Sabine Braun and Kurt Kohn

Towards a pedagogic corpus approach to business and community interpreter training

1 Introduction

Much of Sylvia Kalina’s own research has focused on interpreting pedagogy (cf. e.g. Kalina 2000a and many of her early publications), linked to her interest in interpreting as a strategic process (cf. e.g. Kalina 1998; Kohn/Kalina 1996) and interpreting quality (cf. e.g. Kalina 2002, 2005). Moreover, she has embraced all settings and modes of interpreting including those that have received less attention in the Interpreting Studies community (cf. e.g. Kalina 2000b, 2003, 2009). It seems appropriate then for a contribution to a volume that is dedicated to Sylvia Kalina to focus on interpreting pedagogy in relation to those less explored settings and modes of interpreting, i.e. the various forms of business and community or public service interpreting, which are practiced in a wide range of contexts and language combinations today and in which consecutive and liaison interpreting are the dominant modes.

Higher education programmes for interpreting in such settings have emerged throughout Europe (cf. Niska 2005), but due to the long-standing focus of interpreter training on conference interpreting, available pedagogical resources are often of limited benefit for training in business and community interpreting. The EU speech repository\footnote{http://www.multilingualspeeches.tv/scic/portal/index.html (Retrieved: 19.11.2011).}, for example, a well-established e-learning resource for trainee interpreters, provided by the DG Interpretation (SCIC) with contributions from all European institutions, understandably focuses on material that is most beneficial for the training in conference interpreting as required by these institutions. The SIMON (Shared Interpreting Materials Online) collection for interpreter trainers at the Ecole de Traduction et d’Interprétation in Geneva has been a similar initiative, albeit with restricted access (cf. Seeber 2006). A more recent effort to provide open-access training material is the ORCIT website (Online Resources for Conference Interpreter Training)\footnote{http://www.orcit.eu (Retrieved: 19.11.2011).}. In addition to source text material, it offers interactive tools for trainers and students but is targeted at conference interpreting as well. Different in its aims and composition but geared towards the same group of trainees, the European Parliament Inter-
preting Corpus (EPIC)\(^3\) includes speeches from the European Parliament and their simultaneous interpretations covering English, Italian and Spanish. The subcorpora (source and target texts) have been aligned and annotated for interpreting research purposes and possible pedagogical applications (cf. Bendazzoli/Sandrelli 2005).

What makes these resources interesting beyond their use in conference interpreter training is the approach to the storage and retrieval of the materials. All of the above resources can be searched according to specific criteria including, for example, proficiency level, topic or content, type of exercise, communicative features (cf. Seeber and Zelger 2007; Bendazzoli and Sandrelli 2005). The highest level of detail and linguistic analysis is certainly provided by the EPIC corpus.

Other developments in interpreter training include computer-assisted interpreter training (CAIT) tools (for overviews cf. Berber 2008; Sandrelli and Manuel Jerez de 2007), although these are, once again, mostly geared towards conference interpreting. One exception is Black Box 3.0, which enables interpreter trainers

\[\text{to create exercises to train students in all interpreting modes, i.e. simultaneous, consecutive and liaison interpreting, as well as sight translation. (Sandrelli/Manuel Jerez de 2007: 289)}\]

Authoring tools open up interesting opportunities for supporting interpreter training, but a prerequisite for using them, i.e. filling them with content, would seem to be a good and pedagogically viable collection or corpus of resources, which is what is missing for the area of business and community interpreting. The material collections mentioned above, and especially the EPIC corpus, demonstrate that corpus-based approaches to interpreter training are potentially very useful. However, as said above, they are geared towards conference interpreting. Equally important, research in ‘Applied Corpus Linguistics’, especially in the field of corpus-based language learning, suggests that corpus-based approaches developed for (linguistic) research purposes require adaptation in order to use the full pedagogical potential of corpora (cf. Braun 2005, 2007a, 2010; McEnery/Wilson 1997a, b; Mishan 2004; Campoy/Luzón 2007; Kohn et al. 2010; Kohn 2012). This concerns the content of the corpora as much as the requirements for the retrieval of information and hence the annotation process and tools.

