performance' with 'good interpreter'.

3.4.2. Questionnaire B (stage 1 vs. 2; novice vs. advanced trainees)

The second questionnaire (Table 8) was intended to demonstrate the trainees' understanding of nine terms commonly used to discuss interpreting performance, in both training and professional environments. From the novice groups, at stage one (week 0) I observed huge diversity of understanding and description of many of the proposed attributes was huge. For instance, they variously claimed that '**Cohesion and Coherence**' was about 'making sense', 'connectors', 'sounding fluent', 'grammatical', 'good syntax', 'convincing'. Clearly there was a lack of consensus among the trainees about this attribute. In addition, it was apparent that little distinction was made among different attributes at this stage. **Fluency & Delivery** and **Coherence and Cohesion** were used as if they were synonyms. Many responded to the effect that **Coherence** was about being fluent, smooth, or making sense; in the meantime, **Fluency** was realised by good delivery, no pause, being smooth or flowing target language.

By week 5 (stage two), however, the novices seemed to reach an agreement both cognitively and meta-linguistically, regarding many of the attributes they encountered that had previously caused confusion. Take **Voice**, for instance: many identified 'loudness', 'pleasantness' and 'confidence' as the indicators of quality voice in week 0; while after four weeks of training, the three common features were replaced significantly by 'voice projection', 'clarity of articulation' and 'intonation'. I saw a change in the novices' understanding of this specific attribute, in that volume was no longer major issue and sounding pleasant and confident was no longer a satisfying description.

The attribute of **Coherence & Cohesion** also became a clearer notion for novices after five weeks of training. I saw an emergent consensus regarding this feature. A majority of subjects mentioned ‘discourse structure’ and ‘linking words’ in their responses. The notion of **Register** was unknown to more than half of the novices to start with; while at this second stage I saw some agreement emerge. After four weeks of training (see 3.3.2), **Register** meant ‘appropriateness of language to suit the situations’ and ‘using proper vocabulary’ for some novice trainees. This clearly shows a progress of awareness.

While I found an emergent consensus among novices, the advanced trainees appeared to have shared a common understanding of many of suggested attributes. For instance, all of the returned questionnaires indicated that **Coherence and Cohesion** was about the structure of a discourse. It was realised by the means of linking words/linkage for the benefit of the users to comprehend the message.

Fluency and Delivery, however, was still confusing for even advanced trainees; there was no consensus regarding this attribute. This might result from poor teaching, in that their trainers never addressed these two attributes specifically and clearly. For instance, in 4.4.2, there is evidence to show that trainers and professional interpreters preferred to use ‘easy/difficult to listen to’ when addressing **Delivery** and they did not use **Fluency** much when giving their comments on interpreting performances.

3.4.3. Difficult notions: register, coherence and cohesion

As introduced in 3.3.2, I held a workshop in week 10 to discuss quality issues with the Chinese-English group. I used the LNTO criteria (Figure 18) as a prompt to encourage discussion. It was observable that subjects had become quite critical about the criteria proposed by LNTO.

For instance, when subjects were invited to comment on the criterion below (from the LNTO criteria), all of them criticised it for being too vague and of no significant benefit for them as a benchmark to reflect on their performance.

<p>When you interpret one-way, you must show that:</p> <ol style="list-style-type: none">1. you interpret the meaning of a sustained presentation: precisely and fluently in the target language maintaining a consistently satisfactory performance throughout the assignment
--

Figure 19 Segment of LNTO performance criteria

Criticisms included vagueness, over-generalisation, and unhelpfulness. In addition, the novices at this stage had started to practise both CI and SI and were very aware of the difference between them. Some even suggested that the criteria used to judge interpreting performance of the two modes should be different. They suggested it would be very helpful to have an explicit list of criteria concerning interpreting performance as a reference for self-evaluation and peer-feedback.

In-depth discussion also revealed that some notions of quality interpreting still remained very unclear for this group, even after 10 weeks of training. **Register** was the most difficult for them to explain. The best understanding of 'register' was expressed thus, 'Register is about word choice to make things sound either formal or informal'. Others were rather hesitant but endeavoured to explain it: 'It is very difficult to explain, it can be interpreters' voice', 'It's about style, very mysterious', and 'it's the meaning of phrases'. **Coherence & Cohesion** meant 'linking words' for the novices at this stage. As we saw in Chapter 2, coherence is supposed to be a meaning relationship among text parts, while cohesion is an overt relationship holding text parts by language markers. Yet this distinction was not found from novice trainees' responses at all. This situation corresponds with my data analysis in 4.4.2: even trainers and professionals did not differentiate these two notions clearly. They used **Coherence** (or making sense) to comment on interpreting performance, while used **Cohesion** only very rarely.

It is thus, important to investigate why novices still struggled to comprehend and explain notions like **Coherence & Cohesion** and **Register**. According to them, the two criteria were mentioned frequently during training in comments from the trainers. As such they were not new concepts. Even so, it seems that those notions were not clearly addressed pedagogically and consequently novices were only able to provide sketchy pictures concerning the two vital notions. The workshop at stage 3 in this study led to an explicit explanation of standards concerning interpreting quality. After a thorough discussion of each concept, in which examples were given, the subjects appeared to reach much clearer understandings of the notions in question very quickly.

3.5. Conclusion

I can conclude that the awareness of quality for interpreting performance is a process of evolution. Novices made good progress in both cognition and their capacity to describe what a quality interpreting performance involves. Cognitively, their awareness of quality performance developed from some local features, like

pronunciation, to a more global appreciation with a hierarchical structure. Regarding meta-language, the varied, yet limited, terminology to describe interpreting performance initially not only became more unified but also expanded noticeably over time.

However, this evolution did not happen spontaneously. It is important to note that explicit explanation and exemplification of certain concepts such as register, coherence and cohesion was necessary in order to help clarify confusion and accelerate the formation of consensus.

The conclusion at this stage suggests opportunities for further research in the areas of collaborative learning and trainee autonomy in interpreter training. However this project has begun to address issues in these areas. I identified an urgent need for a tool comprising explicit quality criteria that novice interpreters can benefit from during the development of their interpreting skills. In order to address the pedagogical need and to promote a collaborative learning culture for interpreter training, I devised a feedback grid (see Chapter 4) where most quality attributes regarding interpreting performance were captured and organised hierarchically. This serves as a tool to facilitate and regulate the process of both reflective (self-monitoring) and collaborative (peer-feedback) learning for future interpreters. Moreover, the discussion provoked by the introduction of the tool, and the criteria therein, and the adoption of a shared meta-language for evaluating interpreting performance, led to better interaction between trainers and trainees. This should, in turn, enable trainees to benefit more fully from the experience of expert interpreters.

Chapter 4

Development of the Feedback Grid

From my literature review, it is clear that there is a need for explicit guidance on quality attributes of interpretations, in order to raise and inform trainees' awareness of quality criteria. I developed a tool for them to use not only when they need to reflect on their own learning, but also when they attempt to give constructive feedback to each other. The tool contains explicit and detailed guidance on the key attributes of quality interpretations. This chapter, which presents the tool in detail, is largely based on the report of a joint project between Leeds and Heriot-Watt Universities between December 2002 and December 2003 (Hartley et al., 2004).²⁷

I played an active part in every stage of this joint project. In particular, I was responsible for feedback criteria collection (4.1), the design of the prototype of the feedback grid (4.2), experiment design and implementation of data-collection in Leeds (4.3), revision of the feedback grid (4.6) as well as adaptation of the grid for CI (4.7). As the data were collected at both Leeds and Heriot-Watt, most of the discussion and analysis of the results was done collaboratively (4.4). In addition, it should be noted that, due to the factors of timing and the availability of subjects, the grid was piloted in Heriot-Watt (4.5).

4.1. Feedback criteria collection

I based the initial guidelines for trainees' peer feedback on my previous review of existing sets of criteria in Section 2.2, such as those proposed by profession bodies like AICC and SCIC and those in use at various CIUTI²⁸ training institutions, such as the ETI in Geneva and Trieste in Italy. In order to enrich the guidelines with different perspectives, we involved professional interpreters and trainers (members of AICC and CIUTI) directly in giving comments and feedback to trainee interpreters' performance, in addition to eliciting the needs of the end users of interpreting services. Of course some criteria were proposed by more than one source. For

²⁷ The results are also published at the European Society for Translation Studies 4th Congress, Lisbon in 2004 (Peng et al., 2004).

²⁸ International Permanent Conference of University Institutes of Translators

example, ‘Accuracy’, ‘Coherence’, and ‘Delivery’ were not only raised as criteria in the literature, but were also often mentioned in class by trainers and trainees.

This exercise yielded explicit and detailed feedback criteria for SI, designed to be understandable and used by trainee interpreters in critiquing both their peers’ and their own performance. However, in order that the criteria be used in a principled way, it was necessary to develop a framework in which they would be organised in categories.

4.2. Design of the prototype feedback grid (version 1)

On the basis of all available information, including the literature review on interpreting quality and consultation with experts in the field, I devised a prototype set of criteria (Table 9).

Later, in order to support its use for CI and improve usability, the feedback grid was revised (see 4.7) before introducing it to the trainees (see 5.3).

First of all, each of these criteria was written on a small piece of paper, and they were scattered on a table, waiting to be organised under a hierarchical structure. Figure 20 illustrates the first step of the feedback grid design.

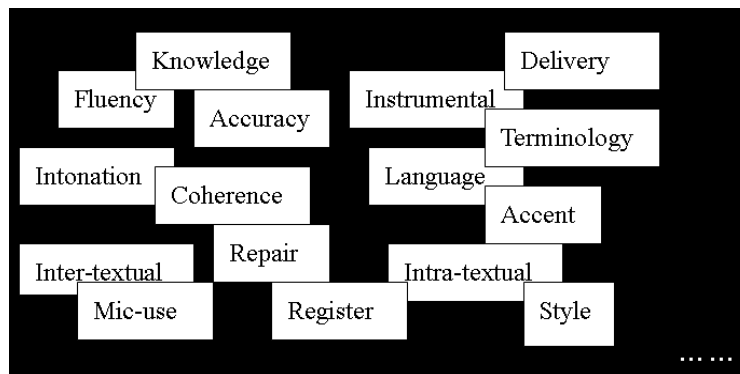


Figure 20 First step of feedback grid design

I adopted the three aspects of evaluation: inter-textual, intra-textual and instrumental, suggested by Shlesinger et al. (1997: 128), as the top level categories of a framework within which to organise the quality criteria I had collected.

In my categorisation, ‘Accuracy’ and ‘Faithfulness’ (also frequently raised) should indicate the relationship between the content of the ST (source text, i.e. source speech) and TT (target text, i.e. interpretations). Both terms come under the heading ‘Inter-textual’ and the subheading ‘Content’. ‘Accuracy’ is about concrete

information, such as figures and names; while 'Faithfulness' is about the degree of distortion in the presentation of statements and arguments.

'Coherence', as an 'Intra-textual – Language - Texture' feature, is about how linguistically coherent an interpretation is as a text, independently of the source speech.

'Fluency' and 'Delivery' are often regarded as synonyms, but we should differentiate between those two concepts. Like 'Coherence', 'Fluency' was assigned as an 'Intra-textual – Language - Texture' feature, meaning how the interpretation as a text flows as a whole. 'Delivery', despite being an 'Intra-textual' attribute, it is about 'Voice' and 'Pace', another two features which are frequently referred to by trainers and trainees.

'Register', despite being a difficult notion for trainees, is included in this grid. It is an 'Intra-textual – Context' feature, meaning that the register of the interpretation as a text should be suitable for its context.

The rest of the attributes all underwent similar analyses, and were assigned positions within this hierarchical structure. The pilot trial of the grid with trainees (Section 4.5) then provided us with an opportunity to fine-tune our prototype and to make the grid more usable.

Inter-textual (ST vs. TT)	Content	Accuracy (figures, names, etc) conveyed correctly)		
		Faithfulness (arguments & statements) conveyed without distortion		
		Completeness (no substantial omissions)		
	Rhetorical force	Speech act (conveys speaker's intention)		
Intra-textual (TT judged as a whole)	Language	Texture	Coherence (making sense, no contradictions)	
			Cohesion (synonyms, pronouns, repetitions, linking words)	
			Fluency	
			Idiomatic expression	
			Grammar (correctness)	
			Vocabulary/ Terminology	
		Structure	Logical link (between messages)	
			Concision (not wordy)	

Intra-textual (TT judged as a whole)	Language		Repair strategy	Error correction		
				Reformulation		
	Delivery		Voice	Articulation	<input type="checkbox"/> Clear <input type="checkbox"/> Mumbled	
				Intonation	<input type="checkbox"/> Flat <input type="checkbox"/> Natural/lively <input type="checkbox"/> Unnatural	
				Accent	<input type="checkbox"/> Native <input type="checkbox"/> Non-native, but comprehensible <input type="checkbox"/> Non-native, and difficult to understand	
				Pace	<input type="checkbox"/> Regular/smooth <input type="checkbox"/> Irregular <input type="checkbox"/> Too fast <input type="checkbox"/> Too slow	
	Context		Register			
			Style			
	Behavioural Skills	Mic. Use	<input type="checkbox"/> Good distance <input type="checkbox"/> Too close <input type="checkbox"/> Too far <input type="checkbox"/> Good direction <input type="checkbox"/> Wrong direction			
		Booth manners	<input type="checkbox"/> Noise management <input type="checkbox"/> Anxiety management			
Supporting Knowledge	Skills		Problem-solving			
			Analysis			
			Reasoning			
	Knowledge		General	Current affairs		
				World knowledge		
Cultural comprehension						
Specific (Subject matter)						

Table 9 Prototype feedback grid for interpreter trainees (version 1)

4.3. Methodology

Data were collected at both sites (Leeds and Heriot-Watt) to maximise the benefit of cooperation. In order to ensure the consistency of data, the protocols of the experiments were discussed, agreed and carried out accordingly by both institutions.

4.3.1. Subjects

To collect authentic feedback about interpreting performances from different perspectives, trainees, trainers, professional interpreters and end-users were recruited from both institutions, covering both French-English and Chinese-English combinations (Table 10). In other words, the focus of this experiment is to generate authentic feedback from different parties by using the interpretations as stimuli. The directionality of the interpretations is therefore not a concern.

Languages	Trainee	Trainer	Professional	Users
French-English	4 x 2 (sites)	1 x 2 (sites)	2	4
Chinese-English	4	1	2	2

Table 10 Participation of subjects

The trainee group had nearly completed their training and were due to take their qualification exam in about a month. French-English trainees were native English speakers: four from Leeds and four from Heriot-Watt. The interpretations were all into the interpreter's A language. Chinese-English trainees, all four from Leeds, were native Chinese speakers. These interpretations were all into the interpreter's B language.

The trainers at both institutions were all experienced in training conference interpreters and have experience in professional practice.

The professionals who we recruited practised conference interpreting in both SI and CI modes: two French-English professional interpreters were recruited by Heriot-Watt; and two Chinese-English by Leeds.

The users were all monolingual English speakers, and therefore were potentially genuine users of interpreting services.

4.3.2. Experiment design

To gather feedback from trainers and self-assessment from trainees, we set up classroom situations for trainees to perform SI on two audio-recorded speeches.

Both speeches lasted for about 5 minutes, and one speech type was instructional and the other argumentative. They were video-recorded in French and Chinese and the speed of delivery was agreed to be 90-120 words per minute in both languages. Notes for speeches on global warming issues were provided in English by Heriot-Watt. Speakers delivered improvised speeches in French and Chinese, using these notes as cues at Heriot-Watt and Leeds respectively. The detailed arrangements of the experiment are explained in Figure 21.

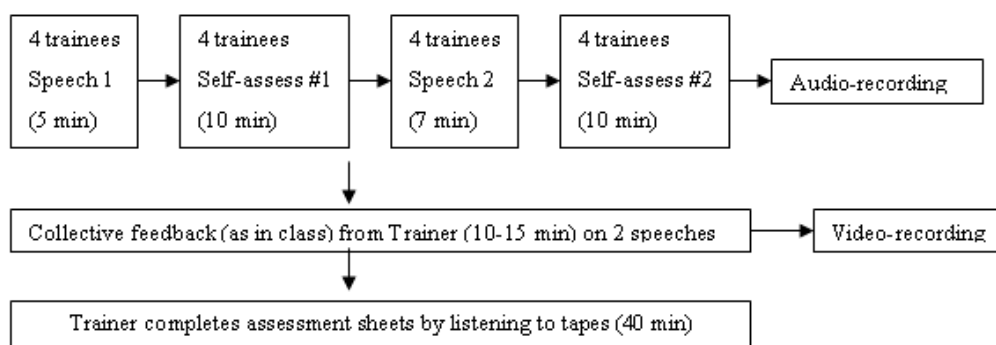


Figure 21 Experiment design for feedback collection

Trainees were asked to interpret two video-recorded speeches in SI and to complete self-assessment sheets (Figure 22) after each performance. They were given approximately 10 minutes for self-evaluation. Listening to the whole performance again would take up to five minutes. Another five minutes would give them time to reflect on their own performance, but not enough time to go back to the tape and focus on specific errors.

Self Assessment

- You are allowed to listen to the tape once without pausing and rewinding. You can make notes if you want. After listening to the tape, take up to 5 minutes to write down your comments on your performance. Please write your comment in English.

Figure 22 Trainees' self-assessment sheet for feedback collection

Trainers, on the other hand, were provided with the speech notes before the experiment, the speech tapes, the evaluation forms and a detailed explanation of the

whole experiment process. After the whole experiment was completed, the trainer gave feedback to the whole group as a whole, as in a normal class situation, for about 10 to 15 minutes for both speeches and exchanged opinions with trainee interpreters. After the collective feedback, the trainer went through the tapes and filled in evaluation sheets below (Figure 23) for each of the four trainees.

Trainee Assessment

- Please summarize your observation on the performances of the interpreter for the benefit of your fellow trainers.

Instructional
Argumentative
Overall and Other Suggestions

Figure 23 Trainers' assessment sheets for feedback collection

The professionals (P) made two vital contributions to this study. They provided SI of the same recorded speeches and they were also invited to assess trainees' (T) performances using the assessment sheet below (Figure 24).

Professional Assessment

- As a professional conference interpreter, please comment on how good the interpretation was, how effective it was for the interpreting purposes and how easy it was to follow.

Figure 24 Professionals' comment sheet for feedback collection

Performances of trainees were picked at random from each of the Heriot-Watt and Leeds groups. One French-English performance was from each institution, and both Chinese-English performances were from Leeds. Detailed set-up is explained in the table below (Table 11).

	Prof.	Recording	Assessing	
Fre/Eng	P1	SI Speech 1 & 2	T1 (HW)	2 dual-track tapes/sheets
			T2 (Leeds)	2 dual-track tapes/sheets
	P2	SI Speech 1 & 2	T1 (HW)	2 dual-track tapes/sheets
			T2 (Leeds)	2 dual-track tapes/sheets
Chi/Eng	P1	SI Speech 1 & 2	T1 (Leeds)	2 dual-track tapes/sheets
			T2 (Leeds)	2 dual-track tapes/sheets
	P2	SI Speech 1 & 2	T1 (Leeds)	2 dual-track tapes/sheets
			T2 (Leeds)	2 dual-track tapes/sheets

Table 11 Professional participation as interpreters and judges

Users of interpreting were involved in this study after the recordings of both professional and trainee interpretations were ready. They were invited to listen to SI performances (mono-track) given by both trainees and professionals. They did not have any prior knowledge of the level of interpreting expertise of the interpreters they listened to. After listening to the performances, they filled in the assessment sheets below (Figure 25).

User Assessment

- You are a delegate at this conference. Please comment on how good the interpretation was, how effective it was for your purposes and how easy it was to follow.

Figure 25 Users' assessment sheet for feedback collection

4.3.3. Qualitative analytic framework

As the dataset was small, we took a qualitative approach towards analysing our results. This analysis was done in joint sessions, at both Leeds and Heriot-Watt with colleagues from both institutions sitting down together. Interesting and important findings emerged which enabled us to improve our prototype feedback grid (Table 9) and to arrive at a set of criteria in which we had reasonable confidence.

4.4. Analysis and results

4.4.1. Input data

The data collected in the manner described in section 4.3 were transcribed and tabulated as necessary and made available for analysis.

4.4.2. Analysis of feedback from trainers, professionals and users

Questionnaires were used to collect feedback from each category of participant in the project about the performance of the trainee interpreters (see Figure 23, Figure 24 and Figure 25). The questionnaires were deliberately open and carefully avoided steering the responses in any particular direction or into pre-conceived performance categories.

In addition, when questionnaires were filled in by trainees, trainers, professionals and users in this experiment, the grid was not made available to them. In other words, the grid did not serve as a guide to evaluation of interpreting performance at this stage. The responses received from the questionnaires did not cover all of the categories in version 1 of the feedback grid.

After the responses were collected from questionnaires, they were then calibrated against version 1 of the feedback grid (Table 9).

4.4.2.1. Categories not used by trainers, professionals and users

It was found that, with very few exceptions, all categories included in the version 1 grid were used by the non-trainee participants, although the terminology was not necessarily the same they used. Among the categories not used at all was **Faithfulness (arguments and statements conveyed without distortion)**. Given the popularity of the other **Content** category, **Accuracy (figures, names, etc conveyed correctly)**, it would appear that the term **Accuracy** is preferred to **Faithfulness**, perhaps on account of the moral overtones of the latter term.

The other significant categories not mentioned in feedback were those under the heading of **Behavioural Skills**, namely **Microphone use (Good distance/too close/too far/good direction/wrong direction)** and **Noise Management** within the category of **Booth Manners**. By contrast, the other rubric under **Booth Manners**, **Anxiety Management**, elicited seven mentions exclusively from the trainers, while the generic behavioural skill of **Persistence/recovery** elicited five mentions.

Problems of **Register** received only three mentions (2 from professionals, 1 from a trainer) and **Style** none at all. This result corresponds with the situation observed in section 3.4.3, that trainee interpreters were not able to explain what ‘register’ was, ten weeks into their interpreting training. As trainers used this attribute very rarely in commenting on interpretations, trainees would then have less exposure to this attribute and subsequently found it confusing.

Halliday explains **Register** as ‘a variety according to use, in the sense that each speaker has a range of varieties and chooses between them at different times’ (1964: 77). In different registers, linguistic elements such as vocabulary, syntax, phonology, morphology, pragmatic rules as well as paralinguistic features such as pitch, volume and intonation may vary. **Style**, on the other hand, is about individuals’ variation in language use, such as formal/informal style.

As half of the professionals and trainers involved in this step of exploration were non-native English speakers, it was not necessarily easy for them to judge whether the many linguistic and paralinguistic features of interpretations were presented at the correct register. Moreover, as the interpreting was done in a class situation, there were not many contextual factors to consider. Consequently, the attribute of ‘Register’ tended to be ignored.

4.4.2.2. Categories used by trainers, professionals and users

An analysis of the distribution of comments by trainers, professionals and users provides some evidence of the main categories of concern to each group. While **Accuracy** elicited 30 mentions (17 from trainers, 13 from professionals), users were largely unable to comment on this category as they only knew English. Having said this, some inaccuracies might be apparent to users knowing the topic but not the source language.

The most used category was **Coherence (making sense, no contradictions)** with 66 mentions. This figure includes both explicit mentions of the category and other comments which were interpreted by the analysts as pertaining to the category. The latter included such terms as ‘clarity’ (two mentions from professionals, 13 from

users), ‘ambiguity’, ‘unconvincing’ (five mentions from trainers), ‘unreliable’ and ‘easy to follow’ (nine users, two trainers). It was noted that the term ‘clarity’/‘clear’ was ambiguous in that it could relate either to **Coherence** or to **Fluency of Texture** (eight mentions – one trainer, two professionals and five users).

In contrast to the apparent concern with coherence, there was a noticeable lack of explicit reference to a category which is often assumed to be popular with trainers, namely **Cohesion (synonyms, pronouns, repetitions, linking words)**. There were just three observations in this category (two from trainers, one from users). It was of course possible, though unlikely, that these figures were due to the lack of any problems of cohesion in the trainees’ output. However, a category **Logical link (between messages)**, which appeared in the Grid under **Structure**, received six mentions. This suggests another terminological preference: that **Logical link** was favoured one and used more than **Cohesion**.

The other **Content**-related category that elicited numerous responses was **Completeness (no substantial omissions)** (eight trainers, nine professionals, four users, despite the inability of the latter to make direct comparison to the source speech). Two trainer comments referred to the generic category of **Rhetorical Force**, while, more specifically, **Speech Acts (conveys speaker’s intention)** elicited 14 mentions, including non-specific terms such as ‘spirit’, ‘tone’, ‘expressiveness’.

There were ten non-specific mentions of problems of **Language** in the trainees’ output (five by trainers, one by a professional, four by users). Despite this relative vagueness, all respondents showed a willingness to be more specific in their categorisation of the **Texture** of the output.

The categories of **Fluency, Idiomatic Expression, Grammar and Vocabulary/Terminology** were also used a lot. **Fluency** elicited eight mentions – five by users but only two by professionals and one by a trainer. **Idiomatic Expression** got 19 mentions – ten by trainers, six by professionals, three by users. **Grammar (correctness)** received six mentions – five by trainers and one from a professional. Interestingly, this category appears to have been of no concern to users. **Vocabulary/Terminology** elicited six mentions from trainers, two from users, yet this category appeared not to have been an issue for professionals.

Seven comments came under the heading of **Repair strategy**, four as **Error correction**, and three as **Reformulation**.

Problems of **Delivery** were of concern to all three participant groups: four trainer comments and two professional comments used this non-specific term, while

the general category of **Voice** (as an immediate hyponym of **Delivery**) elicited 11 mentions (six from trainers, four from professionals, one from a user). The majority of comments under the rubric **Delivery** were, however, more specific: **Clear/unclear articulation** (six mentions by professionals, three by trainers), **Flat intonation** (three mentions by users, two by trainers, one by a professional), **Natural/Lively/Unnatural Intonation** (three mentions by professionals, two by trainers, two by users).

On the other hand **Accent** was little used, limited to three mentions of **Non-native/difficult to understand**. There were 29 mentions of **Pace** (or 'Flow'), including **Regular/smooth** (four by professionals, six by users, one by trainers), **Irregular** (nine by professionals, including the use of such terms as 'halting', 'hesitation', 'ums and ahs', seven by users, three by trainers) and **Too fast** (five by users, one by professional). A much-mentioned comment (eight) by professionals was 'Easy/difficult to listen to', which we classified under the heading of **Delivery**.

Further problems of **Delivery** surfaced in comments which were not explicitly covered in version 1 of the grid. 'Confidence' (four mentions by users, three by trainers) seemed to be a sub-category of **Articulation**. 'Chunking'/'pausing' (five mentions by professionals, four by trainers, three by users) was a problem of **Delivery** but also equally affected the **Coherence** of output. Two user comments also mentioned 'unfinished utterances'. Related to this sub-category are frequent mentions of '**Décalage**'²⁹ and 'problems in keeping up' by trainers (seven mentions) and professionals (six mentions).

Finally there were some mentions (by trainers only) of **Supporting Knowledge and Skills: Problem solving** (one), **Analysis** (four) and **Specific Subject Matter** (two).

A further type of problem, which was not included in version 1 of the grid, but relatively prominent among the non-trainee participants, was 'Interference' (11 mentions by trainers, two by professionals and even two by users – who had no access to the source speech). This was added as an inter-textual attributes in version 2 of the grid.

²⁹ Décalage is the time delay between what the speaker says and the interpreter's output in the target language in SI.

A general problem which we observed in the responses and encountered when trying to classify the comments was that, whereas *effects* can always be identified, *causes* cannot always be attributed with confidence.

4.4.3. Criteria used by trainees for self-assessment

The most frequently cited criterion (with 20 mentions), identified by trainees in the self-assessment feedback was **Delivery**, often further specified as ‘intonation’, ‘articulation’, ‘speed’, ‘pace’ or ‘voice’. **Omission/completeness**, **message/accuracy** and **awkward/natural TL expression** came close behind with 18, 17 and 16 mentions respectively.

Further categories of assessment, which appeared to be significant for this group, and which had not necessarily been mentioned to such an extent by other groups, include **hesitation/excessive backtracking** (14 mentions), **interference/literal translation** (12 mentions) and **sounding confident** (nine mentions). The last point also reflected a more general tendency for the trainees to adopt a user’s perspective, as demonstrated by a number of detailed comments relating to the effects of the performance on the users.

In addition, a significant number of comments reflected attempts at providing explanations for specific errors. This was done by reference to the source text (e.g., deemed too difficult, dense or technical), the interpreting process (e.g., hints at processing capacity overload) and even specific strategies such as modulation of *décalage*, output monitoring or chunking.

4.5. Use of grid by trainees

A sample of seven postgraduate students of translation and interpreting in their first academic term in 2002 in Heriot Watt University was selected. Version 1 of the grid was handed out and explained at the end of a weekly interpreting session. Two trainee recordings and a copy of the grid were made available to the group for individual and peer feedback. Written reports on the perceived usefulness of the grid were requested by a set deadline.

4.5.1. Reports of Usefulness and user-friendliness

In terms of usefulness, all responses were positive. The following reasons were repeatedly given: completeness, coverage of all relevant aspects and range of skills required, usefulness of grid (*vs.* verbal) feedback returned to peers through

identification of specific problems and possible solutions. One caveat was expressed: thorough familiarity with the grid is necessary prior to use.

In general, users were positive about the user-friendliness of the grid. The indented layout, clear division of sections and concise definitions (of categories) were all given as reasons. Although the vertical/horizontal layout was seen as difficult to handle while listening, it does provide a useful reference at the 'macro' level. A few comments were made suggesting that the grid was not user-friendly, these focused on the inability to distinguish 'close' concepts: analysis/reasoning, fluency/pace, cohesion/logical links and error correction/reformulation.

There were only a couple of suggestions for improving the grid: that a box be added on CI-specific presentation skills in order to facilitate peer feedback on CI performance, and that grading boxes be introduced for all criteria listed in the grid – bad, good, very good, excellent.

4.5.2. Actual use by trainees for peer feedback

The trainees in the sample made very few lengthy comments in the boxes provided in the feedback tool. With one notable exception, all students in the sample ticked most boxes (or marked 'good/very good') and on average returned brief comments in just three out of the 29 boxes.

Most of the feedback was returned on the right-hand side, at the most detailed level of the grid. Only two students made minimal use of broader distinctions. Comments are often noted under **Accuracy** or **Faithfulness** (at the top of the grid) and **Vocabulary/Terminology**. Comments on **Delivery** featured on several occasions under **Fluency**; similarly **Idiomatic expression** featured under **Vocabulary/Terminology**. Additional comments were noted where version 1 of the grid invited ticking of boxes – notably under **Articulation** and **Intonation**.

Comments were of two types: general observations such as '*stops in mid-sentence*', '*awkward moments*', '*French sounding*', '*slightly colloquial*', '*noise in booth in second speech*', '*sounds anxious towards the end*' and '*stilted*'; and specific references to output e.g., '*1.30 instead of 2.30*', '*overuse of which*', '*several umhs*', '*agenda sounded like gender*', '*used cabin for guichet*'. There were only two instances of explicitly positive feedback (one general, on good strategies under **Problem-solving**, and one specific, on a particular choice of phrase), and virtually no suggestions of alternatives or solutions.

4.6. Revision of the feedback grid

On the basis of feedback received from all participants, a number of modifications were made, resulting in version 2 of the grid (Table 12).

Fluency was relocated from **Texture** to **Delivery** and specified as including hesitation, regular/irregular delivery, false starts, etc. Because users displayed a desire to include comments under the heading of **Intonation**, space was created for this by including the options in parentheses after the headword. In addition, to improve clarity, a single category, **Accuracy**, was created, subdivided into **Accurate use of fact, figures, etc'** and **Faithfulness to source speech**. Furthermore, a blank box was included in the revised grid for **Miscellaneous Comments**, including positive feedback, suggestions of solutions, analysis of causes in reference to the interpreting process.

