University of Surrey

EPSRC

TIWO – Television in Words

EPSRC GR/R67194/01, January 2002-5

Report on Workpackage 2

Building and Analysing a Corpus of Audio Description Scripts

15 November 2004

FINAL VERSION

Andrew Salway, Elia Tomadaki and Andrew Vassiliou
Department of Computing, University of Surrey

Contact:
a.salway@surrey.ac.uk

TIWO Project web-site:

www.computing.surrey.ac.uk/personal/pg/A.Salway/tiwo/TIWO.htm

Contents

1 Executive Summary

This report describes how the TIWO Audio Description Corpus has been compiled and analysed at the University of Surrey with contributions from Royal National Institute of the Blind (RNIB), ITFC and BBC Audio Description Unit. The reasons for our analyses are on the one hand practical – we want to adapt and apply language technologies to process audio descriptions scripts, and on the other hand theoretical – we want to study how experts put moving images into words and find out whether they use a special language, i.e. is there a 'language of audio description'? Our results may also be of interest to those concerned with implementing and maintaining guidelines for audio description, and those involved in training new audio describers.

At the time of writing, the TIWO Audio Description Corpus comprises a total of approximately 420,000 words, made up mainly of descriptions of 56 feature films, along with a smaller selection of descriptions of television programmes. To try and ensure a representative corpus we consulted two audio description experts and established 9 categories of films in terms of how the experts thought audio description would vary. An overview of results from the analysis of the corpus is given: our results suggest that indeed there is a special language of audio description with idiosyncratic features that distinguish it from general language. Ongoing work is mentioned in which the audio description corpus is being analysed alongside corpora of other text types that are collateral to films, e.g. plot summaries, and screenplays.

Pointers are given to sources of further information including websites, TIWO project documents and academic publications.

2 Description of TIWO Workpackage 2 from Project Proposal Document

A corpus of video data and associated audio description scripts will be gathered. The corpus will include extant descriptions and descriptions elicited from Round Table members in controlled situations following a verbal reporting method. The corpus will be analysed using existing software systems to address the question of how the narrative of a moving image is manifested in text and to adapt language engineering techniques to process audio descriptions, e.g. for style checking; for customisation / personalisation; and, for generating video annotations. In particular Surrey's System Quirk will be used to measure linguistic variance in the corpus and Sheffield's GATE system will be used as an infrastructure for reusing text processing components (e.g. stemming and tagging) and for visualising text structures.
Milestone 2a - month 11: Corpus of digital video data and audio descriptions
Milestone 2b - month 18: Analysis of corpus
Milestone 2c - month 24: Suite of language engineering techniques to process audio description scripts

3 Motivation

The idea of collateral text is emerging as an important consideration for the development of multimedia information systems and it raises interesting questions for artificial intelligence about the link between modalities such as vision and language. In particular, text, which is collateral to moving images (like speech, subtitles and audio description), may be processed into representations of semantic video content to support video retrieval, browsing and summarisation applications. Previous research has investigated the use of speech and subtitles/closed-captions for video indexing and summarisation but little is known about how useful audio description could be – although the fact that it is produced specifically to be informative about the content moving image suggests it should be very useful.

TIWO is concerned with adapting and applying language processing technology to assist the production of audio description and the reuse of audio description, e.g. to automatically generate metadata. As a first step we want to understand more about this language, and more about how trained experts go about putting moving images into words. Because audio description is produced to serve a specific function, and because trained experts who follow guidelines produce it, we expect that it will exhibit characteristics of a special language, i.e. idiosyncrasies that distinguish it from general language. Such idiosyncrasies are of interest because they may give insights into how experts put the moving images into words, and because it may be possible for language processing systems to exploit them. The analysis of a corpus of audio description may also be of interest to audio description professionals, especially those concerned with implementing and maintaining guidelines, and with training new describers.

4 Audio Description

Audio description enhances the enjoyment of television programme for visually impaired viewers including dramas, situation comedies, soap operas, documentaries, as well as feature films. In the gaps between existing speech, audio description gives key information about characters' appearance, scenes, actions, gestures, body language, facial expressions and cinematic techniques in order to enhance the appreciation of television programmes for visually impaired people. In effect, the story told by the moving image is retold in words. In the UK the 1996 Broadcasting Act requires digital terrestrial broadcasters to provide an increasing amount of audio-described programmes (currently up to 10% of their output). Furthermore, there is considerable potential for audio descriptions to be used by the whole television audience; e.g. to 'watch' television on audiocassettes or on WAP devices with little or no visual display.

