Using Rhetorical Structure Theory (RST) to describe the development of coherence in interpreting trainees

Article *in* Interpreting · September 2009

DOI: 10.1075/intp.11.2.06pen

CITATIONS
5

READS
921

1 author:

Gracie Peng
Tunghai University
5 PUBLICATIONS **9** CITATIONS

SEE PROFILE

Using Rhetorical Structure Theory (RST) to describe the development of coherence in interpreting trainees

Gracie Peng
University of Leeds

Making global sense has long been seen as one of the most important criteria for judging the success of a given interpretation. For consecutive in particular, special emphasis is placed on the coherence and structure of the rendition. This study addresses the question of how to investigate coherence in interpreting and observe its development in trainees. We propose Rhetorical Structure Theory (RST), as a framework for exploring how coherence is realised in interpretations produced by professional as well as trainee interpreters. A corpus of 66 consecutive interpretations, by eight novice and three professional interpreters, of three Chinese and three English speeches, was transcribed, segmented into functional units, and mapped into a tree-like RST description. The analyses and results reveal that novices tend to focus on local cohesion while professionals tend to emphasise the global structure of the discourse. This difference can usefully be addressed in training.

Keywords: making sense, consecutive interpreting, Rhetorical Structure Theory (RST), RST trees vs. bushes, coherence profile

1. Introduction

Coherence is an essential attribute of a good interpretation, and indeed of any discourse or text (Beaugrande & Dressler 1981). Yet trainee interpreters often have great difficulty in producing coherent output. Thus for training purposes there is a need to find a way to describe how a given interpretation is more or less coherent and to compare two interpretations of the same source, for example one by a novice and the other by a professional. This need is complicated by the fact that we cannot even expect two interpretations by experienced professionals to be identical: there can be many equally acceptable interpretations of a given source. Our

Interpreting 11:2 (2009), **216–243**. DOI 10.1075/intp.11.2.06pen
ISSN 1384–6647 / E-ISSN 1569–982X

approach must therefore be revealing of coherence without being sensitive to the specific wording of the texts. Moreover, on the premise that conscious knowledge of the differences in coherence between professionals and trainees will eventually help trainees be more coherent, it should be accessible to the trainees themselves.

With this in mind, we experimented with RST as a descriptive framework (Section 3.3). We collected a corpus of source texts and their interpretations by both professionals and trainees, which we annotated in RST terms (Section 4.3). The RST analyses provided the input for an algorithm that assigned a coherence score to each interpretation.

From this data, we derived various parameters for comparing performances (Section 5). First, we observed that interpretations by professionals, although different in length and in wording, were very similar along the derived parameters, not only to one another but also to the source text. On the basis of these parameters, we were able to develop a 'profile' of interpretations by professionals which we then used as the benchmark of a good performance (Section 5.3). Two groups of trainees were recruited for this study. One of them (Test group) received training exposing the trainees to RST and visualisations of their own interpretations. In contrast to the Control group, the performance of the Test group converged quite rapidly towards the professionals' profile (Section 5.6).

2. Coherence in interpreting

In interpreting studies, Seleskovitch's (1978/1994) theory of sense asserts the legitimacy of meaning-based (as opposed to word-based) interpreting and has been accepted and put into practice by interpreters and interpreter-trainers worldwide (Gile 2001). A review of the literature shows that 'making sense', or 'sense consistency' is one of the most frequently proposed attributes to be considered when evaluating the quality of interpretations, both consecutive and simultaneous, in professional settings (e.g. Kurz 1993).

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 two major steps of the interpreter's work as: 'understanding ("making sense of") what has been expressed in a source language, and expressing the ideas grasped, i.e. the "message", in another language so that they "make sense" to the target audience' (2004: 56).

How sense consistency and logical cohesion are actually achieved in interpreting becomes, therefore, an important issue. At this point, it is useful to appeal to a notion of 'texthood' that transcends interpreting studies. Following Halliday and Hasan's definition of a text as 'any passage, spoken or written, of whatever length,

that […] form[s] a unified whole' (Halliday & Hasan 1976: 1), we view interpretations as texts, or at least as products that aspire to texthood. Shlesinger observes that a text is held 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).

Coherence, as Blum-Kulka describes it, '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). Hatim and Mason suggest that successful consecutive interpreting, in particular, should show 'a clear outline of the way a text is structured' (2002: 262). In short, for an interpretation to 'make sense', it needs not only to be clearly signposted with cohesive links, but also structured at the macro level so that it is easy for the target audience to comprehend.

It is, not easy, however, for trainees to demonstrate their grasp of this concept in their interpretations. As a trainer, I often have heard colleagues complain that while trainees might be able to preserve most of the information they receive from the speech, they still fail to deliver the message coherently. Such informal observations, together with what has been suggested in the literature, motivate the need to investigate the possibilities of describing coherence in interpreting in more formal terms. Specifically, it seems important to compare how coherence is manifested differently in professional and novice interpretations.

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 & Souza 1990; Shlesinger 1995). Beaugrande and Dressler (1981) suggest that a text, whether oral or printed, should serve as a communicative discourse, mediating between the intentions of the speaker and the needs of the listeners. If the text is not comprehensible for the listeners, it does not fulfil its communicative function. In this sense, in the case of texts which are translated or interpreted, it might be said that two-fold mediation takes place. Shlesinger observes that 'successful translation, after all, will depend on whether target text recipients can achieve second-degree interpretation with minimal extra processing effort' (1995: 209).

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). This ease of understanding has been shown to depend, more specifically, on 'cognitive relationships such as contrast, equivalence, cause and consequence, and temporal sequence, which present and organise information in a logical manner' (Higgins et al. 1999: 347). Moreover, Reinhart believes that a coherent ('ideal') text needs to be 'connected', that the clauses of a text should be formally connected, such that each 'adjacent pair is either referentially linked, or linked by a semantic

connector'(Reinhart 1980 cited in Sanford & Moxey 1995: 162). Also each sentence needs to be logically consistent with the previous one, and sentences need to be 'relevant' to both the discourse topic and to the context of the utterance.

Specifically, with respect to interpreting, Ficchi observes that the discourse of poor consecutive interpretations by trainee interpreters can be confusing and imprecise, 'lacking coherence and cohesion', with sentences not linked but juxtaposed (1999: 202).