It became apparent during the feedback process that detailed training in use of the grid prior to self/peer study by trainees would be advisable. For example, in a session with the trainer, the grid could be filled in by the whole group, to provide feedback on using existing student performances, and then discussed.

Inter-textual (ST vs. TT)		Content	Accuracy	Accurate (fact, figures, etc)			
				Faithfulness to source speech.			
			Completeness (no substantial omissions)				
		Grammar		Interference			
		Rhetorical force		Intention (conveys speaker's speech act)			
				Emotion (conveys speaker's attitude)			
		Décalage		<input type="checkbox"/> Too far behind			
				<input type="checkbox"/> Too close			
		Intra-textual (TT judged as a whole)		Language	Texture	Coherence (making sense, no contradictions)	
						Concision (not too wordy)	
Cohesion (synonyms, pronouns, repetitions, linking words)							
Idiomatic expression							
Grammatical correctness							
Vocabulary/ Terminology							
Structure	No unfinished utterances						
	Chunking signalled by intonation and pauses						
	Logical links between chunks						
Repairs	Error correction						
	Reformulation						

Intra-textual (TT judged as a whole)	Delivery		Voice	Articulation	<input type="checkbox"/> Clear <input type="checkbox"/> Unclear <input type="checkbox"/> Confident <input type="checkbox"/> Hesitant
				Intonation (Flat / Lively; Natural / Unnatural)	
				Accent	<input type="checkbox"/> Native <input type="checkbox"/> Non-native, but comprehensible <input type="checkbox"/> Non-native, and difficult to understand
				Quality	<input type="checkbox"/> Pleasant <input type="checkbox"/> Unpleasant Other:
			Pace (fast/slow)		
			Fluency (Hesitant, regular, irregular, false start, etc.)		
	Context		Register		
Style					
Behavioural Skills		Microphone use	<input type="checkbox"/> Good distance & direction <input type="checkbox"/> Too close <input type="checkbox"/> Too far <input type="checkbox"/> Wrong direction		
		Booth manners	<input type="checkbox"/> Noise management <input type="checkbox"/> Anxiety management <input type="checkbox"/> Other:		
		Grit	Staying power Recovery		
User friendliness user perception		Clarity	Important points / Secondary points Conviction/(confidence inspiring) Convincing / Unconvincing		
		Clear / Ambiguous			
		Relevance / Salience / Priority			
Supporting Knowledge	Skills	Problem-solving			
		Reasoning			
		Analysis			
	Knowledge	General	Current affairs World knowledge Cultural comprehension		
		Specific (Subject matter)			
Miscellaneous (not covered by categories above)					

Table 12 Revised feedback grid (version 2)

4.7. Adaptation for consecutive interpreting

The implementation of a tool to help trainees have a better understanding of quality attributes in interpreting performance, thus facilitating better judgement of performance, is one of the key objectives of the thesis.

Version 1 of the grid was designed to evaluate SI performances. This is also true of the revised grid - version 2. Based on grid version 2 (Table 12), I made some adjustments in order to accommodate features of CI, the mode of interpreting on which this thesis focuses, and also to make it easier to use. I deleted attributes which are not relevant to CI, and added features which are. This resulted in version 3 of the grid (Table 15).

Inter-textual (ST vs. TT)	Content	Accuracy	Accurate (fact, figures, etc)	
			Faithfulness to source speech.	
		Completeness (no substantial omissions)		
	Grammar	Interference		
	Rhetorical force	Intention (conveys speaker's speech act)		
		Emotion (conveys speaker's attitude)		
	Décalage	<input type="checkbox"/> Too far behind		
		<input type="checkbox"/> Too close		

Table 13 Inter -textual criteria (version 2)

For instance, I deleted ‘décalage’ and attributes such as ‘microphone use’ and ‘booth behaviour’ from grid version 2 and added attributes like ‘eye-contact’ and ‘posture’ to grid version 3. In addition, for ease of use, I simplified the organisation of the whole grid by flattening the hierarchy of criteria, so that four levels of attributes (as in Table 13) were flattened into one level (as in Table 14) to avoid confusion.

☉ **How closely does the interpretation match the speech?**

Accuracy of facts	<input type="checkbox"/> Mostly accurate <input type="checkbox"/> Partially accurate <input type="checkbox"/> Rarely accurate
Accuracy of figures	<input type="checkbox"/> Mostly accurate <input type="checkbox"/> Partially accurate <input type="checkbox"/> Rarely accurate
Accuracy of names	<input type="checkbox"/> Mostly accurate <input type="checkbox"/> Partially accurate <input type="checkbox"/> Rarely accurate
Completeness of content	<input type="checkbox"/> No omissions <input type="checkbox"/> Minor omissions <input type="checkbox"/> Major omissions
Conveys the speaker's intention	<input type="checkbox"/> Fully <input type="checkbox"/> Partially <input type="checkbox"/> Not at all
Conveys the speaker's emotion	<input type="checkbox"/> Fully <input type="checkbox"/> Partially <input type="checkbox"/> Not at all

Table 14 Inter-textual criteria for CI (version 3)

The language was also modified to make it more user-friendly. The top category ‘Inter-textual (ST vs. TT) in Table 14 was changed into a heading ‘How closely does the interpretation match the speech?’ By providing this heading, trainee interpreters can easily understand the major theme of the section. I also removed the text box after each attribute for descriptive comments. Instead I provided options that trainee interpreters can easily tick. The multiple choice options are kept as consistent as possible across the different attributes to avoid confusion. In order to avoid restricting responses to the prescribed options, I added a text box for further comments at the end. Version 3 of the feedback grid (Table 15) was then made available to our trainee interpreters to use as guidance to reflect on their own performances and comment on each others’ CI performances. Detailed arrangements for the introduction of the feedback tool to trainees and administration of its use in the MAITS programme at Leeds University will be explained in Section 5.3.



Self & Peer Feedback Grid for Tomorrow's Interpreters

Interpreter	Language: _____ → _____
Evaluator	Speech Topic
Date/Time	Source of Speech: <input type="checkbox"/> live speech <input type="checkbox"/> audio recording <input type="checkbox"/> video recording

● **How closely does the interpretation match the speech?**

Accuracy of facts	<input type="checkbox"/> Mostly accurate <input type="checkbox"/> Partially accurate <input type="checkbox"/> Rarely accurate
Accuracy of figures	<input type="checkbox"/> Mostly accurate <input type="checkbox"/> Partially accurate <input type="checkbox"/> Rarely accurate
Accuracy of names	<input type="checkbox"/> Mostly accurate <input type="checkbox"/> Partially accurate <input type="checkbox"/> Rarely accurate
Completeness of content	<input type="checkbox"/> No omissions <input type="checkbox"/> Minor omissions <input type="checkbox"/> Major omissions
Conveys the speaker's intention	<input type="checkbox"/> Fully <input type="checkbox"/> Partially <input type="checkbox"/> Not at all
Conveys the speaker's emotion	<input type="checkbox"/> Fully <input type="checkbox"/> Partially <input type="checkbox"/> Not at all

● **How good is the language of the interpretation?**

Appropriate use of synonyms	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Appropriate use of pronouns	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Appropriate use of linking words	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Making sense	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Natural expression	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Grammatical correctness	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Concision	<input type="checkbox"/> Concise and clear <input type="checkbox"/> Too wordy <input type="checkbox"/> Too many repetitions
Vocabulary/Terminology	<input type="checkbox"/> Rich & appropriate <input type="checkbox"/> Varied <input type="checkbox"/> Basic
Complete sentences	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Frequency of self-correction	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
Clarity of self-correction	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
Interference	<input type="checkbox"/> No interference from source language <input type="checkbox"/> Some, and causes confusion for users <input type="checkbox"/> Major interference impedes communication
Level of formality	<input type="checkbox"/> appropriate <input type="checkbox"/> inappropriate

● **How good is the delivery of the Interpretation?**

General delivery	<input type="checkbox"/> Confident & natural <input type="checkbox"/> Hesitant <input type="checkbox"/> Nervous & unnatural
Clarity of articulation	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
Liveliness of intonation	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
Pauses in right places	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely

Gracie Peng
Centre for Translation Studies
University of Leeds



Self & Peer Feedback Grid for Tomorrow's Interpreters

☉ How good is the delivery?-continued

Accent	<input type="checkbox"/> Near-native <input type="checkbox"/> Non-native, but comprehensible <input type="checkbox"/> Non-native, and difficult to understand
Voice quality	<input type="checkbox"/> Pleasant <input type="checkbox"/> Neutral <input type="checkbox"/> Unpleasant
Pace	<input type="checkbox"/> Just right <input type="checkbox"/> too fast <input type="checkbox"/> too slow
Fluency	<input type="checkbox"/> Consistent <input type="checkbox"/> Inconsistent
Appropriate use of style	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low

☉ How good is the interpreter?

Eye-Contact	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Appropriate Posture	<input type="checkbox"/> Mostly <input type="checkbox"/> Partially <input type="checkbox"/> Rarely
Stamina	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
Problem-solving skills	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
Reasoning	<input type="checkbox"/> Clear logic <input type="checkbox"/> Confusing at times <input type="checkbox"/> Very confusing

Cultural comprehension	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
World knowledge	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
Knowledge of current affairs	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low
Knowledge of subject matter	<input type="checkbox"/> High <input type="checkbox"/> Medium <input type="checkbox"/> Low

☉ Further Comments

Gracie Peng
Centre for Translation Studies
University of Leeds

Table 15 Feedback grid for CI (version 3)

4.8. Summary

This was a relatively small-scale exploration, involving a small dataset from groups of participants from just two institutions. Our findings could usefully be tested against the outcomes of large scale studies. In particular, it would be useful to test the grid again on more advanced trainees who appear to have conceptualised a better ‘hierarchy’ of attributes of quality interpreting.

One encouraging finding was that trainees were uniformly positive about the prototype grid. The reasons cited included completeness of coverage of criteria and lasting usefulness compared to verbal feedback. Other more critical responses to the feedback grid from trainees, trainers and experts informed its revision and the development of version 2. This was further revised in order to accommodate CI. The result was version 3, the final version of the grid.

The suitability of a performance as a source of relay interpreting was one of the most frequently mentioned techniques for evaluating quality among experts. Due to constraints of time and research scope, this real-life criterion has not been included in the grid version 2.

This feedback grid fulfilled my second research goal:

2. To abstract and organise systematically the performance criteria for conference interpreter training

In addition, the grid provided important guidance to raise trainees’ awareness of quality issues and to aid self-reflection on their performances. In the long run, I intend to observe trainees’ progress in both producing coherent interpretations and their ability to give appropriate judgement on interpretations. These developments (see Chapter 6) relate to my final research purpose:

4. To investigate the development of awareness of these criteria in trainees and its impact on their judgement of their peers and on their own performances

Chapter 5

Methodology

In previous chapters (Chapter 2, 3 and 4), I addressed the first two research goals of this thesis. In this chapter, I will explain the research steps and arrangements involved in order to answer the other two goals.

To establish a framework to capture coherence of conference interpreting in such a way that we can make comparative and qualitative judgements about interpretations produced by professional and trainee interpreters.

To investigate the development of awareness of these criteria in trainee interpreters and its impact on their judgement of their peers and on their own performances.

In 5.1, I will explain the approaches I took to collect interpretations from both trainee and professional interpreters, including the composition of the Trainee (Control and Test groups) and Professional (5.1.1) subject groups, and the setup of the experiment (5.1.2), including the selection of speeches and arrangements of the recording.

In 5.2, I discuss my adoption of RST as the framework for data annotation. I introduced two new RST relations, Coda and Repair. Coda is used to mark the end of speeches. Repair is used to describe the occurrence of self-correction (see Section 2.4.3). I observed in the interpretations by both trainees and professionals (see Section 5.2.2.1). Next I apply Marcu's algorithm (2000), which is used to give scores to the tree-like text structures produced by RST annotation (RST tree). By doing so, I am able to compute the scores of all the RST trees of interpretations I collected (see Section 5.2.2.2). In Section 5.2.3, I explain the principles I defined and adopted in choosing explicit markers from both Chinese and English texts.

In 5.3, I will describe the introduction of the Grid to the trainees. After the trainees had used this Grid for about six months, I set up another experiment to explore trainees' judgement (Test group) on coherence as the key attribute of the quality of interpretation (Section 5.4).

5.1. Data collection

The difficulty of collecting authentic interpreting performances is widely reported in the field of interpreting studies due to the very special nature of this profession. No speech would be interpreted twice in one language by more than one interpreter. Some argue that an experimental setup to collect interpretation is always less than ideal, and might fail to give a genuine reflection of reality. For example, the lack of real audience in some experiments, thus a lack of some contextual clues, is likely to affect interpreters' performance. The register of interpretations would vary when working for different users.

As discussed in Chapter 2, coherence, on the other hand, as discussed, is an inherent feature of texts, in this case interpretations, and should be regardless of contexts. Therefore, it is reasonable to expect that interpretations from professionals for this experiment would also display coherence.

On the other hand, producing coherent interpretation remains a goal for trainees to work towards. Since comparing the coherence displayed in professional and in trainee interpretations is one of the major aims of this thesis, I am confident that I have collected a valid dataset to conduct further analysis and discussion.

5.1.1. Subjects

To collect CI performance from professionals and trainees, I recruited three professional interpreters based in London who work into both Chinese and English, and trainee interpreters from the MA programme in Interpreting and Translation Studies (MAITS) at Leeds University.

The three professionals all have been active in interpreting for more than ten years and are recognised conference interpreters. Two of them are originally from Mainland China and the other is Chinese-born British, with near-native standard Chinese.

Students at Leeds University are mainly from Mainland China with some from Taiwan. All have Mandarin Chinese as their mother tongue (A language) and English as their second language (B language). None of the trainees in this study had any previous professional training or experience in either CI or SI. Both groups of students were recruited by the same recruiting standards and procedures in two consecutive years.

In 2003, I recruited a 'Control group' of eight students. Six students were recorded for each speech. Four recordings of each speech were used for data analysis. The selection of recordings to be used was made according to practical considerations. Technical problems inevitably arose during the recording process. A few tapes became inaudible and thus impossible to transcribe.

In 2004, the following year, I recruited the 'Test group', which comprised six students at first, but later two decided to withdraw from the experiment. As a result, four students stayed for all recordings. Two were from China; the other two were from Taiwan. Trainees in both groups received essentially the same training in conference interpreting, from same team of trainers, following the same curriculum.

I considered recruiting trainees from the same year and dividing them into two groups (Test and Control) for different treatments for research purposes. In this way, I could guarantee that the subjects being recruited under the same context and therefore with fewer variables. However this would necessarily mean that some students would not benefit from access to the training in and use of the grid, which I will demonstrate had significant positive impact on training outcomes. I was keen to assure that no students should receive preferential treatment, when compared to their peers from the same cohort. As things stand, the only major difference between the two groups is that the Test group received explicit guidance (i.e. the use of feedback grid) to focus their attention on quality attributes of conference interpreting, and on coherence in particular. The Control group also received guidance during the training process, but the approach was less explicit and, in particular, they had no access to the grid itself. Details on the introduction of the feedback grid to the Test group are explained in Section 5.3.

5.1.2. Experiments

5.1.2.1. Materials

In order to collect interpretations from my subjects (two groups of trainee interpreters and one group of professionals), I collected six speeches (three in Chinese and three in English) of various styles but with similar lengths, each lasting for five minutes on average. The speeches were mainly from the training materials used in MAITS.

I delivered the Chinese speeches with some prepared bullet points acting as speech notes. English Speech 1 was also delivered as a live speech under the same conditions. English Speeches 2 and 3 were audio recordings of two live speeches by the same native English speaker in two MAITS mock conferences.

All the speeches had quite clear text structures and required no specific background knowledge. Features of the six speeches are detailed in Table 16.

Speech	Details	Speech 1	Speech 2	Speech 3
Chinese	Length	3.5 minutes	5 minutes	6 minutes
	Topic	English education in Taiwan	False travel documents	Climate change
	Type	Informative	Informative	Argumentative
	Note	Without note-taking	Note-taking	Note-taking
	Delivery	Live	Live	Live
	Source	China Times (2002) ³⁰	Euro-China Meeting: Illegal immigration ³¹	Speech notes from Isabelle Perez (2002)
English	Length	4 minutes	5 minutes	5.5 minutes
	Topic	Tiredness	Immigration & Asylum seekers	Climate change
	Type	Informative	Informative	Informative
	Note	Without note-taking	Note-taking	Note-taking
	Delivery	Live	Audio recording	Audio recording
	Source	Boots Family Health Book ³²	MAITS mock conf. recording: 26/11/03	MAITS mock conf. recording: 19/02/03

Table 16 Details of source speeches

Take the notes for Chinese Speech 2, for example (Figure 26). They form the basis of a very structured speech with an introduction of the topic, the setting and scope of the problem, suggestions from the Chinese delegation, and a conclusion. The other speeches used in the experiment (see Appendix A), likewise, are all clearly structured.

³⁰ Lin, C. (林照真) (26.11.2002). I'm sorry, I was wrong (對不起, 我錯了). China Times.

³¹ EU-China Meeting: False Travel Documents. (3-4 Nov 2003) the Hague.

³² Boots Family Health Book (1997: 156-157)

<p>Introduction</p> <p>Setting</p> <ul style="list-style-type: none">- EU-China Dialogue on false travel document <p>Scale of the problem</p> <ul style="list-style-type: none">- illegal immigration- human trafficking- call for cooperation <p>The case in China</p> <ul style="list-style-type: none">- use of false documents to get to China (some examples)- trafficking Chinese to the world <p>Chinese government's action</p> <ul style="list-style-type: none">- example (travel document check by customs) <p>China's suggestions to share</p> <ul style="list-style-type: none">- 1 high-tech to make document forge difficult- 2 staff training- 3 technical support (e.g. ultraviolet machine for passport barcodes)- 4 global network (example) <p>Conclusion</p>
--

Figure 26 The structure of Chinese Speech 2

As Chinese was the mother tongue of all subjects, the three Chinese speeches I prepared were diverse in two major aspects, the length of the speech (from 3.5 minutes to 6 minutes long) and the speech types (informative and argumentative). English speeches, compared to Chinese ones, were more consistent in both aspects. They were all informative speeches of about 5 minutes.

5.1.2.2. The recording of interpretations

Since all of the six speeches (in Chinese and English) were of approximately similar length of about five minutes, the interpreters were not given a break. The subjects were not allowed to take any notes during the recording of their interpretation of Speech 1, in both Chinese and English. This type of practice, according to Gile, 'is very useful for the purpose of demonstrating to the students how memory works, and in particular the fact that if they listen carefully and understand the logic of the speech its content will be stored in their memory' (2005: 131).

In MAITS, note-taking for CI is usually introduced after the first four weeks of memory training and public speaking (see Section 3.3.2 for training details). By the time trainees interpreted Speech 2 (week 9), they had been practicing CI with note-taking for four to five weeks. Trainees' interpretations of Speech 3 were recorded after their end-of-term exam, when they were supposed to perform successful CI with note-taking. To maintain the realistic nature of the training and to reflect the trainees' progress fully, the trainee subjects were allowed to take notes in both Speech 2 and Speech 3.

The professional interpreters were recruited and recorded individually. Firstly, they were informed of the topics of the speeches and the arrangement of the recording in advance. The recording conditions were as consistent as possible with those in which the trainees were recorded. When meeting up before the actual recording, they were reminded again of the instructions. Note-taking was not allowed for Speech 1 in both English and Chinese, while it was allowed for Speeches 2 and 3. Their notes were not collected afterwards. Also, each speech was delivered without any break. They were given two minutes' break between each speech.

In brief, trainee interpretations were recorded in three sessions (one for each speech) over five months. As explained above, these sessions followed by points in the progress of their training. The professionals were recorded individually, with each of the six speeches being interpreted in a single session. The recording arrangements are summarised in Table 17.

	Speech 1	Speech 2	Speech 3
Professional	Six speeches were delivered and interpreted in one go with 2 minutes of break between each speech and instructions given beforehand.		
	3 interpreters x 3 speeches x 2 languages = 18 interpretations		
Novice	week 4/term 1	week 9/term 1	week 2/term 2
Training progress	4 weeks into memory training	4 weeks into note-taking	After term exam on consecutive
	8 interpreters x 3 speeches x 2 languages = 48 interpretations		

Table 17 Details of the recording of interpretations

After recording, all interpretations were transcribed and then annotated.

5.2. Data annotation

I adopted a range of approaches for analysing the core phenomena of message, coherence and cohesion, and identify the relations between them. I define as the structural realisation of content, as such the messages is not simply the content itself but also, crucially, the way in which fragments of content are combined and related to each other.

Core phenomenon	Indicator/framework	Details	Tool/reference
Message	Text	English (total word count)	Microsoft Word
		Chinese (total word count)	NEUCSP
Coherence	RST	RST relations (types & tokens)	RSTTool
		RST tree weight	Marcu's algorithm
Cohesion	Explicit markers	English (generic and specific)	Halliday and Hasan (1976)
		Chinese (generic and specific)	Chan et al (2000)

Table 18 Data annotation

5.2.1. Sense to texts

66 interpretations by three professionals and eight trainees were transcribed and then segmented according to their functional units (Mann & Thompson, 1986), either at sentence or clause level, for later construction of RST tree-like representations (henceforth RST trees).

English word counts can easily be obtained from Microsoft Word. However, space does not appear between Chinese words. Neither are words in Chinese comprised of a fixed number of characters. For instance, 'water' in Chinese is one character '水 [shui3]'; 'tired' can be expressed as the single-character word '累' [lei4] or by the two-character word '疲倦' [pi2 juan4]; and 'kindergarten' in Chinese would be a three-character word '幼稚園 [you4 zhi4 yuan2]'. I used a Chinese word segmenting tool, called NEUCSP³³, to facilitate word-counting.

5.2.2. Text to RST trees

As mentioned previously, the recorded interpretations were annotated as RST trees. RST annotations are sometimes criticised for their lack of objectivity. Indeed, each annotator's analysis is likely to be distinct due to individual interpretation. Den Ouden et al. (1998), however, carried out a series of studies and proved that there is high inter-coder reliability for some aspects of RST analyses. Among the analyses produced by trained annotators, the segmentation and attribution of nuclearity

³³ Downloaded from <http://www.nlplab.cn/cipsdk.html> (Natural Language Processing Lab of Department of Computer Science of Northeastern University, China)

revealed high compatibility. The identification of individual relations, by comparison, did not reveal such level of agreement. Significantly, as we shall see in Section 5.2.2.2, the segmentations and nuclearity of relations are the only factors considered when assigning scores, or ‘weights’, in our RST trees.

At the initial stage of data annotation, I consulted academic colleagues to validate my RST coding and in particular, to check my segmentation and attribution of nuclearity. I was thus reassured that my annotation was acceptable and I should carry on in the same manner. To ensure the consistency of my RST coding, the data were reviewed and annotated twice after all the data were assembled. This enabled me to identify some inconsistency in the early annotation which I was subsequently able to correct.

I used O’Donnell’s RSTTool³⁴ to produce RST tree-structures for all my texts. The tool can handle both Chinese and English, and it provides statistics about the variety and occurrence of the RST relations used in annotating each text. Below is a series of screenshots showing the four major functions of the RSTTool.

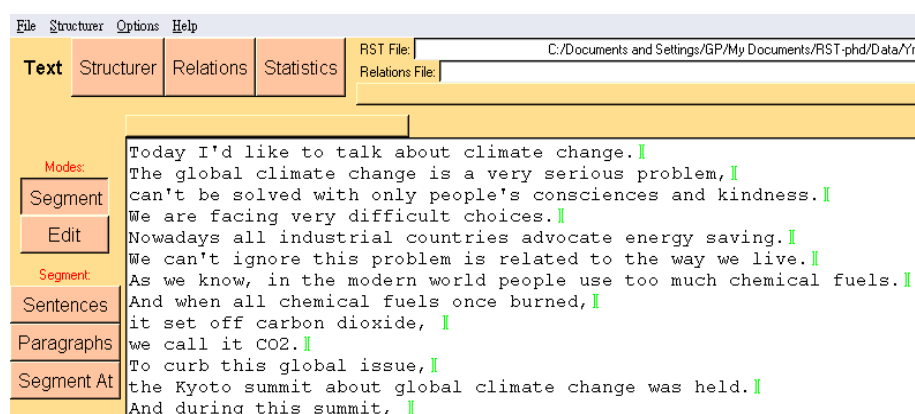


Figure 27 RSTTool-Text segmenting in English

³⁴Mick O'Donnell's RST tool is used and downloaded from

<http://www.wagsoft.com/RSTTool/section2.html> on 11 August 2004.

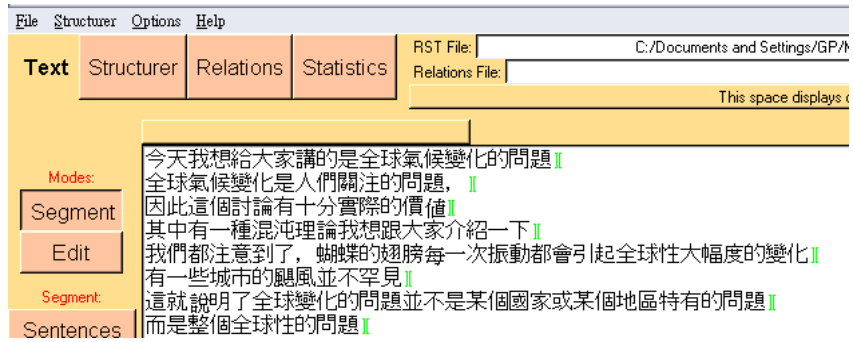


Figure 28 RSTTool-Text segmenting in Chinese

First of all, the RSTTool supports text segmentation in both Chinese and English (Figure 27 and Figure 28). For English texts, automatic segmentation into sentences or paragraphs is reliable. Chinese texts need to be segmented manually as the tool cannot recognise Chinese sentences and paragraphs. After segmenting a text, I annotated it by attributing nuclearity and assigning RST relations to the text spans with the RSTTool (Figure 29).

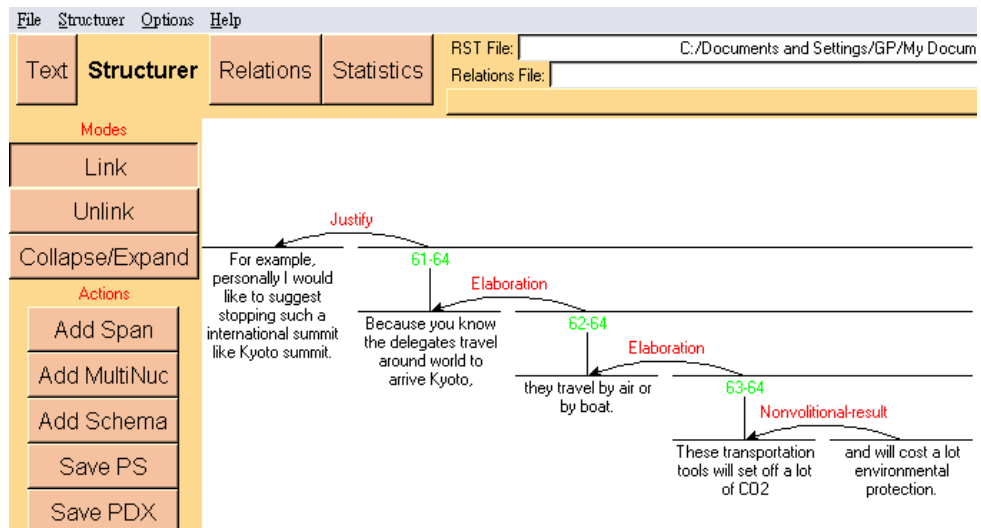


Figure 29 RST tree – text annotation

As explained in Section 2.4.3, an RST relation is the rhetorical relation that holds two spans of a text: a nucleus, the core part, which contributes to the rhetorical goal of a text, i.e. the intention of the speaker) and a satellite, which supports the nucleus. The construction of rhetorical relations will then turn a text into a hierarchical organisation.

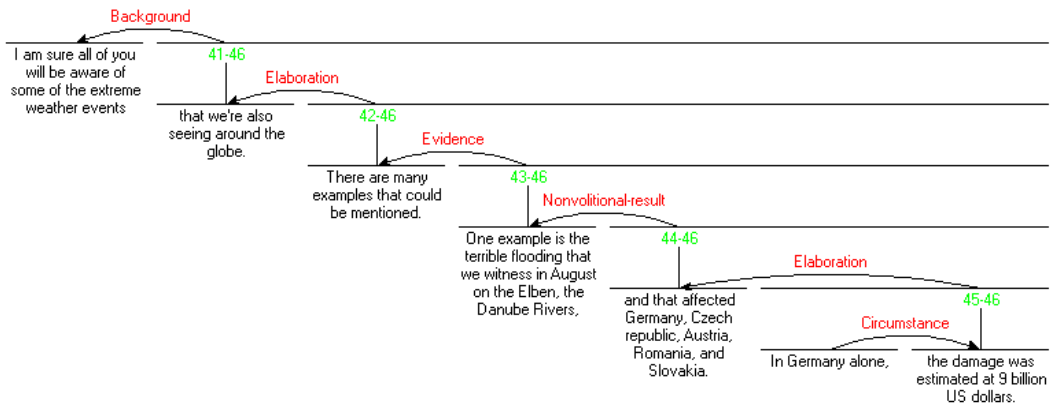


Figure 30 Example of RST annotation of English Speech 3

Figure 30 shows a part of an English speech (Speech 3), which was delivered for the subjects to interpret. First of all, segment 40 is the nucleus (*I am sure all of you will be aware of some of the extreme weather events*), which is central to whole text, and the rest of the text (41-46) is its satellite. The RST relation holding between these two text parts is that of ‘Background’. Segment 42 (*There are many examples that could be mentioned*), is a nucleus, and the satellite includes segments 43-46, which supports the nucleus with ‘Evidence’. The last two segments (45 and 46) also represent an RST relation. Segment 45 (*In German alone*) is a satellite, which supports the nucleus, segment 46 (*the damage was estimated at 9 billion US dollars*) as a ‘Circumstance’.

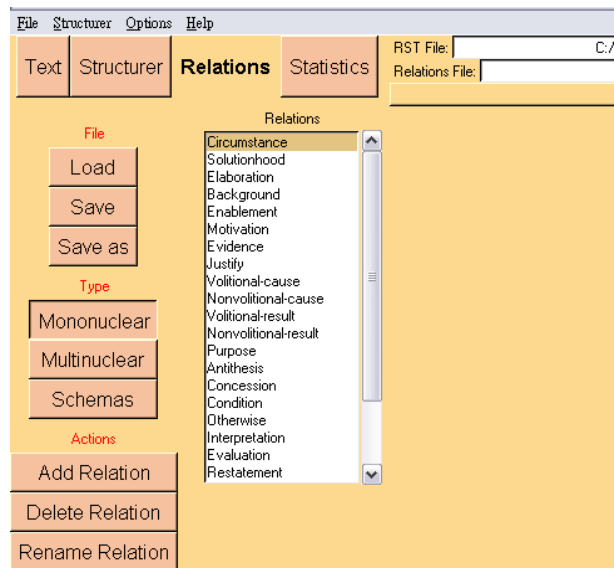
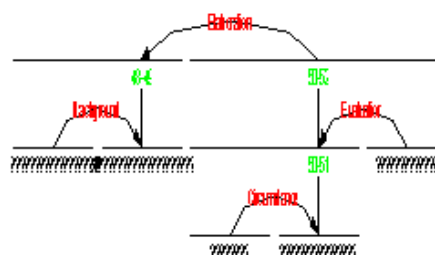


Figure 31 RST relations in the RSTTool

Figure 31 lists the RST relations which were used in my data annotation. The RSTTool allows me to manage RST relations in different ways. I am free to add or delete RST relations to suit my annotation needs. I adopted the classic set of relations, but also introduced a new relation, Repair (discussed later in 5.2.2.1), to annotate the

phenomenon of self-correction in my data. The tool allows me to ‘save’ the set of relations I have been using to help keep my data annotation consistent between sessions.