Whilst the work with genuine source text material, e.g. speeches that have been interpreted for real-life purposes, is clearly a valuable exercise, not all such speeches may lend themselves to classroom practice (or self-study), because it may be difficult for a student to re-contextualise them, i.e. there may be problems with what Widdowson (2003: 66, 115) has referred to

Towards a pedagogic corpus approach

as ‘authentication’. A pedagogically viable corpus of resources should therefore follow other – pedagogically motivated – criteria for material selection. It would, for example, seem reasonable to include texts that exhibit common and characteristic interpreting problems. For practical reasons, it would also be useful to include material with a greater lifetime than many political speeches. The collection of such material can be challenging. Equally important, once a suitable selection has been made, the question arises as to what kind of retrieval options will be useful for interpreter trainees and trainers, bearing in mind that they are not normally linguistic or research experts.

To return to the resources required for business and community interpreter training, these are often bilingual dialogues, which are at the heart of many interpreter-mediated business and public service encounters, and/or spontaneous speech. Such materials are not only harder to come by than prepared ‘conference’ speeches, but they also pose different challenges than the materials (i.e. source texts) conference interpreters mostly work with. One initiative that is currently developing resources for training in business and community interpreting is the European Lifelong Learning project ‘Interpreting in Virtual Reality’ (IVY). The project combines the use of spoken audio/video corpora with the use of a 3D virtual environment to simulate interpreting in business and community settings. The 3D environment is used for role plays, while the corpora are used to fill the 3D environment with content such as monologues and bilingual dialogues, to support the practice of interpreting in the students’ own time and to create exercises to train individual skills such as listening comprehension and note-taking.

This paper will focus on the use of spoken corpora in this context. As pointed out above, ‘Applied Corpus Linguistics’ has produced a growing body of research into the use of corpora in language pedagogy, with most recent work focusing on spoken and multimedia corpora for language teaching. We will argue that interpreter training for business and community settings can benefit immensely from this research and we discuss how these approaches can be adapted to suit the needs of business and community interpreter training. Section 2 provides further background to contextualise the idea and the concept of corpus-based interpreter training. Sections 3 and 4 outline a discourse processing model of interpreting and a range of source text related challenges of interpreting as a framework for developing appropriate annotation categories. Section 5 presents initial ideas for the design of a pedagogical corpus for interpreter training. Section 6 concludes the paper by highlighting how this approach is integrated into the wider context of the IVY project and its aim to support business and community interpreter training.

2 Background: from IVY to BACKBONE and back

The corpus-based research and development work currently conducted in the IVY project is based on prior initiatives to develop pedagogically viable corpora. An early initiative was the development of the ELISA corpus5 (English Language Interview Corpus as a Second-Language Application) in 2003/04, which was a small collection of video-recorded narrative interviews with speakers from different walks of life who gave accounts of their professional life, their career, educational background, future plans etc. (cf. Braun 2005, 2006, 2007a). The aim of this corpus was to have a resource of spoken professional English, intended first and foremost for language learning purposes. The wider aim of the creation of the ELISA corpus was, however, to develop a methodological solution for the use of corpora in language learning and teaching, based on the observation that traditional corpora, created for linguistic research, were not necessarily appropriate for pedagogical purposes. The approach adopted a corpus design sensitive to pedagogical needs. This includes initial ideas on the pedagogical annotation of such a corpus and the development of ‘pedagogical enrichment’ materials (e.g. exercises) to facilitate ‘pedagogical mediation’ (cf. Widdowson 1991, 2003: 102-103), i.e. to support learners and teachers in exploring the corpus.

A key element in this concept was the division of the interviews into topic-based sections (of approximately two to three minutes in length) and their annotation with topic keys, allowing the users (learners and teachers) to retrieve similar sections from interviews with different speakers. It was assumed that this would give learners, who are not normally familiar with using corpus techniques such as concordancing, a more convenient ‘unit of text’ to explore than a set of concordance lines. Working with similar sections from different interviews would, for example, allow learners to compare and acquire the various lexical and grammatical means of expression that different speakers used in similar contexts.