Audio description is produced by trained professionals who typically follow guidelines (be they in-house, national or international). It may take 60 hours and many viewings to produce descriptions for a 2-hour film whereas a 30 minute soap opera which is almost full of dialogue and has familiar scenes may take only 90 minutes. Audio description is scripted before it is recorded – our corpus comprises audio description scripts, i.e. written text rather than speech.

For research into multimedia computing audio description is a particularly interesting kind of collateral text because it is intended to be a substitute for the moving image with enough information so that the human audience can reconstruct the narrative told by the moving image. The amount of information it can provide is constrained by the fact that audio description must fit within existing dialogue and important sounds effects and music.

Websites

- For an overview of audio description in television as well as other entertainment and leisure activities, see the RNIB's Talking Images project at: www.rnib.org.uk/talkingimages.
- The ITC Guidelines on Standards for Audio Description can be found at: www.itc.org.uk/divisions/eng_div/subtitle/Audio_Description
- For links to global audio description information, see the Audio Description International WebPages: www.adinternational.org

5 Building and Analysing the TIWO Audio Description Corpus

Corpus is the Latin word for 'body'. In linguistic terminology, a corpus is a body of texts and, in this perspective, a selection of texts that should "represent a language or some part of the language" (Biber et al, 1998). Text corpora are built to study how language is used. They may comprise samples of general 'everyday' language, or they may focus on texts produced by specific groups of language users and their 'special' language. One way of studying a special language is to compare a corpus of texts produced by the specialists (e.g. audio description scripts) with a corpus of general language (e.g. British National Corpus). Systematic differences between the two can highlight idiosyncrasies in the special language.

5.1 Corpus Building

When building a corpus it is important to make it "as representative as possible" (Biber et al, 1998). In order to establish what categories of film we needed audio description samples for and in order to have a representative corpus, the following questions were asked during a two-hour meeting with two audio description experts – Denise Evans (RNIB) and James O'Hara (ITFC):

1. How can you categorise different kinds of film based on variations in the language you use to describe them?
2. How do you think the language you use in audio description for films does vary? For example does it vary in number / length of descriptions, choice of vocabulary, choice of grammatical constructions, etc.
3. To what extent do you think describers will vary in their language for the same film? Is there significant individual variation? Is there significant variation between different organisations?
4. Is there significant variation between English audio description in different countries?
5. Do you think that different sets of guidelines may lead to variations in the language used for audio description?

This meeting led us to sample audio description for films from 9 categories: action, children's animation, children's live-action, comedy, dark, miscellaneous, period drama, romantic and thriller. Another important outcome of the meeting was that we learnt that audio description is likely to vary within English speaking countries, notably between the UK and the US, so we must be clear that we are analysing a British English corpus.

For more information about the building of the TIWO Audio Description Corpus please see Elia Tomadaki's MPhil-PhD transfer report 'Integrating Information from Collateral Media' (2003) which is available on the TIWO website.

Reference

Biber, D., Conrad, S. and Reppen, R., 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.

5.2 TIWO Audio Description Corpus

At the time of writing the corpus comprises 420,000 words of audio description scripts from three prominent producers of audio description in the UK: ITFC, Royal National Institute of the Blind (RNIB) and the BBC Audio Description Unit. The majority of the corpus is audio description of feature films – 56 films in the 9 categories, for details please see Appendix 1. The remainder of the corpus consists of audio description for television programmes including documentaries, drama series and soap operas, for details please see Appendix 2. To complement the audio description corpus, we have also gathered a 15,500 word corpus of plot summaries for 114 films and a 1,930,000 word corpus of screenplays for 71 films.

5.3 Corpus Analysis

Standard methods of corpus-based linguistics have been applied in order to characterise the language used by audio describers at lexical, morphological, syntactic and semantic levels (Ahmad and Rogers 2001).

Reference
Ahmad K. and Rogers M., 2001, 'The Analysis of Text Corpora for the Creation of Advanced Terminology Databases.' *In: Wright, S.E. and Budin, G., The Handbook of Terminology Management.* Amsterdam: John Benjamins.