The RSTTool allows me to link, unlink, collapse and expand text spans. I used the ‘Add MultiNuc’ function to deal with multi-nucleus relations. English diagrams can also be saved and printed. Unfortunately this facility does not work for Chinese. Although the RST structure and relations remain intact, the printed text itself is not readable (Figure 32). Each character appears as a question mark. Having consulted Mick O’Donnell, the developer of the RSTTool, I still have not found a solution for this technical problem.



Error! Bookmark not defined.

Figure 32 Export of Chinese RST structure

The statistics function provided by the tool is also useful (Figure 33). It records the number of times each relation occurs in a text. This information was essential when I later compiled data to compare the coherence features of different texts. It also gives the number of the “tops” (second from bottom in Figure 33). It thus records the number of independent RST tree structures there are in the text. The more “tops” there are in a text, the less connected the various parts of the text are with one another. In Figure 34, for instance, there are seven “tops” for the whole speech, meaning that there are seven major parts of the speech. In short, despite the inability to print annotations of Chinese texts, the RSTTool is very well suited to my annotation need and analysis.

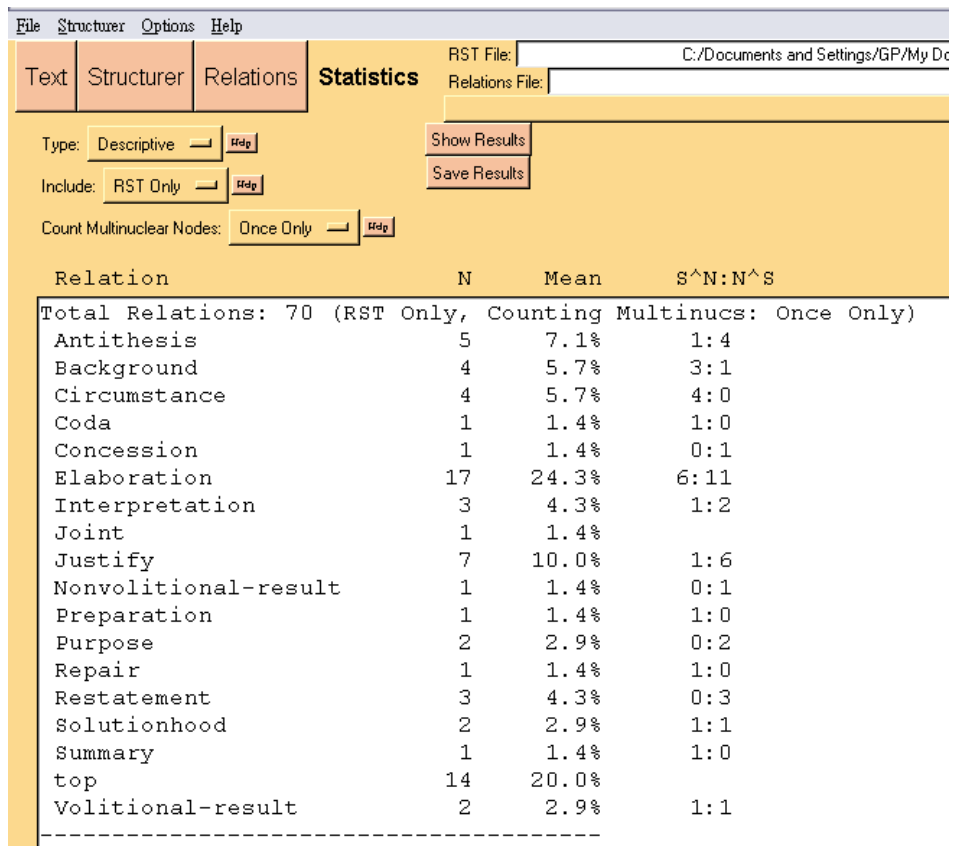


Figure 33 Statistics function in the RSTTool

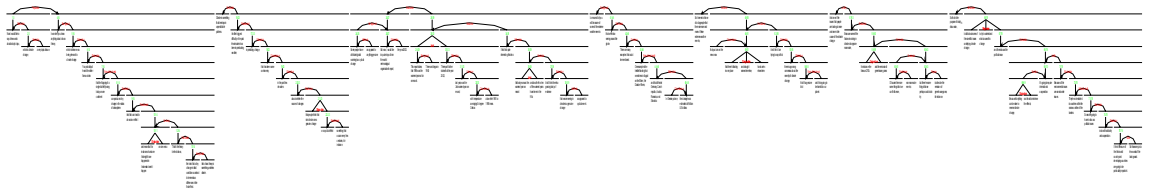


Figure 34 Tops-English Speech 3

5.2.2.1. Repair: introduction of a new relation

As discussed previously, I introduced a new RST relation: Repair. This relation accounts for the few occurrences of self-correction in the interpretations in my dataset. I provide a definition of Repair in Table 19.

Relation Name	Constraints on either S or N individually	Constraints on N + S	Intention of Speaker
Repair	on S: Listeners (L) won't comprehend S sufficiently before hearing N. on N: N always occurs to the right of S.	N repairs the insufficiency of S. S is normally insufficient or faulty and thus not understood by L.	R's mis-comprehension of S is repaired by hearing N

Table 19 Definition of Repair

As a natural occurrence in everyday communication, self-correction has been extensively explored and discussed from psycholinguistics and neurolinguistics. As Levelt states, 'speakers monitor what they are saying and how they are saying it. When they make a mistake, or express something in a less felicitous way, they may interrupt themselves and make a repair' (Levelt, 1989: 497). Interpreters are no exception. 'Interpreters — like speakers — have access to a monitoring function during speech production. This control mechanism allows them to edit their own output and possibly attend to some trouble and, in so doing, produce a self-modification, also known as a repair' (Petite, 2005: 28). According to one piece of corpus-based research, 'self-repairs occur in about 10% of spontaneous utterances' (Nakatani & Hirschberg, 1994). In Levelt's earlier work (1983), he analysed a large corpus of spontaneous self-repairs in natural speech and came up with classification of reasons, types and impacts of self-correction. He also found that, speakers 'can monitor almost any aspect of their speech, from content to syntax to the choice of words to properties of phonological form and even articulation' (Levelt, 1989: 297). Elsewhere, it has been noted that disfluency happens more often when speakers refer to something new than when referring to information that has already been given (Arnold & Tanenhaus, in press).

Some state that repairs or restarts cause disfluencies and 'do not contribute to the meaning of the spoken utterance' (Honal & Schultz, 2003). Levelt asserts that they pose a 'continuation problem for listeners' who have to edit out those disfluencies to understand the speaker's utterance (1989). Others still, however, state that disfluency, of which repair is a type, does not always hinder comprehension (Brennan & Schober, 2001; Fox Tree, 1995). For instance, Brennan and Schober (2001: 275) claim that listeners are able to 'make the appropriate parsing decisions,

solve the continuation problem, and interpret speakers' intentions without much apparent difficulty'.

Although much of the research on self repair focuses on everyday communication, the subject has also attracted attention in the field of interpreting studies. Petite (2005) addresses repair mechanisms in SI. She concludes that 'interpreters not only repair errors, but take time to attend to their outputs for different reasons' (2005: 27). For example, interpreters repair to 'achieve greater resemblance with the original input', or they correct themselves to reach 'greater relevance by maximizing the effect of (their) output and minimizing the effort in producing and receiving it' (ibid: 44). My investigation of repair in CI makes an additional contribution to research in this field. The figure below (Figure 35) contains occurrences of repairs in trainee interpretations in both Chinese and English.

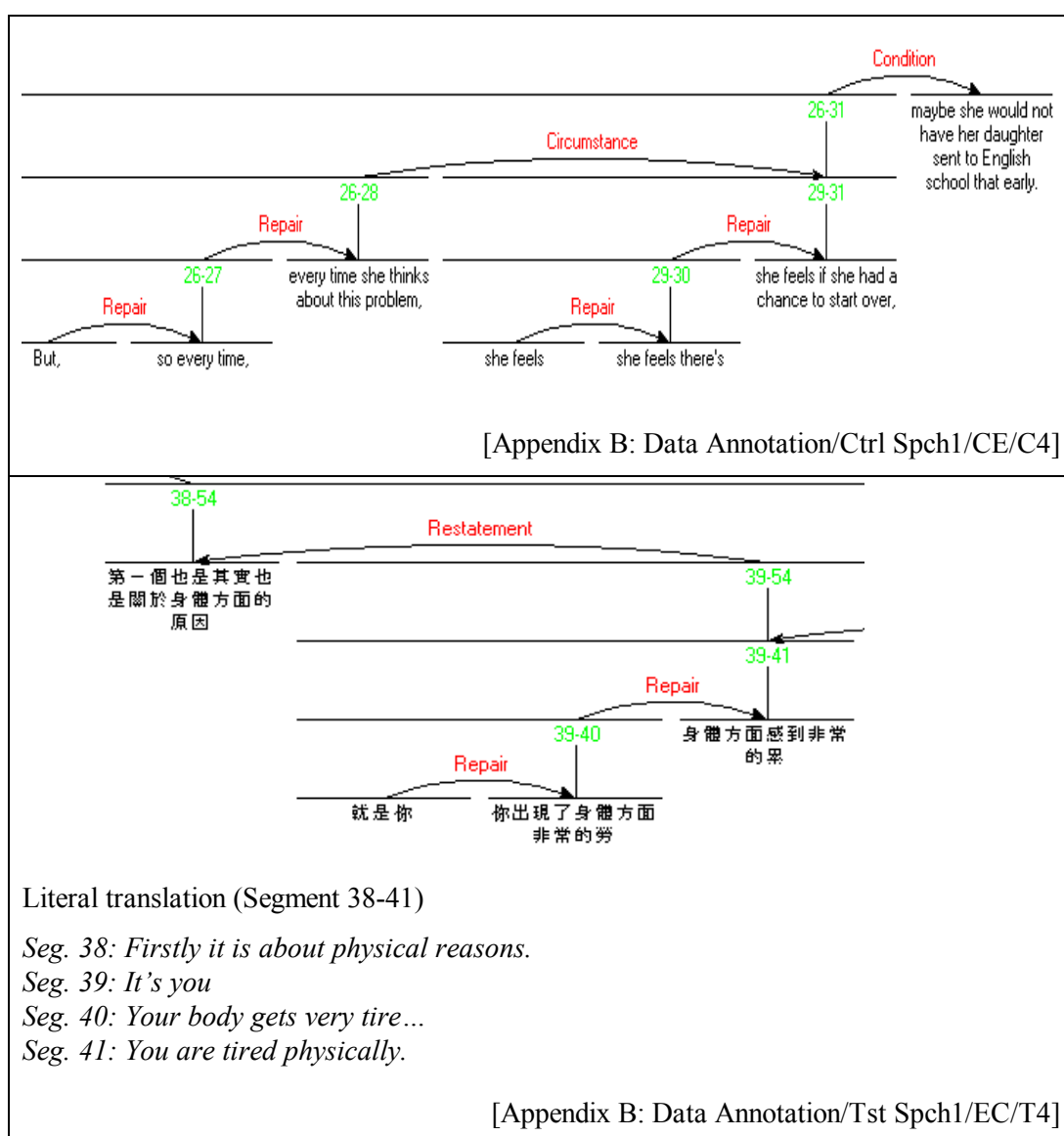


Figure 35 Examples of repairs (A)

5.2.2.2. Weighing RST trees

Annotations of my texts with the RSTTool produced tree-like discourse structures (RST trees). I then used Marcu's algorithm (Equation 1) to assign a weight to each RST tree, thus facilitating comparison of different trees.

The algorithm was designed to assign salience weightings to text spans in order to provide a principled basis for summarisation by the progressive deletion of less salient spans. It favours right-branching structures, following the observation that the best discourse trees, are 'often those that are skewed to the right' (Marcu, 2000: 137). The higher its score, the better a tree is deemed to be. This measure of quality depends on the structure of the tree but not on the identity of the particular relationships. Focusing solely on well-formedness accommodates the fact that there can be several different but equally acceptable interpretations of a single source speech.

$$w(tree) = \begin{cases} 0 & \text{if } isLeaf(tree) \\ w(leftOf(tree)) + ww(rightOf(tree)) & \text{otherwise} \\ + depth(rightOf(tree)) - depth(leftOf(tree)) & \end{cases}$$

Equation 1 Marcu's algorithm (2000: 139)

The algorithm is explained by Marcu as follows:

'The weight function w , ..., is computed recursively by summing up the weights of the left and right branches of a text structure and the difference between the depth of the right and left branches of the structure. Hence, the more skewed to the right a tree is, the greater its weight w is.' (2000:139)

5.2.3. Explicit markers and overtly marked relations

As discussed in Section 2.3.2, research shows that conjunctions are useful in 'conveying relationships between ideas' in a text (Ballester and Jimenez, 1992). It is then reasonable to investigate the use of conjunctions as explicit markers in interpretations.

The use of conjunctions in both English and Chinese interpretations were observed and recorded. Table 20 gives the classification of English explicit markers according to Halliday's Summary Table of Conjunctive Relations (1976: 242). The

rows with bold fonts identify generic markers. The rest are treated as specific markers.

Additive	General	And, and also
		Or, or else
	Emphatic	Furthermore, in addition, besides, moreover, what's more
		in particular, particularly, especially,
		incidentally, by the way, another thing
	Apposition	that is, I mean, in other words, this means
		for instance, for example, i.e. like, such as, say
	Comparison	likewise, similarly, in the same way
By contrast		
Adversative	General	But, yet, although, though, only, even though
	Emphatic	however, nevertheless, on the other hand, at the same time, despite this, while, whereas
	Avowal	in fact, actually, as matter of fact
	Correction	instead, rather, on the contrary, I mean, at least
	Dismissal	in any case, anyhow, at any rate, however it is
Causal	General	So, then, in this way
		for, because, cause, because of this
	Emphatic	consequently, hence, therefore, thus
	Reason specific	for this reason, on account of this, on this basis, the reason why, on this note
	Result specific	As a result, in consequence, arising out of this, so that
	Purpose specific	for this purpose, with this in mind, to this end, in order to, in aim of
	Conditional	then, under the circumstances, as long as, in doing so, once
		otherwise, in that case
Respective	In this respect/regard, regarding, in terms of	
	otherwise, in other respects, aside from this	
Temporal	Sequential	then, next, after that,
		at first, first of all, secondly, thirdly
	Simultaneous	just then, at the same time, meanwhile,
	Past/Present/Future	previously, before that, up to now
		at this point/moment, here from now, at once, here
		from now on, soon, after a time
Conclusive	finally, at last, in conclusion, in the end	
	to sum up, in short, briefly, in summary, to summarise	

Table 20 English explicit markers from Halliday (1976: 242)

Table 21 displays the Chinese explicit markers grouped according to Chan et al's classification (2000: 13). The generic markers are those in bold.

causality	由於/鑑於	disgression	無論/不管
	因為 (because)		不管如何
	因此	disjunction	或者, 或是, 還是 (or)
	所以 (so, therefore)	illustration	比如說/例如/譬如/像
	為了		包括
concession	雖然但是 (although, but)		舉例來說/舉..來說
	即使	restatement	就是/也就是說 (which means)
	然而		換句話說/換言之
	而/而是		我的意思是說
condition	如果...那麼 (if...then)	sequence	首先/一開始/第一
	只有...才		其次/然後
	假如/假使/假設		接下來/以下/下面
conjunction	以及和 (and)		第一種/第二種/第三種
	還有/也有 (also)		最後
	另外/此外/再說/除了		同時
	並且/而且	specification	尤其/特別是
contrast	卻	summary	總之/總體來說/廣義來說/以上
	但是 (but)		簡而言之
	並非/也不是		一般而言
	事實上	condition	這樣的話

Table 21 Chinese explicit markers from Chan et al. (2000: 13)

I set five basic rules to distinguish valid use of explicit markers, henceforth good markers, from the inappropriate uses in which they sometimes occurred in the interpretations.

1) Good explicit markers should correspond with the intended rhetorical relations. A causal relation should be marked by causal markers rather than adversative markers.

2) A good marker should lead a complete rendition or a successful repair relation. In other words, if a sentence is started with a marker and later the sentence is abandoned, the marker in this case is not counted as a good marker.

3) A good marker is recognised regardless of the diversity of its forms and its frequency. No matter how frequently a marker is used in the discourse, or how diverse the form of the marker may be, all markers that conform to the first and the second rules are considered to be good markers.

4) Redundant markers are not counted. This rule is set to manage a common mis-use of English markers by Chinese speakers. In Chinese both cause (因為: because) and effect (所以: so) in a causal relation are marked. Similarly, antithesis is marked through both 雖然 (although) and 但是 (but). As was the case for the trainee interpreters, Chinese speakers are inclined to combine these conjunctions in English.

To apply the fourth rule, I only count one English marker in each relation instead of two.

5) The last rule distinguishes noise, blank fillers and sentence ‘starters’ from proper explicit markers. Noise such as ‘and’ and ‘so’, blank fillers such as ‘actually’, and ‘you know’ and sentence starters such as ‘hmm...yes’ and ‘ok’, are found in both professional and novice interpretations. These ‘empty’ markers normally take place at the beginning of a sentence, followed by the real discourse markers. They do not count as good markers.

Below are two examples of ‘And’ as empty markers observed from my data. In example A, ‘And’ was redundantly used with ‘also’ and ‘in addition’ to represent an elaboration relation, and thus was not be counted as a good marker. In example B, ‘And’ was used with ‘on the other hand’ to indicate a contrast relation. In this case, it is ‘on the other hand’ which marks the relation, not ‘and’. Thus, ‘and’ is regarded as an empty marker.

And also in addition we had someone coming from the neighbouring countries in the south. And in addition, in China also we had this criminal groups who had engaged in taking out, taking a lot of people overseas as illegal immigrants.

[Appendix B: Data Annotation/Prof/CE/P1_2CE]

And on the other hand, there are some people say this should be practical. And of course they are the benefit, they are the people who benefit from this regulation.

[Appendix B: Data Annotation/Tst Spch3/CE/T3]

Figure 36 Example of empty markers

In addition to identifying the explicit markers (conjunctions) in my Chinese and English texts, I also noted which of the RST relations in each text were signalled by such markers. I went through all of the RST tree annotations and picked out the RST relations which were marked by explicit markers. Figure 37 is an example of how I recorded the occurrence of overtly-marked relations. This spreadsheet was used to record the overtly-marked relations in English interpretations by the three professionals. The highlighted row gives the total number of overt-marked relations in each interpreting performance.

	A	B	C	D	E	F	G	H	I	J	K	L
1		spch1ce				spch2ce				spch3ce		
2		gz	jl	pc		gz	jl	pc		gz	jl	pc
3	antithesis	4	1	1						3	1	4
4	background	2		1								
5	circumstance					1	1					
6	concession		1	1							2	3
7	condition		1									
8	elaboration	4	7	5		13	5	7		7	6	4
9	evaluation	1	1			3	4	3		1	4	1
10	evidence			2		4				2	2	
11	interpretation	1	2							3	2	1
12	justify	2		1		2		3		3	1	
13	non-volitional cause											
14	non-volitional result	1					1			1	1	
15	restatement	1	1							2	1	1
16	solutionhood									1		
17	summary	0	0	1			1	1		1		
18	volitional cause		3									
19	volitional result	1	1			4	1			1		
20												
21	conjunction		1	1		3	2					
22	contrast	1	1								1	
23	disjunction			1		1	1			1		
24	joint	1	1	1		1	1	3		3	1	
25	sequence	2	1			4	3	1		1		
26		21	22	15		36	20	18		30	22	15
27		18.5					24.67			18.5		
28	type	12	13	10		10	10	6		14	11	6
29		11.5					8.7			8.5		

Figure 37 Example of recording overtly-marked RST relations

5.2.4. Coherence Profile

As discussed in Section 2.3.1, the coherence of a text is not a tangible item that can be measured as entries as the length of a text. Comparison of the coherence of different texts represents a challenge. Some studies suggest that simply knowing the length of a text, and the number of rhetorical relations within it, is not sufficient to establish the degree of coherence. The balanced combination of these factors creates the coherence of a text (Scott & Souza, 1990: 56).

Following this observation, I identified five major factors which together could enable us to construct a picture of coherence:

- 1) the length of a text (total words),
- 2) the use of explicit markers (markers),
- 3) the number of RST relations (relations),
- 4) the number of overtly-marked relations by explicit markers, and
- 5) the weight of RST trees.

These five major factors will form the basis of my discussion about the coherence of interpretations. I do not claim to produce a complete picture of coherence, but I believe that, in combination, these factors are sufficient to give us a general profile of the coherence of a discourse.

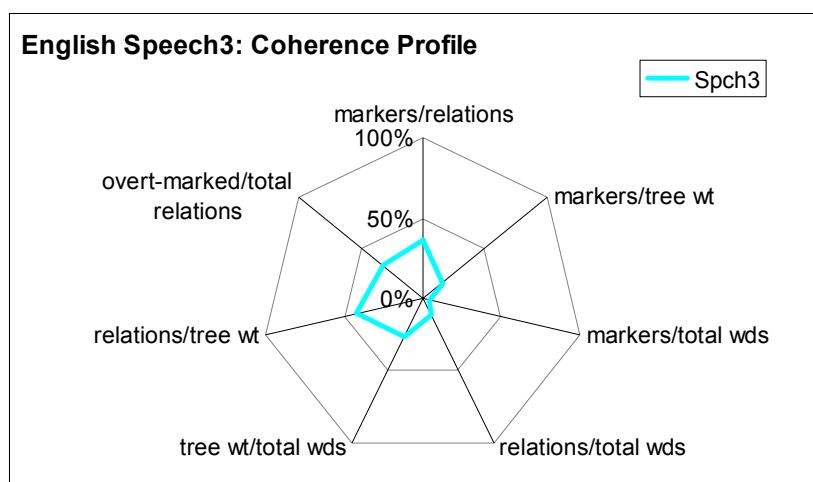
The ‘coherence profile’ which can describe one of more speeches or interpretations comprises a series of seven ratios (Table 22) between the five factors.

Parameter	Abbreviation and rationales	
Explicit markers :: RST relations	Markers/relations	To see how the use of explicit markers contributes to the total RST relations, the RST tree weight and the total number of words of a text.
Explicit markers :: RST tree weight	Markers/tree wt	
Explicit markers :: total words	Markers/total wds	
RST relations :: total words	Relations/total wds	To see RST relations and RST tree weight in relation to the total words of a text.
RST tree weight :: total words	Tree wt/total wds	
RST relations :: RST tree weight	Relations/tree wt	To see how RST relations contribute to the tree weight of a text.
Overtly-marked :: RST relations	Overt-marked/RST relations	To see how explicitly RST relations are marked.

Table 22 Seven ratios as parameters of coherence profile with rationales

I plotted the coherence profiles as radar charts to make them easier to analyse. The coherence profile of English Speech 3 was shown in the radar chart in Figure 38. This visual representation of the coherence profile also enabled us to compare and contrast interpretations by the different interpreter groups.

The shape presented in the radar chart does not directly represent the coherence of a text per se. However, the chart facilitates comparison of several interpretations of a single speech.



English Speech 3	
markers/relations	36.5%
markers/tree wt	15.3%
markers/total wds	4.1%
relations/total wds	11.2%
tree wt/total wds	26.6%
relations/tree wt	42.0%
overt-marked/total relations	32.4%

Figure 38 Coherence profile of English Speech 3

In summary, my data processing involved several major steps: text transcription, RST annotation, RST tree weight calculation, explicit marker extraction, identification of marked relations and the calculations of the coherence profile. This resulted in a comparable evaluation of professional and trainee performances.

5.3. Feedback grid introduction and administration

Having developed a feedback grid (see Section 4.7) for the trainee interpreters, it was essential to ensure that they received the training they needed to draw maximum benefit from it. In order to achieve this, I organised three introductory

sessions to introduce all 45 trainee interpreters working in ten languages, in the MAITS programme after the first recording session in October 2004.

Firstly, I explained the content of the feedback grid. I paid special attention to explaining the terms used in the criteria, in order to avoid the confusion reported in 4.4.1 and 4.4.2. I explained that the grid provides comprehensive coverage of important evaluation criteria from professional and training domains, and that it is flexible enough to use for both self evaluation and peer feedback. Importantly, I also explicitly addressed the rationales for using the tool, explaining how it would help trainees to reflect effectively on their performances. I emphasised the significance of reciprocal processes in learning and of the role of constructive feedback. In hands-on mode, I demonstrated how the grid could easily be used and gave examples of how suggestions and observations can be noted down in the boxes provided.

After the introductory sessions, copies of feedback grid were made available in the interpreting facilities where trainees practise outside class. Trainees were given storage space to deposit the grids that they completed in different modes: self-evaluation, peer-feedback and trainers' comments. In the best cases, by the end of the term, trainees had compiled substantial progress portfolios, from which they could identify scope for further improvement.

Since the use of feedback grid was not made compulsory to begin with, it proved to be difficult to collect sufficient results for further analysis. Most trainees enthusiastically adopted the grid as a tool to help them comment properly on each other's performances and reflect on their own interpretations. However, some were less interested in taking part in what they might see as extra work. This unevenness made it difficult to conduct a quantitative analysis.

However, the results in Chapter 6 show clear benefits for those who made use of the grid. It is evident that the tool reinforced trainees' awareness of quality attributes. As a result, trainees became more reflective learners. In turn, their interpreting performances showed significant improvements.

5.4. Trainee Judgement of Peer Performance

In the following sections, I will describe the steps involved in investigating trainees' judgement of peer performance. The results will be reported and discussed in Section 6.5.

After introducing the feedback grid to the trainee interpreters, they were given time and resources to use the tool when commenting on interpreting performances. It is interesting to explore how much impact this feedback grid had on their ability to comment on interpretations.

Quality attributes of CI interpretation are many and extensive. I was not able to cover them all. I chose to focus on ‘coherence’ to investigate how the trainee interpreters from the 2004 cohort judged this feature. The development of coherence in trainees’ performance is also the major focus of this thesis.

Moreover, I was interested in exploring the development of trainees’ awareness of coherence. If trainees prove themselves to be able to give objective evaluations of interpreting performances, I can plausibly assume that they can provide proper feedback to their colleagues. This would lead me to explore opportunities for a more systematic adoption of peer feedback and self assessment in training interpreters in the future. On the other hand, by comparing the individual evaluation given by trainees, I can validate the suitability of RST as a framework for the analysis of coherence in interpretations.

5.4.1. Subjects

To collect trainees’ judgement, I originally recruited eight trainees from the 2004 cohort, including the four from the Test group who also contributed performances for analysis (described in 5.1.1). All of them have Chinese (A) and English (B). Later, two more joined. Thus I had ten subjects altogether for the experiment on trainees’ judgment. This group of trainees had been using the feedback grid for nearly six months both to evaluate their own performances, and to comment on those of their peers. All subjects had finished their postgraduate training in interpreting by the time of the experiment.

5.4.2. Materials

Since the mother tongue of the subjects is Chinese, it is reasonable to believe that they are fully competent to make judgements on Chinese interpreting performance with respect to coherence. I chose four Chinese interpretations from the two groups of trainee interpreters: two from the Control group (labelled as A interpretations), and the other two from the Test group (B interpretations). A1 means the Chinese interpretation of Speech 1 from a trainee interpreter in the Control group, and B3 means the Chinese interpretation of Speech 3 by a trainee interpreter in the Test group.

The interpretations I chose as experiment materials (A1, A3, B1 and B3) approximately represent the average group performance. I validated this by looking at the ratio of RST tree weight to the total number of words in a text (RST tree weight/total words), as the research results show that this ratio is an important parameter for textual coherence (Section 6.1.1). Figure 39 demonstrates the representiveness of the chosen interpretations.

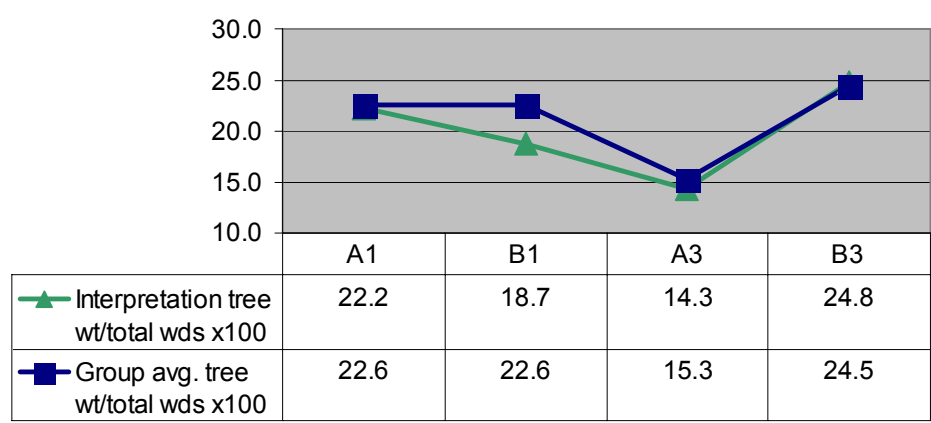


Figure 39 Chosen interpretations vs. group average

The RST tree weight/total words ratios of the selected interpretations, A1, A3 and B3 are the closest to the group average. However, the interpreter whose interpretation I selected for B3 also gave B1 interpretation, which was closest to the group average. To avoid using two interpretations by a single interpreter, I chose the interpretation of B1 with the ratio which was second closest to the group average.

To collect trainees' judgement on the interpretations, I played recordings of them to the trainee judges. After listening to each interpretation, they were asked to give scores (1 = worst and 5 = best) for seven attributes on coherence and cohesion in the Trainee's judgement sheet (Figure 40). 'Frequency of self-correction' was scored differently from the others (1 = best and 5 = worst). A low score means few occurrences of self-correction in the interpretation. The seven coherence features were derived from the feedback grid for CI (version 3) that they had been using for six months.

Evaluator
<input type="checkbox"/> A1 <input type="checkbox"/> A3 <input type="checkbox"/> B1 <input type="checkbox"/> B3

1. You'll be listening to a Chinese interpretation.
2. Please give a score from 5 (*the best*) to 1 (*the worst*) as appropriate for the features as follows of the performance.

In terms of...	Good-----Poor
1. Making Sense	5 4 3 2 1
2. Appropriate use of linking words	5 4 3 2 1
3. Natural expression	5 4 3 2 1
4. Concision (concise/too wordy /too many repetitions)	5 4 3 2 1
5. Complete sentences	5 4 3 2 1
6. Frequency of self-correction (the lower the better)	1 2 3 4 5
7. Clarity of self-correction	5 4 3 2 1

Thank you!

Figure 40 Trainee's assessment sheet

5.4.3. Experiment setup

Before listening to an interpreting performance, each subject was given the assessment sheet and briefed about the session. To avoid distraction, and to focus subjects' attention on the coherence of an interpreting performance, I played the interpretations to them without providing the source speech.

Each of the ten subjects listened to two different interpretations (but not two interpretations of the same speech). In total I have four results for interpretation B1, five results for A3 and B3, and six results for A1. The detailed arrangement of the experiment is described in Table 23. The arrangement was designed to maximise the

objectivity of the results by generating responses from as many perspectives as possible.

Time	Subjects/Interpreting	A1	A3	B1	B3
10.30	1	X			
10.40	1		X		
10.40	2		X		
10.50	2			X	
10.50	3			X	
11.00	3				X
11.00	4				X
11.10	4	X			
11.20	5				X
11.30	5			X	
11.30	6			X	
11.40	6		X		
11.40	7 & 8		XX		
11.50	7 & 8	XX			
11.50	9 & 10	XX			
12.00	9 & 10				XX

Table 23 Experiment setup for trainee judgement session

In short, this experiment explored several major aspects of trainees' judgements. Having received training and gained six months experience in using the grid, trainees should have developed the ability to give proper judgements not only on the interpretation as a whole, but also on specific aspects of coherence and cohesion. In 6.5.2, we will see that judgements by the trainees who were guided by the grid correspond well with the results of the RST analysis of coherence.