The ELISA corpus was a forerunner of the EU Lifelong Learning project BACKBONE (Corpora for Content and Language Integrated Learning, 2009-10; co-ordinator: University of Tübingen, Germany)6, which produced video interviews with native speakers of British and Irish English, French, German, Polish, Spanish and Turkish, as well as with non-native speakers of English (English as a Lingua Franca corpus). In addition, it further developed and implemented pedagogical corpus annotation and search tools available from the EU Minerva project SACODEYL (European Youth Language, 2005-2008; co-ordinator: University of Murcia, Spain)7. Apart from

---

offering a wide range of pedagogical materials to support second language learning activities in secondary, higher and vocational education, the BACKBONE corpora are also accompanied by generic interpreting exercises that can be applied to any of the video interviews in the BACKBONE corpora. This was a first step to use the corpora for interpreter training. In the IVY project, the BACKBONE and ELISA audiovisual content will form the core of the IVY content.

The objective is to take the BACKBONE pedagogic corpus approach, first developed for second language learning and teaching (cf. Kohn et al. 2010; Kohn 2012), and to adapt it to the needs and requirements of interpreter training. The video interviews will be analysed with regard to the challenges that they would pose as source texts in business and community interpreting. The results of this analysis will then be used as a basis for developing an annotation that is specifically geared towards interpreter training.

The key issue that needs to be addressed, i.e. how the video-recorded interviews can be annotated and stored in such a way that interview sections with relevant and interesting interpreting challenges for business and community interpreting, involving consecutive and liaison interpreting, can be found and retrieved, will be discussed in section 4. As a basis for this, a model of discourse processing in consecutive and liaison interpreting will be specified in section 3. It will provide a framework for identifying source text characteristics as indicators of interpreting challenges. A cognitive discourse processing perspective will be adopted for this model, enabling us to distinguish between source text induced challenges, which the suggested annotation is likely to be able to cover, and other challenges, e.g. challenges arising from an individual lack of knowledge, which are more difficult to capture with any annotation scheme. The cognitive discourse perspective will, however, support predictions about likely problems for interpreting on the basis of both source text features and required knowledge.

3 A cognitive discourse processing perspective

There can be no doubt that successful business and community interpreting (involving consecutive and liaison interpreting) requires specialized knowledge and skills that are commonly acquired through intensive professional training. But where is the actual challenge? Interpreting, like translating, involves understanding a source text and, on the basis of what one has understood, producing a target text. Understanding and producing spoken or written texts, however, involve skills that we all have developed during first language acquisition and that we have further refined with each additional foreign or second language. When it comes to understanding and producing texts, we should thus be sufficiently competent. And here is the snag. With
the acquisition of our first language and of each further language, our compre-
prehension and production skills have been specialised and geared to meet
the conditions and requirements of monolingual processing. Interpreting,
however, involves understanding and producing texts under bilingual condi-
tions. In order to better understand the specific challenges of interpreting,
and translating for that matter, we thus need to take a closer look at how
interpreting creates bilingual processing conditions under which our rou-
tinely practised comprehension and production skills are on unfamiliar terri-
tory and tend to operate somewhat out of tune (cf. Kohn 1990, 2004; Braun
2004).

According to a cognitive and pragmatic discourse model of interpreting
Albl-Mikasa 2007), the interpreter’s understanding of a source text can be
depicted as the creation of a ‘mental model’ of what the text is about (cf.
Brown/Yule 1983; Johnson-Laird 1983); or as the pragmatic engagement in a
meaning negotiating ‘discourse’ event, of which the text is a mere product
(cf. Widdowson 2004: 8). The cognitive-pragmatic unfolding of discourse
from text involves strategic processes of continuous and cyclic utterance
meaning formation (including monitoring and revision) based on linguistic
as well as world-related knowledge. Even before the first words have been
uttered and perceived, the interpreter has formed inferential meaning expec-
tations depending on the relevance-guided activation of previously available
world knowledge.8 These expectations provide ‘top-down’ guidance for an
intuitive linguistic ‘bottom-up’ analysis of the incoming speech signals. This
way, the text-based formation of meaning inferences can be seen as the re-
sult of a continuous interaction between ‘bottom-up’ and ‘top-down’ proc-
essing (cf. Brown/Yule 1983). The formation of utterance meaning is cyclic
in so far as the understanding of previous text passages feeds into ‘top-
down’ expectations for the understanding of succeeding passages which, in
turn, may give rise to backward revision processes. Meaning formation is
strategic because of its intentional goal-orientation: as soon as the interpreter
is satisfied with the comprehension result achieved, the meaning formation
effort is reduced, freeing capacities for target text production.