Advice on remedying such shortcomings is offered in various pedagogic works. Jones (1998), for example, strongly recommends paying special attention to distinct types of relationships in order to analyse the 'links', and adopting a different strategies for handling the different types of connectors, depending on the nature of the textual link.

We have explained the dual significance of 'making sense' for interpreters. First, they try to identify coherence relations in the incoming discourse in order to grasp the speaker's intended message. In turn, to facilitate their listeners' comprehension of the outgoing interpretation, interpreters have a duty to signal the structure of their own discourse with linguistic markers of cohesion.

It appears that 'there has not been very much research on coherence related to interpreting and translation' (Ahn 2005: 699).We are aware of only one study (Shlesinger 1995) that investigates the shifts of cohesive devices in translation by comparing the source text (speech) and the target text (simultaneous interpretation).

In the absence of any suitable framework in interpreting studies to describe systematically how coherence is displayed in interpretations and how trainees progress in this respect, we set out to establish just such a means of making comparative qualitative judgements about interpretations produced both by individuals and also by groups of interpreters. Assuming we can make generalisations about the structural features of interpretations produced by professionals, we might then see how interpretations by trainees compare. Our aim is to move beyond intuitions to quantifiable observations based on a rigorous linguistic analysis.

In subsequent sections, we analyse the consecutive interpretations produced by both professional and trainee interpreters, with a view to capturing any differences in their handling of textual coherence. First, however, we discuss the potential of Rhetorical Structure Theory (RST) to serve as a suitable framework.

3. Rhetorical Structure Theory (RST)

Rhetorical Structure Theory (RST) is specifically designed to show 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. It enables analysts to represent the coherence relations between the parts of a text in a hierarchical structure. This facilitates comparison of texts that are similar but different, such as various interpretations of the same source text. Moreover, RST confines itself to the text itself. In Mann and Taboada's (2005) words, 'RST is intended to describe texts, rather than the processes of creating or reading and understanding them.'

3.1 RST relations and definitions

While it was originally used primarily to analyse monolingual written texts, from its early days RST has also been used to draw comparisons across languages. In particular, Mann and Thompson (1987) cite its application in a contrastive study of rhetoric in Chinese and English essays. More recently, RST has proved useful in describing the structure of spoken discourse (Tappe & Schilder 1998) and dialogues (Taboada 2004). Significantly, texts of any size or type, including ill-formed speeches, can be analysed using RST.

The 'building blocks' of RST are spans of text. Adjacent spans are linked by one of a set of relations defined by reference to rhetorical goals which 'correspond with the intentions of the speaker' (Bateman & Delin 2006: 590). Typically, one of the spans is identified as the 'nucleus' and the other as the 'satellite'. Bateman and Delin further explain that the identification of a nucleus is determined by its contribution to the rhetorical goals of the text as a whole:

> 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).

In other words, if the satellites are deleted from a text, it still tends to make sense though it will lose some of its content, while deleting the nuclei from a text destroys its texthood by rendering it incoherent. In terms of nuclearity, 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 & Delin 2006: 590).

Figure 1 shows that the result of RST analysis can be represented as a tree-like structure of relations. The first two spans (segment 1–2) — *'She picked up the phone'* and *'She dialled the number'* — form a symmetric multi-nuclear relation of 'Sequence', while the third span *'in order to call the airline'* is the satellite in an asymmetric relation of 'Purpose'.

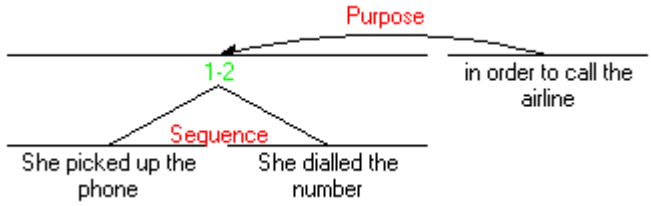


Figure 1. Sample of RST structure (Peng & Hartley 2006)

Mann and Thompson (1987) provide formal definitions for all of the rhetorical relations. Each definition of an asymmetric relation has 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' is defined as follows:

Table 1. Definition of RST relation: Purpose (Mann & Thompson 1987: 64)

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

Mann and Taboada's website on RST (http://www.sfu.ca/rst/) provides a table which summarizes the relationships between the Nucleus and the Satellite in most RST relations. To illustrate this relationship, entries for three selected RST relations are reproduced in Table 2.

Table 2. RST: Nucleus vs. Satellite (Mann & Taboada 2005)

Relation Name	**Nucleus**	**Satellite**
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

Multinuclear relations are defined in similar terms, though clearly there are no satellites on which constraints are placed. The definition of Sequence given by Mann and Thompson is reproduced in Table 3.

Table 3. Definition of RST relation: Sequence (Mann & Thompson 1986: 73)

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

As with the asymmetric relations, the relationship between spans in multinuclear relations is described in a table on Mann and Taboada's website (2005). Entries for the multinuclear relations used in the annotation of our data are reproduced in Table 4.

Table 4. RST multinuclear relations (Mann & Taboada 2005)

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

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 & Delin 2006: 589). They collected and defined about 25 relations, now known as 'classical RST'. This set is reported to cover most of the relations in English texts (Hovy 1990: 19). RST has been successfully applied to the description of text organisation in languages other than English, such as Dutch, Chinese, French, Portuguese, German and Spanish (Bateman & Delin 2006: 589).

Although Mann and Thompson explicitly stated that this list of relations was open-ended, Bateman and Delin note that 'it has in fact proved very stable over the years' (2006: 589). We adopted the set of classical RST relations to annotate data, and added two of our own: Coda and Repair. A Coda is often used to mark the end of a conference speech, in wordings 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 the corpus data.

Figure 2 presents an example of RST analysis of an interpretation from our corpus and illustrates its suitability as a framework for representing the textual coherence of interpretations.