Chapter 6

Development of Coherence in Conference Interpreting

It is widely agreed that successful interpretation should be ‘coherent’. This is an important part of making sense and therefore is essential to the communicative function of interpreting. As discussed in my literature review, the importance of coherence has been emphasised by professional organisations of conference interpretations, prestigious training programmes around the world, professional conference interpreters themselves and even users of such services. Undoubtedly, it is one of the most vital attributes to consider when evaluating trainee interpreters’ progress. I have collected interpretations from both professional and trainee interpreters and used both qualitative and quantitative analysis to look for signs of coherence.

To address the features of textual coherence of the interpretations, I adopted RST as a framework for data annotation and analysis. The results of RST analysis will be presented in Section 6.1. I investigate the phenomena of ‘explicitation’ through the use of conjunctions in interpretations in Section 6.2. In Section 6.3, I will describe the occurrence of repairs in the interpretations and the impact of them on textual coherence. In Section 6.4, I use the ‘Coherence Profile’ (described in Section 5.2.4) to construct a more comprehensive picture of coherence and to compare different versions of interpretations. Finally, in Section 6.5, I report on trainees’ judgements of interpretations by the peers, and compare the results of such human judgements with the results of RST analysis.

It should be noted that, due to the small number of subjects in each group (4 in the Control group, 4 in the Test group, and 3 in the Professional group), the results presented and discussed here were not tested for statistical significance. In the long run, it would be useful to conduct a larger study to test whether these results are replicable.

6.1. RST analysis

In Section 2.4, I set out my motivations for adopting Rhetorical Structure Theory (RST) as a framework for studying features of textual coherence in interpretations. With RST, I can describe texts as tree structures which represent rhetorical relations between text segments.

As discussed in Section 5.2, the relations I use for annotating my texts are the set of ‘classical RST’ relations (Mann and Thompson, 1986). Although Mann and Thompson intended that the set of around 20 relations be open-ended and subject to addition to suit different annotation needs, ‘it has proved very stable over the years’ (Bateman & Delin, 2005: 2). To better describe my data, I added two more relations to describe the texts of consecutive interpretations I collected: Coda and Repair (see Section 5.2.2.1).

Coda, used to describe the ending of a speech, is a very common rhetorical feature of spoken texts, especially of speeches in formal settings, such as conferences. Repair, as discussed and defined in Section 5.2.2.1, also occurs often, not only in natural speech but also in interpretations. In Section 6.3, I will discuss the phenomenon of repair, also known as self-correction, in regard to both trainee and professional interpretations. I will report some prominent features of the two sets of interpretations.

6.1.1. RST trees and tree weight

As discussed above, one of the most significant indicators of coherence from RST analysis in this study is the weight of the RST tree of each text which I calculated using Marcu’s algorithm. From my data, I observed a strong and direct correlation between the RST tree weight and the total words of a text. In other words, the longer the text, the heavier the RST tree. Moreover, I found that the ratio of RST tree weight to the total number of words in a text (tree wt/total wds) was higher.

To further explore this phenomenon, I averaged the ratios for the interpretations produced by each group (Professional, Control and Test). The following figures give the average group ratios of RST tree weight to total words in both Chinese (Figure 41) and English (Figure 42) interpretations of the three speeches.

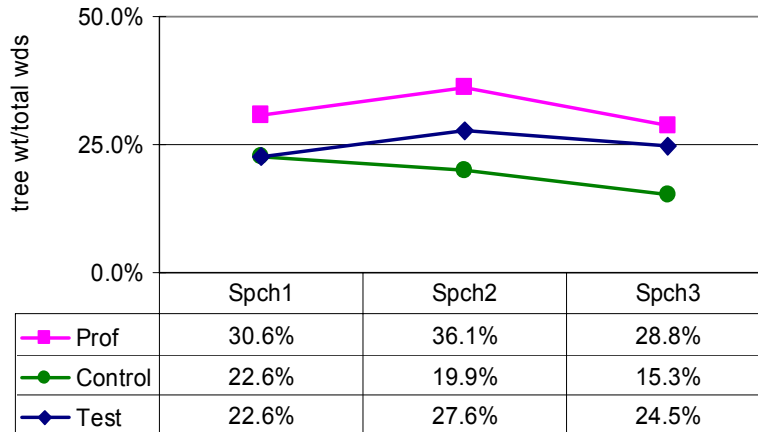


Figure 41 E>C interpretations: RST tree weight vs. total words

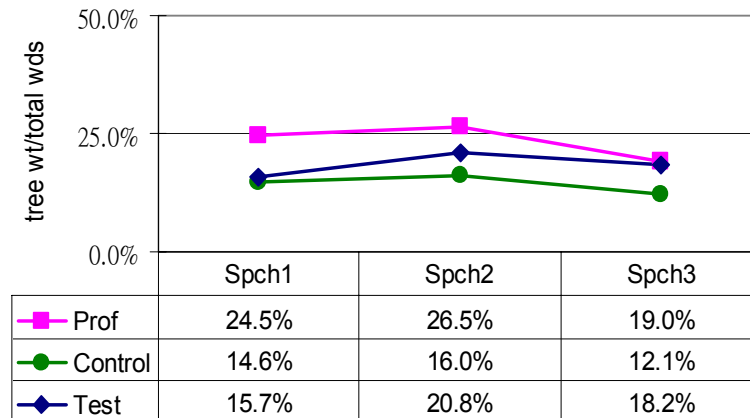


Figure 42 C>E interpretations: RST tree weight vs. total words

In both figures above, the ratios (tree wt/total wds) of interpretations by the Professional are higher than those of the trainee groups (Control and Test) across all three speeches. This means that, with the same length of text, interpretations by the Professional group are heavier, according to Marcu's algorithm, than those of trainee interpreters.

Marcu's algorithm makes use of two major elements of RST annotations: the depth of the discourse structure and the number of RST relations in a text. In other words, the deeper the structure and more relations involved in a text, is deemed to be and the larger its score. The depth of a text structure relies on how globally related a text. Yet the ratios between the number of relations and the length of a text is rather stable (see detailed discussion in Section 6.1.3 and Figure 44).

The RST tree weight alone, however, is not sufficient to represent the level of coherence of a text structure. This score needs to be considered in relation to the total length of the text. For example, a tree weight of 80 for a shorter text suggests a

higher degree of coherence than would be the case if the same weight were achieved by a longer text.

This ratio (RST tree weight/total words) will be taken as an important indicator of the textual coherence of the interpretations in this study. The figures above suggest that the difference between interpretations by Professional and trainee groups appears to lie in the depth of their text structures. Detailed discussion and illustrations of this finding will follow in Sections 6.1.3, 6.1.4 and 6.1.5.

6.1.2. Trees and bushes

In addition to the figures above, the RST trees of the interpretations produced by professionals and trainees revealed very different discourse structures. In general, the RST trees produced by annotating professional interpretations are deeper and broader than those of student performances. All the spans in the text tend to be related to a single or very few root nodes and the internal structure of the discourse reveals complex, nested relations. In other words, the performances by professional interpreters appear to achieve global coherence. By contrast, the tree structures derived from trainee interpretations look more like ‘bushes’, exhibiting only local coherence with no single root.

Figure 43 gives an impression of the difference between the RST representations of interpretations of the same speech by a trainee and by a professional. For this illustration, I deliberately chose a genuine, if rather weak, interpretation by a trainee in order to emphasise the contrast with the professional performance. The top half represents the interpretation given by a trainee interpreter from the control group (C1). The bottom half is a representation of an interpretation of the same speech by a professional (P2)³⁵.

³⁵ From Peng & Hartley (in press).

Error! Bookmark not defined.Error! Bookmark not defined.

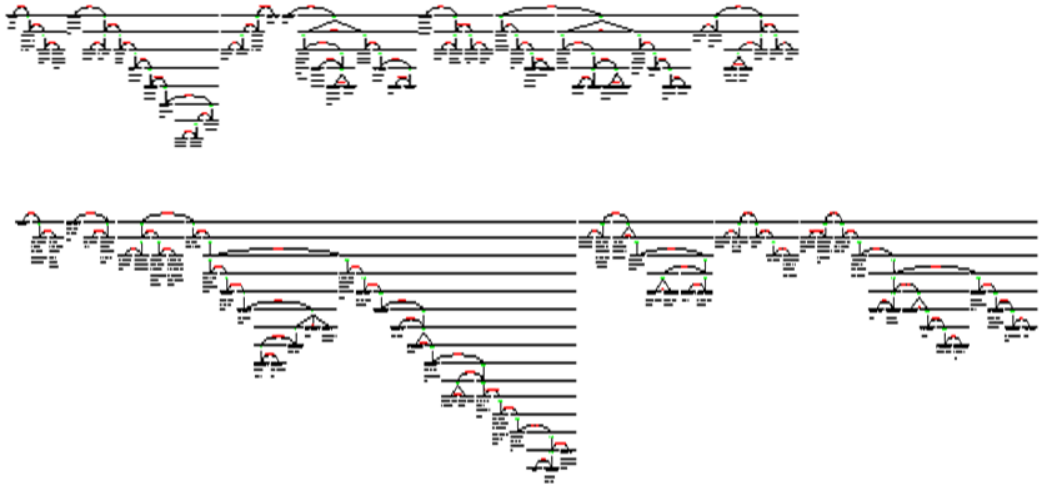


Figure 43 RST bushes and trees.

While the number of RST relations is also a major element in determining the weight of an RST tree using Marcu's algorithm (2000), the right-branching principle and depth of the structure are even more important (see discuss in Section 5.2.2.2). Greater depth means, ultimately, that a greater number of spans are subsumed under a single root relation. Significantly, in Figure 43, the RST tree of the professional's performance reaches a maximum depth of 15, while that of the trainee's performance reaches a maximum depth of only eight.

The relative depth of the tree structure of the professional interpretation (shown in Figure 43) will contribute to the total score of the RST tree. The number of relations in the two tree structures in Figure 43 is similar: 52 for the trainee and 59 for the professional. However, the score of the RST tree for professional interpretation reaches 156 and that of the trainee interpretation only is just 41.

To clarify this point further, I investigated the ratio of the number of RST relations to total number of words in the text (relations/total wds). I observed that this ratio stayed fairly consistent across speeches, language combinations and different interpreter groups, with around 1% variance overall (Figure 44).

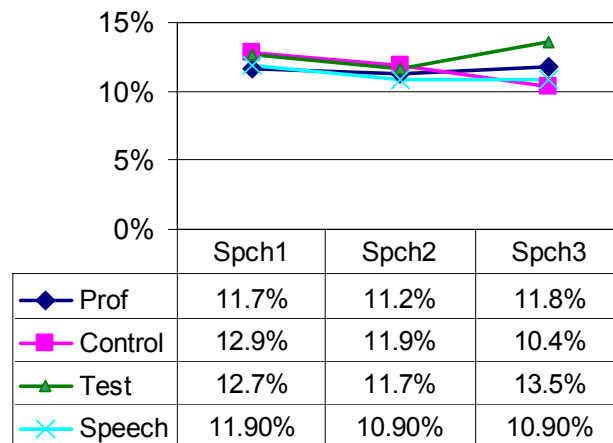


Figure 44 C>E interpretations: RST relations vs. total words

Thus, from my data, the difference between professional and trainee interpretations appears not to lie in the amount of information being retained in terms of text spans, but in the representation of this information in terms of coherence. In other words, professional interpretations do not necessarily contain more information than trainee interpretations, but the parts of the message are more richly related as a whole and more explicitly signposted (Figure 45).

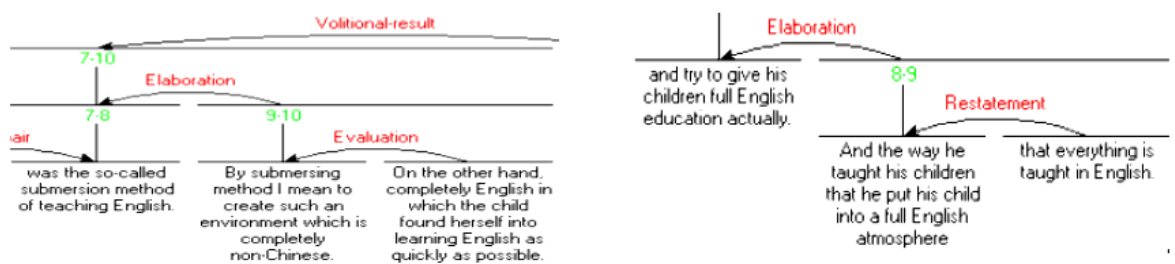


Figure 45 Explicit signposting: Professional vs. trainee interpretations

In summary, it is plausible to claim that the difference between professional and trainee interpretations lies in how information is represented with coherence.

6.1.3. 'Heavier' RST trees in Chinese

From Figure 41 and Figure 42, it is clear that the ratios of RST tree weight to total words (tree wt/total wds) are generally higher in Chinese interpretations than in English interpretations. Let us first consider trainee performances. I observed that the ratios (tree wt/total wds) were consistently higher in Chinese interpretations (Figure 41) than in English interpretations (Figure 42). In Chinese interpretations, the average ratios in both Test and The Control trainee groups in Speech 1 were 22.6%. The same ratios were around 15% for the English interpretations of the Chinese

Speech 1 by both trainee groups. There was a difference of around 7% between the two language directions. The average ratios for Chinese interpretations are therefore about 50% higher than those for the English interpretations.

In Speech 1 (Chinese and English), both groups of trainee interpreters had only received four weeks of training on memory and public speaking. Moreover they had had little interpreting practice. These results indicate that at this early stage trainee interpreters were more capable of conveying coherence in Chinese than in English. In other words, trainees were better in conveying coherence in their A Language than in their B language.

As in the case of the trainee interpretations, the ratios for professional Chinese interpretations (Prof E>C) were consistently higher than those for English interpretations (Prof C>E) as shown in Figure 46. For Professional interpretations, the difference in the ratios for Speech 2 and Speech 3 was about 10% in better cases and about 5% for Speech 1.

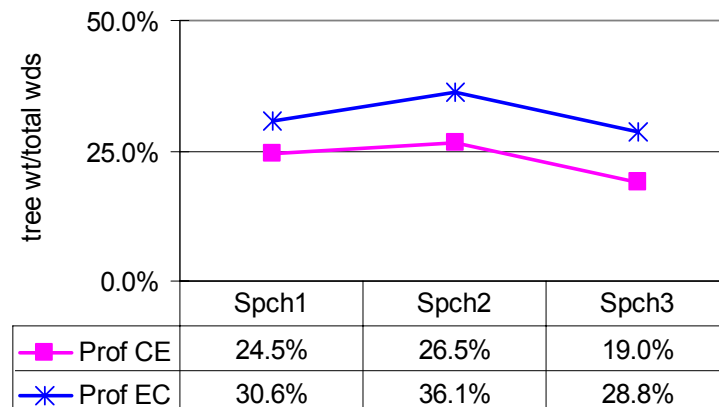


Figure 46 Prof. C>E & E>C interpretations: RST tree weight vs. total words

Initially, I suspected that this might be due to inherent differences between Chinese and English. However, Figure 47, which shows ratios for the source speeches, suggests that language difference was not the main reason for the constant gap I observed in Figure 46.

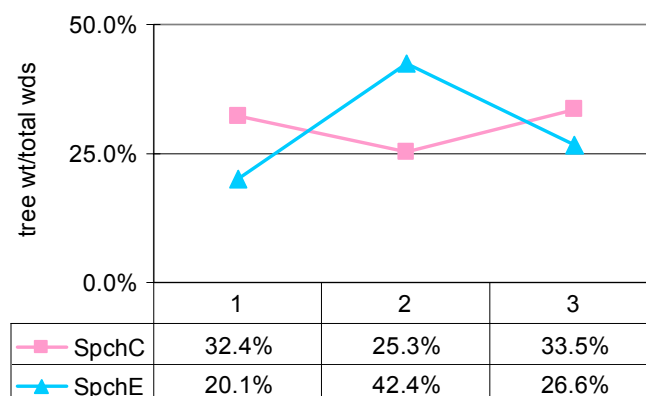


Figure 47 Chinese and English speeches –RST tree weight vs. total words

In Figure 47, the ratios (tree wt/total wds) of the Chinese speeches (SpchC) are not consistently higher than those of the English speeches (SpchE) as was the case for the Chinese interpretations. The ratio of RST tree weight to total words in an English text is sometimes higher than that in a Chinese text.

Thus, the difference between the ratios for English and Chinese interpretations can perhaps be understood as a feature of interpreted text, and vice versa. The interpreters in my experiment (both professionals and trainees) were better at conveying coherence in their mother tongue, Chinese, than in their B language, English.

This finding from both trainee and professional interpretations supports one of the major suggestions on language combinations by professional bodies such as SCIC and AIIC. Professional interpreters are advised to work into their A language (native language), where possible, rather than into their B or other passive (C) languages.

The professional norm corresponds with the feelings of the interpreters themselves. Bartłomiejczyk (2004) conducted a survey of interpreters' point of view on the directionality of SI. She reports that professional interpreters feel more confident of working into their A language. Trainee interpreters, on the contrary, often feel more at ease when working into their B language. This way there are fewer comprehension problems.

My findings substantiate the point of view of professional interpreters: interpretations into the A language convey better textual coherence.

6.1.4. Trainee interpreters start at same level

A second important observation is that the ratios of RST tree weight to total words of the text in the interpretations of both trainee groups (Control and Test)

were initially very similar. This is not surprising. The Test group had not yet being exposed to the feedback grid and had not yet had their attention explicitly drawn to the significance of coherence. At this point, both trainee groups had received the same treatment.

In English interpretations (Figure 42) the average ratios for The Control and Test groups for Speech 1 are 14.6% and 15.7% respectively. The average ratio for both groups in Chinese interpretations of Speech 1 is exactly 22.6% (Figure 41). This similarity of these ratios shows that both trainee groups started their training with equal capabilities of conveying coherence. Thus, in addition to the controlled procedures for recruiting trainee interpreters, such as the requirement about language proficiency and skills (Chapter 5 Methodology), these figures further validate my selection of research subjects.

6.1.5. 'Heavier' RST trees of professional interpretations

A third finding is that professional interpretations have higher ratios of RST tree weight to total words than trainee interpretations into both Chinese (Figure 41) and English (Figure 42). This follows from the fact that the weight of an RST tree representing a professional interpretation is consistently greater than that of trainee interpretations.

From Figure 49 (C>E interpretations: RST tree weight vs. total words – Professional interpretations as benchmark) I observe that the difference in the ratios of the interpretations of the trainee (Control group) and professional group was almost constant. Trainee interpretations might have the same length as professional interpretations, but the weighting of the RST trees of their interpretations only accounted for about 60% of that of the professional interpretations. Moreover, this situation does not improve through the period of training. When working into Chinese (Figure 41), the ratio gap between the professional and the trainee (Control) groups was narrower in Speech 1 but became wider in both Speech 2 and Speech 3. This does not necessarily mean that the trainee interpretations were more coherent prior to training (with tree wt/total wds ratios closer to those of the professionals). In section 6.2.3., alternative, substantial reasons are presented.

Nevertheless, the widening gap between professional and trainee interpretations of Speech 2 and Speech 3 in Chinese drew my attention. For one thing, as discussed previously in 6.1.3, I believed that interpreters should be able to work better into their mother tongue, Chinese in this case. Yet the big ratio gap between professional and trainee interpretations into Chinese intrigued us. I investigated further and found

several possible explanations. Firstly, trainee interpreters' comprehension of the source speeches in English (Speech 2 and Speech 3) might not be as complete as that of professionals.

Seleskovitch points out that, 'absence of comprehension results in immediate oblivion, whereas comprehension is synonymous with retention' (1994: 32). Therefore it is plausible that, even with note-taking, trainee interpreters still failed to catch as much information from the source speech as professionals. As a result, even in their mother tongue, they were not able to reproduce the coherence of the source speech as fully as were the professionals. What they managed to capture and convey, however, were small fragments of coherence and local cohesive features. Professionals, on the other hand, produced a more global coherence. This is reflected in the higher tree wt/total wds ratios of their interpretations.

In short, when interpreting into English, trainee interpreters have a 'comprehension bonus' (Donovan, 2004) and as they would have little difficulty in comprehending the speeches in Chinese, their A language. Yet comprehension of the source speech does not guarantee good interpretation. Comprehension does not necessarily lead to deep processing of information and individuals have different abilities in this aspect. Also, even with the same level of information processing, individuals with different language proficiency would represent the message differently. For instance, interpretations in English by trainees and professionals could be both acceptable regarding the accuracy and completeness of message, but they are likely to differ greatly in terms of discourse structure. As I saw in the RST trees and bushes in Figure 43, what distinguishes professional from trainee interpretations was the connectedness and the depth of the discourse structure, i.e. the coherence.

When working into Chinese, professional interpreters are still better in conveying coherence than trainees. Professionals are able to comprehend the speeches in English better and are also likely to have better awareness of the subject matter than trainees. Thus professionals have an advantage over trainees from the first step of the interpreting process. Consequently, it is no surprise that professional interpretations into Chinese are better connected and more logically structured than those of trainees.

6.1.6. RST relations vs. RST tree weight

The ratio of the number of RST relations to RST tree weight also caught my attention. This ratio (relations/tree wt) gives an indication of the significance of the contribution of the number of RST relations to the weighting of an RST tree. Yet it is

important to remember that to calculate the weight of an RST tree, both the number of RST relations and the depth of the tree structure need to be taken into account. In other words, a higher ratio in this case indicates less contribution from the depth of the structure to the overall weight of the RST tree.

This is also an important indicator to distinguish professional and trainee interpretations. In professional interpretations, the weight of the RST trees comes from both the number of the relations and the depth of the tree structures, while in trainee interpretations, the RST tree weight is largely depending on the number of relations. Later, in Section 6.4 (Coherence profile), we will see that one of the most striking differences between professional and trainee interpretations lies in this ratio.

In Figure 48, I present the ratios of the number of RST relations to the RST tree weight in Chinese-English interpretations. I observe that in the Control group, the ratios across the three speeches were constantly high, at around 90% in all three speeches. The Test group, by comparison, varied much more. In Speech 1, the ratio is 82.3%, which as we would expect was rather close to that of the Control group. In Speech 2, it dropped to 57%, which was close to that of the Professional group. In Speech 3, the ratio rose to 77%. The ratios of the Professional group, on the contrary, were constantly lower than both trainee groups, but show a similarly-shaped pattern to the Test group. In Speeches 1 and 2, both ratios were lower than 50% and in Speech 3, the ratio reached around 60%. There was a difference of around 50% between the Professional and the Control groups in the first two speeches. In Speech 3, the gap narrowed to around 27.7%.

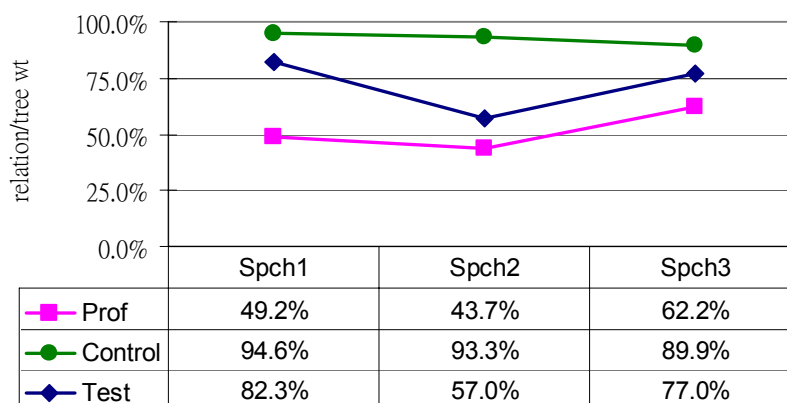


Figure 48 C>E interpretations: total relations vs. RST tree weight

The ratios here reflect the phenomenon of ‘trees and bushes’ that I reported previously (6.1.1.). For one thing, they show why trainees produced bush-like structures while professional interpretations made trees. The bulk of the weight of the RST trees of trainee interpretations, especially those of the Control group, comes

from the number of RST relations. In other words, the depth of the structure did not make much contribution to the weight of the RST tree. The depth of RST structure of professional interpretations, by contrast, contributed a significant part of the total score of their RST trees.

Secondly, this set of ratios also demonstrates different approaches to conveying coherence by professional and trainee interpreters. It is clear that, in trainee interpretations, discourse was structured at a local level: the ratio of the number of RST relations to the total words (relations/total wds) was high and the corresponding RST annotation resembled bushes rather trees. The RST structure of professional interpretations, on the other hand, was defined by both the number of RST relations and the depth of the structure: the same ratio (relations/total wds) was about 50% lower than that of the trainee interpretations (Control group).

Thirdly, the Test group's ratios are also interesting. Compared to Speech 1, the ratio dropped dramatically in Speech 2. For Speech 2, the difference between the Test group and the Professionals is relatively small (approximately 14%) when compared with the difference between the Test group and the Control group. The gap between the Test group and the Professionals remained steady in Speech 3 (15%). In other words, the RST structures of interpretations by the Test group are more globally connected as trees in both Speeches 2 and 3, but still not yet as well-formed as those of the Professional interpretations.

It is no surprise that the ability to convey coherence for trainee interpreters, like other skills, is acquired over time. However, the Control group made much less progress over the same period. Therefore I can conclude that the process of skill acquisition can be facilitated (as happened for the Test group) by explicitly drawing attention to the significance of key features, such as coherence.

6.1.7. Quality awareness facilitates performance

I can find further support for my claim, that trainees progressed faster in conveying global coherence if their attention was drawn to it explicitly, in the trends in the RST tree weight/total words ratio over time between Test and The Control groups. As I have just seen, the weighting captures both relatedness and complexity; combining this score with total word count allows us to directly compare speeches of different length.

I normalised the ratios of RST tree weight to total words by setting the professional ratios to 100 and adjusting the ratios of trainee interpretations according

to this benchmark. Let us first consider Chinese-English interpretations. Figure 49 shows a constant gap between the professional group and the Control group.

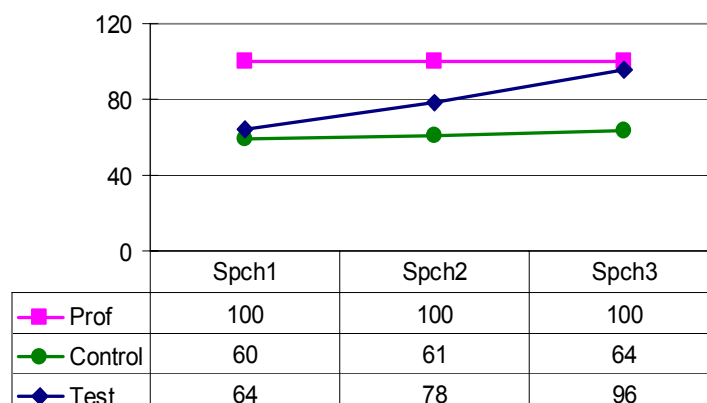


Figure 49 C>E interpretations: RST tree weight vs. total words – Professional interpretations as benchmark

The score for interpretations by the Control group achieved around 60% of the Professionals. However, the Test group showed a clear trend towards converging with the Professional profile. In Speech 1, both trainee groups achieved a similar score, while in Speech 2 the Test group showed a marked improvement, which continued in Speech 3 – the most argumentative of all, with a rather complex discourse structure – where the score was very close to that of the Professional group.

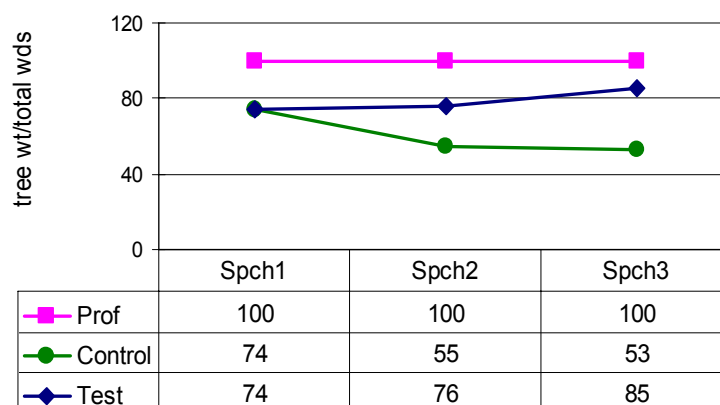


Figure 50 E>C interpretations: RST tree weight vs. total words – Professional interpretations as benchmark

In English-Chinese interpretations (Figure 50), likewise, the gap between Professional and the Control groups was very steady, in particular in Speech 2 and Speech 3. Speech 1 was meant to be very straightforward, and interpreters did not need any special preparation to comprehend the speech. Therefore, the interpretations by both The Control group and Test group were not too far away from the professional interpretations. In Speech 2 and Speech 3, I saw the difference

between the two trainee groups. The noticeable gap between the performances by the Control group and the Professional became wide, while the Test group showed a rise towards the Professional. In Speech 3, the most complicated and challenging speech among the three, the Test group still managed to reach a score of 85.

As described in Chapter 5, the major difference between the two trainee groups lies in the introduction of the feedback tool and specific attention drawn to the realisation and significance of coherence of their interpretations. Apart from these two conditions, the Test group was recruited and trained in the same way as the Control group. I suggest that the introduction of feedback tool and attention given to coherence explained the significant improvement of the Test group.

Carrying out a long-term study of the positive impact of the introduction of the feedback grid and the overall progress of interpretation was beyond the scope of the present project. However, from my RST analysis, I witnessed the development of coherence in interpretations in the Test group. I suggest that their understanding of quality criteria and of coherence in particular, was improved as a result. My results also show that awareness of coherence leads to better interpretations. Of course I do not claim that trainees can develop into professionals after a few months of training. My analysis does show, however, that to produce interpretations with a degree of coherence similar to that of professional can be developed, when trainees are explicitly guided by the peer feedback tool (Hartley et al., 2004).

6.2. Overtly-marked relations with explicit markers

Despite the fact that RST relations are not always marked explicitly, I observed that professional interpretations in both Chinese and English generally had higher ratios of overtly-marked relations to total RST relations (overt ratios) than the corresponding trainee interpretations (Figure 51 and Figure 52).

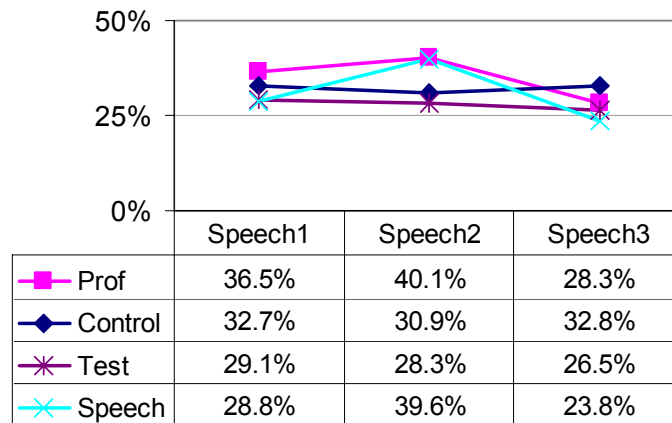


Figure 51 Overt ratios of C>E Interpretations

The overt ratios of professional interpretations were also higher than those of source speeches. Indeed this is a very common feature of translated and interpreted text. As explained in Chapter 5 (Methodology), I chose to study the use of conjunctions as explicit markers. The use of other means which also contribute to the cohesion of a text such as reference, substitution and ellipsis, were not considered.

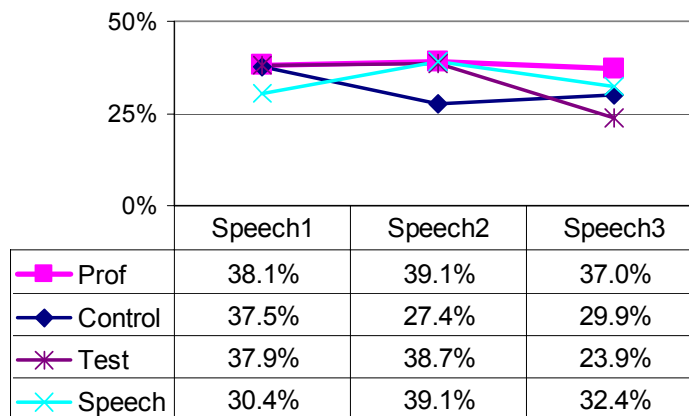


Figure 52 Overt ratios of E>C Interpretations

6.2.1. Overt ratios in English interpretations

In English interpretations of the first two Chinese speeches (Speech 1 and Speech 2), professional performances had higher overt ratios than both groups of trainee interpretations (Figure 51). The overt ratios in professional interpretations also fluctuated from speech to speech. Unlike professional performances, the overt ratios of trainee interpretations were rather flat, without much variation. Moreover, they did not follow the general trend of the overt ratios of the source speeches. This might be due to their limited proficiency in English, their B language: trainees' use of English conjunctions for overt marking appeared to be less efficient than the professionals.

