EVIVA
Evaluating the Education of Interpreters and their Clients through Virtual Learning Activities

Report to accompany Deliverable 5.1
Annotated source text material

Petra Hoffstaedter
Kurt Kohn
[petra.hoffstaedter@stw.de]
[kurt.kohn@stw.de]

December 2014

Project coordinator: Sabine Braun

D 5.1 ANNOTATED SOURCE TEXT MATERIAL

1. Summary

The present report accompanies the set of annotated monolingual and bilingual source text content used in the BACKBONE and IVY 3D interpreting practice sessions of the evaluation studies. The report outlines the approach taken. It also provides an overview of the annotation categories and of the annotated content consisting of 32 monolingual narrative interviews in English, German, Greek, Polish and Spanish and of 85 bilingual dialogues in 11 language combinations.

The annotated monolingual content can be accessed from the BACKBONE corpus search site [http://purl.org/backbone/searchtool]; the bilingual dialogues are available in the practice mode of the IVY 3D environment.

2. Content creation approach

One of the aims of the EVIVA evaluation is to gain insights into the practice performance and learning processes interpreting students engage in when using the prepared monolingual and bilingual content in the BACKBONE and IVY 3D environments, respectively. In this connection, difficulties related to the source text are expected to have a crucial impact on the students' performance and learning outcomes. The evaluation thus requires an approach that makes it possible to analyse students' interpreting performance in the light of such difficulties and in relation to features of the relevant VLE.

With this in mind, the sources text material involved in the evaluation was analysed and annotated with regard to potential interpreting difficulties. In a first step, the procedure concerned the video-recorded monolingual narrative interviews available from the BACKBONE environment. Based on the cognitive discourse model of interpreting challenges developed in Braun & Kohn 2012, a system of source text related annotation categories was established. A careful analysis of the BACKBONE content selected to be used in the evaluation helped to further refined and adapt the original system (see chap. 3 below).

Using the BACKBONE Annotator, category trees of interpreting challenges were created and assigned to the different language corpora of monolingual narrative interviews (see Adaptation of Video-Based Environments, Deliverable 4.3). Selected narrative interviews of these corpora were analysed with regard to interpreting challenges and potential difficulties and annotated with the help of the respective category tree. This involved a total of 32 annotated narrative interviews across 5 languages, which also formed the basis for the creation of 83 bilingual dialogues (see chap. 4 below). The bilingual dialogues are indirectly annotated since their creation was based on the respective underlying narrative interviews. (cf. "Bilingual dialogues" in chap 4 below; also see chap 1 of the report accompanying Deliverable 5.2)

3. Categories for the annotation of interpreting challenges

The annotation categories created for the EVVA evaluation cover a wide range of source text related interpreting challenges including in particular lexical, terminological, grammatical, pragmatic, cultural, genre-specific, and paralinguistic ones. In the following overview, the categories used are listed along with a brief comment and (where feasible) an illustrating example. They are organised in five broad processing types: *discourse*, *utterance* and *lexis* as the main linguistic processing dimensions of source text comprehension, *memory* as a key mental processing dimension, and *delivery* behavioural performance.

	Annotation categories	Comments	Examples
Discourse	Connectors	Conjunctions joining sentences, including idiomatic phrases that behave as a unit with the same function, for example "as well as", "provided that".	
	High information load	Ratings based on (interpreter trainer) intuition: a section with a lot of content and e.g. specialist lexis and/or enumerations.	
	Reference to context	Reference to context	"And you break it at this joint here. Then you break it here. Then you take your crack. And you make a couple of little cracks to it. And then you can take the meat out." (Ann: "How to crack a lobster")
	Weak coherence	e.g. lack of connectivity, unclear structuring of information	
	Backward reference		
Utterance	Complex sentence structure	Hypotactic structures	
	Insertion	Insertions with different functions, e.g. asides providing background information, giving additional explanations etc.	"But it was quite — when my parents did live there – it was quite useful ..." "And luckily – it's a very long story and it's a lovely story but I haven't got time now – we actually got the commission."
	Reformulation	Meaning related: repetition of an idea, say something in a different way; may also include summarising sentences	
	Grammatical inconsistency	Includes elliptical structures, incomplete sentences, syntactic blends, sloppy grammar	"That contrasts with our high streets" (instead of "That contrasts with our high street bookshops")