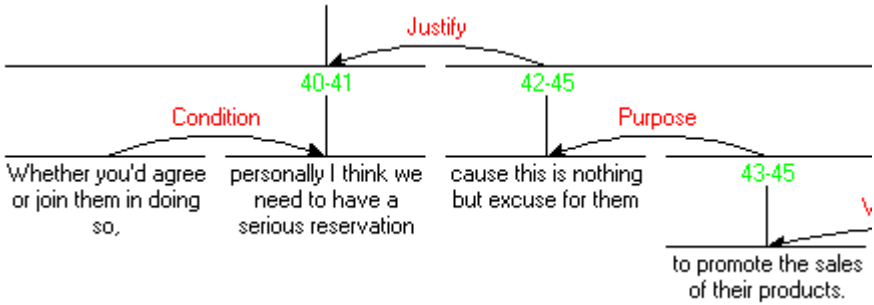


Figure 2. Example of Justify relation

The span containing segments 40–41 (*'Whether you'd agree or join them in doing so, personally I think we need to have a serious reservation'*) is an opinion presented by the speaker (Nucleus), and the span containing segments 42–45 (*'cause this is nothing but excuse for them…'*) supports the speaker's opinion (Satellite). It is clear that the cohesive device *'cause'* (because) in segment 42 explicitly marks the relation of 'Justify'. As we shall see, in this regard it is particularly useful in the kind of analysis undertaken in the present study.

3.2 Applications of RST

In addition to its application in text generation, over the last decade or so RST has been used in novel ways.

Significantly for our own research, Marcu in his work on automatic text summarisation (2000) designed an algorithm (Equation 1) to assign salience weightings to text spans in order to provide a principled basis for summarisation by the progressive deletion of less salient spans.

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

Equation 1. Marcu's algorithm (2000: 139)

It makes use of two major elements of RST annotations: the depth of the discourse structure as visualised in the branching structure of the RST diagrams above, and the total number of RST relations in a text. For our purposes, Marcu's algorithm has the useful property of assigning a weight to each complete RST tree. In essence, the deeper the branching structure and the greater the number of relations present in a text, the higher the weighting. The depth of branching of a text struc-

ture is particularly telling because it reflects the complexity and connectedness of the relations in a text.

In sum, this algorithm facilitates the comparison of texts in terms of global coherence, without regard to the specific relations that are present.

3.3 RST for comparing interpretations

Given the time and other constraints interpreters face, they often have to prioritise incoming information, possibly omitting items judged less salient or redundant or re-ordering items to produce a more coherent narrative. Such strategies partly explain why there can be many equally acceptable interpretations of a given source text. In the absence of a single gold standard against which to judge a number of interpretations, Marcu's algorithm enables us to abstract away from the wording of the text and capture in quantitative terms the degree to which it hangs together as a whole. Despite this abstraction, it appears to distinguish reliably between good and poor interpretations; and thanks to this abstraction, it provides a striking visualisation of the differences between coherent and incoherent texts, as we soon show.

4. Methodology

4.1 Subjects

In addition to six source speeches, the corpus data consisted of consecutive interpretations by eight trainee and three professional interpreters of three Chinese and three English speeches, giving a total of 66 texts. All subjects had Chinese as their A language and English as their B language. All three professionals had been active for more than ten years and are regarded as established conference interpreters.

The two groups of trainee interpreters were recruited from the one-year MA programme in Interpreting and Translation Studies (MAITS) at Leeds University from two consecutive years. Both groups of students were recruited by the same recruiting standards and procedures. They were mainly from mainland China, with some from Taiwan. None of the trainees in this study had any previous professional training or experience in interpreting before joining the MA programme.

At the start of the first year, eight students were recruited on a voluntary basis for the 'Control group'. Six students were recorded for each speech. Four recordings of each speech were used for data analysis. Recordings were selected according to practical considerations. Technical problems inevitably arose during the recording process. A few recordings were inaudible and thus impossible to transcribe.

In the following year, six students were recruited for the 'Test group', but later two decided to withdraw from the experiment. As a result, four students were recorded for each speech.

The only difference between the two groups was that the Test group had its attention drawn more systematically to issues of coherence and cohesion than the Control group. Broadly speaking, two instruments contributed to this. Firstly, a grid was introduced to encourage structured peer feedback (Hartley, Peng, Mason, & Perez, 2004). Secondly, the Test group were given sessions on RST, not only to learn about the difference between sound and poor discourse structures but also to facilitate the visualisation of analysis of the various hierarchical structures representing different levels of coherence in interpretations.

4.2 Data collection

4.2.1 *Speech preparation*

Six speeches were prepared: three in Chinese and three in English. While the topics varied, none required special or technical knowledge. All speeches were relatively short and clearly structured. Specific features of the six speeches are detailed in Table 5.

Table 5. Details of source speeches

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
	Notes	Without note-taking	Note-taking	Note-taking
	Delivery	Live	Live	Live
	Source	China Times (2002)[1]	Euro-China Meeting: Illegal immigration[2]	Speech notes from Isabelle Perez (2002)
English	Length	4 minutes	5 minutes	5.5 minutes
	Topic	Tiredness	Immigration & Asylum seekers	Climate change
	Notes	Without note-taking	Note-taking	Note-taking
	Delivery	Live	Audio recording	Audio recording
	Source	Boots Family Health Book[3]	MAITS mock conf. recording: 26/11/03	MAITS mock conf. recording: 19/02/03

The Chinese speeches were delivered by the same speaker working from bullet points that she had prepared for herself as lecture cues. English Speech 1 was also delivered live under the same conditions. English Speeches 2 and 3 were audio recordings of two live speeches by the same native English speaker in two mock conferences held in Leeds.

4.2.2 *Collection of interpretations*

In both Chinese and English, the subjects were not allowed to take notes during the recording of their interpretation of Speech 1. 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 the interpreting programme in Leeds, note-taking for consecutive interpreting is usually introduced after the first four weeks of active listening and training in public speaking. By the time trainees interpreted Speech 2 (week 9), they had been practicing consecutive interpreting with note-taking for four to five weeks. Trainee interpretations of Speech 3 were recorded after their end-of-term exam, by which time they were expected to be able to perform successful consecutive interpreting with note-taking. In order to maintain the realistic nature of the training and 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 recording arrangements in advance. Conditions were as consistent as possible with those under 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 either English or Chinese, while it was allowed for Speeches 2 and 3. Notes were not collected afterwards. Also, each speech was delivered without a break. A two minute' break was given between each speech.

In brief, in each year trainee interpretations were recorded in three sessions over five months. As explained above, these sessions took place at certain points in the course of their training (Table 6). The professionals were recorded individually, with each of the six speeches being interpreted in a single session.