In Speech 1, all interpreters worked without taking notes. The source speech was easy to comprehend and its content was not technical. The interpretations by the professional and the two trainee groups were all more overtly-marked than the source speech. In Speech 2, on the contrary, when note-taking was allowed but the speech contained some technical facts, trainee interpreters struggled and failed to reproduce some of the major arguments of the source speech. Speech 3, by comparison, had a more complex discourse structure. The professional interpretations of Speech 3 were more overtly-marked than the source speech.

Its RST tree description comprised a mixture of asymmetric relations (Nucleus/Satellite) and multinuclear relations in a very deep structure. Interpretations of Speech 3 by the Control group, however, had an even higher overt ratio than the Professional group. The interpretations from this group are more marked by conjunctions. This is likely to be the result of transferring some implicit transitions of arguments to explicit twists of discourse. In addition, other cohesive devices such as reference, substitution and ellipsis of the speech may have been transferred as conjunctions. It should be noted that, despite being higher than that of the speech, the overt ratio of professional interpretations followed the general features of the source speech more closely than did the trainee interpretations.

6.2.2. Overt ratios in Chinese interpretations

In English-Chinese interpretations (Figure 52), the overt ratios of professional performances were consistently higher than those of the trainees in all three speeches. The overt ratios of professional interpretations were also consistently higher than those of source speeches. It should also be noted that, professional interpretations in Chinese had a fairly consistent overt ratio across the three speeches (38.1%; 39.1% and 37.1%). This could be related to the similarity of the three English speeches. All three speeches were informative with clear structural organisation. Therefore professional interpreters consistently adopted a similar level of explicit markers to reproduce the explicitness of the text. In addition to the similarity of the source speeches, professionals with many years of practice as conference interpreters, might have developed a stable approach to use Chinese conjunctions as explicit markers to signpost their interpretations.

Trainee interpretations, on the other hand, showed a different picture. Both groups had almost the same ratios as the professionals for Speech 1. Then the ratios diverged in both Speech 2 and Speech 3. The interpretations of the Control group, for instance, had the lowest ratio in Speech 2. The ratio for this group picked up in Speech 3. The Test group, despite having a very close ratio to that of the

professional interpretations in Speech 2, had the lowest ratio of all groups in Speech 3. This dramatic fluctuation of ratios reflects some important issues. For one thing, trainee interpretations were simply not as clearly and explicitly marked as professional interpretations. Although Chinese is their native language, the trainees were not as experienced as the professionals in presenting arguments in Chinese in response to an English speech. In other words, they were still acquiring interpreting skills, which include handling explicit markers properly even in the mother tongue. As a result, their interpretations were not as explicitly marked as professional performances, and therefore had lower overt ratios.

In short, trainee interpretations in Chinese, the trainees' native language, varied greatly in terms of overt marking. Professional interpretations, on the other hand, are consistent in both use of overt marking and in following the general trend of the ratios of source speeches.

6.2.3. Higher overt ratios in trainee interpretations at the beginning

It is interesting to observe that, both groups of trainee interpretations had higher overt ratios than those of first source speech (Speech 1) in both Chinese and English. In both English and Chinese, the content of Speech 1 was of a general nature and could be easily understood without any specific background knowledge. All interpreters were asked to memorise the speech without taking notes. They produced their interpretations from memory.

As both groups of trainees had had four weeks of training on memory and public speaking before being recorded for Speech 1, they were already able to memorise the major structure and arguments of a speech. They were trained to reproduce this structure with the aims of producing coherent interpretations. They had also acquired other basic skills: as such they were similarly capable as professionals in terms of handling non-technical speeches such as Speech 1.

6.2.4. Double conjunctions in English

One noticeable feature of trainee interpreters' lower proficiency in English is the occurrence of 'double conjunctions'. This is a form of Chinese language interference which can be observed in the interpretations of both trainee groups.

As Leung notes, in Chinese, when a subordinator is used to introduce a concessive clause, 'it is required to introduce a balancing clause with another matching conjunction' (2005: 12). For instance, 雖然 [sui1ran2] (although) and 但是 [dan4shi4] (but) often appear together in a Chinese sentence, but such structures are

grammatically wrong in English. Other combinations of double conjunctions include ‘because...therefore/so’, ‘if...then’, ‘even if...still’, ‘since/as long as/as soon as...then’, ‘although...but/nevertheless’. This type of syntactic structure in Chinese can help Chinese speakers to process related discourse relations, such as concession, cause and effect, conditions, etc.

Leung also notes that ‘dual/double conjunctions are ‘a common feature of Chinese speakers’ English’ (2005: 27). This mother-tongue interference means that, ‘non-native speakers directly transfer the conjunctions and connectors from their first language (in this case, Chinese) to their second language (English)’, and can bring about ‘overuse of connectors’ (Chan, 2004 cited in Leung, 2005: 25).

In trainees’ English interpretations, I observe such interference of Chinese double conjunctions. Some examples from my data are presented in Table 24 below.

Source speech	Literal translation	Interpretations
現在在台灣，不到三歲的小孩子都在開始拚命地學英文了， <u>因為</u> 家長希望孩子不要輸在起跑點上	Now in Taiwan, children younger than 3 years old have started learning English, <u>because</u> their parents do not want them to lose at the start point.	<u>Because</u> in Taiwan// parents don't want their children to lose at the beginning,// <u>so</u> they want them to learn more. (-Spch1ce C4)
<u>因為</u> 有不少人透過假造文件偷渡到中國來	<u>Because</u> quite a lot of people by using false document smuggled themselves to China.	<u>Because</u> they got the illegal documents from some criminals//, <u>so</u> they have the access to emigrate to China. (-Spch2ce T3)
大家都知道，在京都高峰會上面，我們首度訂定了國際的目標，要來降低二氧化碳的排放量。當然，這一步是踏對了方向。 <u>但是</u> 我們可以發現在京都高峰會上面所設定的目標，其實低的可憐	I all know, in Kyoto Summit, I set up a goal to reduce CO2 emission for the first time. Of course it was a right move. <u>But</u> I find that the goal I set in the Summit was pathetically low.	<u>Though</u> I have signed the Kyoto Treaty//and that was the first treaty// protocol// to do the target// to reduce the emission of CO2.// It is a good way//a good direction.// <u>However</u> this kind of goal of the emission is very unpractical.(-Spch3ce T2)

Table 24 Examples of dual-conjunctions from trainee English interpretations

From the table above, it is clear that the interference did not come directly from the source text, as there were no dual conjunctions in those source text segments.

The trainees, however, after comprehending the Chinese text, reproduced the message in English using the Chinese syntactic structure, by inserting double conjunctions into their English interpretations. It is very likely that the trainee interpreters intended to mark the discourse relations to make their language comprehensible.

The ungrammatical use of double conjunctions to mark discourse relations might cause confusion for English speakers. Moreover, as noted in Section 5.2.3, when both ‘although’ and ‘but’ are used at the same time, it makes RST annotation and counting ‘good’ explicit markers problematic. The span marked by “although” could be the satellite and the nucleus at the same time, and the relation holding the two spans can be seen either Concessive or Antithesis. In 5.2.3, I set five basic rules to choose ‘good explicit markers’. The fourth rule is relevant here. It states, 4) *Redundant use of markers will only be counted once.*

Moreover, although it may be problematic, annotating texts containing double conjunctions with RST is possible. It should be remembered that, what connects the text spans (nuclear and satellites) are the rhetorical relations between them. These are often, but not always, signalled by discourse markers like conjunctions. Conversely, despite the presence of double conjunctions, rhetorical relations still hold between spans.

In order to decide which one of the double conjunctions to keep, I first assigned the rhetorical relation according to context and the apparent logic of the discourse; this assignment effectively selected which conjunction would count. For instance, I observed that ‘although...but’ was used mostly to indicate a concessive relation in my data, and I could assign Concession as the RST relation holding between the two spans marked by ‘although’ and ‘but’. In this case, ‘but’ became redundant and was ignored. Figure 53 gives the RST tree-representation of the ‘although...but’ example I presented earlier in Table 24.

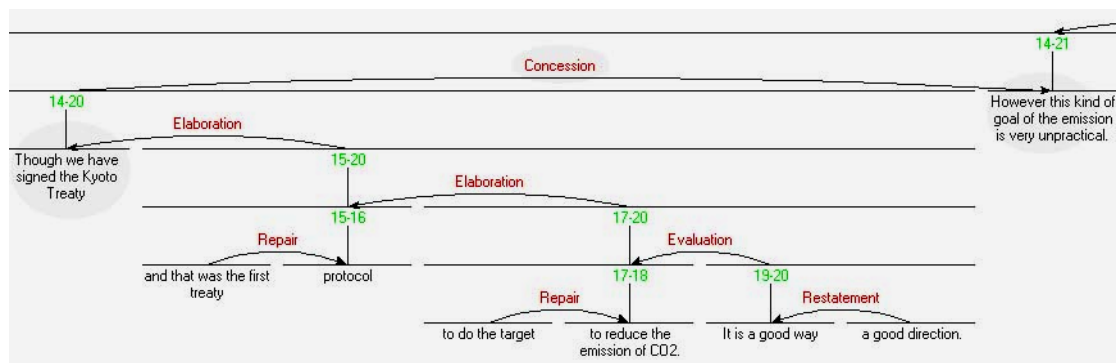


Figure 53 RST tree of double-conjunctions of ‘although...but’

Comprehension of speeches which comprise such flawed structures might not be easy, but it is possible. As we saw in Chapter 2, when people try to understand a speech, they make effort to make inferences based on their knowledge and experience. They may also filter out elements, such as redundant conjunctions, to understand what is being said.

Nevertheless, comprehending English speeches which contain double-conjunctions takes extra effort, and such grammatical interference from Chinese should be avoided.

My analysis showed that the use of double conjunctions in my data was not as prevalent as I had anticipated. By manually checking all the transcripts of all the interpretations of the three speeches (33 texts in total), I found just 14 occurrences of double conjunctions: 11 ‘because/so’, 2 ‘although/but’ and 1 ‘if/then’. Table 25 describes the distribution of these occurrences across speeches.

	Speech 1	Speech 2	Speech 3	Total	
Because/...so	8	2	1	11	14
Although/...but	0	0	2	2	
If/...then	1	0	0	1	

Table 25 Occurrences of double-conjunctions in trainee interpretations

In other words, the interference of double conjunctions from Chinese appeared to be rather noticeable at the initial stage of training, but not as widespread in the end. Leung’s corpus-based study of the use of English conjunctions in writing by Chinese university students in Hong Kong also shows that the use of double conjunctions is not a frequent phenomenon in the writing of competent English users (2005: 27). In short, as their English language competence improves and by monitoring their own performances carefully, trainee interpreters can deal with the interference of double conjunctions and this will help to produce more readily understood interpretations.

6.3. Repair

As discussed in Section 5.2.2.1, repairing is a common feature of natural utterance, and it is a mechanism for speakers to edit their output and fix problems (Petite, 2005: 28). It occurs in about 10% of spontaneous utterances (Nakatani & Hirschberg, 1994). It also occurs in interpretations.

The figure below (Figure 54) contains two examples of repairs observed from trainee interpretations in both Chinese and English.

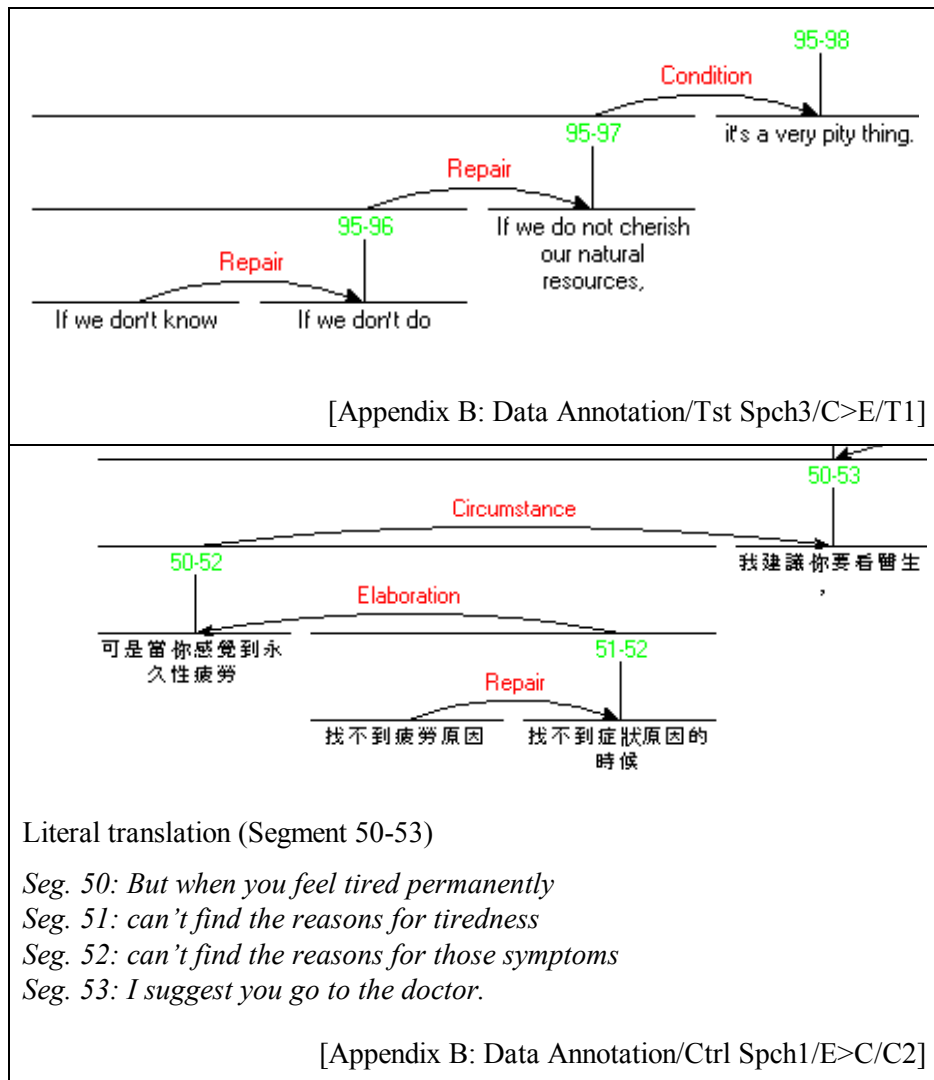


Figure 54 Examples of repairs (B)

Other literature suggests that, despite causing disfluency, repairs in speeches might not always hinder listeners' comprehension of the utterances. It might nonetheless cause a certain degree of disruption to textual coherence. Thus, it is important to take this phenomenon into consideration when analysing relevant factors regarding the coherence of a text.

Counting how many repair relations there are in a text is not sufficient to account for the impact of self-correction on coherence. It is more sensible to look at the occurrences of repair relations against the total relations in a text (Table 26 and Table 27).

E > C (Prof)	Spch1			Spch2			Spch3		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
Relations	65	59	38	62	77	74	76	64	70
Repair	5	0	2	1	1	1	4	0	5
Repair/total rel	7.7%	0.0%	5.3%	1.6%	1.3%	1.4%	5.3%	0.0%	7.1%

E > C (Control)	Spch1				Spch2				Spch3			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
Relations	36	55	71	45	57	49	49	47	55	47	49	70
Repair	1	1	0	2	0	0	3	0	1	0	0	0
repair/total rel	2.8%	1.8%	0.0%	4.4%	0.0%	0.0%	6.1%	0.0%	1.8%	0.0%	0.0%	0.0%

E > C (Test)	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
	Relations	86	64	76	81	71	47	63	45	47	52	88
Repair	8	3	6	11	12	9	6	5	7	2	12	6
repair/total rel	9.3%	4.7%	7.9%	13.6%	16.9%	19.1%	9.5%	11.1%	14.9%	3.8%	13.6%	11.1%

Table 26 ‘Repair’ vs. total relations in English-Chinese interpretations

C > E (Prof)	Spch1			Spch2			Spch3		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
Relations	66	59	37	73	69	43	78	84	74
Repair	8	5	2	3	7	3	2	4	4
repair/total rel	12.1%	8.5%	5.4%	4.1%	10.1%	7.0%	2.6%	4.8%	5.4%

C>E (Control)	Spch1				Spch2				Spch3			
	C1	C2	C3	C4	C1	C2	C3	C4	C1	C2	C3	C4
Relations	52	66	61	69	50	43	41	52	54	70	45	50
Repair	4	7	6	15	6	5	6	3	1	1	2	1
repair/total rel	7.7%	10.6%	9.8%	21.7%	12.0%	11.6%	14.6%	5.8%	1.9%	1.4%	4.4%	2.0%

C>E (Test)	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
	Relations	60	57	69	73	62	39	60	48	100	89	87
Repair	12	7	10	8	20	7	9	7	14	5	10	8
repair/total rel	20.0%	12.3%	14.5%	11.0%	32.3%	17.9%	15.0%	14.6%	14.0%	5.6%	11.5%	14.0%

Table 27 ‘Repair’ vs. total relations in Chinese-English interpretations

According to the two tables, I averaged the ratios of Repair to total relations (repair ratio) for each group in each speech and produced two figures (Figure 55 and Figure 56) to present general pictures of the occurrence of Repair in my three groups of interpreters: Professional, trainee (Control and Test) groups.

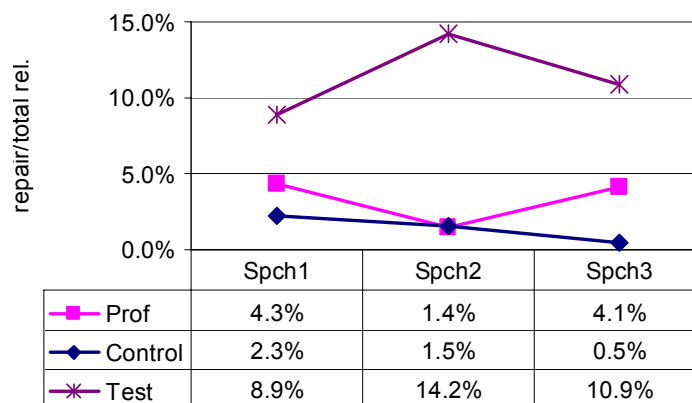


Figure 55 E>C Interpretations: Repair vs. total RST relations

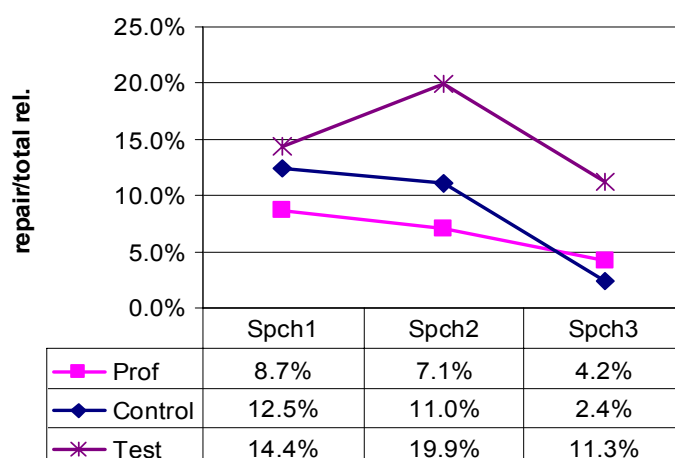


Figure 56 C>E Interpretations: Repair vs. total RST relations

6.3.1. More repairs in English than in Chinese interpretations

From these figures, I observed first of all that there were more occurrences of repair in English than in Chinese, in both professional and trainee interpretations. The repair ratios of the professional interpretations into Chinese over the three speeches were well below 5% (Figure 55). For the Control group, the repair ratios of the Chinese interpretations of the three speeches were below 2.3%, but the repair ratios went as high as 12.5% for Speech 1 when these trainees worked into English. It is understandable that they made more effort to monitor themselves when working into English, their B language. This does not mean that they did not monitor themselves at all when working into Chinese. For the professional interpretations, the repair ratios were rather stable, suggesting that self-monitoring and self-correction was in operation constantly in their interpretations, regardless of which language they worked into. In addition, with experience and expertise acquired over time, the professional interpreters did not have as many errors to correct as the trainees and so the repair ratios of professional interpretations remained low.

6.3.2. Trainee difference in repair

I observed that the two trainee groups were very different in terms of repair. Overall, the repair ratios of the Test group were much higher than those of the Control group. When interpreting into Chinese (Figure 55), the Control group had even less self-correction than the Professional group, while the Test group maintained a very high occurrence of repair in their interpretations. I also observed that after the feedback grid had been introduced (after recording Speech 1), the repair ratios of the interpretations in both Chinese and English by the Test group rose noticeably. This is perhaps because the Test group had been reminded of the significance of self-monitoring for quality assurance in interpretations. In Chinese interpretations, the repair ratio for the Test group reached at its highest at 14.2% for Speech 2. In English interpretations of Speech 2, the ratio reached 19.9%. In Speech 3, the Test group's repair ratios in both Chinese and English interpretations dropped to 10.9% and 11.3% respectively.

Interpretations by the Control group, on the other hand, showed a very different trend regarding self-correction. In English interpretations (Figure 56), the repair ratios kept decreasing over the three speeches and reached a low of 2.4% in Speech 3, lower even than that of the Professional group. One possible explanation for the drop in ratio is that trainees in the Control group had become more confident over time and, as a result, made fewer self-corrections.

6.3.3. Fewer repairs vs. better coherence

Utterances with fewer disruptions, such as repair, might contribute to better fluency, but not necessarily to better coherence. So far I have explored only one of many forms which may cause disfluency. I have not explored other forms of disfluency, e.g. filled pauses such as 'ums' and 'uhs' (Kormos, 1999; Schachter et al., 1991), and others described in detail elsewhere in the literature (Honal & Schultz, 2003; Kormos, 1999; Levelt, 1983; Nakatani & Hirschberg, 1994). Of course, I do not suggest that trainee interpreters become fluent, or even that they become more fluent than professional interpreters, simply by reducing the occurrence of repair.

In addition, when counting the occurrences of repair in interpretations, it is important to distinguish three different possible situations: (a) the interpretations contain few errors, and therefore fewer repairs; (b) the interpretations contain more errors, but these are not repaired, so there are still few repairs, and (c) the interpretations contain more errors, but these are repaired. In my dataset, it was observed that situation (a) occurred more often in professional interpretations.

Situation (b) occurred more in trainee groups. Yet, when such situations occurred (errors not repaired), the textual coherence would be affected, as connections between text spans might be lost. For instance, in the following figure (Figure 57), the un-repaired error meant that there was no rhetorical relation with neighbouring text spans.

Situation (c) was more prevalent in the Test group, and some trainees would correct themselves more than once for a single error. Thus, the occurrence of repairs is a lot higher in the Test group than in the Control group.

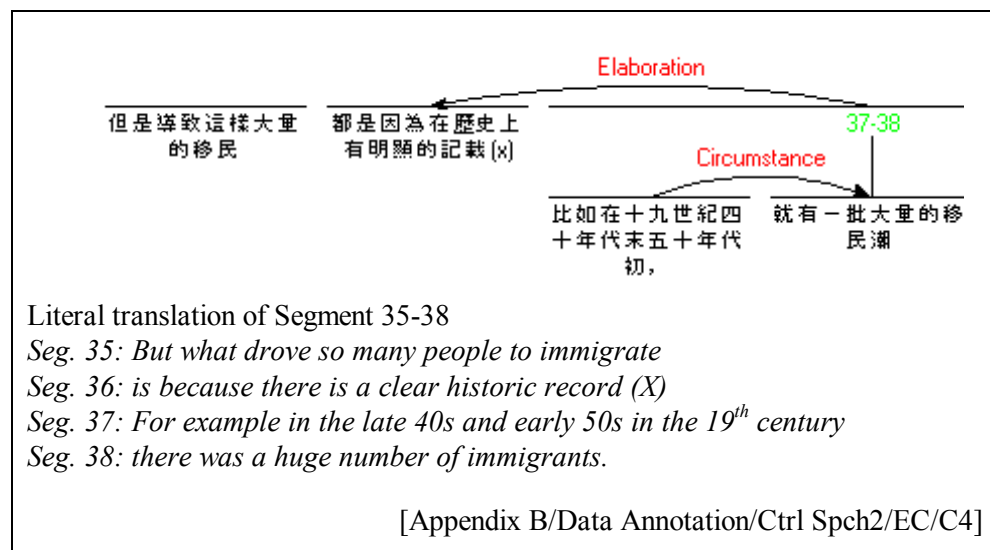


Figure 57 Example of an error not repaired

Nevertheless, it is important to recognize occurrences of repair when investigating the coherence of an interpretation, and to examine whether they have any significant impact on getting messages across in interpretations. I observed that using fewer self-corrections does not contribute significantly to better coherence. Indeed, people can talk nonsense for a long while without stops and self-corrections. This does not guarantee that their utterances are coherent. On the other hand, if people correct themselves too often with constant filled pauses, false starts, semantic, grammatical and syntactical alterations, and so on, their listeners are likely to be distracted and lose the thread of messages as a result. Thus, a high frequency of repair definitely impedes the flow of messages and, is very likely to jeopardise listeners' perception of the textual coherence as a result.

As discussed in the previous session on RST trees (6.1.1), the weight of a RST tree is determined by several factors: the numbers of relations, the depth of the discourse structure as well as the right-branching flow of a text. Repair relations are always annotated as left-branching structures at a given depth. In this way, the occurrence of repair is penalised when computing the RST tree weight of the text.

Therefore, the tree weight of a text with recursive repetition of repair relations will be heavily penalised.

Figure 58 and Figure 59 show the relationship between Repair and RST tree weight in both Chinese and English interpretations by three groups of subjects. We can see that the ratios of repair relations to total tree weight in the interpretations by the Test group were consistently higher than those of both the Control group and the Professional group. In Chinese interpretations (Figure 58), the repair relations of the interpretations by the Control group had little impact on the RST tree weight; but the Test group had consistently higher ratios of repair relations to the RST tree weight, at around 6% for all three speeches.

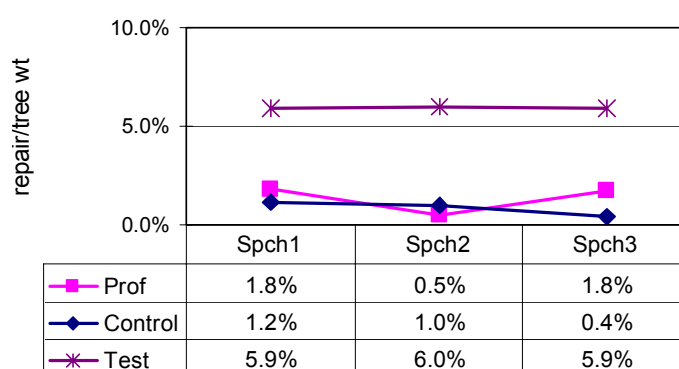


Figure 58 E>C interpretations: Repair vs. RST tree weight

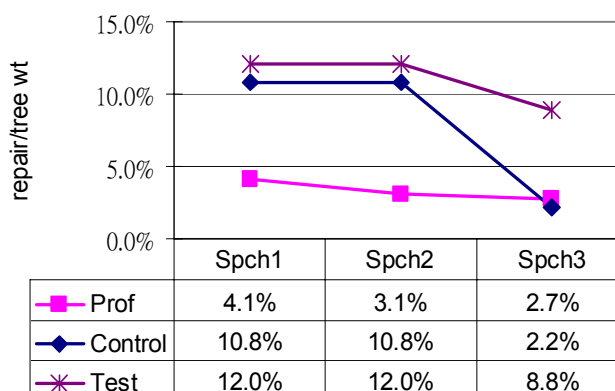


Figure 59 C>E interpretations: Repair vs. RST tree weight

In contrast to the Chinese interpretations, I observed that both trainee groups had much higher ratios of repair to the total tree weight for the English interpretations of Speech 1 and Speech 2 than the professional group (Figure 59). In other words, the weight of RST tree representations of these trainee interpretations had been heavily penalised due to the higher occurrence of repair relations. Compared to trainee interpretations, professional interpretations had fewer obvious interruptive repairs in English interpretations, and thus the weight of RST tree of the

interpretations was not greatly affected. In addition, I noted that the influence of repair relations on the RST tree weight of professional interpretations fluctuated very little across speeches. Figure 58 shows that the fluctuation of ratios (repair/RST tree weight) was less than 2% in Chinese interpretations, and about 2.5% in English interpretations.

To summarise, self-correction (represented in the relation of Repair) is commonly observed in all natural human discourse, and interpretations are no exception. Some view it as a sign of being disfluent, while others claim that we are able to process and comprehend such utterances. For interpreters, fluency is one of the major quality criteria, but has a lower priority than coherence and ‘making sense’. It is evident that a low level of disfluency of this kind does not impede the textual coherence of an interpretation. The stable repair ratio found in professional interpretations may even help listeners focus on new information, instead of on “given” messages (Brennan & Schober, 2001: 280). On the other hand, a fluent utterance with few repairs (in Chinese interpretations by the Control group) does not necessarily guarantee its coherence. All in all, the level of self-correction in professional interpretations is noticeably lower and more stable than that of trainee interpretations. This indicates that the mechanism of self-monitoring is in operation and helps in professional interpretations to achieve both a fluent and coherent output.

6.4. Coherence profile

As already explained in Section 5.2.4, I believe that a combination of the following factors is sufficient to present a ‘coherence profile’ for one or more speeches or interpretations:

- 1) the length of a text (total words),
- 2) the use of explicit markers (markers),
- 3) the number of RST relations (relations),
- 4) the number of overtly-marked relations by explicit markers, and
- 5) the weight of RST tree

These five factors are parameterised as a set of seven ratios which are reproduced in the table below.

Parameter	Abbreviation and rationales	
Explicit markers :: RST relations	Markers/relations	To see how the use of explicit markers contributes to the total RST relations, the RST tree weight and the total words of a text.
Explicit markers :: RST tree weight	Markers/tree wt	
Explicit markers :: total words	Markers/total wds	
RST relations :: total words	Relations/total wds	To see RST relations and RST tree weight in relation to the total words of a text.
RST tree weight :: total words	Tree wt/total wds	
RST relations :: RST tree weight	Relations/tree wt	To see how RST relations contribute to the tree weight of a text.
Overtly-marked :: RST relations	Overt-marked/RST relations	To see how explicitly RST relations are marked.

I use radar charts to present the coherence profile of a text. This approach facilitates comparison of different interpretations of the same speech.