Production of the target text is, first of all, based on how the interpreter is
and has been able to understand the source text. The same source text, un-
derstood in different ways, gives rise to different discourses and may thus
require different target texts. The following “Rocky” text, frequently cited to
illustrate the inferential nature of human comprehension (cf. Brown/Yule
1983: 139), aptly illustrates the one-to-many comprehension relation between

---

8 ‘World knowledge’ is understood here in a broad sense including e.g. professional,
cultural, fictional or everyday knowledge.
Rocky slowly got up from the mat, planning his escape. He hesitated a moment and thought. Things were not going well. What bothered him most was being held, especially since the charge against him had been weak. He considered his present situation. The lock that held him was strong but he thought he could break it. (cf. Anderson et al. 1977: 372)

Depending on whether this text is understood as referring to Rocky being locked-up in jail or fighting in a wrestling match, certain lexical ambiguities have to be resolved differently. With, for instance, German as the target language especially the underlined words and phrases are affected.

Target text production, however, does not only depend on whether the source text has been successfully understood or not. It is also crucially guided and influenced by considerations of purpose and adequacy (cf. Reiß/Vermeer 1984; Nord 1991), i.e. the interpreter’s attempt to tailor the interpretation to certain expectations and requirements reflecting e.g. target audience characteristics or genre changes.

4 Interpreting challenges and source textual indicators

The cognitive discourse processing perspective enables us to sketch out a model of interpreting challenges and related source textual indicators. These challenges and indicators can be identified along three dimensions of text processing: source text understanding, source text retention, and target text production.9

4.1 Source text understanding

Understanding a source text can be challenging because of certain textual properties whose ‘bottom up’ exploitation requires advanced or specialised linguistic knowledge [interpretation challenge A1].10 This may concern e.g. speaker accent, lexis and phraseology, the use of language for special purposes (LSP), complex grammatical structures, discourse markers, or cohesive ties. Comprehension problems resulting from such challenges can usually be successfully dealt with provided the interpreter possesses the necessary linguistic knowledge as well as the topic-related background knowledge [interpretation challenge A2] required for successful top-down processing in interaction with the linguistic bottom-up clues.

Source text 1: Regional accent (“Alemannischt”, orthographic approximation) combined with a few terminological expressions [IC A2]; interview with a fisherman at Lake Constance, Germany.

---

9 The following account is a further elaboration of the conceptual clarifications in Kohn 1990, 2004.
10 See overview in section 4.4.
Denn fahr’mer zu unserer Felchennette, die ha’mer am Abend vorher, so um 17 Uhr, ausgelegt, und die sind die ganze Nacht im See drin. We’mer die Netze erreicht hand, weret die Netze gehob; es sind Stellnetze. Des kann ich nachher mal erkläre, Unterschied Stellnetz, Zugnetze. Die Stellnetze weret g’hobe, des isch rein Handarbeit, also es isch Muskelkraft gefragt, weil wir zurzeit im Rhein fischet und der Rhein isch ziemlich tief, 20 bis 40 Meter, wo’mer fischet, und des wird durch reine Muskelkraft – weret die Netze angehobe, die Fisch’ weret ausg’löst. Es sind ungefähr 500 Meter Netz und fünf Meter hoch, die ganze Netzfläche.

Mere availability of the required linguistic and thematic knowledge is not always sufficient; knowledge processing constraints can lead to serious comprehension challenges as well. Fast speed, high lexical density, complex textual organisation, heavy information load, or weak ‘bottom up’ support (i.e. indirectness), for instance, may prevent the interpreter from successfully accessing and retrieving knowledge that would be easily available under less strenuous conditions [interpretation challenge A3]. Source text listening and comprehension processes are further compounded by the fact that the interpreter is usually planning and internally formulating a suitable target text at the same time.

A fourth type of interpretation-related comprehension challenge concerns cases in which the depth and precision of understanding supported by the source text are not sufficient for making adequate target text related decisions [interpretation challenge A4]. Pertinent source text characteristics include, e.g., vague lexis, confused and contradictory terminology, convoluted sentences, lexical, structural or pragmatic ambiguities, weak textual organization and lack of coherence. Any of these source text characteristics may result in a lack of clarity that cannot be resolved despite the interpreter’s bottom-up and top down effort.