Table 6. Details of the recording of interpretations

	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		

4.3 Data annotation

After recording, each speech and all interpretations were transcribed and then annotated. These were then segmented into spans of text (e.g. a clause which serves as a 'building block' in the text), and manually mapped into a tree-like RST description (RST trees) with RSTTool.[4]

RST annotations are sometimes criticised for their lack of objectivity. Indeed, each annotator's analysis is likely to be distinct due to subjective concerns. 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 analysis. Among the analyses produced by trained annotators, the segmentation and attribution of nuclearity revealed much higher compatibility than did the identification of individual relations.

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

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. The screenshots in Figures 3, 4 and 6 illustrate the major functions of the RSTTool.

In total this resulted in 72 transcribed and annotated texts (6 original speeches, 18 professional interpretations and 48 trainee interpretations).

4.4 RST trees and tree weight

Figure 3 shows a part of an English speech (Speech 3) which was delivered for interpreting purposes.

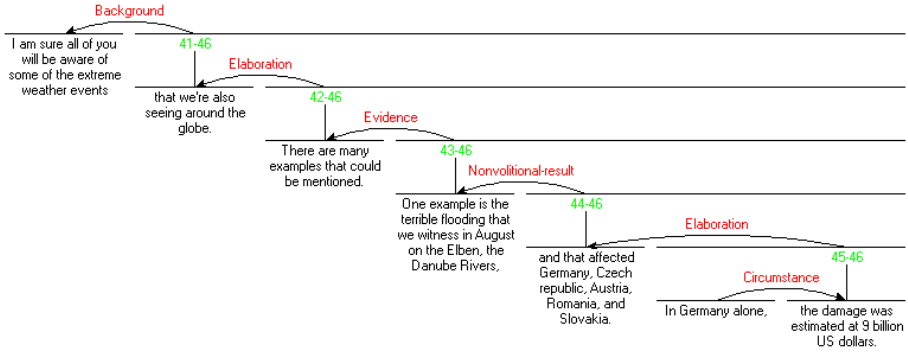


Figure 3. Example of RST annotation of English Speech 3

First of all, segment 40 (*'I am sure all of you will be aware of some of the extreme weather events'*) is the nucleus, which is central to the whole text, and the rest of the text (41–46) is its satellite. The RST relation holding between these two text spans is that of 'Background'. Segment 42 (*'There are many examples that could be mentioned'*) is a satellite in relation to segment 40–41. On the other hand, segment 42 is also a nucleus, and its satellite includes segments 43–46, which provides supporting 'Evidence'. The last two segments (45 and 46) also participate in 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'.

We then used Marcu's algorithm (Equation 1 in Section 3.2) to assign a weight to each complete RST tree, thus facilitating comparison of different trees. The algorithm favours right-branching structures, following Marcu's observation that the best discourse trees, are 'often those that are skewed to the right' (Marcu 2000: 137), which also corresponds to the trend of a natural text which unfolds in linear fashion.

The higher its score, the better a tree is deemed to be. Significantly, this measure of quality depends on the structure of the tree and not on the identity of any particular rhetorical relationships. Focusing solely on rhetorical well-formedness, the measure is able to accommodate the fact that there can be several different but equally acceptable interpretations of a single source speech.

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 overall weight were recorded by a longer text. The ratio of RST tree weight to total words will therefore be taken as an important indicator of the textual coherence of the interpretations in this study.

5. Results

5.1 Trees vs. bushes

The RST trees of the professional and trainee interpretations revealed very different discourse structures. In general, the RST trees describing professional interpretations are deeper and broader than those describing student performances. In the professional interpretations, all of the spans in the text tend to be related to a single root node 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 rather more like 'bushes', exhibiting only local coherence with no single root.

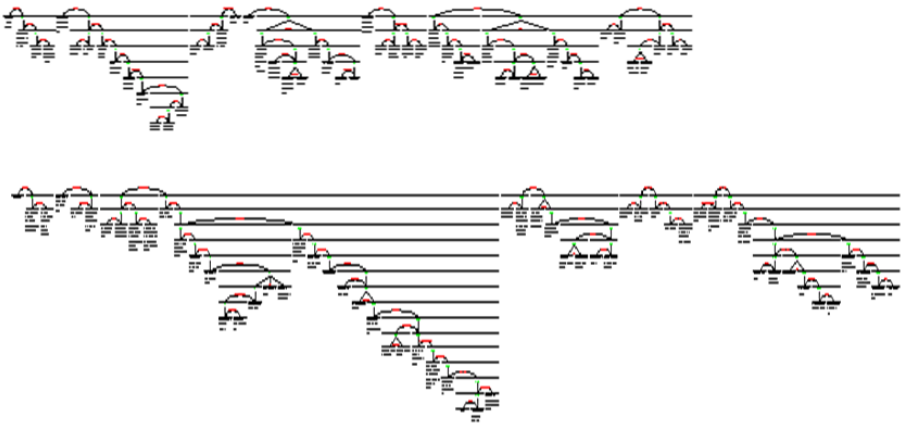
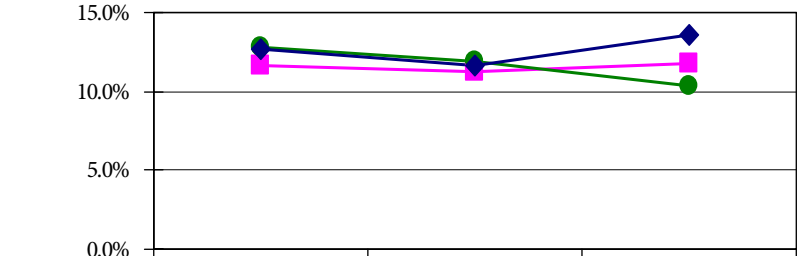


Figure 4. RST 'bushes' and 'trees'

Figure 4 is intended to give a gestalt impression of the difference between the RST representations of interpretations of the same speech by a trainee and by a professional. For this illustration, we 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. The bottom half represents an interpretation of the same speech by a professional.