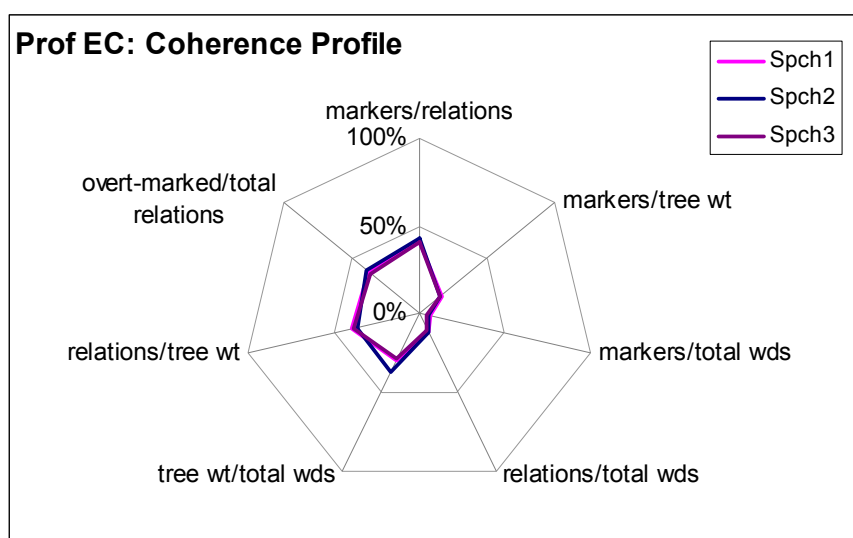
6.4.1. Benchmarking and benchmark validation

In order to compare different coherence profiles, I need to establish a benchmark. I took it as axiomatic that the output of my professional interpreters was “good” and would serve as a benchmark against which to compare trainee interpretations. According to my initial observations, the coherence profiles of professional interpretations present two noticeable characteristics: 1) professional

interpretations are rather stable without much deviation across speeches; 2) the coherence profiles had a close match with the source speech. To support my observations on these two characteristics, I examined the coherence profiles of professional interpretations on radar charts, firstly across speeches and among individuals, and secondly in comparison with the source speech.

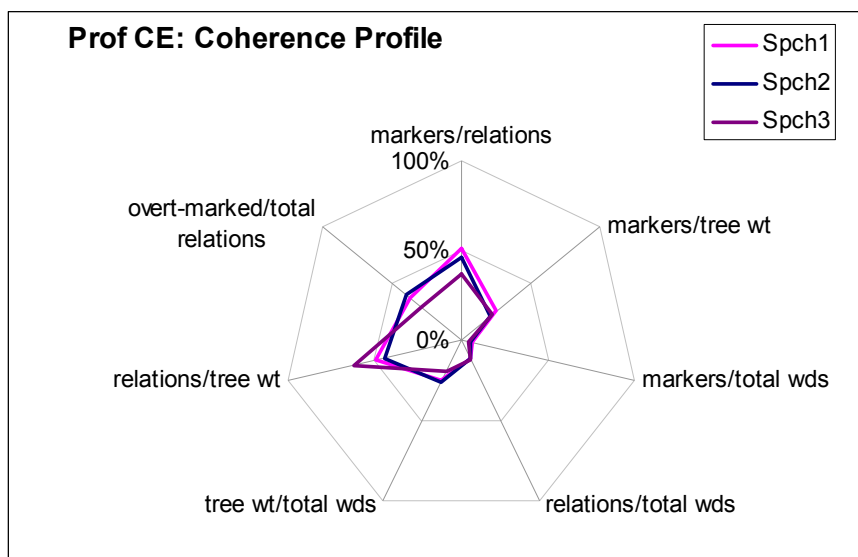
6.4.1.1. Professional interpretation on its own

First of all I plotted the coherence profiles of professional interpretations of all three speeches in both Chinese and English onto radar charts (Figure 60 and Figure 61). From the radar chart of English-Chinese interpretations (Figure 60), I noted that the coherence profiles of the professional interpretations of the three speeches were all quite similar. Indeed the specific ratios given on the table below the chart shows that the value of each parameter did not vary much across speeches.



Prof E>C	Spch1	Spch2	Spch3
markers/relations	40.9%	43.5%	40.5%
markers/tree wt	16.0%	15.5%	15.4%
markers/total wds	6.0%	5.1%	4.3%
relations/total wds	12.0%	11.6%	11.1%
tree wt/total wds	30.6%	36.1%	28.8%
relations/tree wt	39.7%	36.9%	39.0%
overt-marked/total relations	38.1%	39.1%	37.0%

Figure 60 Coherence profiles of professional English-Chinese interpretations



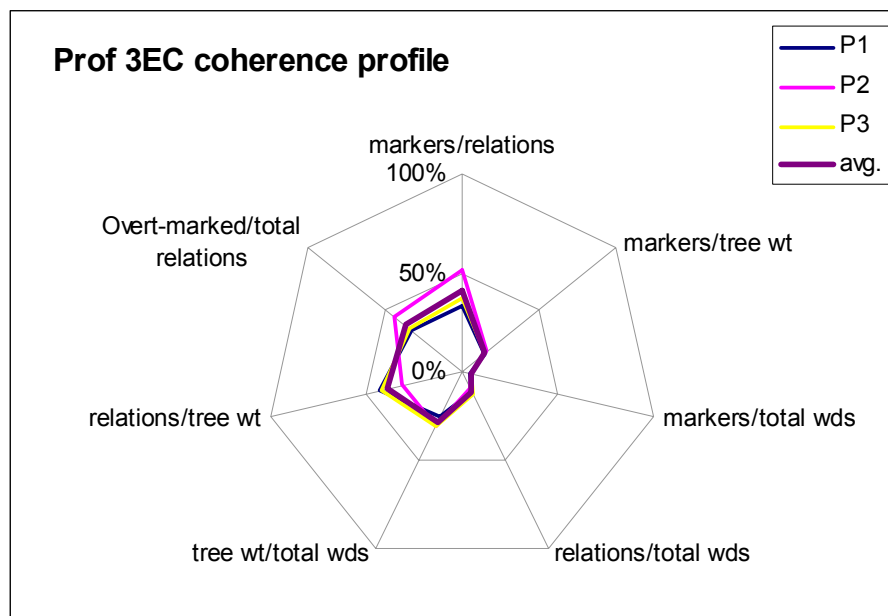
Prof C>E	Spch1	Spch2	Spch3
Markers/relations	50.6%	45.7%	36.4%
Markers/tree wt	25.6%	20.0%	22.6%
Markers/total wds	6.0%	5.1%	4.3%
Relations/total wds	11.7%	11.2%	11.8%
tree wt/total wds	24.5%	26.5%	19.0%
Relations/tree wt	49.2%	43.7%	62.2%
overt-marked/total relations	36.5%	40.1%	28.3%

Figure 61 Coherence profiles of professional Chinese-English interpretations

In Chinese-English interpretations (Figure 61), the coherence profiles for Speech 1 and Speech 2 were also similar. Speech 3, on the other hand, showed a rather different profile from the other two, with higher ratios on the parameter of relations/tree wt. and slightly lower ratios on makers/relations and overt-marked/total relations. These two parameters, markers/relations and overt-marked/total relations, were supposed to have a strong connection, as overt marked relations were characterised by the use of explicit markers in the forms of conjunctions. As a result, the more markers there were in a text, the more likely there would be more occurrences of overt-marked relations in a text, and vice versa. In addition, Speech 3 was an argumentative speech, which was very different from Speech 1 and Speech 2 which both had simpler structural organisation.

Apart from the deviations I noted in the Speech 3 interpretations in Figure 61, the striking consistency of coherence profiles for the Professional group across different speeches prompted us to investigate the performance of individual professionals further to see whether consistency existed among individual professional interpretations.

I therefore reviewed the statistics that make up the coherence profiles for individual interpreters' performances. I found that subject to minor deviations, profiles remained quite consistent. I took interpretations of Speech 3 in English by my three professionals as an example (Figure 62).



Prof 3E>C	P1	P2	P3	avg.
markers/relations	32.9%	51.6%	37.1%	40.5%
markers/tree wt	14.2%	16.3%	15.7%	15.4%
markers/total wds	3.7%	4.9%	4.7%	4.4%
relations/total wds	11.2%	9.6%	12.7%	11.1%
tree wt/total wds	25.8%	30.3%	30.1%	28.8%
relations/tree wt	43.2%	31.5%	42.2%	39.0%
overt-marked/total relations	32.9%	43.8%	34.3%	37.0%

Figure 62 Coherence profile deviations in professional interpretations

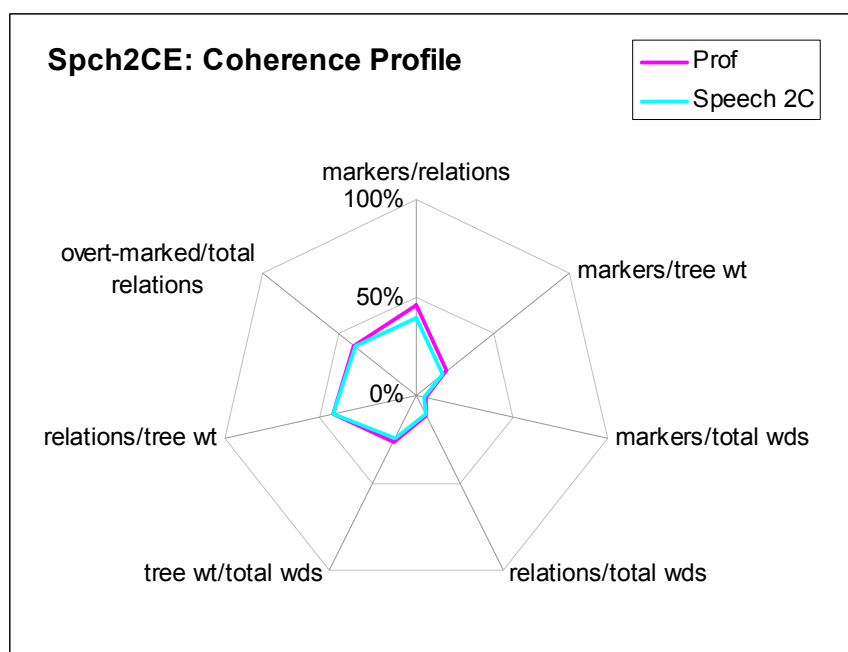
I found that there was little deviation in most of the profile parameters. From the radar chart, it was clear that the interpretations given by P1 and P3 had very similar profiles, and they also fit with the group profile rather well. P2's profile, by comparison, showed higher ratios for markers/relations and overt-marked/total relations but a lower ratio for relations/tree wt than other two professionals and the group average. The reason for the higher ratios was the strong connection between the use of explicit markers and overt-marked relations, as previously explained. The lower ratio of total number of RST relations to RST tree weight indicates that this interpretation had a deeper text structure. More 'depth' attracted a greater tree weight.

According to my analysis above, it was evident that there existed a good degree of stability of coherence in professional interpretations. It was also observed that,

despite the fact that variations occurred between individual performances as a natural phenomenon, there appeared to be a level of agreement among the coherence profiles of the group. In short, I can conclude that the consistency of coherence of interpretations within the Professional group and across speeches made professional performance a suitable benchmark to compare and contrast other trainee interpreters' performances.

6.4.1.2. Professional interpretation vs. source speech

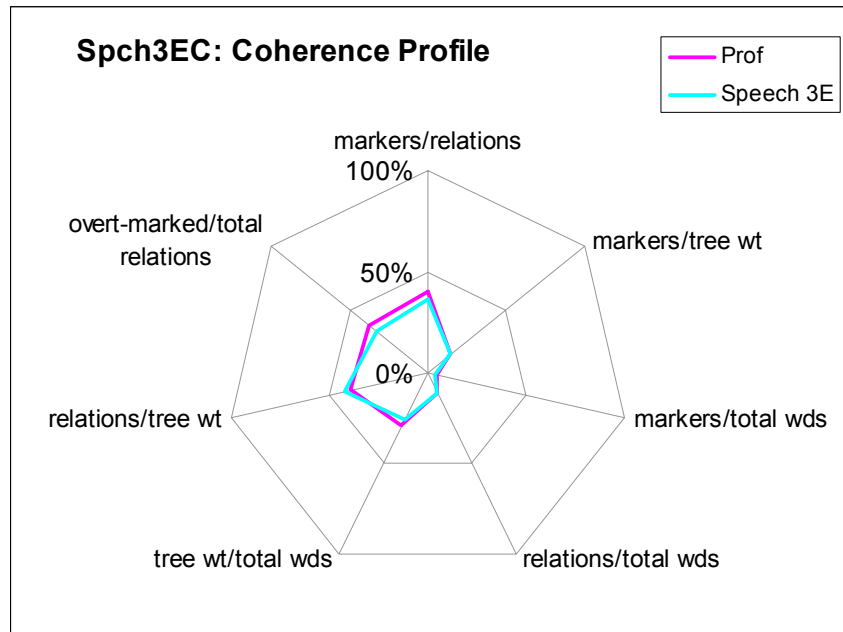
Having established the internal consistency of coherence in professional interpretations, I looked into the relation between professional interpretations and source speech. The coherence profiles of the professional interpretations and the source speeches were very similar. Moreover, the detailed figures show very little difference in the ratios for most parameters. This striking correspondence of coherence profiles was true for both language directions. This suggests that my coherence profile is suitable for languages as different as English and Chinese.



Spch2C>E	Prof	Speech
markers/rerelations	45.7%	39.6%
markers/tree wt	20.0%	17.0%
markers/total wds	5.1%	4.3%
relations/total wds	11.2%	10.9%
tree wt/total wds	26.5%	25.3%
relations/tree wt	43.7%	42.9%
overt-marked/total relations	40.1%	39.6%

Figure 63 Coherence profiles of professional interpretations and Speech 2C

The radar chart (Figure 63) shows the coherence profiles of professional interpretations into English and of the source speech in Chinese (Speech 2). In Figure 64, the radar chart shows the coherence profiles of professional interpretations of English Speech 3 into Chinese.



Spch3E>C	Prof	Speech
markers/relations	40.5%	36.5%
markers/tree wt	15.4%	15.3%
markers/total wds	4.3%	4.1%
relations/total wds	11.1%	11.2%
tree wt/total wds	28.8%	26.6%
relations/tree wt	39.0%	42.0%
overt-marked/total relations	37.0%	32.4%

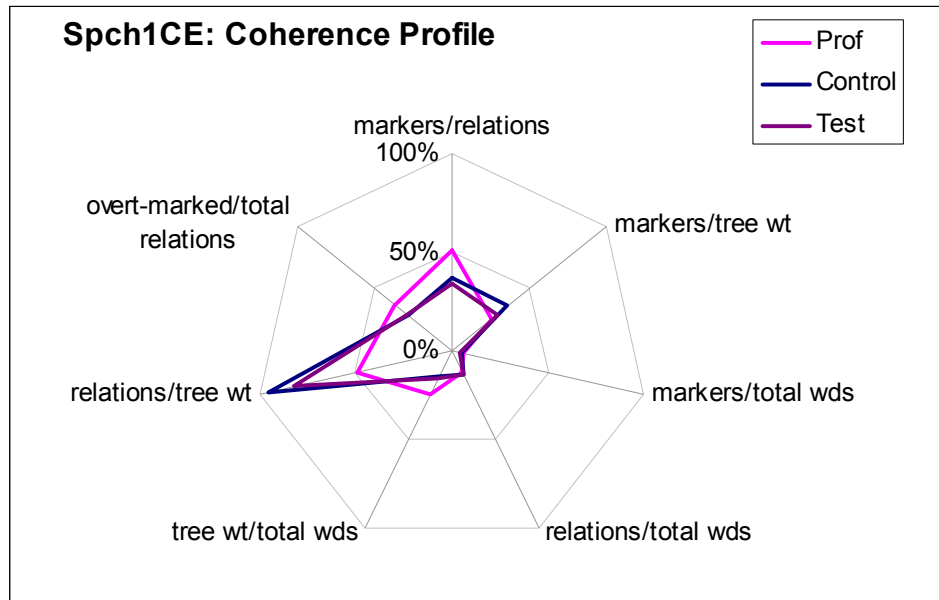
Figure 64 Coherence profiles of professional interpretations and Speech 3E

As is commonly observed in written translation - a translated text tends to be more explicitly marked than the source text – for the uses of explicit markers in professional interpretations here are noticeably higher than those of the source speeches. In Speech 2 (Figure 63), the markers/relations ratio of professional interpretations is 45.7%, while that of the speech is 39.6%. In Speech 3 (Figure 64), the same explicitation ratios for professional interpretations and the source speech were 40.5% and 36.5% respectively.

In short, having examined the coherence profiles of professional interpretations, the consistency of performances and the striking similarity of the profiles of interpretations with their source speeches, I conclude that the professional interpretations make a satisfactory benchmark.

6.4.2. Trainees' coherence profile

I compared the coherence profiles of trainee interpretations with the benchmark provided by the professional interpreters. The radar chart in Figure 65 illustrates some salient features of trainee interpretations and the way they differ from my benchmark in terms of coherence.



Speech 1C>E	Prof	Control	Test
markers/relations	50.6%	36.9%	34.4%
markers/tree wt	25.6%	35.5%	28.6%
markers/total wds	6.0%	4.5%	4.3%
relations/total wds	11.7%	12.9%	12.7%
tree wt/total wds	24.5%	14.6%	15.7%
relations/tree wt	49.2%	94.6%	82.3%
overt-marked/total relations	36.5%	28.3%	29.1%

Figure 65 Coherence profiles of professional vs. trainee interpretations

Both trainee groups have very similar profiles. This might reflect the similar level of language proficiency and interpreting skills of both groups.

Secondly, both groups have very high ratios of relations to RST tree weight: 94.6% and 82.3% for the Control and the Test groups respectively. This compares with just 49.2% for the professional interpretations and constitutes one of the most distinctive features of the coherence profile of trainee interpretations. As discussed previously in 6.1.6, a high score for this parameter indicates that the RST tree weight of the text came mainly from the number of relations the text comprises, rather than from the depth of the structure.

The phenomenon of self-correction might be an additional factor which contributed to the difference between the ratios of RST relations to RST tree weight for the professional and trainee interpretations. As discussed in 6.3, self-correction (Repair) appears to be more prevalent yet less stable in trainee interpretations than in professional interpretations, especially in English interpretations. If we were to take repair into consideration, the ratio of good relations (total relations minus repair) to RST tree weight would also be rather striking.

The trainees' markers/tree weight ratio was also high, with 35.5% and 28.6% for the Control and the Test groups respectively. The professionals scored 25.6% for this parameter. This ratio indicates the number of explicit markers to the total weight of RST trees. This does not necessarily mean that interpretations by trainee interpreters contained a higher number of explicit markers than professional interpretations. After all, their RST trees tend to use weigh less than those of professionals.

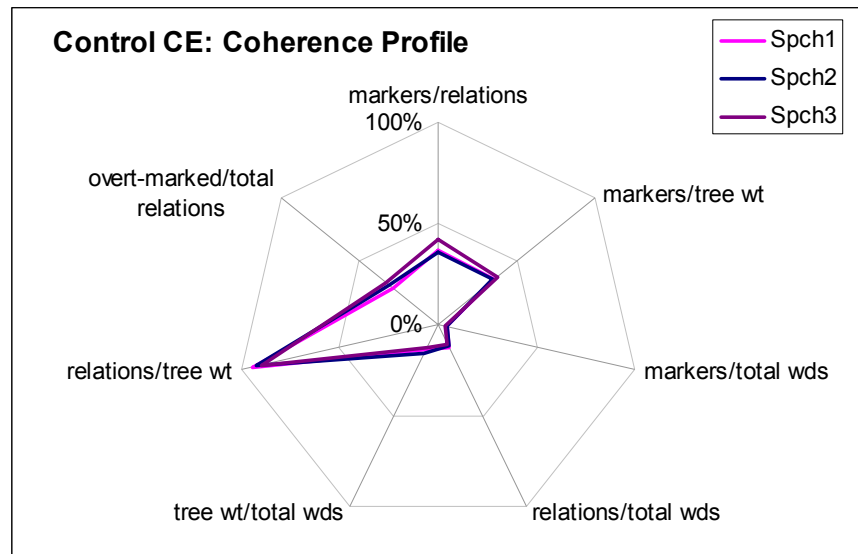
Thirdly, parameters from two other categories had lower ratios than those of the Professional group. One category concerned the explicitness of RST relations (markers/relations and overt-marked/total relations). The other concerned the ratio of RST tree weight to the length of a text (tree weight/total words). The ratio of explicit markers to total relations (markers/relations) indicates the prevalence of explicit markers compared with the total number of RST relations. The overt ratio (as discussed in 6.2) describes the 'explicitness' of RST relations. The lower ratios of markers/relations and overt-marked/total relations in trainee interpretations suggest that trainees were not as good in using explicit markers as professionals in this particular performance.

The ratio of RST tree weight to the total words of a text is also an important indicator of coherence. Trainee interpretations from both groups had lower ratios (14.6% and 15.7%) for this parameter than professional interpretations (24.5%). According to my discussion in 6.1.1, it is clear that the formation of RST tree weight comes from both the depth and width of a text structure. A text which branches to the right will also have higher tree weight. Right-branching is a natural tendency of textual flow. A text that grows to the left, on the other hand, will be penalised. In sum, I have shown previously, RST tree weight is a major indicator of coherence of a text.

6.4.2.1. Chinese-English interpretations

When the trainees interpret from Chinese into English, they should not have much difficulty in comprehending the source speech, but are likely to struggle to

reproduce the text in English, their B-language. This is, however, not an insurmountable obstacle to achieving textual coherence (see discussion in 6.2.3).

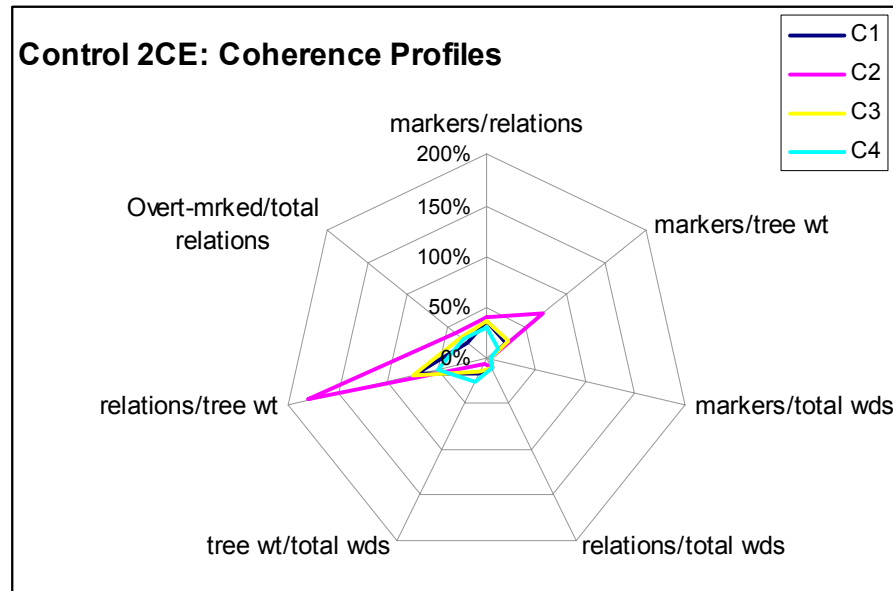


Control C>E	Spch1	Spch2	Spch3
markers/relations	36.9%	35.7%	41.8%
markers/tree wt	35.5%	34.6%	37.2%
markers/total wds	4.5%	4.2%	4.0%
relations/total wds	12.9%	11.9%	10.4%
tree wt/total wds	14.6%	16.0%	12.1%
relations/tree wt	94.6%	93.3%	89.9%
Overt-marked/total relations	28.3%	30.9%	32.8%

Figure 66 Coherence profiles: Control group C>E interpretations

The coherence profiles of the Chinese-English interpretations by the Control group (Figure 66) are very consistent over the three speeches. Such consistency over speeches indicates that the interpretations by the Control group did not vary much in terms of textual coherence. In other words, the ability of this group of trainees in conveying coherence and cohesive devices did not progress much.

I also investigated whether this consistency of coherence profiles across speeches existed among interpretations by different interpreters within the group. I observed that the consistency did not exist among individual interpretations of Speech 2 (Figure 67).



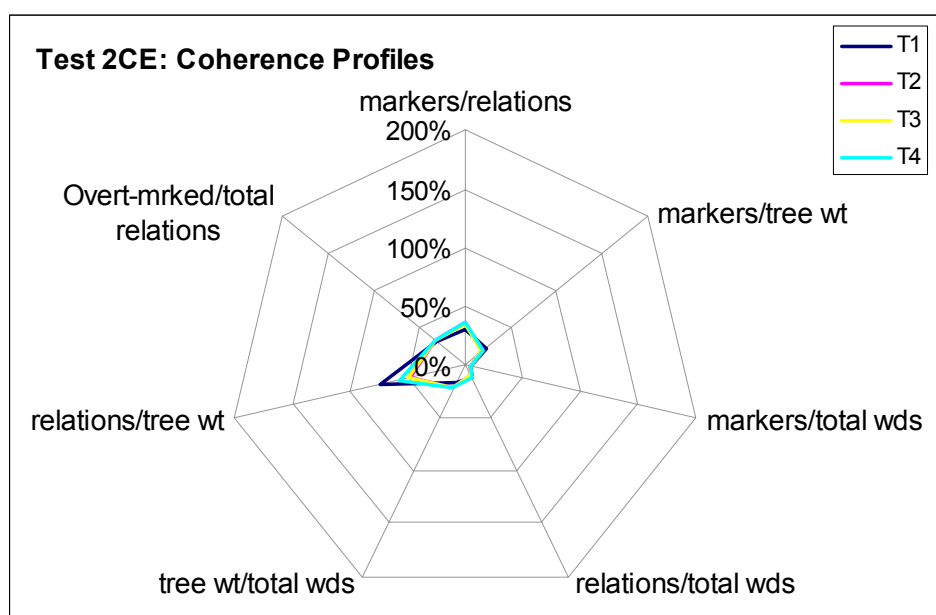
Control 2C>E	C1	C2	C3	C4
markers/relations	36.0%	39.5%	36.6%	30.8%
markers/tree wt	25.4%	70.8%	27.3%	15.1%
markers/total wds	4.30%	4.50%	4.40%	3.70%
relations/total wds	12.0%	11.3%	12.1%	12.1%
Tree wt/total wds	17.0%	6.3%	16.2%	24.6%
relations/tree wt	70.4%	179.2%	74.5%	49.1%
overt-marked/total relations	24.0%	37.2%	31.7%	30.8%

Figure 67 Individual coherence profiles of Control group (Speech 2C>E)

On the contrary, the coherence profiles varied greatly. Take the relations/tree wt ratios for instance, the interpretation by C2 scored astonishing 179.2% for this ratio. The interpretation by C2 contains 43 relations yet only scores 24 for the RST tree weight. Also, as a result of the very low tree weight, the ratio of explicit markers to tree weight rose to 70.8%. This means that the performance by this trainee interpreter (C2) was rather disjointed. In this case, the high ratio of markers/tree wt. did not mean that the interpretation was more explicitly marked than the others, but simply showed that the RST tree weight was lower. Evidence for this was found in another ratio (tree weight/total words: 6.3%). In short, individual trainee interpreters from the Control group appeared to develop differently in terms of conveying coherence when interpreting into English after several weeks of intensive training. But, if the average profile of the group remains similar across speeches, while

individuals vary, surely this suggests that the individuals make little progress in this area.

The Test group, by contrast, appeared to do better. Firstly, the coherence profiles of individual interpretations seemed to correspond to the group average. The radar chart in Figure 68 shows the coherence profiles of English interpretations of Speech 2 by the Test group.



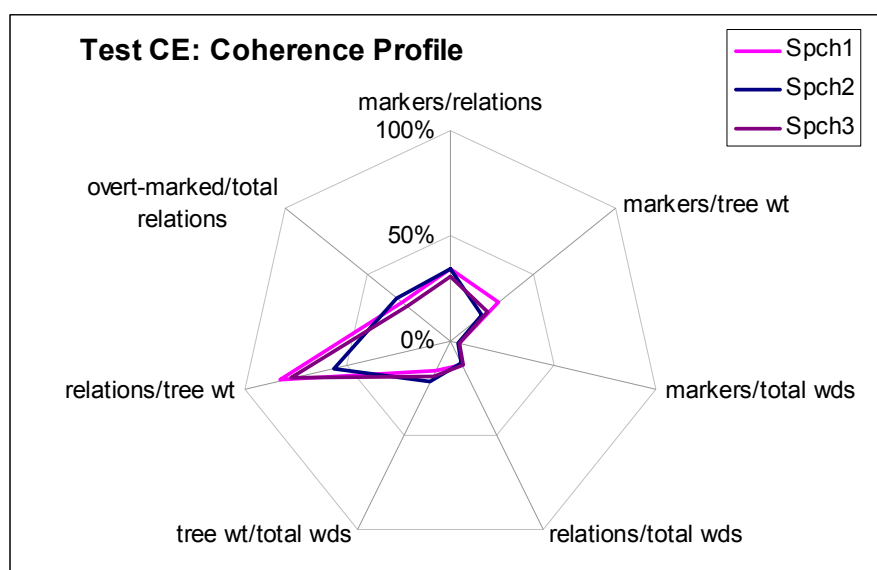
Test 2C>E	T1	T2	T3	T4
markers/relations	30.6%	35.9%	35.0%	35.4%
markers/tree wt	22.4%	17.5%	17.5%	20.0%
markers/total wds	3.80%	3.90%	3.80%	4.40%
relations/total wds	12.6%	10.8%	11.0%	12.3%
Tree wt/total wds	17.2%	22.1%	21.9%	21.9%
relations/tree wt	72.9%	48.8%	50.0%	56.5%
Overt-marked/total relations	30.6%	33.3%	33.3%	33.3%

Figure 68 Individual coherence profiles of Test group (Speech 2C>E)

It is noticeable that the individual profiles (T1 to T4) did not vary to the same extent as was the case for the Control group. The most extreme case in the Test group was T1, but T1 was definitely not as extreme as C2 in the Control group. In fact, in terms of the selected ratios (relations/tree weight and tree wt/total wds), T1's profile is close to that of both C1 and C3, the better two of the Control group.

Significantly, the coherence profiles show that the RST structures of the interpretations by the Test group should be properly related as trees rather than the small bushes which suggest only local coherence.

Furthermore, unlike the coherence profiles of the Control group, the average coherence profiles of the Test group (Figure 69) show some variation across different speeches. In this group, trainees' ability to convey coherence improved more noticeably over time. To be more specific, it was clear that in Speech 2, the coherence profile was very different from that of Speech 1 or 3. The ratio of the number of RST relations to RST tree weight (relations/tree wt) was lower than both Speech 1 and 3 by more than 20%, yet at the same time, with more RST relations were overtly-marked by explicit markers (32.7%).



Test C>E	Spch1	Spch2	Spch3
markers/relations	34.4%	34.2%	30.6%
markers/tree wt	28.6%	19.3%	22.4%
markers/total wds	4.3%	3.8%	4.0%
relations/total wds	12.7%	11.7%	13.5%
Tree wt/total wds	15.7%	20.8%	18.2%
relations/tree wt	82.3%	57.0%	77.0%
Overt-marked/total relations	29.1%	32.7%	26.5%

Figure 69 Coherence profiles of Test group

The comparison in Figure 70 gives a compelling illustration of the development of trainee interpreters in conveying coherence.