Understanding is always geared towards a certain purpose. While in most situations of monolingual understanding a certain lack of clarity does not create any problem for the listener (or reader) and, more often than not, may not even be noticed, this can be quite different under interpreting conditions. If neither the bottom-up signals nor the available top-down knowledge are sufficient for the interpreter to reach the required depth and precision of source text understanding, basically three strategic options are available in consecutive and liaison interpreting:

- a ‘risk-taking’ strategy by which one of the possible meanings is favoured over the others when deciding on the target text ‘equivalent’;
- a ‘playing-it-safe’ strategy by which the passage is rendered in such a generalised fashion that the vagueness or uncertainty is kept;
- a ‘meaning negotiation’ strategy by which the interpreter asks the speaker for more information.
Obviously, a ‘risk-taking’ strategy would be the least preferable solution since it may easily result in a meaning decision that is not coherent with what comes up in later source text passages. A ‘meaning negotiation’ strategy is normally possible in consecutive (especially ‘short consecutive’) and liaison interpreting. In this connection, it should be noted that ‘meaning negotiation’ is part and parcel of ‘ordinary’ monolingual communication; it is a collaborative process of strategic importance for communication success. In some cases, however, a ‘playing-it-safe’ strategy may be the only feasible option for the interpreter; although it tends to water down the interpretation result.

4.2 Source text retention

Because of the time lag between source text comprehension and target text production in consecutive and liaison interpreting, successful target text production not only depends on depth and precision of source text understanding but also quite significantly on the interpreter’s ability to memorize and keep track of what s/he understood previously both in terms of meaning and form. Retention of what the interpreter has understood is particularly important when s/he is faced with long source text passages, complex thematic text organisation, or high information loads [interpretation challenge B1]. Retention of certain phonetic/phonological, lexical or syntactic means of expression (i.e. ‘bottom up’ signals used in the source text) may help the interpreter to reanalyse the meaning of a previous passage if required [interpretation challenge B2]. This might be the case in situations where, for instance, insufficient processing and retention of meaning or speaker errors necessitate a retrospective correction. Elements of expression without much context such as names, numbers and dates may also require special retention techniques.

Source text 2: Enumeration of fish names [IC B2] combined with regional accent (“Alemannisch”, orthographic approximation); interview with a fisherman at Lake Constance, Germany.

Mir hand weniger Cypriniden, also’s heißt weniger Weißfische, da sind die ganze Karpfe drin, die Rotaug, die ganze Aale, Schleie, alles, was Cyprinide sind, geht zurück, und die Fische wie der Saibling, die Seeforelle, dene gefällt’s wieder sehr viel besser, die hand wieder e’n Aufschwung, a Renaissance. Es gibt immer wieder Verschiebungen. Der Felche im Freiwasser, der wird sich auch gut behaupte könne’. Barsche wisser’mer no nicht genau, weil die ernähret sich hauptsächlich wieder von junge Rotaug.
4.3 Target text production

In addition to adequate source text comprehension and retention, the creation of a suitable target text depends on the availability of potential target language equivalents in the interpreter’s lexical and grammatical target language knowledge. This concerns first of all the level of advanced or specialised linguistic knowledge required for successful production, e.g. lexis and phraseology, terms and terminological phrases (LSP), grammatical constructions, functional expressions [interpretation challenge C1]. In particular non-native speakers of the target language may encounter interpreting problems due to insufficient target language knowledge. Other target text production challenges arise in cases where there is a principal lack of target language equivalents due to lingua-cultural differences (‘realia’) between the two languages. This concerns in particular language-specific domains or cultural systems such as politics and society, education, health, or folklore [interpretation challenge C2].