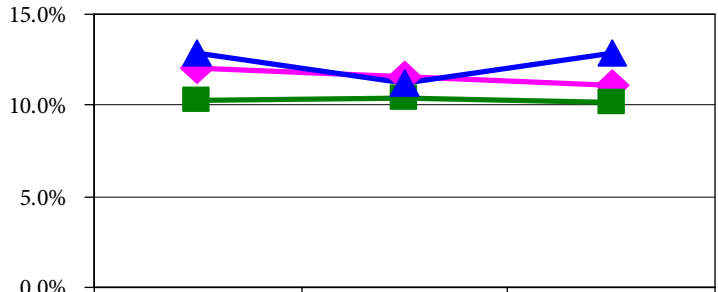
While the number of RST relations is also a major element in determining the weight of an RST tree using Marcu's algorithm (2000), the right-branching principle and depth of the structure are even more important. Greater depth means, ultimately, that a greater number of spans are subsumed under a single root relation. Significantly, in Figure 4, the RST tree of the professional's performance reaches a maximum depth of 15, while that of the trainee reaches a maximum depth of only 8.

The relatively deep tree structure of the professional interpretation (shown in Figure 4) makes a significant contribution to the total score for the RST tree. The number of relations in the two tree structures in Figure 4 is similar: 52 for the



	Spch1	Spch2	Spch3
Prof	11.7%	11.2%	11.8%
Control	12.9%	11.9%	10.4%
Test	12.7%	11.7%	13.5%

Figure 5a. C>E interpretations: RST relations vs. total words



	Spch1	Spch2	Spch3
Prof	12.0%	11.6%	11.1%
Control	10.3%	10.4%	10.1%
Test	12.9%	11.3%	12.9%

Figure 5b. E>C interpretations: RST relations vs. total words

trainee and 59 for the professional. However, the RST tree representing the professional interpretation scores 156 while that of the trainee interpretation scores just 41.

To clarify this point further, we investigated the ratio of the number of RST relations to total number of words in the text (relations/total wds) and found that it remained fairly constant across speeches, language combinations and different interpreter groups, with around 1% variance overall (Figures 5a and 5b).

Thus, from the corpus data, one difference between professional and trainee interpretations appears to lie not 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 (Figures 6a and 6b).

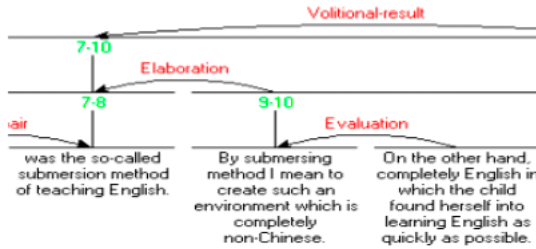


Figure 6a. Explicit signposting: professional interpretation

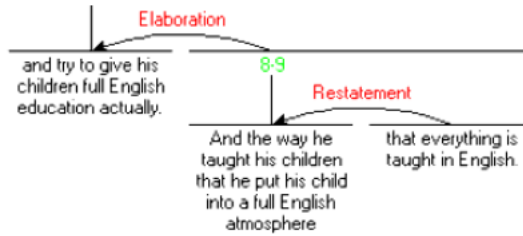


Figure 6b. Explicit signposting: trainee interpretation

In summary, it is plausible to claim that the difference between professional and trainee interpretations lies in the degree of coherence with which information is expressed. This would seem to provide evidence in support of the pre-theoretical observation noted above.

5.2 Coherence profile

Comparison of the coherence of different texts represents a challenge, since it is the combination of factors such as the length of a text, the number of rhetorical

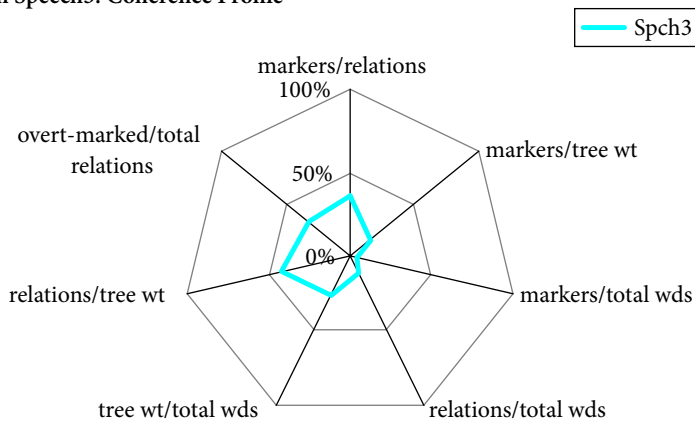
relations within it, the number of marked relations, and many others that contributes to the coherence of a text (Scott & Souza 1990: 56).

That said, after annotating the RST relations in six source speeches by both professional and trainee interpreters, we identified five major parameters that could plausibly enable us to characterise the coherence of a text: (1) the length of a text (*total wds*); (2) the use of explicit markers (*markers*); (3) the number of RST relations (*relations*); (4) the number of relations marked by explicit markers (*overt-marked*), and (5) the weight of RST trees (*tree wt*). These parameters were combined in seven ratios (Table 7), intended to reveal a 'coherence profile' for the performance of individuals or sets of individuals. While these factors do not produce a complete picture of coherence, they are sufficient to give a general profile of the coherence of a discourse (Peng 2006).

Table 7. Seven ratios as parameters of coherence profile with rationales

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.

The seven ratios were calculated for individual performance first, and then averaged for each group for further comparison. The coherence profiles were plotted as radar charts for each group to make them easier to analyse and compare. The coherence profile of English Speech 3 is shown in the radar chart in Figure 7.



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 7. Coherence profile of English Speech 3

5.3 Benchmark validation

In order to compare different coherence profiles, we must first establish a benchmark, preferably not the source speech, since translations (interpretations in the present study) are quite different from source texts. For instance, many studies have confirmed that translations tend to be more explicitly marked (Pym 2005). For the purposes of the present study, we considered it reasonable to use the output of our professional interpreters as a benchmark against which to compare student interpretations. Indeed the radar charts clearly show a strong similarity between the coherence profile of professional interpretations and that of the original speech. This significant finding is illustrated by Figure 8, showing the match between the professionals' interpretation (Prof) and the original speech in Chinese (Speech 2, see Table 5). In contrast, the coherence profiles of the two groups of trainee interpreters, the 'Control' and 'Test' groups, present a very different picture (Figure 9). Table 8 gives the average percentage for each ratio for all the three groups of interpreters of the same Chinese speech (Speech 2C).