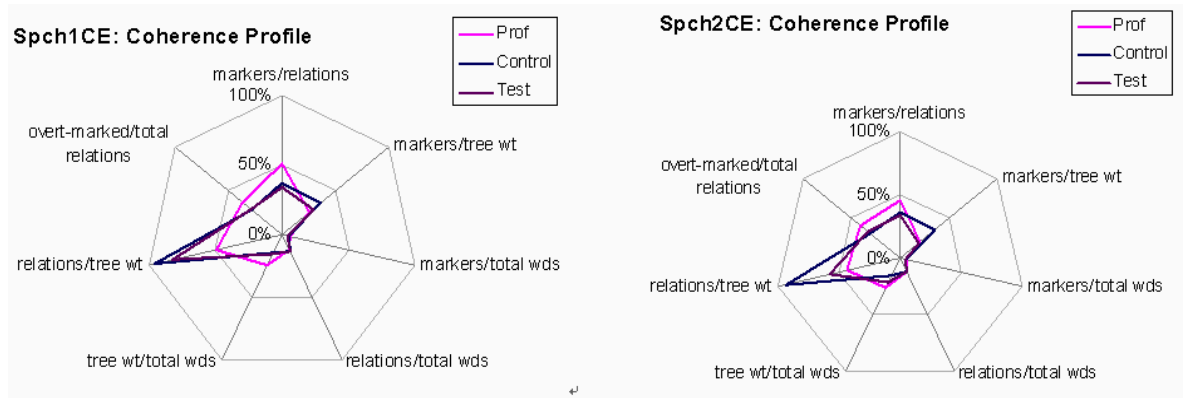
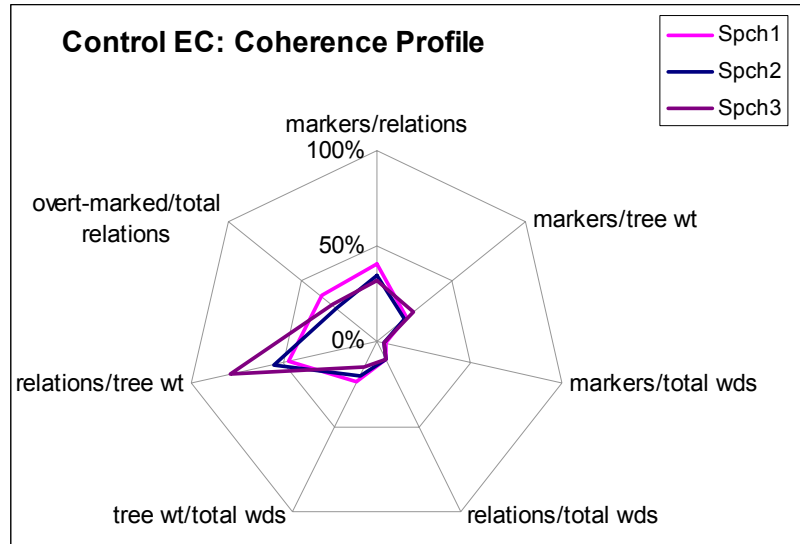


Figure 70 Development of coherence-English interpretations Speech 1 & 2

All in all, compared to the professionals, trainee interpretations in English were not as connected globally (both trainee groups had a very high relations/tree wt ratio), and neither were they as explicitly marked (they had a low overtly-marked to total relations ratios) as the professional interpretation. Yet both groups of trainees started with similar profiles as shown by the radar chart on the left. Both show a similar ability to convey coherence in Speech 1. The radar chart on the right, by comparison, showed some variation. The coherence profile of the Test group in Speech 2 changed, and moved towards to the profile of professional interpretations. In short, despite the fact that trainee interpreters' English was not as proficient as professionals', their ability to give coherent interpretations in English progressed noticeably. This is further evidence that the specific guidance they received in quality criteria led these trainees to progress more rapidly.

6.4.2.2. English-Chinese interpretations

When working into Chinese, trainee interpreters had more flexibility in language use, and thus in conveying coherence in their interpretations. Comprehending the source speech in English, however, was not as easy as in Chinese and sometimes might cause information distortion. In turn this might jeopardise coherence.



Control E>C	Spch1	Spch2	Spch3
markers/relations	40.5%	34.8%	31.8%
markers/tree wt	18.9%	18.5%	24.7%
markers/total wds	4.1%	3.6%	3.2%
relations/total wds	10.3%	10.4%	10.1%
tree wt/total wds	22.6%	19.9%	15.3%
relations/tree wt	46.7%	54.5%	78.8%
Overt-marked/total relations	37.5%	27.4%	29.9%

Figure 71 Coherence profiles: Control group E>C interpretations

For Speech 1, the speech without many difficult concepts or technical facts, trainee interpreters delivered coherent interpretations successfully into their mother tongue. For Speech 2 and 3, however, the radar chart of the coherence profiles of trainee (Control group) interpretations into Chinese (Figure 71) show that the profiles were neither as ‘centralised’ as they had been for Speech 1, nor were they as consistent as that of the Professional group (Figure 60).

To be more specific, the relation/tree wt. ratio for Speech 2 and Speech 3 rose to 54.5% and 78.8% respectively. This is a sign of interruption of the global coherence. Reasons for this regression could be the failure to comprehend the source text.

The Test group, however, seem to have progressed from the initial trainee profiles gradually towards professional ones. Figure 72 shows that for Speech 1 (radar chart on the left), several ratios in the coherence profiles of the two trainee groups were very similar. Yet in Speech 2 (radar chart on the right), the coherence profile of the Test group converges towards that of the professional group. The coherence profile of the Control group, on the contrary, diverges from those of both the Test and the Professional groups. This figure, in short, demonstrates how

coherence in English-Chinese interpretations developed differently in the two trainee groups.

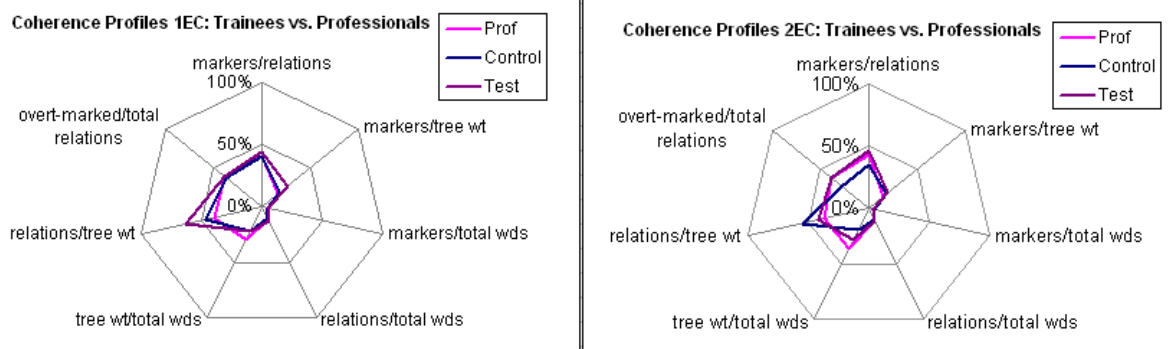


Figure 72 Development of coherence-Chinese interpretations Speech 1 & 2

6.5. Trainees' judgements of interpretations

Trainees' awareness of quality appears to have been raised by using the feedback tool for self evaluation and peer feedback (discussed in Section 5.3). My discussion in 6.4.2 suggested that the development of coherence in interpretations was also stimulated by the introduction of the feedback tool. To validate the development of quality awareness, I explored whether trainees had developed the ability to give reliable judgements on interpretations. I also thought it desirable to investigate whether human judgements corresponded to what I learned from RST annotation and analysis. Detailed information about the design of the experiment discussed in this section is given in Section 5.4.

6.5.1. Judgement of coherence

Trainee subjects were asked to score interpreting performances from 1 (the worst) to 5 (the best) on seven coherence attributes I exported directly from the feedback grid. I collected 20 judgements from 10 subjects commenting on different interpretations into Chinese. To get an overview of trainee judgement, I averaged all the scores.

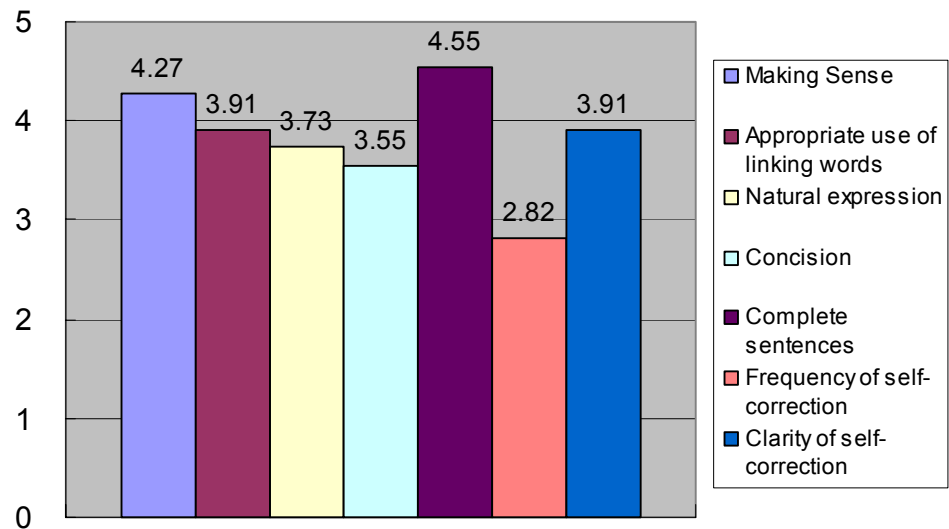


Figure 73 Trainee judgement on coherence

The results (Figure 73) show that the highest score my subjects gave was for ‘Complete sentences’ (4.55), followed by ‘Making sense’ (4.27) and ‘Linking words’ as well as ‘Clarity of self-correction’ (both at 3.91). The lowest score was for ‘Frequency of self-correction’ (2.82) and the second lowest was for ‘Concision’ (3.55). ‘Complete sentences’ were not only easy for the subjects to identify, this measure enshrined a simple principle for trainee interpreters to bear in mind from the start of their training. This awareness seems to have made an impression, as incomplete sentences only occurred very rarely in my data. ‘Frequency of self-correction’ was also easy to evaluate. Here, a low score is clearly preferable. The score for ‘Frequency of self-correction’ corresponded with my RST analysis on Repair. Even when interpreting into Chinese, the trainees corrected themselves. This became especially common once trainee interpreters in the Test group started to monitor their own performances closely with the introduction of the feedback grid (6.4.2.2).

‘Clarity of self-correction’ is also important. This attribute, as explained to the subjects before their assessment sessions, measures whether the self-correction was successful and clear. ‘Clarity of self-correction’ did not score as highly as ‘Complete sentences’ and ‘Making sense’. It seems that trainee judges could be rather critical when attention was drawn to commenting on ‘Self-correction’. ‘Concision’ and ‘Natural expression’ both scored lower than ‘Clarity of self-correction’ and the other attributes. ‘Concision’ was the weakest feature of all. This measure relates to how natural the language sounds as well as the occurrence of self-correction.

Overall, the trainee interpreters recruited from the Test group’s cohort were able to give specific judgements on each of these attributes of coherence. Details of

the scores from the subjects for each interpretation (A1, B1, A3 and B3) are presented in Figure 73.

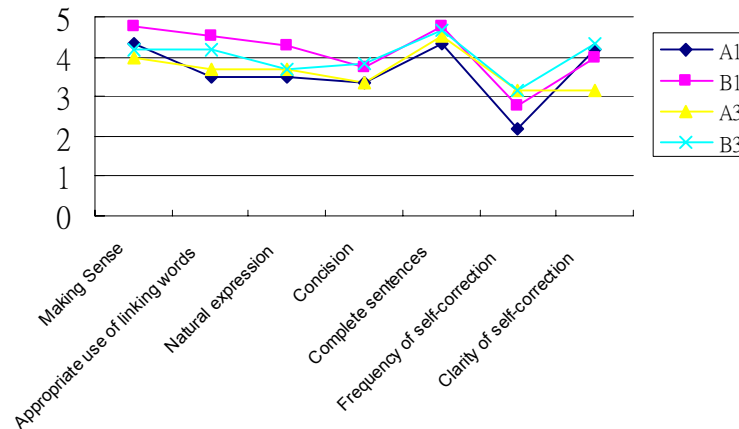


Figure 74 Trainee judgements on coherence attributes

Figure 74 shows the general trend for the Chinese interpretations. B1 was judged to be the best of the four, while A3 scored lower than the others in most attributes. To achieve a general score for the performance, I added up the scores (subtracting the score for ‘Frequency of self-correction’) from each participating trainee subject and produced an average score for each interpretation. An ideal interpretation would score 29: 5 (the best) for six of the coherence attributes and 1 for ‘Frequency of self-correction’.

I took the ideal interpretation as the benchmark and came up with a normalised value for each of the interpretations (Table 28).

Interpreting	Ideal interpretation	A1	B1	A3	B3
Original score	$29 = (6 \times 5) - 1$	21	23	19	22
Score at 100	100	72	80	66	75

Table 28 Scores from trainee judgements

The results in the table agree with the general trend I noted from Figure 74, that B1 scored the highest with 80 and A3 scored just 66 being the lowest of the four.

6.5.2. Judgement results vs. RST tree weight

To validate the reliability of trainee judgements, I compared the results with those of RST analysis of textual coherence (Figure 75). I took the value of RST tree weight of each interpretation (A1, B1, A3 and B3) as a general indicator of textual coherence.

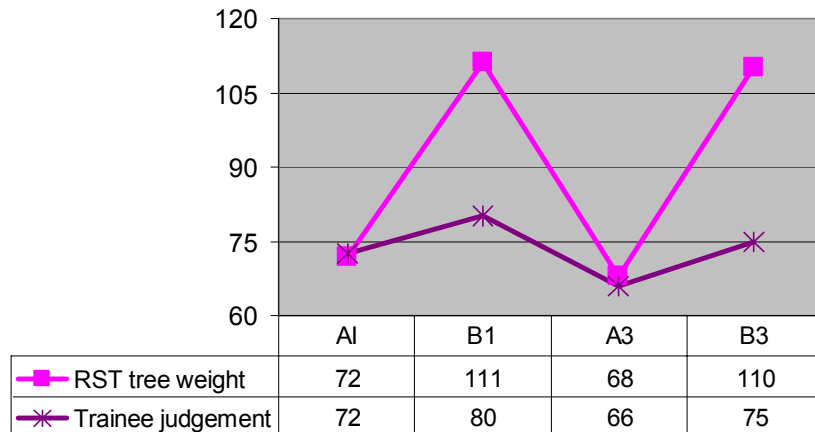


Figure 75 Trainee judgement on coherence vs. RST tree weight

It was encouraging to see that the trainee judgements of coherence agreed with the general trend reflected in the RST tree weight. B1 scored the highest in both trainee judgement and RST tree weight with values of 80 and 111 respectively. A3, on the country, had the lowest score in both RST tree weight and trainee judgement. The rank order of the scores and values from trainee judgement and RST tree weight were both the same: B1>B3>A1>A3. In other words, there was clear agreement between the results from trainee judgement and the analysis of RST annotation.

6.5.3. Analysis of RST-awareness

RST tree weight has proved to be a useful indicator of the coherence of a text, but it is a very costly approach. Calculating the tree weight of a text involves annotating the text, assigning relations and computing Marcu's algorithm. It is therefore difficult in practical terms to use it as a normal practice for pedagogical purposes. In order to benefit trainees and keep them aware of the importance of coherence, it is important to devise a practical approach to make this type of analysis more feasible. For instance, when in class, I conducted informal sessions on RST and use the metaphor of trees and bushes to explain the difference between coherent and incoherent discourse structures. I identified some successful and unsuccessful segments of their own interpreting performance on the spot, and analysed them in the form of RST trees on the board to illustrate my point.

Moreover, although the use of the feedback grid and the adoption of RST as a framework for data analysis do not share any direct link, it appears that both contribute to trainee interpreters' development.

My results support the view that, with guidance and a proper tool, trainees can be reliable judges of the coherence of an interpretation. The experiment shows that trainees were able to distinguish different levels of coherence, which corresponded to

the results using RST tree weight as the judgement indicator. I recognise that it is not feasible to introduce the full Rhetorical Structure Theory to all trainees, but I believe it is sensible and helpful to provide them with systematic guidance (in my case the feedback grid) to develop their ability to give proper judgement on interpreting performances.

Acquaintance with quality criteria and the use of the feedback grid had positive impacts on the trainee subjects in judging interpretations based on explicit criteria. I do not have sufficient evidence to prove that their improvement of performance was directly related to the use of the feedback grid, but such a claim is supported also by the clear improvement of trainee's interpretations over time. Many other factors may have contributed to this progress, but by comparing the two groups of trainees, the Control group and the Test group, I conclude that the awareness of quality criteria promotes self-monitoring which in turn helps to produce sound performances.

As I have already established, coherence is a vital feature of successful interpreting performance. With time, I witnessed noticeable progress on coherence profiles of trainee interpreting performances. Some might argue that the development of coherence for trainee interpreters is only a natural result of learning. Yet I also noted that by raising awareness of quality attributes with the feedback grid, the development of the trainees in the Test group was enhanced and the learning results improved overall (Figure 49 and Figure 50).

6.6. Summary

In this chapter, I explored the development of coherence in interpretation by trainees. I adopted RST as the framework for analysis of coherence in interpretations. One striking difference between professional and trainee interpretations was made clear by the RST annotation. Professional interpretations resembled a tree which comprises both 'leaves' (the number of spans) and 'trunks and branches' holding the leaves together (depth of the structure). By contrast, trainees' performance resembled bushes due to the lack of depth of the discourse structure and global coherence.

In addition, my discussion of overtly-marked relations showed that professional interpretations were more explicitly marked in both English and Chinese than either the source speeches or the trainees' interpretations. Trainee interpreters, on the other hand, did not show much stability in marking their discourses with explicit markers. When they fully comprehended the source speech in English, their interpretations in Chinese were as coherent as those of professionals. However, their use of double

conjunctions, a sign of Chinese mother-tongue interference, can cause confusion rather than giving clear signposts when interpreting into English.

The introduction of the new RST relation of Repair also elicited some interesting results. I noted that the occurrence of Repair was a common self-monitoring mechanism and was used to handle problems of both fluency and textual coherence. The occurrence of self-correction in professional interpretations was noticeably lower and more stable than it was in trainee interpretations. After the introduction of the feedback grid to evaluate their interpretations, the trainee interpreters in the Test group corrected themselves much more often than their counterparts in the Control group in both Chinese and English. All in all, from the professional interpretations, I observed that a low level of self-correction is quite natural and does not impede the global coherence of the text. From interpretations of the Control group, it is clear that speaking fluently without much repair does not guarantee text coherence.

From RST tree weight, which I have shown to be an important indicator of textual coherence, and the coherence profiles I compiled from several significant parameters, I observed the progress of the trainees in conveying coherence in their interpretations. I noted that the coherence profiles of trainee interpretations (both groups) were very different from those of the professionals at first. Those of the Test group later gradually converged with the professional profiles toward the end of the training. Furthermore, I noted some major differences between the two trainee groups (Control and Test) regarding their progress in terms of coherence.

The introduction of feedback guidance also helped raise trainees' awareness of quality issues in interpretations. When compared to trainees in the Control group, those in the Test group were more cautious about the quality of their interpretations. This is evidenced by the number of occurrences of self-correction in both Chinese and English.

In short, using RST and the coherence profiles I based on RST annotation, I explored various features of coherence in both trainee and professional interpretations, and observed significant progress in trainee interpretations as their awareness of interpreting quality grew.

Chapter 7 Conclusion

7.1. Synopsis

The evaluation of conference interpreting and training for quality performance are very topical issues in interpreting studies. They are also the ultimate concerns of the present thesis. In order to address these concerns, I formulated four research goals. In this section I summarise the findings in relation to each of these.

To address the first statement (*To explore the basis for judgement about quality for conference interpreting*), I reviewed the state of the art of quality issues in conference interpreting from different perspectives in Chapter 2. I noted that quality has always been an important issue and has been discussed from a number of perspectives. I showed that the criteria currently used by professional bodies (e.g. AIIC) and training institutions (e.g. EMCI) to judge interpreting quality are too vague to be useful in training.

The second statement (*To abstract and organise systematically the performance criteria for conference interpreter training*) established the aim of addressing the inadequacies of existing sets of criteria. I conducted a literature review and collected criteria from the existing schemes. I then devised a prototype feedback grid in which relevant criteria were organised in a hierarchical structure (see section 4.2.4). With colleagues in the CILT project (reported in Chapter 4), I carried out a small-scale experiment to validate and improve the feedback grid. We mapped undirected comments on interpretations collected from users, trainees, trainers and professionals onto the criteria in the grid. We also piloted the feedback grid with trainees (see section 4.4). They reported that the grid was useful and easy to use. With their comments, the feedback grid was further revised (see section 4.5). After this, the feedback grid was adapted for use in CI (see 4.6). The grid gave explicit guidance for trainees to critique each other and assess their own performances. In turns, this also raised trainees' awareness of quality and subsequently facilitated the development of coherence in their interpretations (6.5).

I have shown that in most of the literature on quality in conference interpreting, coherence is regarded as a vital attribute of successful interpretation. I adopted Rhetorical Structure Theory (RST) to address my third research statement (*To establish a framework to capture coherence of conference interpreting in such a way*

that I can make comparative and qualitative judgements about the interpretations by professional and trainee interpreters). My results and analysis show that there are several key differences between the coherence of professional and trainee interpretations.

RST annotations of typical professional interpretations will resemble trees, with roots on which globally coherent structures grow. This represents the globally coherent textual structures that professional interpreters produce. Trainee interpretations look more like bushes. The coherence here is localised (6.1). In addition, professional interpretations were consistently more explicit than source speeches in both Chinese and English. The use of explicit markers (conjunctions) by professionals was more stable than that of trainees (6.2). It was also observed that professional interpreters repaired their interpretations in both Chinese and English. This shows that they use self-monitoring. The trainees in the Test group also repaired their interpretations. The Control group, on the other hand, seemed to use less self-monitoring (6.3). I also devised a 'coherence profile' to facilitate comparison of the coherence in different interpretations (6.4). The coherence profile of professional interpretations corresponded well with those of the corresponding source speeches. Trainee profiles did not initially correspond with the source speeches. They lacked consistency and stability. In time, however, the profiles of the Test group converged with those of the professionals.

The last statement (*To investigate the development of awareness of these criteria in trainee interpreters and its impact on their judgement of their peers and on their own performances*) was explored in two different chapters. Chapter 3 addressed the first half of the statement. The meta-language used by trainees to discuss quality interpretations showed that their awareness of quality was vague and inconsistent. This issue was explicitly addressed in the training given to the Test group. Iterative discussion of the feedback grid, as outlined in 5.4, reinforced the importance of using appropriate terminology when evaluating interpretations. By the end of their training, their use of meta-language showed that their awareness of quality was more comprehensive and consistent. Their ability to discuss quality was also much more developed. The results in 6.5 show that trainee interpreters who had used the feedback grid were able to give reliable judgements regarding the textual coherence of interpretations. Moreover, their judgements corresponded with the results from RST analysis.

All in all, trainees' awareness of quality and, significantly, their interpreting performances demonstrated noticeable development over the course of their training. Before training and the introduction of the feedback grid, their awareness was

sketchy and their performances only displayed local coherence. By the end, they proved to be reliable judges of interpretations and their interpretations became more globally coherent.

7.2. Limitations

As with all such research, the scope of the present PhD study was limited. There were inevitable constraints on the availability of resources, including people, time, equipment, technical support and institutional arrangements. Such limitations posed challenges.

From the start, I was concerned about the feasibility of recruiting second and third RST analysts to validate my data annotations. However, it was financially impossible to hire trained analysts and impractical to train new analysts from scratch. I therefore had to be satisfied with the two-pass, single annotator strategy described in 5.2.2.2. Secondly I was aware that, in order to avoid subjectivity, the data should ideally be annotated blind, i.e. without knowing whether a particular interpretation had been produced by a professional or a trainee. Again, this was impossible for practical reasons. Of necessity, the researcher carried out every step of the process, from data-collection, through transcription and annotation to analysis on her own. In order to remove possible influence from the rhetorical structures of the source speeches, and therefore to avoid biased annotation, the interpretations were analysed first and the speeches afterwards. Of course my intention was to compare the features of professional and trainee interpretations. I had no interest in proving that professionals were better than trainees. I believe the bias played no significant part in the annotations. Therefore I had nothing to gain from bias in the annotation.

I faced similar challenges in relation to data collection. Wider participation from professional interpreters would have resulted in more comprehensive data. Yet recruiting professional interpreters for empirical research has never been easy in interpreting studies. Firstly, it was rather costly to recruit professional interpreters to take part in experiments. Secondly, from experience I knew that few professional interpreters would be willing to join studies of this kind, as being reluctant to have their performances analysed for fear of damage to their professional reputations. To address this challenge, I made the best possible use of those interpreters who I was able to recruit. I recorded six performances from each of the three professional interpreters I recruited and therefore obtained 18 performances in total, covering both Chinese and English interpretations.

In introducing the peer feedback tool I faced some institutional challenges. I was not able to make use of the tool compulsory, and not all of the trainee subjects used the tool systematically and consistently. As a result, I was not able to collect enough completed grids to conduct further analysis. Having said this, some trainees, such as the subjects in the Test group, were extremely interested and cooperative. They used the feedback grid in both class time and their private practice. The results in Chapter 6 show the clear benefits of such practice. If further investigations are to be conducted to study the influence of using the grid on trainees' interpretations, it would be advisable to make it a compulsory part of the curriculum.

In addition, as reported in 6.5.3, when RST was introduced in class I used trainees' performance as authentic examples to demonstrate the significance of coherence and to raise their awareness of their performance. I believe that the use of an electronic whiteboard, and access to the necessary technical support, could make a significant contribution to both training and research by facilitating more dynamic presentation of examples, real-time annotation and, importantly, their recording for subsequent analysis.

7.3. Methodology review

Subjectivity is a common concern in research methodology. Using RST as the framework to analyse and discuss coherence in interpretations produced positive results. It also drew my attention to a degree of subjectivity in data annotations. The RST annotation carried out as part of this study involved two major steps, text segmentation into function units and assigning RST relations. As has been mentioned, both steps were necessarily carried out by the researcher.

According to several studies cited by Bateman and Delin (2005: 7), however, the segmentation and attribution of nuclearity in RST tend to be consistent when carried out by trained annotators. In other words, RST annotations by different annotators vary little regarding these two aspects, which are precisely the parameters taken into account by Marcu's algorithm in assigning scores to RST tree structures. Different judgements about the assignment of RST relations would have no effect in the score resulting from the algorithm. In addition, my analysis involved a series of neutral indicators such as word counts and the occurrences of explicit markers. Identification of these objective indicators involved little human judgment and thus helped to balance any subjectivity in the annotations. To ensure the consistency of analysis, the data were reviewed and annotated twice after all the data were ready.

7.4. Contribution

In spite of the constraints described above, my results contribute to the research literature in interpreting studies in several ways. First of all, the feedback grid has proved to be useable and useful for trainee interpreters. Existing professional standards of conference interpreting are vague and unsystematic. It is difficult for trainee interpreters to comprehend and follow them as guidelines to evaluate their own performances. Similarly in training institutions, the exam criteria to which trainees have access tend to be descriptive and based largely on impressionistic judgment. As such they are not ideal guidelines for trainees to follow. The feedback grid devised in this thesis comprises most existing criteria used not only in professional and training organisations of conference interpreters, but also by practicing trainers and professional interpreters when they comment on interpreting performances. Significantly, the hierarchical organisation of the criteria provides trainee interpreters with a structured approach and explicit guidelines to conduct both self evaluation and peer feedback. Most important of all, I witnessed noticeable progress in trainees' performance with the introduction and use of the feedback grid during their training.

What is more, my novel adoption of RST as a framework for describing and analysing interpreted texts has proved successful. According to my data analysis and research results, RST is a sound framework to describe coherence across languages, in my case, Chinese and English. In addition, although there is no single ideal interpretation, RST enables researchers to compare coherence across different interpretations. I believe that it is worth introducing RST analysis (or at least an RST-aware analysis) to interpreters during their training, for such analyses appear to enable them to grasp sense relations better in the incoming speech and represent them in their subsequent interpretations.

7.5. Future work

I have identified several topics which can usefully be taken further in future work. First of all, further development of the peer feedback grid could make a significant contribution to conference interpreter training. At the moment, the self-assessment and peer-feedback criteria have been implemented in paper-based format. I plan to develop a hypertextual implementation, with easy access to a precise definition of each criterion and instructions for use. This will allow trainees to make flexible use of the tool. It will also make it straightforward for trainers to collect and

analyse the information captured. As such it would form a key role in distance learning of conference interpreting skills.

With regard to the development of coherence in trainee interpretations, my next step is to explore how interpreters render relations in a way that more closely approximates the coherence profile of the source speech. I would also like to investigate the explicit marking of rhetorical relations, as favoured by professionals.

In short, this thesis demonstrates that my exploratory approach offers interesting findings and implications for interpreter training, as well as offering directions for further research in both the conference interpreting and RST communities.