Source text 3: Educational realia (underlined) [IC C2]; interview with a secondary school teacher, Germany


As with source text comprehension, availability of knowledge is not all there is to successful target text production. Under monolingual conditions the speaker/author usually enjoys a considerable degree of semantic authority: it is the speaker who determines the direction in which the meaning of a text goes. This authority is exercised in a process of continuous meaning negotiation and development which begins with the first word uttered (or rather before) and ends with the last word. Heinrich von Kleist’s remark about “the gradual production of thoughts while speaking” (Kleist 1986) needs to be seen in the light of this insight, which is fully in line with cognitive models of discourse processing. The common concluding statement “this is what I wanted to say” thus needs to be reconsidered and changed to “this is what I have said and I endorse it”. Semantic authority constitutes a key condition for successful spoken or written production; and the strategic processes

---

11 We use the term ‘equivalent’ in the sense of Neubert/Shreve’s (1992) communicative and pragmatic notion of ‘textual equivalent’.

12 Kohn 1990, 2004 used the terms “semantic autonomy” and “semantische Autonomie”, respectively.
speakers have developed for the activation and deployment of knowledge for production purposes have been adapted to this condition.

Under interpreting (and translating) conditions, however, the interpreter’s semantic authority is usually severely curtailed. With reference to translation, Eugene Nida fittingly speaks of the challenge “to express one’s own creativity through someone else’s creation” (Nida 1976: 58). The linguistic means of expression the interpreter activates have to fit a more or less ‘given’ meaning. Unlike in monolingual production, there is thus little freedom for bringing the unfolding meaning in line with the means of expression already activated. This explains why production under interpreting conditions is marked by retrieval problems and traces of translationese. Interpreters thus need to learn to adapt their knowledge activation and deployment strategies to the processing constraints that come with a reduced degree of semantic authority [interpretation challenge C3].

A notorious interpretation-specific challenge for target text production is caused by the continuous presence of at least some of the phonetic/phonological, lexical and grammatical means of expression activated and exploited as bottom-up signals in the source text [interpretation challenge C4]. Activation of adequate means of expression, already constrained by the lack of semantic authority, is further hampered by negative transfer and interference from source text means of expressions. To the extent that they continue to be active in the interpreter’s active memory, or are strategically retained for retrospective monitoring, means of expressions from the source text tend to get in the way during retrieval of means of expression for the target text under construction. A particular sentence structure may thus be retained in the target text, although a different word order or a re-organization in several sentences might have been more appropriate. Lexical, syntactic or functional translationese might be the result. While this source text induced production challenge is a general phenomenon, it is also strongly influenced by language-pair specific characteristics: the same source text does not necessarily have the same conflict potential for the production of target texts in different languages.

4.4 Summary of interpreting challenges

The discourse model of potential interpretation problems sketched out above distinguishes and integrates three dimensions: source and target text processing, corresponding interpreting challenges, and source text indicators of these challenges. These dimensions are mapped out in Table 1.

---

13 This holds true even under the most liberal conditions of translational adaptation (cf. the notion of translation adequacy, Reiß/Vermeer 1984).
Table 1: Text processing tasks, ensuing interpreting challenges and their textual indicators

The model addresses the issue of why interpreting (compared to monolingual text processing) can be difficult in the first place, and why it is necessary for interpreters to develop specifically adapted comprehension and
production skills and strategies. It should be noted, however, that the challenges discussed are not absolute; they should rather be thought of as parameters whose settings are sensitive particularly to source text properties, linguistic and lingua-cultural differences between source and target language, speaker and interpreter characteristics, especially the interpreter’s individual knowledge and skills, and the specific interpreting situation.

5 A pedagogic source text corpus for business and community interpreter training

Guided by the discourse model of potential interpretation problems and based on the corpus annotation and search tools available from the European project BACKBONE\footnote{All BACKBONE development work was supported by the EU Lifelong Learning Programme, KA2: Languages (2009-11); also see http://www.uni-tuebingen.de/backbone (Retrieved: 19.11.2011).} (Kohn 2012; Kohn \textit{et al.} 2010), it is possible to use the BACKBONE and ELISA corpora (see section 2) as pedagogic source text corpora for business and community interpreter training purposes. The key component of these corpora are the narrative interviews from BACKBONE and ELISA, which cover topics and ways of speaking that are relevant for business and community interpreting situations. Another crucial feature is a pedagogic annotation and enrichment to facilitate interpreter training activities, with special emphasis on source text related challenges and problems. Furthermore, a range of online search modes has been integrated including pedagogically motivated search function from ‘browse’ and ‘section search’ to ‘co-occurrences’, ‘concordances’ and ‘lexical lists’.