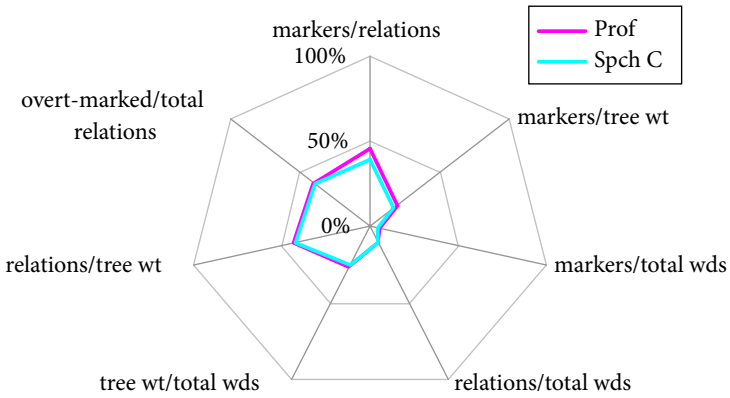


Figure 8. Coherence profiles of professional interpretation and Speech 2C

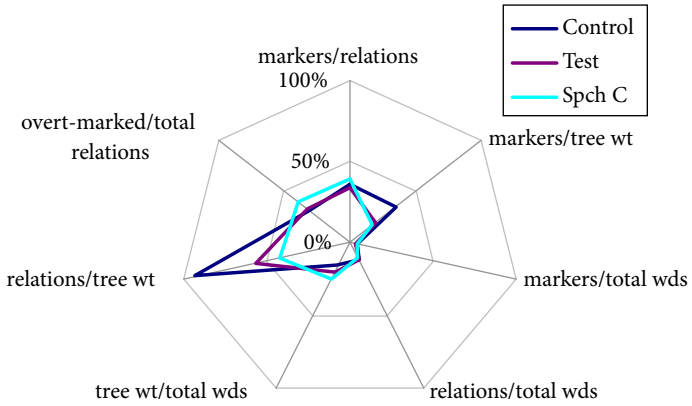


Figure 9. Coherence profiles trainee interpretations and Speech 2C

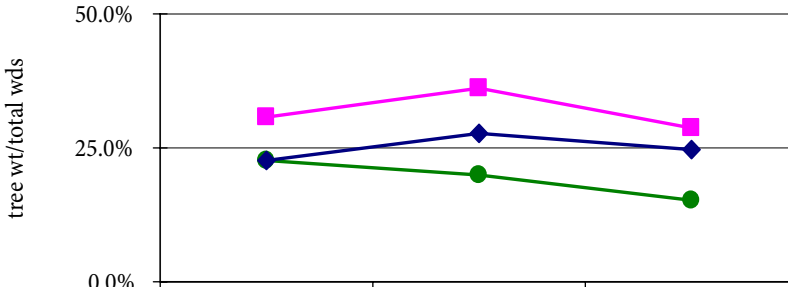
Table 8. Coherence profile data for Speech 2C & its interpretations

Spch2C>E	Speech	Prof	Control	Test
markers/relations	39.6%	45.7%	35.7%	34.2%
markers/tree wt	17.0%	20.0%	34.6%	19.3%
markers/total wds	4.3%	5.1%	4.2%	3.8%
relations/total wds	10.9%	11.2%	11.9%	11.7%
tree wt/total wds	25.3%	26.5%	16.0%	20.8%
relations/tree wt	42.9%	43.7%	93.3%	57.0%
overt-marked/total relations	39.6%	40.1%	30.9%	32.7%

5.4 'Heavier' RST trees in Chinese

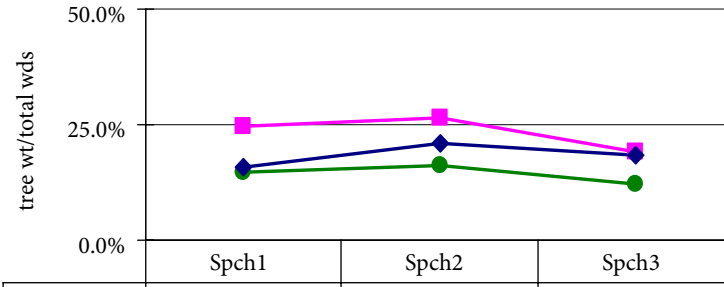
As discussed in 4.3, one of the most significant indicators of coherence is the weight of the RST tree for the text as a whole, which is calculated using Marcu's algorithm. The corpus data show a strong and direct correlation between the RST tree weight and the total words of a text: the longer the text, the heavier the RST tree. Moreover, the ratio of RST tree weight to the total number of words in a text (*tree wt/total wds*) was also higher in longer texts.

To further explore this phenomenon, we 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 10) and English (Figure 11) interpretations.



	Spch1	Spch2	Spch3
Prof	30.6%	36.1%	28.8%
Control	22.6%	19.9%	15.3%
Test	22.6%	27.6%	24.5%

Figure 10. E>C interpretations: RST tree weight vs. total words



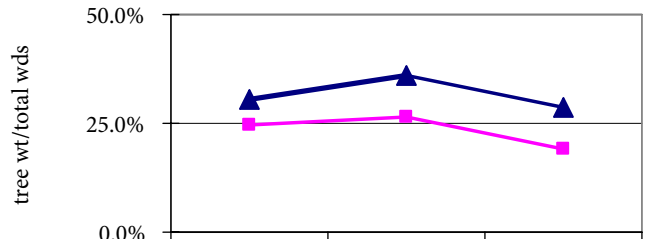
	Spch1	Spch2	Spch3
Prof	24.5%	26.5%	19.0%
Control	14.6%	16.0%	12.1%
Test	15.7%	20.8%	18.2%

Figure 11. C>E interpretations: RST tree weight vs. total words

First of all, from Figure 10 and Figure 11 it is clear that the ratios of RST tree weight to total words (*tree wt/total wds*) are generally higher in the Chinese interpretations than in the English interpretations. Now let us consider the ratios for each group of interpreters in turn. Taking the trainee performances first, the average ratios (*tree wt/total wds*) were consistently higher in Chinese interpretations (around 7%) than in English interpretations. In the Chinese interpretations, the average ratios in both Test and Control trainee groups for English Speech 1 were 22.6% vs. 15% for the English interpretations of the Chinese Speech 1 by both trainee groups.