5.3.1 Language Engineering Tools

The following tools were used in our analyses:

- **System Quirk** - a language engineering workbench developed by researchers at the University of Surrey, and available as freeware from: www.computing.surrey.ac.uk/SystemQ/
- **GATE –** General Architecture for Text Engineering, developed at the University of Sheffield, and available as freeware from: http://gate.ac.uk/
- **EngCG** – English Constraint Grammar tagger, advocated by a team of computational linguists in Helsinki, Finland, and available on-line from: http://www.ling.helsinki.fi/~avoutila/cg/demo/
- **WordNet –** a lexical database for the English Language, developed at Princeton University, and available on-line from: http://www.cogsci.princeton.edu/cgi-bin/webwn
- **CYC –** a common sense knowledge base, developed by Cycorp, and available from: http://www.opencyc.org/
- **ThoughtTreasure** - a commonsense knowledge base and architecture for natural language processing, developed by Signifrom, and available from: http://www.signiform.com/tt/htm/tt.htm
- **Unitex** - Unitex is a corpus processing system, based on automata-oriented technology developed at LADL (Laboratoire d'Automatique Documentaire et Linguistique). Available from: http://www-igm.univ-mlv.fr/~unitex/

5.3.2 Summary of Results to Date

Results to date have provided a variety of evidence to suggest that there is indeed a 'special language of audio description':

- Of the 100 most frequent words in the audio description corpus, 41 are open-class words. In general language corpora there are usually only 2-3.
- Most of the frequently occurring verbs refer to 'material processes', i.e. things that can be seen happening (as opposed to characters thoughts and other abstract processes).
- Many words for specifying temporal information, i.e. when things are happening, are unusually frequent in the audio description corpus.

- Words that are suggestive of the emotions that characters are experiencing are frequent in the audio description for many films, e.g. instances of characters experiencing FEAR and DISTRESS are very common.
- The most frequent open class words in the audio description corpus refer to characters (*man, men, woman*), acting in relation to space (*walk, go* etc) or moving in a certain way (*turn*) or looking at something in a certain way (*look, stare, watch*)
- We have found a few collocations of words that seem to be characteristic of the language of audio description (as well as screenplays), e.g. 'looks at' (22 times per AD film on average), 'turns to' (11 times per AD film on average), 'opens door' (5 times per AD film on average).

For a more information about our corpus analysis work please see the following –all available on the TIWO website:

Vassiliou, Salway and Pitt (2004), 'Formalising Stories: sequences of events and state changes', IEEE Conference on Multimedia and Expo, ICME 2004.

Salway and Graham (2003), 'Extracting Information about Emotions in Films', ACM Multimedia 2003, Berkeley, 2-8 November 2003.

Salway, Graham, Tomadaki and Xu (2003), 'Linking Video and Text via Representations of Narrative', AAAI Spring Symposium on Intelligent Multimedia Knowledge Management, Palo Alto, 24-26 March 2003.

Salway and Tomadaki (2002), 'Temporal Information in Collateral Texts for Indexing Moving Images', LREC 2002 Workshop on Annotation Standards for Temporal Information in Natural Language.

Elia Tomadaki's MPhil-PhD transfer report 'Integrating Information from Collateral Media' (2003).

Andrew Vassilou's MPhil-PhD transfer report 'Representing Narrative in Multimedia Systems' (2004).

Appendix 1: TIWO Audio Description Corpus for Films

Category	Film	Describer	Organisation	Text Format
Action				
	Apocalypse Now (2 parts)	Tom Whalley, Tom Lonney, John Wolskel	ITFC	**xml**
	Armageddon	Di Langford	RNIB	**On paper**
	Die Hard with a Vengeance	Peter Wickham	RNIB	**On paper**
	End of Days	William Roberts	RNIB	**On paper**
	Gone in 60 Seconds	Di Langford	RNIB	**txt**
	Indian Fighter	Lonny Evans and Aimee de Larrabeiti	ITFC	**xml**
	Spiderman (6 parts)	James O'Hara	ITFC	**xml**
	The Great Escape	Peter Wickham	RNIB	**doc**
Children - animation				
	Atlantis	Di Langford	RNIB	**txt**
	Dinosaur	Di Langford	RNIB	**txt**
	Emperor's New Groove	David Banks	RNIB	**txt**
	Lady and the Tramp	Peter Wickham	RNIB	**txt**
	Lilo and Stich (5 parts)	Lonny Evans, Adam English, Tom Whalley	ITFC	**xml**
	Monsters Inc.	Peter Wickham	RNIB	**doc**
Children - live action				
	Harry Potter	Di Langford	RNIB	**doc**
	Scooby Doo (6 parts)	Mark Levesley, Aimee de Larrabeiti	ITFC	**xml**
	Spy Kids	Clare Le May	RNIB	**doc**
Comedy				