	Factual speaker errors	with or without correction	
	Unclear meaning	Incomplete, inconsistent, ambiguous, vague	
	Quoting and reporting		
	Humour		
Lexis	Specialist lexis	In a broader sense (not only terminology)	
	Idiomatic phrases	Including collocations and figurative expressions	
	Lingua-cultural concepts		
	Acronyms		
	Compound adjective		"topic-based", "energy-wise", "programming-type (work)"
	Related lexical items	Annotate at section level NB two types of related lexical items: (1) e.g. 'shopping' vs 'retail'; 'regeneration' vs 'reconstruction'; 'supported' vs 'assisted' (2) derivatives e.g. 'legislative' vs 'legislature' vs 'legislator' → same category or different category	
	Small phrases	discourse marker, hedges	(e.g. "it has to be said", "you know", "obviously")...
	Colloquialisms	For instances in the ST where there is an unexpected shift to a more colloquial register	"The business plan was the *sticky part* for this group."
Memory	Enumerations	Including lists	
	Numbers and dates		
	Proper names	Incl. their use in adjectives ("Cheshire accent")	
	Quotation		
Delivery	Accent		
	High speed		
	Unclear pronunciation		
	Slips of the tongue	only those slips that are not corrected	
	Nonverbal communication	Facial expressions, gestures etc.	
	Tone of voice	Expressing irony etc.	

4. Overview of annotated content

The categories were used to annotate potential interpreting difficulties in the source text material available for the EVIVA interpreting practice sessions. The identification of difficulties and the subsequent assignment of annotation categories involved a complex procedure combining a thorough source text analysis with experience-based assumptions about difficulties likely to be encountered by students when rendering the texts in another language. The procedure was carried out by a team of annotators cross-checking each other's analyses in a collaborative effort of reaching a consensus.

Narrative Interviews

For the creation of the prepared BACKBONE content, 32 video-recorded monolingual narrative interviews were selected from the English, German, Greek, Polish and Spanish corpora and annotated with regard to interpreting challenges.

In the lists below, the five language groups of narrative interviews are represented by the topics dealt with. A language group is further characterised by an annotation example, each of which illustrates search results for different kinds of interpreting challenges (cf. Fig. 1 to Fig. 5).

The 32 video-recorded monolingual narrative interviews can be accessed from the BACKBONE corpus search site and are available for focused interpreting practice with prepared material.

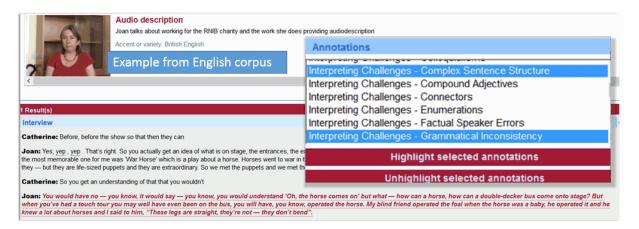| **11 interviews from the IVY English corpus** |
| --- |
| • Baby unit |
| • Community Dietician |
| • The Permaculture Credit Union |
| • Lawyer |
| • Bishops Wood Centre |
| • Primary School |
| • Audio Description |
| • Birmingham |
| • Bristol |
| • Santa Fe Chamber of Commerce |
| • Web designer |

Fig. 1: Annotation examples from the English corpus"

11 interviews from the German corpus

- Ein Berliner im Schwabenland
- Die Schulische Ausbildung von Werkzeugmechanikern
- Das deutsche Schulsystem
- Die Grundschulausbildung
- Die Arbeit an einer Seehundstation in Schleswig-Holstein
- Ein Wattführer aus Schleswig-Holstein
- Die Entwicklung von Kraftstoffanlagen
- Pläne für den Bau einer Konzerthalle
- Einblick in eine Arztpraxis
- Klinikmanagement
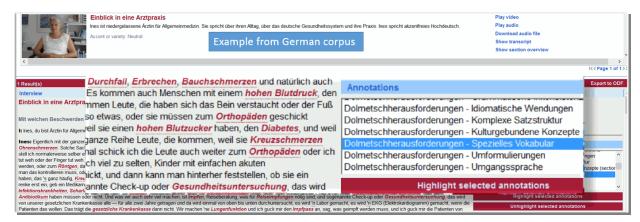- Nachhaltige Fischerei am Bodensee