At the time of interpreting the first speech in both language directions, each group of trainee interpreters had received only four weeks of training on memory and public speaking, and had had little interpreting practice. The results suggest that, at this early stage, trainee interpreters were more capable of conveying coherence in Chinese (their A language) than in English (their B language).

As illustrated in Figure 12, the ratios for professional Chinese interpretations (Prof E>C) were also consistently higher than those for English interpretations (Prof C>E). 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.



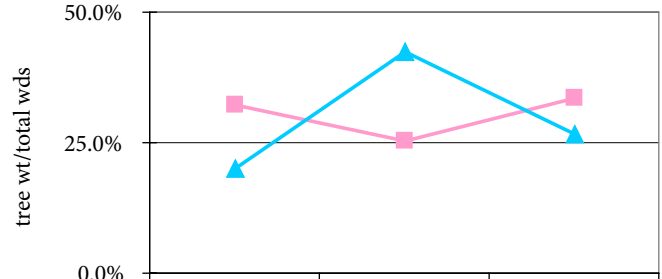
	1	2	3
Prof C>E	24.5%	26.5%	19.0%
Prof E>C	30.6%	36.1%	28.8%

Figure 12. Prof. C>E & E>C interpretations: RST tree weight vs. total words

While it is tempting to suspect that this is due to inherent differences between Chinese and English, Figure 13, which shows ratios for the source speeches, suggests that language difference was not the main reason for the pattern observed in the interpretations.

In Figure 13, the ratios (*tree wt/total wds*) for the Chinese speeches (Spch C) are not consistently higher than those of the English speeches (Spch E): 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 texts. In particular, it



	1	2	3
SpchC	32.4%	25.3%	33.5%
SpchE	20.1%	42.4%	26.6%

Figure 13. Chinese and English speeches –RST tree weight vs. total words

would seem that the interpreters in this experiment (both professionals and trainees) were better at conveying coherence in their mother tongue, Chinese.

Bartłomiejczyk (2004) conducted a survey of interpreters' views on the directionality of SI. She reports that professional interpreters feel more confident working into their A language. Trainee interpreters, on the contrary, often feel more at ease when working into their B language; in this direction, there are fewer comprehension problems. The findings here, whereby interpretations into the A language convey better textual coherence, substantiate the point of view of professional interpreters.

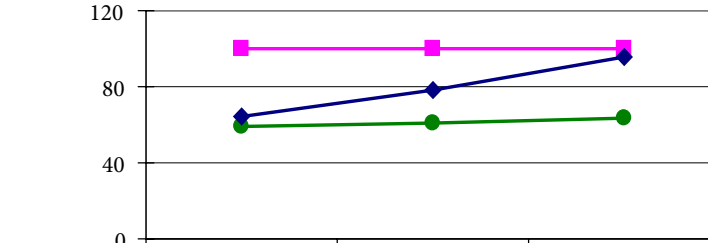
5.5 'Heavier' RST trees of professional interpretations

Another important finding is that professional interpretations have higher ratios of RST tree weight to total words than trainee interpretations into both Chinese and English. This follows from the fact that the weights of the RST trees which represent the professional interpretations are consistently greater than those of trainee interpretations.

In order to facilitate comparison, the ratios of RST tree weight to total words were normalised by setting the professional ratios to 100 and adjusting those of trainee interpretations proportionally. Let us first consider Chinese-English interpretations.

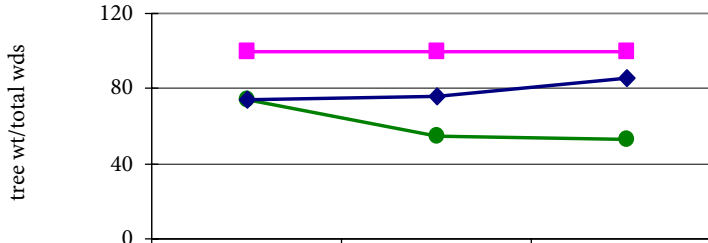
Figure 14 shows a consistent gap between the Professional group and the Control group. The interpretations produced by trainees in the Control group may have the same length as professional interpretations, but the total weighting of RST trees is around 60% that of professional interpretations of the same speeches.

Moreover, this situation does not improve through the period of training. When working into Chinese (Figure 15), the gap between the professional and the



	Spch1	Spch2	Spch3
Prof	100	100	100
Control	60	61	64
Test	64	78	96

Figure 14. C>E interpretations: RST tree weight vs. total words — Professional interpretations as benchmark



	Spch1	Spch2	Spch3
Prof	100	100	100
Control	74	55	53
Test	74	76	85

Figure 15. E>C interpretations: RST tree weight vs. total words — Professional interpretations as benchmark

trainee (Control) groups was relatively narrow 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. In fact, their early success may be due to the four weeks of training on memory and public speaking that they had been given prior to being recorded for Speech 1. This enabled them to memorise the major structure and arguments of a speech. They were trained to reproduce this structure with the aim of producing coherent interpretations. They had also acquired other basic skills. As such their ability to handle simple speeches, such as Speech 1, was similar to that of the professionals.

Nevertheless, the widening gap between professional and trainee interpretations of Speech 2 and Speech 3 into Chinese warrants further investigation, given

that interpreters should be able to work better into their A language.

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' (1994/1986: 32). Therefore it is plausible that, even with note-taking, trainee interpreters still failed to grasp 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.

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

5.6 Quality awareness facilitates performance

It is worth noting, however, that the gap between the Test group and the Professional narrows with training in both Chinese and English interpreting. This result supports our claim that trainees progressed faster in conveying global coherence when their attention was drawn to it explicitly.

We can see from Figure 14 that the score for interpretations by the Control group was around 60% that 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.

In English-Chinese interpretations (Figure 15), likewise, the gap between Professional and the Control groups was conspicuous, particularly 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 far removed from those of the professionals. In Speech 2 and Speech 3, the difference between the two trainee groups was clear. The noticeable gap between the performances by the Control group and the Professional grew, while the Test group showed a rise

towards the Professional. In Speech 3, the most complicated and challenging of the three, the Test group still managed a score of 85.