Appendix 1: TIWO Audio Description Corpus for Films

	Robin Hood: Men in Tights	Tom Lonny and Alan	ITFC	**xml**
	See no Evil Hear no Evil	John Wolskel	ITFC	**xml**
	Some Like it Hot	Peter Wichkam	RNIB	**txt**
	The Full Monty	Veronika Hyks	RNIB	**On paper**
Dark				
	A Murder of Crows	Mark Levesley	ITFC	**xml**
	Midnight in the Garden of Good and Evil	Matthew Vickers and Clare Le May	ITFC	**xml**
	The Others	David Banks	RNIB	**doc, pdf**
	The Postman (2 parts)	Jim Apted, Adam English, Eliza Langland	ITFC	**xml**
	The Silence of the Lambs	James O'Hara	ITFC	**xml**
	The Sixth Sense (a) The Sixth Sense (b)	Peter Wickham Alice Austin	RNIB WGBH Boston	**txt** **On paper**
	Unbreakable	Peter Wickham	RNIB	**txt**
Miscellaneous				
	Iris	Di Langford	RNIB	**doc**
	Green for Danger	Jim Apted, Mark Levesley	ITFC	**xml**
	Hear my Song	Veronika Hyks	RNIB	**On paper**
	Jason and the Argonauts	Louise Fryer	ITFC	**xml**
	Jerry Maguire (2 parts)	Diana Speed	ITFC	**xml**
	Tea with Mussolini	Aileen Downey	ITFC	**xml**
	The Shipping News	Di Langford	RNIB	**doc**
	The Amazing Howard Hughes (2 parts)	Tom Whalley, Adam English	ITFC	**xml**
	One Hot Summer Night	Aimee de Larrabeiti	ITFC	**xml**
	One True Thing	John Wolskel	ITFC	**xml**

Appendix 1: TIWO Audio Description Corpus for Films

	Submarine X1	John Wolskel, Louise Fryer	ITFC	**xml**
Period drama				
	Emma	Peter Wickham	RNIB	**On paper**
	Howard's End	Di Langford	RNIB	**On paper**
	Mill on the Floss	Louise Fryer	RNIB	**On paper**
	Persuasion	Louise Fryer	RNIB	**On paper**
	Wings of the Dove	William Roberts	RNIB	**On paper**
Romantic				
	Captain Corelli's Mandolin	William Roberts	RNIB	**txt, doc**
	Chocolat	Di Langford	RNIB	**txt**
	English Patient (a) English Patient (b) (2 parts)	Di Langford Louise Fryer and Michael Baker	RNIB ITFC	**txt, doc** **xml**
	Pretty Woman	Veronika Hyks	RNIB	**On paper**
	Runaway Bride	Veronika Hyks	RNIB	**On paper**
	The Horse Whisperer	Peter Wickham	RNIB	**txt**
Thriller				
	Enigma	William Roberts	RNIB	**doc**
	Oceans 11	Aimee de Larrabeiti, Adam English	ITFC	**xml**
	Road to Perdition (8 parts)	James O'Hara	ITFC	**xml**
	The Pelican Brief (2 parts)	Lonny Evans	ITFC	**xml**

Appendix 2: TIWO Audio Description Corpus for Television Programmes

Category	Film	Describer	Organisation	Text Format
Documentaries				
	Jellikins Nature	Gloria Hatrick	ITFC	**xml**
	Journey to the Bottom of the Sea	Lonny Evans	ITFC	**xml**
	Secrets of the Ancients	Adam English	ITFC	**xml**
	The Beast must Die	John Wolskel	ITFC	**xml**
Drama				
	Canterbury Tales: A Knight's Tale		BBC	**htm**
	Miller's Tale		BBC	**php**
	Wife of Bath		BBC	**htm**
Recipes				
	Taste of Barbados	Tom Whaley	ITFC	**xml**
Series				
	Ally Mc Bill	Tom Whaley	ITFC	**xml**
	Anna Karenina	Lonny Evans	ITFC	**xml**
	Friends	Jim Apted	ITFC	**xml**
	Jonathan Creek	Mark Levesley	ITFC	**xml**
Soap operas				
	Eastenders		BBC	**txt**