Fig. 2: Annotation examples from the German corpus

5 interviews from the IVY Greek corpus

- Η Κυπριακή διάλεκτος (The Cypriot-Greek dialect)
- Η κρίση χρέους της Ελλάδας (The Greek debt crisis)
- Αρχαιολογικές ανασκαφές και Μουσείο Ακρόπολης (Archaeological excavations and the Acropolis Museum)
- Τα νέα αναλυτικά προγράμματα εκπαίδευσης (The new national curricula
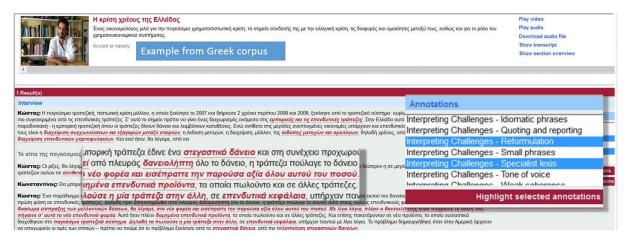- Η "Αραβική άνοιξη" (The "Arab spring")

Fig. 3: Annotation examples from the Greek corpus

4 interviews from Polish corpus
• Środowisko. Ruch uliczny a zanieczyszczenia (The environment. Traffic and pollution)
• Rząd i polityka. Polityka regionów, władze regionalne (Government and politics. Regional policy and regional authorities)
• Pomysły na podróże (Travel ideas)
• Ekonomia. Mobilność pracy, praktyki zawodowe (The economy. Professional mobility, internships)

Fig. 4: Annotation examples from the Polish corpus

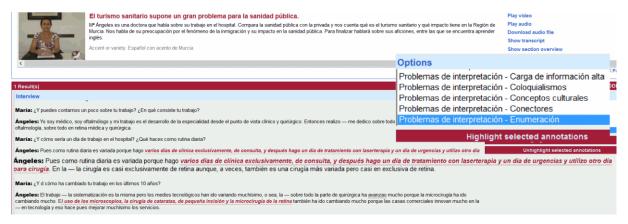| **2 interviews from the Spanish corpus** |
|---|
| • El turismo sanitario |
| • Las nuevas tecnologias |

Fig. 5: Annotation examples from the Spanish corpus

Bilingual dialogues

The annotated 32 monolingual narrative interviews in English, German, Greek, Polish and Spanish also served as a source for the creation of bilingual dialogues implemented in the IVY 3D practice environment. The bilingual dialogues were created by cutting and editing the monolingual narratives and by inserting specifically

scripted and recorded question passages in various languages (English, German, Greek, Polish and Spanish) into or between the original answer turns. This approach is described and illustrated in the report accompanying deliverable 5.2 (see Deliverable 5.2, Chap 1). Due to this procedure, 85 annotated bilingual dialogues in different combinations of English, German, Greek, Polish, and Spanish were available for the evaluation sessions with students (cf. Fig. 6).

The 85 bilingual dialogues can be accessed in the practice mode of the IVY 3D environment and are available for focused interpreting practice with prepared material.

ANNOTATED PREPARED MATERIALS						
		Question turns				
		English	German	Greek	Polish	Spanish
Answer turns	English	10	10	10	10	2
	German	11	11	11	11	
	Greek	5	5	5		
	Polish	4	4		4	
	Spanish	2				2

Fig. 6: Annotated monolingual narratives (grey fields) and bilingual dialogues available in the different languages and language combinations

5. Deployment of the annotated material

The annotation of interpreting challenges served as predictors for potential interpreting difficulties. They provided a strong basis for the selection of suitable material for the evaluation sessions and thus made it possible to select monolingual narrative interviews and bilingual dialogues covering a critical mass and variety of different interpreting challenges.

The annotated interpreting challenges were also used as benchmarks for analysing students' interpreting performance in their evaluation session as well as in the reflective sessions in which students provided introspective comments on the problems encountered. The annotations helped to analyse and better understand how students coped with different kinds of challenges in the different virtual learning environments (see, for instance, the analysis of students' renditions in Case Study 1in chap. 8.1 of Evaluation report, Deliverable 6.1).

The deployability of the annotated source text material is not limited to the EVIVA evaluation. The monolingual narrative interviews and bilingual dialogues offer a rich source for future practice sessions outside the specific evaluation phase. Students are able to use the annotations for selecting and practicing particular interpreting challenges regarding, for instance, grammatical inconsistency, quoting and reporting, numbers and dates, quotations, unfamiliar accents or high speed of delivery. And last but not least, the annotation approach can be used as a model for the further expansion and customisation of the available practice content.