As described in 4.1, the major difference between the two trainee groups lies in the introduction of the feedback tool and in the specific attention drawn to their realisation of the significance of coherence as described in informal sessions on RST. Apart from these two conditions, the Test group was recruited and trained in the same way as the Control group. We suggest that the introduction of the feedback tool and the attention given to coherence account for 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 study. However, based on our RST analysis, we have witnessed the development of coherence in interpretations in the Test group. We suggest that their understanding of quality criteria and of coherence in particular, improved as a result. The results also show that awareness of coherence leads to better interpretations. Of course we do not claim that trainees can develop into professionals after a few months of training. Our analysis does show, however, that it is possible for trainees to produce interpretations with a degree of coherence similar to that of professionals, when they are explicitly guided by the peer feedback tool.

6. Conclusions

Like any research project, this study was subject to certain constraints in terms of time and materials.

However, every care was taken to ensure the validity of the not inconsiderable body of data under analysis. On the other hand, this study is the first use of RST with interpreted texts. Not only has RST proved to be very suitable for describing coherence across languages, it has also provided a useful framework for the comparison of coherence across different interpretations. Using RST, the improvement of performance over time and between groups can also be easily visualised.

Given the empirical evidence presented here in support of the notion that what distinguishes professional interpretations from those by trainees is not the amount of information, but the coherence with which it is expressed, coherence would seem to be an obvious issue to tackle in training. The introduction of peer feedback also shows great potential in raising trainees' awareness of quality. This in turn further improves their performances. The study demonstrates that this innovative approach offers interesting findings and implications for interpreter training, especially in terms of collaborative learning, as well as directions for further research in both the conference interpreting and RST communities.

Acknowledgement

We would like to acknowledge the valuable comments and suggestions of the anonymous reviewers while assuming responsibility for any remaining errors or infelicities.

Notes

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

2. EU-China Meeting: False Travel Documents (3–4 Nov 2003), The Hague.

3. *Boots Family Health Book* (1997: 156–157).

4. Mick O'Donnell's RST tool can be downloaded from http://www.wagsoft.com/RSTTool/section2.html

References

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.

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* . Amsterdam/Philadelphia: John Benjamins, 239–249.

Bateman, J. & Delin, J. (2006). Rhetorical Structure Theory. In K. Brown (Ed.), *Encyclopedia of language and linguistics* (2nd edn), vol. 10. Amsterdam: Elsevier, 558–296.

Beaugrande, R. de & Dressler, W. (1981). *Introduction to text linguistics*. New York: Longman.

Blum-Kulka, S. (1986/2000). Shifts of cohesion and coherence in translation. In L. Venuti (Ed.), *The translation studies reader*. London/New York: Routledge, 298–313.

den Ouden, J. N., van Wijk, C. H., Terken, J. M. B. & Noordman, L. G. M. (1998). Reliability of discourse structure annotation. *IPO Annual Progress Report*, 33. Technische Universiteit Eindhoven, The Netherlands: IPO Centre for Research on User-System Interaction, 129–138.

Ficchi, V. (1999). Learning consecutive interpretation: An empirical study and an autonomous approach. *Interpreting* 4 (2), 199–218.

Gile, D. (2001). Interpreting research: What you never wanted to ask but may like to know. http://www.aiic.net/ViewPage.cfm/article229.htm (accessed 9 May 2005).

Halliday, M. A. K. & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hartley, A., Peng, G., Mason, I. & Perez, I. (2004). Peer and self-assessment in conference interpreter training. Sponsored by CILT. http://www.lang.ltsn.ac.uk/prf.aspx#lang1 (accessed 9 May 2005).

Hatim, B. & Mason, I. (2002). Interpreting: A text linguistic approach. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader*. London/New York: Routledge, 255–265.

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.
Hovy, E. H. (1990). Unresolved issues in paragraph planning. In R. Dale, C. Mellish & M. Zock (Eds.), *Current research in natural language generation*. London: Academic Press, 17–45.
Jones, R. (1998). *Conference Interpreting Explained*. Manchester: St Jerome.
Kurz, I. (1993). Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter* 5, 13–21.
Mann, W. & Thompson, S. A. (1987). *Rhetorical Structure Theory: A theory of text organization*. Marina del Rey, CA: USC Information Sciences Institute.
Mann, W. & Taboada, M. (2005). *An introduction to Rhetorical Structure Theory (RST)*. http://www.sil.org/~mannb/rst/rintro099.htm (accessed 21 February 2005).
Marcu, D. (2000). *The theory and practice of discourse parsing and summarization*. Cambridge, MA: The MIT Press.
Peng, K. (2006). *The development of coherence and quality of performance in conference interpreter training*. Unpublished PhD thesis, University of Leeds.
Peng, G. & Hartley, A. (2006). Rhetorical Structure Theory for grasping sense in conference interpreting. In M. Lederer (Ed.), *Le sens en traduction*. Paris: École Supérieure d'Interprètes et de Traducteurs, 247–257.
Pym, A. (2005). Explaining Explicitation. http://www.tinet.org/~apym/on-line/translation/explicitation_web.pdf (accessed 8 April 2009).
Pöchhacker, F. (2004). *Introducing interpreting studies*. London/New York: Routledge.
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*. Amsterdam/Philadelphia: John Benjamins, 161–187.
Seleskovitch, D. (1978/1994). *Interpreting for international conferences: Problems of language and communication* (trans. S. Dailey & E. N. McMillan). Washington, DC: Pen and Booth.
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*. London: Academic Press, 47–74.
Shlesinger, M. (1995). Shifts in cohesion in simultaneous interpreting. *The Translator* 1 (2), 193–214.
Taboada, M. T. (2004). *Building coherence and cohesion: Task-oriented dialogue in English and Spanish*. 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.

Author's address

Gracie Peng
Centre for Translation Studies
University of Leeds
Leeds, LS2 9JT
U. K.

k.peng@leeds.ac.uk

About the author

Gracie Peng is a Lecturer in interpreting at the Centre for Translation Studies in the University of Leeds. Before, during and since her PhD project on improving interpreter training, Gracie Peng has been active as both a trainer and practitioner of conference interpreting. She is particularly interested in applying research to drive pedagogic innovation with a focus on collaborative learning and performance evaluation.