The suggested annotation focuses in particular on topics that are relevant for successful top-down processing and on the source textual indicators of potential interpretation challenges described in section 4. In addition, relevant source text passages may be further enriched with supporting training resources. The corpora can be consulted by interpreter trainers and trainees using queries that are based on flexible combinations of thematic and textual annotation categories as well as lexical items. The spoken source text passages which are returned by the queries lend themselves to a variety of interpretation practice activities from source text comprehension and retention to target text production.

The technological kingpin is the BACKBONE suite of corpus processing tools for transcription and time-stamping, collaborative annotation, linking of enrichment resources, and online search. All BACKBONE development has adopted the principles of open source software and open access content; the corpus tools and corpora are thus available free of charge on the project website. In addition, all BACKBONE corpus tools have been designed with a
'do it yourself' approach in mind to facilitate deployment for new and different pedagogic corpora.

These ‘do it yourself’, open source and open access qualities make it possible to exploit the BACKBONE tools and contents for the creation of IVY source text corpora for interpreter training. For an experimental phase, a sample set of interviews exhibiting textual indicators of relevant interpretation challenges has been selected from the BACKBONE and ELISA corpora, and has been annotated using relevant thematic categories and categories relating to the source text-related interpreting challenges.15

The thematic annotation includes the following main annotation categories (sub-categories are added where appropriate): world of work and business, education, environment and healthcare, cultural, legal, social and political issues, media and technologies, places and regions. For the purposes of the annotation of interpreting challenges, the source text indicators identified in section 4 have been re-grouped according to levels of text description. For an experimental phase, the following categories and sub-categories of source textual indicators are proposed:

Performance
- high speed, sloppy pronunciation
- messy syntax
- confused textual organisation
- accent and dialect
- American, Australian, Irish, Scottish, etc.
- English as a lingua franca (ELF)

Lexis
- terminology
- lexical chunks
- abstract concepts
- metaphorical language
- lexical density
- lack of TL equivalents
- lingua-cultural differences, realia

Syntactic structures
- e.g. structures that are difficult to process (e.g. certain types of embedding)

Utterance meaning
- ambiguities
- vagueness, indirectness
- irony, humour

Towards a pedagogic corpus approach

Pragmatic fluency
- gambits
- forms of address
- politeness
- communicative functions

Discourse
- discourse markers
- lack of cohesive markers (> implicitness)
- need for non-verbal signals
- lack of coherence
- high information load
- afterthoughts or reformulations

Need for interpretation-specific strategies
- anticipation
- factual speaker errors and slips of the tongue

Memory challenges
- long lists, enumerations
- numbers, figures
- proper names
- quotations

The thematic annotation categories and the source textual indicators have been implemented as an annotation category tree in the BACKBONE Annotator tool. Categories from this tree are used to annotate interview sections in a drag & drop fashion. Words and phrases to which a category applies can also be marked. In addition, instructions for interpreting practice activities can be linked to interview sections, using the BACKBONE Virtual Resource Pool and its Resource Sheet function.

In the BACKBONE Search Interface, the annotated interviews can be accessed in five complementary search modes: ‘Browse’, ‘Section Search’, ‘Co-Occurrences’, ‘Concordances’ and ‘Lexical Lists’. In ‘Browse’ all interviews are presented for scrolling, including transcript, section overview, sound and video; ‘Section Search’ gives access to all interview sections including their sound and video files that satisfy a certain combination of thematic and interpretation-specific textual annotation categories; ‘Co-Occurrences’ lists interview sections that deal with a selected topic and at the same time contain a certain free combination of words; ‘Concordances’ displays lines of concordances with the search word(s) in the middle and a short co-text to the left and right; “key word in context” (KWIC) and combines this with a thematic filtering option; and finally ‘Lexical Lists’ produces lists of words and annotated words and phrases again with thematic filtering.
In the IVY project, this new corpus resource will support one of the IVY working modes. This mode has been entitled ‘learning activity mode’ and gives students the opportunity to practise individual interpreting skills such as listening comprehension and note-taking, to cope with specific challenges such as the source text related challenges discussed in this paper and to prepare for an interpreting assignment, e.g. for one of the role plays offered by the ‘interpreting practice mode’ of the IVY project (in the 3D environment).