Appendix A Text of source speeches

Appendix A.1 Chinese Speech 1

各位女士，各位先生
今天很高興能跟大家談談一些英語學習的亂象
在台灣呢，這個英語學習已經帶來了很多問題了
所以這就是今天我要討論的重點
現在在台灣，不到三歲的小孩子都在開始拚命地學英文了
因為家長希望孩子不要輸在起跑點上。
我有個朋友
她是個幼教專業的學者
她在美國留學
那她回台灣之後
當然覺得要給自己的小女兒最好的英語教育
所以她開始在小女兒身上進行實驗
進行所謂的「浸泡式學習法」
希望小孩呢可以學得一口好英文
這個所謂的「浸泡式學習法」就是把不到三歲的小孩，
從幼兒教育開始，就送到全美語的英語環境中學習
回家的時候
小朋友只有說英文
有時候也說說中文
但是看的電視看的書全都是英語的
直到小朋友到別人家裡看電視
才知道小美人魚也會說中文
她才發現這個問題有點奇怪了
這種方式就是把英文不是用「教」的
而是用「泡」的
希望在這個環境裡面
小孩子不用移民也可以把學英文學得好
但是這樣的學習給小女兒帶來很大的文化衝擊
因為第一個，在幼稚園的老師都是外國人

跟這些外國人相處久了
小孩子的個性變得非常地外向
肢體上喜歡擁抱
還要叫大人說「SORRY」
那這個和父母的價值觀有很大的不同
另外呢，小朋友自言自語時說的都是英文
講中文也是怪腔怪調
全部都是英式中文
她會說：「念我一本書，好嗎」
或是說「你好嗎？今天」
這樣的中文讓我的朋友非常擔心
另外呢，這些全美語教學的小孩，他們從小學一年級才開始學中文
比其他的小朋友晚了三年了
所以在閱讀方面，
小朋友開始學英文時，他們的閱讀也是英文
所以和中文比較下來
他們的母語程度顯然差得很多
因此上小學的時候
小朋友覺得非常痛苦
看到中文就想哭
看到英文就覺得很親切
更嚴重的是，她不覺得自己是中國人
而對於西洋的文化如數家珍
對於端午節元宵節，則一點興趣也沒有
所以現在我的朋友只要一談到台灣的英語學習的熱潮
心裡當然不是很好過
因為很多這些美語補習班或是美語幼稚園根本不符合政府的規定
但政府也不管
另外家長對於這些幼兒園漫天要價的行為
收費很高
也都盡量配合
因此對於師資方面，家長也不要求
家長只要看到這些老師都是外國人
就會覺得這些師資都是好的

完全不管這些老師是不是合格的老師
那這些家長看到這些老外老師
逃得比誰都還快
因為這種心態
簡直就是把外國人高舉在上的殖民心態
所以這些問題我們現在不認真考慮的話
只怕我們會培養出一堆會說流利英語的孩子
完全不認同自己的文化自己的語言
也失去了很多在地學習的機會跟學習的能力
這樣子又有什麼意義呢
我今天就講到這兒，謝謝。

Appendix A.2 Chinese Speech 2

女士們先生們大家好
中國代表團很高興有這個機會來到海牙
參與這次的中國歐盟會議
一起與大家來討論假造旅行文件的問題
這個議題自然與非法移民與人口買賣有直接的關係
也給我們社會帶來了很多的負面影響
也花費了很多的社會成本。
而這兩個問題是全球性的問題
需要我們世界各國一起努力
大家也許不清楚
中國也很關切非法移民的問題
因為有不少人透過假造文件偷渡到中國來
有的是由朝鮮過來
有的是透過沿海地區進來
不少非洲黑人就是這樣
也有的是從南邊的鄰國進來
也很多非法集團將中國人大批地偷渡到世界各地
中國政府為了打擊犯罪
投入了很多的心力
也展開了一些具體項目

比如說，所有乘客在要搭機出國之前
我們的海關人員不僅檢查護照真偽
也查驗有沒有合法的簽證進入目的地的國家
等於是為其他國家先把一道關
今天我就這個假造旅遊文件的問題
給大家報告我們中國的意見及想法
首先，旅遊文件不給假造的第一個方法，就是提高製造技術及加強防偽措施。
這樣一來，這些非法集團在偽造上面就困難許多
再說，人員的訓練很重要
特別是站在第一線的邊防人員
要讓他們能夠一眼就辨認出手上的護照是真是假
需要的就是非常完整的訓練及知識
另外也要配合使用相關的設備
來幫助他們做出最有效率也最準備的判斷
比如說有的護照上的條碼需要紫外線判讀
在機器下一閃就可以知道是真是假
所以人員的訓練及機器的配合
在辨別旅遊文件的真偽特別重要
除此之外我方相信，這個問題還需要有個全球網絡
讓各國間的資料流通更快速。
比如說，今天有人非法出境了
我立即通知其他可能的入境國家
讓他在下了飛機後就馬上原機遣返。
這樣才能有效打擊人口買賣及非法移民的問題
希望中方的這些意見可以給大家帶來一些想法
也希望這次會議中
我們能就這個問題多多交換意見
在會議結束後
能有具體可行的方法
可以合力一起打擊假造文件的問題
打擊全球非法移民及人口買賣的犯罪

Appendix A.3 Chinese Speech 3

各位女士先生

很感謝各位給我這個機會

來跟大家談談有關全球氣候變遷的一些問題

我們要正視這些問題，

需要的並不只是良心跟善意

我們要面對的是非常非常困難的選擇

這就是今天我想和大家一起討論的重點

無庸置疑地，目前每個工業國家都積極地竭盡所能要節約自然資源

因為我們不能繼續去假裝

全球有這麼多極端的氣候現象

跟我們平常呢習以為常的生活習慣跟生活方式沒有關係

我想有個現象，可以非常清楚地來闡釋這兩者之間的密切關係

也就是現在我們所使用的化石燃料

造成了很多溫室氣體的排放

大家都知道，在京都高峰會上面

我們首度訂定了國際的目標

要來降低二氧化碳的排放量

當然，這一步是踏對了方向

但是我們可以發現在京都高峰會上面所設定的目標，其實低的可憐

根據專家指出，

二氧化碳的排放量需要降低 60%到 80%

才能防止氣候型態的大幅度改變

在談到降低二氧化碳排放量的時候，

要來達到這個國際目標的時候

我們想說要要求每個國家的人民都要做出一樣的貢獻

這種說法，在道德上面說不過去，

在政治上的考量來講，也不是非常正確

我認為，各個國家在過去跟現在所耗用的資源都要納入考量

近來最受推崇的解決方法，

就是以二十五年為期，

持續地去降低二氧化碳的排放量

以目前而言呢，二氧化碳的人均排放量，是在每年每人十噸以下左右

但根據我的估算，
這個數據應要降到一噸以下才行
因此這樣做來，我們每一項使用化石燃料的活動，都會受到非常嚴格的監控跟監督
無疑的，這是非常嚴厲的，也是非常極端的做法
當然就會出現非常多反對的聲音及論調
不過有些人這是沒有問題，
這樣做不會有問題
是因為他們是所謂的既得利益者
以一些歐洲汽車．製造廠來講，
他們同意在十年內，
使新車的效能增加 25%
這些公司常常非常沾沾自喜地認為
他們對生態環保都盡了心力
我們應該要加入他們的行列嗎
我非常難以苟同！
他們這種看似環保的形象
其實是掛羊頭賣狗肉的
他們這些作為，為的不過就是要增加自己車子的銷售量
進一步地，又增加密集性能源的使用
我們別忘了，他們一直以來就是這麼成功
所以我們的路空交通才會在過去三十年來
有這麼驚人的發展
也許在座的人，有人會這麼想
特別是一些來自美國的代表
認為這些陸空交通都非常的重要
也沒有其他可以兼顧環保的替代方案
我們誰有時間坐船去橫渡大西洋呢
會這樣講
就表示你已經認為，使用這些汽車飛機是理所當然
完全不管會給我們的環境帶來多大的破壞
但請別忘了，二氧化碳排放量有四分之一左右，都是來自於這些陸空交通工具
因此，這些交通工具呢在節約能源方面，一定也有舉足輕重的影響
這些人群的移動交通問題，

其實都是非常困難的選擇
我今天的結論也非常的困難
因為，這個困難的決定，有賴你我
也就是這次國際研討會的代表們
大家，一起來面對
同意我今天這場談話的人
應該一起以行動來表示
比如說，像今天這樣的國際會議不能再理所當然地辦下去
怎麼說呢？想一想，飛機來回一趟倫敦跟佛羅里達之間
會排放三噸左右的二氧化碳
相當於人均排放量的二倍之多
然而，若是今天的會議也是最後一次
那會讓人覺得很遺憾
但更重要的是，如果我們存心故意地去糟蹋我們的環境
剝奪下一代享受相同生活品質的權利的時候
我想那是最令人遺憾的了！
我今天就說到這兒，謝謝

Appendix A.4 English Speech 1

Ladies and gentlemen, good morning.

I hope you feel energetic and fine at the moment, not very tired, not too sleepy.

Today this is the topic I am going to talk about: tiredness.

What is tiredness?

Tiredness is a lack of energy and a feeling of exhaustion usually resulting from overwork or lack of sleep.

It generally disappears after a good night's sleep.

Tiredness can also be a symptom of an underlying disease.

Some of these can be self-treated

while others may require medical treatment.

So what causes tiredness?

Basically we can divide tiredness into two categories.

One is natural tiredness

and the other one is pathological tiredness,

which means it's very persistent and it lasts for quite a long time.

Let's talk about natural tiredness first.

There are three major causes for natural tiredness.

The first reason for you to feel tired is that you have intensive physical exercise, particularly something you're not very used to.

For example, you might feel really tired after hiking to the mountains.

Secondly, after you have a very busy day at work or at home, you feel very tired.

The third reason for you to feel tired is probably a poor night's sleep.

Insufficient or disturbed sleep is almost certain to cause tiredness.

Difficulty in falling asleep

or waking regularly

might be a sign of sleep problems.

So apart from the natural tiredness,

let's talk about pathological tiredness now.

It's permanent tiredness without any of the natural causes,

which persists even after plenty of sleep,

may be a symptom of an underlying disease, as I said before.

Physical causes include infections, anaemia,

an underactive or overactive thyroid gland

or drugs you are taking for another condition.

Even in moderate quantities,

regular alcohol consumption can have a depressant effect, causing tiredness.

Other than physical causes,

we have psychological causes for this persistent tiredness.

It includes anxiety, depression and stress,

generally through loss of sleep.

I mentioned anemia just now.

What is that? You might ask.

Anemia is an iron-deficiency problem.

This is especially common in women

who suffer from heavy blood loss
during their period or childbirth.

The elderly, pregnant women and people whose diet is low in iron are also prone to
this type of anaemia.

Iron deficiency anaemia can be treated with iron supplement
or by increasing the iron in your diet.

Deficiencies of some of the B vitamins can also cause anaemia.

So if your doctor thinks your tiredness is due to anaemia or an infection,
then he or she may do a blood test to help sort out the cause.

If you are anaemic,

your doctor will usually recommend or prescribe tablets
to replace the deficient substance that is causing the anaemia.

So when should we see the doctor if we feel really tired?

If your tiredness is prolonged

and does not respond to changes in lifestyle

and remedies that you have bought from your pharmacy

you may have an underlying medical problem

and should see your doctor.

If you think your tiredness could be caused by a drug you are taking,
you should consult your doctor.

Never stop any medication

without discussing it with your doctor first.

So that's it for today

and I hope this talk is helpful for you.

Thank you very much.

Appendix A.5 English Speech 2

Thank you very much all of you.

I look forward to all those contributions

and perhaps from my personal point of view particularly to hearing the Greek's
perspective on immigration

which I have some familiarity.

But before we begin,

I would like to a few words about the causes of immigration.

I like to turn first of all to the question of asylum seekers,

and examine statistics over the past ten years

for asylum seekers entering the EU.

Over the past ten years

more than a half of the asylum seekers entering the EU

came from former Yugoslavia, Iraq, Romania, Sri Lanka, Iran, the Democratic
Republic of Congo, Afghanistan, Turkey, Bosnia and Somalia.

And I like first to think for a moment what these countries have in common.

Why is it that so many asylum seekers are coming from these regions?

I think what these countries have in common is perhaps not so much poverty
or an increasing population

or a low life expectancy,

but some kind of conflict.

So that maybe a civil war in some of these countries,
or discrimination against minorities,
I'm thinking about Romania in particular,
or cases of human right abuse.
Sometimes what pushes people to leave their countries of origin is this type of
conflict. Moving away from asylum seekers to immigrants,
the reasons why people choose to immigrate,
emigrate, I should say,
are varied.
But I like to turn first of all to economic issues.
Obviously very often people choose to leave their countries of origin
in the hope of escaping from grinding poverty at home.
But if you look at it from the other perspective,
not from the country of origin's perspective,
but from host country's perspective,
there are also reasons
why some countries might welcome immigrants.
And I like to take the example of UK.
Following world war two,
there was a labour shortage in the UK.
So the government actually went out looking for immigrants
to try to fill posts.
So 157,000 Poles came into the UK,
that was partly because certain ties had being formed
during the war between Poland and the UK.
A numbers of Italians came in,
and also many West Indians.
And in fact to choose a historical landmark, if you like,
it was when the ship called the Empire Wind Rush docked at Tilbury on 22 of June,
1948,
that the era of mass immigration into the UK began.
So it was at the end of the 40s,
that really in the 50s mass immigration became a phenomenon in the UK.
And those are economic reasons if you like,
tie to the labour market,
but to be more generic
and explaining the reasons why people leave the country,
I said often it's because they're seeking for a better life elsewhere.
And that might be in economic terms
and it might be for other reasons.
So some people might be seeking home somewhere where they can find,
for instance, better housing,
or a better education system,
opportunities for their children, perhaps.
And superior medical care,
or a rather broader term,
a better quality of life.
And if you look now at the EU,
you'll find that there are people leaving the UK, leaving the EU,
to emigrate to places like New Zealand, or Canada, or even Spain with the EU,

seeking a better life in somewhere.

And I think what I'd like to do now is to turn to our speakers
and hear their insights into the causes of immigration
and indeed asylum application.

So I'd like to begin first of all with Eric, please.

Appendix A.6 English Speech 3

Now I would like to say a few words about today's topic,
which is climate change,
a very topical issue.

I wonder if you know anything about chaos theory,
which strikes me as being relevant to climate change.

You probably all heard the rather famous idea
that the flapping of a single butterfly's wing today on one continent
can produce a tiny change in the status of atmosphere.

And that can lead to a knock-on effect
and mean that for instance a hurricane that might have happened in Indonesia doesn't
happen
or vice versa.

That's the theory behind chaos,
the idea that a tiny change in initial conditions can lead to tremendous differences in
the final effect.

And chaos theory is something underlies climate.

Climate is something that develops in unpredictable patterns.

And the biggest difficulty in the past few decades has been in predicating weather,
in predicting change,

that has been some controversy in the past few decades
about whether the source of changes
that people think that is in climate are a genuine change
or a cyclical effect,

something that occurs every few centuries, for instance.

Some people have attributed global warming to a cyclical change
as opposed to anything genuine.

But now, I would like to quote to you from the world meteorological organisation's
report,
the year 2002.

This report states, that 1998 was the warmest year so far on record.

The record began in 1960.

The report further states that the year 2002,
last year, was the 2nd warmest year on record,
with temperature averaging .5 degree Celsius
above the 1961 to 1990 mean.

I find that quite interesting this idea that last year was the warmest year on record
and also the 9 or ten of the warmest years have been in the 90s.

And I think there's a growing body of evidence
hat we are seeing in climate is a genuine change
as opposed to cyclical event.

I am sure all of you will be aware of some of the extreme weather events

that we're also seeing around the globe.
There are many examples that could be mentioned.
One example is the terrible flooding
that we witness in August on the Elbe, the Danube Rivers,
and that affected Germany, Czech republic, Austria, Romania, and Slovakia.
In Germany alone,
the damage was estimated at 9 billion US dollars.
So it seems to those of us laypeople that there are more and more of these extreme
weather events,
that you turn on the news now
that there's flooding in one place
and drought somewhere else, hurricane's elsewhere.
And I think I am trying to say is that there is a growing consensus about the severity
of climate change, that it's a genuine fact
and this is going to be harmful to our planet.
And one of the issues that people are looking at more and more is the cause of the
climate change.
Because some of the factors involving in climate change are man-made,
for instance the famous CFCs
and the emission of greenhouse gases.
Of course there are some things that we can't influence,
non man-made events,
but there're some things that we perhaps could do to try
to alleviate the emission of greenhouse gases, for instance.
So that is the purpose of today's discussion,
to talk about some of the scientific issues underlying climate change,
to try to understand what causes this change
and then discuss the political issue.
Because if anything can be done to reverse climate change,
or at least to minimise the effects,
it's going to require international cooperation.
Because all the environmental issues are cross-border issues.
They're not restricted to countries within the narrow confine of the borders.
So we are going to have to discuss political issues
to do with solidarity and cooperation.
I think the issue of the third world country and developing countries are going to be
particularly important.
So those are just a few words of the background.

Appendix B RST annotations and RSTTool

The attached CD-Rom contains all the RST annotations of the six source speeches in Chinese and English, as well as the interpretations produced by the Professional and Trainee (Control) and (Test) groups discussed in this thesis.

In conformance with the terms of its licence, Mick O'Donnell's RSTTool is also included.

Bibliography

- Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting*, 6(1), 43–67.
- Ahn, I.-K. (2005). Pedagogical Considerations of Perspective Coherence Problems in Simultaneous Interpreting as a Result of Linguistic Structure, Illustrated by German-Korean Examples. *Meta*, 50(2), 696-712.
- AIIC. (2004). *Setting up a Conference Interpreting Training Programme*. AIIC. Available: http://www.aiic.net/ViewPage.cfm?article_id=27&plg=1 [2005, March].
- Aktins, M. J., Beattie, J., & Dockrell, W. B. (1993). Assessment issues in higher education. London: Employment Department, Future & Higher Education Branch.
- Anderson, J. R. (1995). *Cognitive Psychology and its Implications*. New York: W.H. Freeman.
- Arnold, J. E., & Tanenhaus, M. K. (in press). Disfluency effects in comprehension: How new information can become accessible. In E. Gibson & N. Perlmutter. (Eds.), *The processing and acquisition of reference*. Cambridge, MA: MIT Press.
- Ballester, A., & Jimenez, C. (1992). Approaches to the teaching of interpreting: mnemonic and analytic strategies. In C. Dollerup & A. Loddegaard (Eds.), *Teaching Translation and Interpreting* (pp. 237-244). Amsterdam/Philadelphia: John Benjamins.
- Bartłomiejczyk, M. (2004). Simultaneous interpreting A-B vs. B-A from the interpreters' standpoint. In G. Hansen, K. Malmkjær & D. Gile (Eds.), *Claims, Changes and Challenges in Translation Studies* (Vol. 50, pp. 239-249). Amsterdam/Philadelphia: John Benjamins.
- Bateman, J., & Delin, J. (2005). Rhetorical Structure Theory, *Encyclopedia of Language and Linguistics* (2 ed.). Oxford: Elsevier. (to appear).
- Blum-Kulka, S. (2000). Shifts of Cohesion and Coherence in Translation. In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 298-313). London and New York: Routledge.
- Brennan, S. E., & Schober, M. F. (2001). How Listeners Compensate for Disfluencies in Spontaneous Speech. *Journal of Memory and Language*, 44(2), 274-296.
- Buhler, H. (1986). Linguistic (semantic) and extralinguistic (pragmatic) criteria for the evaluation of conference interpretation and interpreters. *Multilingua*, 5(4), 231-235.

- Campbell, S., & Hale, S. (2003). Translation and interpreting Assessment in the Context of Educational Measurement. In G. Anderman & M. Rogers (Eds.), *Translation Today: Trends and Perspectives* (pp. 205-224). Clevedon: Multilingual Matters.
- Caron, J. (1997). Toward a procedural approach of the meaning of connectives. In J. Costermans & M. Fayol (Eds.), *Processing Interclausal Relationships. Studies in Production and Comprehension of Text* (pp. 53-74). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chan, S. W. K., Lai, T. B. Y., Gao, W. J., & T'sou, B. K. (2000). *Mining Discourse Markers for Chinese Textual Summarization*. Available: <http://acl.ldc.upenn.edu/W/W00/W00-0402.pdf> [2005, 6/1].
- Choi, Jung-Yoon. (2004). A metacognitive approach to evaluating consecutive interpretation for novice learners. *Conference Interpretation and Translation*, 6(2), 169-185.
- Clifford, A. (2001). Discourse Theory and Performance-Based Assessment: Two Tools for Professional Interpreting. *Meta*, 46(2), 365-378.
- Collados Ais, A. (2002). Quality Assessment in Simultaneous Interpreting: The Importance of Nonverbal Communication. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies Reader* (pp. 327-336.). London/New York: Routledge.
- Crystal, D. (1991). *A Dictionary of Linguistics and Phonetics* (3rd ed.). Oxford: Basil Blackwell.
- Dam, H. V. (1998). Lexical Similarity vs. Lexical Dissimilarity in Consecutive Interpreting. *The Translator*, 4(1), 49-68.
- De Beaugrande & W. Dressler. (1981). *Introduction to Text Linguistics*. New York: Longman.
- Delin, J., Hartley, A., Paris, C., Scott, D., & Linden, K. V. (1994). Expressing Procedural Relationships in Multilingual Instructions, *Proceedings of 7th International Workshop on Natural Language Generation (IWNLG 7)* (Vol. 1, pp 61-70). Kennebunkport, Maine.
- Donovan, C. (2004). Report: European Masters Project Group: Teaching simultaneous interpretation into a B language: Preliminary findings. *Interpreting*, 6(2), 205-216.
- Elliott, N., & Higgins, A. (2005). Self and peer assessment - does it make a difference to student group work? *Nurse Education in Practice*, 5(1), 40-48.
- Ericsson, K. A. (2000). Expertise in interpreting: An expert-performance perspective. *Interpreting*, 5(2), 189-222.
- Ficchi, V. (1999). Learning consecutive interpretation: An empirical study and an autonomous approach. *Interpreting*, 4(2), 199-218.

- Fox Tree, J. E. (1995). The Effects of False Starts and Repetitions on the Processing of Subsequent Words in Spontaneous Speech. *Journal of Memory and Language*, 34(6), 709-738.
- Gaiba, F. (1999). Interpretation at the Nuremberg Trial. *Interpreting (Special Issue on History of Interpreting)*, 4(1), 9-22.
- Gentile, A., Ozolins, U., & Vasilakakos, M. (2001). *Liaison Interpreting: A Handbook*. Australia: Melbourne University Press.
- Gernsbacher, M. A., & Givón, T. (Eds.). (1995). *Coherence in Spontaneous Text*. Amsterdam/Philadelphia: John Benjamins.
- Gile, D. (2000). The History of Research into Conference Interpreting: A Scientometric Approach. *Target*, 12(2), 297-321.
- Gile, D. (2001). *Interpreting Research: What you never wanted to ask but may like to know*. Available: <http://www.aiic.net/ViewPage.cfm/article229.htm> [05 September 2005]
- Gile, D. (2001a). Quality Assessment in Interpreter Training Courses (L'Evaluation de la qualite de l'interpretation en cours de formation). *Meta*, 46(2), 379-393.
- Gile, D. (2001b). *The Role of Consecutive in Interpreter Training: A Cognitive View*. AIIC. Available: <http://www.aiic.net/ViewPage.cfm/article262> [05 September 2005]
- Gile, D. (2002). *Recent trends in research into conference interpreting*. Paper presented at the International Conference on Translation and Interpreting Studies, Hankuk University of Foreign Studies Graduate School of Interpretation and Translation, Korea.
- Gile, D. (2005). Teaching conference interpreting. In M. Tennynt (Ed.), *Training for the New Millennium* (pp. 127-151). Amsterdam/Philadelphia: John Benjamins.
- Greene, J., & Coulson, M. (1995). *Language Understanding: Current Issues*. Buckingham & Philadelphia: Open University Press.
- Halliday, M.A.K. (1964), Comparison and translation. In M.A.K. Halliday, M.McIntosh and P. Strevens. *The linguistic sciences and language teaching*, London: Longman.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Harris, B. (1992). Teaching interpreting: a Canadian experience. In C. Dollerup & A. Loddegaard (Eds.), *Teaching Translation and Interpreting* (pp. 259-268). Amsterdam/Philadelphia: John Benjamins.
- Hartley, A., Peng, G., Mason, I., & Perez, I. (2004). *Peer and self-assessment in conference interpreter training*. Sponsored by CILT. Available: <http://www.lang.ltsn.ac.uk/prf.aspx#lang1>.

- Hatim, B., & Mason, I. (2002). Interpreting: A Text Linguistics Approach. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies Reader* (pp. 254-265). London: Routledge.
- Herbert, J. (1978). How Conference Interpretation Grew. In D. Gerver & H. Sinaiko (Eds.), *Language Interpretation and Communication* (pp. 5-10.). New York: Plenum Publishing.
- Higgins, J. J., Lawrie, A. M., & White, A. G. (1999). Recognising coherence: the use of a text game to measure and reinforce awareness of coherence in text. *System*, 27, 339-349.
- Honal, M., & Schultz, T. (2003). *Correction of Disfluencies in Spontaneous Speech using a Noisy-Channel Approach*. Available: <http://www-2.cs.cmu.edu/~tanja/Papers/Euro03-HonalSchultz.pdf> [2005, 21/07/05].
- Hovy, E. H. (1990). Unresolved Issues in Paragraph Planning. In R. Dale & C. Mellish & M. Zock (Eds.), *Current Research in Natural Language Generation* (pp. 17-45). London: Academic Press.
- Jones, R. (1998). *Conference Interpreting Explained*. St. Jerome. Available: http://dzibanche.biblos.uqroo.mx/cursos_linea2/azanier/trad_iv_u1.htm#tres [2006, 31 May].
- Kahane, E. (2000). *Thoughts on the Quality of Interpretation*. AICC. Available: <http://www.aicc.net/ViewPage.cfm/page197.htm> [2002, 6 November].
- Kalina, S (2002). Interpreters as Professionals. *Across Languages and Cultures*, 3 (2), 169-187.
- Kiraly, D. (2000). *A Social Constructivist Approach to Translator Education: Empowerment from Theory to Practice*. Manchester: St. Jerome.
- Knott, A. and Dale, R. (1992). *Using Linguistic Phenomena to Motivate a Set of Coherence Relations*. HCRC: Edinburgh.
- Knott, A., & Sanders, T. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, 30(2), 135-175.
- Kopczynski, A. (1994). Quality in Conference Interpreting: Some Pragmatic Problems. In M. Snell-Hornby, F. Pöchhacker & K. Kaindl (Ed.), *Translation Studies. An Interdiscipline* (pp. 189-198). Amsterdam/Philadelphia: John Benjamins.
- Kormos, J. (1999). The effect of speaker variables on the self-correction behaviour of L2 learners. *System*, 27(2), 207-221.
- Kurz, I. (1993). Conference Interpretation: Expectations of Different User Groups. *The Interpreters' Newsletter*, 5, 13-21.

- Kurz, I. (2001). Conference Interpreting: Quality in the Ears of the Users. *Meta*, 46(2), 394-409.
- Leung, C. (2005). *A contrastive comparison of the use of major English conjunctions by American and Hong Kong university students (Using the HKUST corpus, HKBU corpus and the ICLE corpus of American English)*. Unpublished Bachelor thesis, Lunds University.
- Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, 14(1), 41-104.
- Levelt, W. J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, Mass.; London: MIT Press.
- Lotriet, A. (2002). Can short interpreter training be effective? The South African Truth and Reconciliation Commission experience. In E. Hung (Ed.), *Teaching Translation and Interpreting 4* (pp. 83-98). Amsterdam/Philadelphia: John Benjamins.
- Mackintosh, J. (1999). Interpreters are Made not Born. *Interpreting (Special Issue on History of Interpreting)*, 4(1), 67-80.
- Mann, W., & Taboada, M. (2005, February 2005). *An Introduction to Rhetorical Structure Theory (RST)*. Available: <http://www.sil.org/~mannb/rst/rintro099.htm> 21/2/05].
- Mann, W., & Thompson, S., A. (1986). *Rhetorical Structure Theory: A Theory of Text Organization*. California: USC Information Sciences Institute.
- Marcu, D. (2000). *The Theory and Practice of Discourse Parsing and Summarization*: The MIT Press.
- Moser-Mercer, B. (1997). Skill Components in Simultaneous Interpreting. In Y. Gambier & D. Gile & C. Taylor (Eds.), *Conference Interpreting: Current Trends in Research* (pp. 123-148). Amsterdam/Philadelphia.: John Benjamins.
- Moser-Mercer, B. (2000). The rocky road to expertise in interpreting: Eliciting knowledge from learners. In M. Kadric & K. Kaindal & F. Pöchhacker (Eds.), *Translationswissenschaft: Festschrift für Mary Snell-Hornby zum 60 Geburtstag* (pp. 339-352). Tübingen: Stauffenburg.
- Moser-Mercer, B. (2003). *Training Workshop for Simultaneous Interpreting Trainers*. Porto: AIIC 32nd Assembly.
- Moser-Mercer, B., Frauenfelder, U. H., Casado, B., & Kunzli, A. (2000). Searching to Define Expertise in Interpreting. In B. E. Dimitrova & K. Hyltenstam (Eds.), *Language Processing and Simultaneous Interpreting* (pp. 107-131). Amsterdam/Philadelphia: John Benjamins.
- Moser, P. (1996). Expectations of Users of Conference Interpretation. *Interpreting*, 1(2), 145-178.

- Nakatani, C. H., & Hirschberg, J. (1994). A corpus-based study of repair cues in spontaneous speech. *Journal of the Acoustical Society of America*, 95(3), 1603-1616.
- Niska, H. (1999, 1 April 1999). *Text Linguistic Models for the Study of Simultaneous Interpreting*. Stockholm University. Available: <http://www.geocities.com/~talk/lic/LIC990329.htm> [2006, 19 May].
- Pearl, S. (1999). The Other Three Eights and the Four F's. Finiteness, Fallibility, Freedom of Speech and Fair Competition in the simultaneous interpretation environment. *The Interpreters' Newsletter*, 9, 3-28.
- Peng, G. (2004, Feb. 25-27). *Emergence of Notions of Expertise: Novice Interpreters' Perceptions of Interpreting Quality*. Paper presented at the 4th Conference on Quality in T&I – Academic & Professional Perspectives, Madrid. Available: http://www.uem.es/web/fil/invest/publicaciones/web/EN/autores/peng_art.htm.
- Peng, G., & Hartley, A. (2005, 2-3 June). *Rhetorical Structure Theory for Grasping Sense in Conference Interpreting*. Paper presented at the International Conference on Sense in Translation, E.S.I.T, Paris.
- Peng, G., Hartley, A., Mason, I., & Perez, I. (2004). *Self and Peer-Feedback for Simultaneous Interpreter Training*. Paper presented at the European Society for Translation Studies 4th Congress, Lisbon.
- Petite, C. (2005). Evidence of repair mechanisms in simultaneous interpreting: A corpus-based analysis. *Interpreting*, 7(1), 27–49.
- Pöchhacker, F. (1993). From Knowledge to Text: Coherence in Simultaneous Interpreting. In Y. Gambier & J. Tommaola (Eds.), *Translation and Knowledge* (pp. 87-99): Centre for Translation and Interpreting, University of Turku.
- Pöchhacker, F. (2001). Quality Assessment in Conference and Community Interpreting. *Meta*, 46(2), 410-425.
- Pöchhacker, F. (2004). *Introducing Interpreting Studies*. London and New York: Routledge.
- Reiman, A. J. (1999). The evolution of the social roletaking and guided reflection framework in teacher education: recent theory and quantitative synthesis of research. *Teaching and Teacher Education*, 15(6), 597-612.
- Reiss, K., & Vermeer, H. (1984). *Grundlegung einer allgemeinen Translationstheorie*. Tübingen: Niemeyer.
- Riccardi, A. (2002). Evaluation in interpretation: macrocriteria and microcriteria. In E. Hung (Ed.), *Teaching Translation and Interpreting 4* (pp. 115-126). Amsterdam/Philadelphia: John Benjamins.

- Rozan, J.-F. (2003). *Note-Taking In Consecutive Interpreting*, [Website]. Tertium. Available: <http://interpreters.free.fr/consecnotes/rozan.htm> [2005, 8/11/05].
- Russo, M., & Salvador, P. (2004). Aptitude to Interpreting: Preliminary Results of a Testing Methodology Based on Paraphrase. *Meta*, 49(2), 409-432.
- Sanders, T. J. M., & Noordman, L. G. M. (2000). The Role of Coherence Relations and Their Linguistic Markers in Text Processing. *Discourse Processes*, 29(1), 37-60.
- Sanford, A. J., & Moxey, L. M. (1995). Aspects of coherence in written language: a psychological perspective. In M. A. Gernsbacher & T. Givon (Eds.), *Coherence in Spontaneous Text* (pp. 161-187). Amsterdam/Philadelphia: John Benjamins.
- Sawyer, D. B. (2004). *Fundamental aspects of interpreter education: curriculum and assessment*. Amsterdam/Philadelphia: John Benjamins.
- Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech Disfluency and the Structure of Knowledge. *Journal of Personality and Social Psychology*, 60(3), 362-367.
- Schjoldager, A. (1996). Assessment of Simultaneous Interpreting. In C. Dollerup & V. Appel (Eds.), *Teaching Translation and Interpreting 3* (pp. 187-195). Amsterdam/Philadelphia: John Benjamins.
- Scott, D. R., & Souza, C. S. d. (1990). Getting the Message Across in RST-based Text Generation. In R. Dale & C. Mellish & M. Zock (Eds.), *Current Research in Natural Language Generation* (pp. 47-74). London: Academic Press.
- Secara, A. (2005). Translation Evaluation- a State of the Art Survey, *eCoLoRe/MeLLANGE Workshop* (pp. 39-44).
- Seleskovitch, D. (1977). Why interpreting is not tantamount to translating languages. *The Incorporated Linguist*, 16(2), 22-23.
- Seleskovitch, D. (1978a). *Interpreting for International Conferences*. Washington, DC: Pen and Booth.
- Seleskovitch, D. (1978b). Language and Cognition. In Gerver and Sinaiko (Eds.), *Language Interpretation and Communication* (pp. 333-341). New York and London: Plenum Press.
- Seleskovitch, D. & Lederer, M. (1986). *Interpréter pour traduire*. Paris: Didier Erudition.
- Seleskovitch, D. & Lederer, M. (2002). *Pédagogie raisonnée de l'interprétation*. 2nd Edition. Brussels: Office of Official Publications of the European Communities and Didier Erudition.

- Seleskovitch, D. (1994). *Interpreting for International Conference: Problems of Language and Communication* (S. Dailey & E. N. McMillan, Trans. 2 ed.). Washington, D.C.: Pen and Booth.
- Setton, R. (1994a). Experiments in the Application of Discourse Studies to Interpreter Training. In C. Dollerup & A. Lindegaard (Eds.), *Teaching Translation and Interpreting 2* (pp. 183-198). Amsterdam/Philadelphia: John Benjamins.
- Setton, R. (1994b). Training Conference Interpreters with Chinese- Problems and Prospects. *Literary Studies*, 8, 55-66.
- Shlesinger, M. (1995). Shifts in Cohesion in Simultaneous Interpreting. *The Translator*, 1(1), 193-214.
- Shlesinger, M., Dejean le Feal, K., Kurz, I., Mack, G., Nilsson, A., Niska, H., Pöchhacker, F., & Viezzi, M. (1997). Quality in Simultaneous Interpreting. In Y. Gambier & D. Gile & C. Taylor (Eds.), *Conference Interpreting: Current Trends in Research* (pp. 123-131). Amsterdam/Philadelphia: John Benjamins.
- Tappe, H., & Schilder, F. (1998). *Coherence in Spoken Discourse*. Paper presented at the COLING-ACL conference, Université de Montréal, Québec, Canada.
- Viaggio, S. (2003). *The Tribulations of a Chief Interpreter*. AIIC. Available: <http://www.aiic.net/ViewPage.cfm/article1031.htm> [09/05/05]
- Victori, M., & Lockhart, W. (1995). Enhancing metacognition in self-directed language learning. *System*. 23(2), 223-234.