To practise individual skills, students can, for example, select an entire interview (based on its topic) or retrieve relevant interview sections, using the thematic annotation. To learn how to cope with specific interpreting challenges, students can use the source text related annotation categories and can retrieve relevant instances. In preparation of an interpreting assignment involving a speaker of American English, it would, for example, be useful for the student to be able to retrieve several examples of speech with an American accent. Similarly, it may be useful to retrieve a number of instances of ambiguities. This can be linked with an exercise that asks students to explore the source of the ambiguity in each case and then to reflect on appropriate interpreting strategies. The availability of a whole series of instances will enable students to see possible ‘patterns’, i.e., to make links between the source of a problem and an appropriate strategy for its resolution. To prepare for an interpreting assignment, students can use the various search modes provided by the BACKBONE search interface.

6 Conclusion

This paper has outlined an approach to using corpus-based resources to support interpreter training in business and community interpreting contexts. It was highlighted that resources are scarce in this area and that a corpus-based approach can in principle provide a useful solution but that traditional corpus-based approaches known from corpus linguistics and translation/interpreting research need to be adapted in order to exploit the full pedagogical potential of corpora for interpreter training purposes.

The approach suggested in this paper is based on a discourse processing model of (consecutive/liaison) interpreting that is capable of highlighting source text induced interpreting challenges and their textual indicators. It was discussed how these indicators can be used to develop corpus annotation categories that would support interpreter trainers and trainees in retrieving corpus material that can be used to practice individual interpreting skills, prepare for interpreting assignments and explore specific interpreting challenges.
As a concrete application of this approach, the IVY project, which focuses on business and community interpreter training, was introduced and its link with prior initiatives to create pedagogically relevant corpora (ELISA and BACKBONE) was explained. The IVY project combines the use of the BACKBONE/ELISA corpora and the use of a 3D virtual environment to support a) the practice of live interpreting (in simulations), b) interpreting with ready-made materials (from the BACKBONE and ELISA corpora), and c) the practice of individual interpreting skills, preparation of assignments and exploration of interpreting challenges. It was explained how the BACKBONE corpus approach and the tools developed in the BACKBONE project can be adapted to the needs of interpreter training and in particular to support aspect (c) of the IVY project.

Future research will focus on the pedagogical and technical embedding of BACKBONE and ELISA contents and tools into IVY. This also includes the further analysis and exploration of interpreting challenges and their source textual indicators as well as the specification and pedagogic evaluation of corpus-based interpreter training activities.

References

Braun, Sabine (2005). From pedagogically relevant corpora to authentic language learning contents. ReCALL 17.1, 47-64.


Sabine Braun: Senior Lecturer in Translation Studies at the University of Surrey. She holds an MA in Translation Studies (Heidelberg) and a Dr Phil in Applied English Linguistics (Tübingen). Her research focuses on new forms of translation and interpreting, especially remote and videoconference interpreting. She has led several European projects in this area, including AVIDICUS I and II (Assessment of Videoconference Interpreting in the Criminal Justice System) and IVY (Interpreting in Virtual Reality). She is furthermore interested in the use of video-based corpora in language and interpreter training. She has developed a video corpus of spoken professional English (ELISA) and was a member of the EU project BACKBONE (Corpora for Language and Interpreter Training). She teaches interpreting theory and practice, and applied linguistics, and has developed several MA programmes in Business and Public Service Interpreting.

Kurt Kohn: Chair Professor of Applied English Linguistics at the University of Tübingen and director of the Steinbeis Transfer Centre Sprachlernmedien/Language Learning Media. Research and teaching: second language learning and teaching, English as a lingua franca, translation and interpreting. Since the early 1990s, involvement in European projects on computer-assisted language learning and teaching within a communicative-constructivist approach. Co-developer (with Petra Hoffstaedter) of the language learning and authoring software ‘Telos Language Partner’ and co-author of a large number of multimedia and web-based language learning packages mainly for English, French and German. Recent EU project activities concern the development of pedagogic corpora and web collaboration scenarios for language learning and teaching purposes and interpreter training.