MCQ-Creation Methodologies Workshop

http://www.testcentres.co.uk/MCQ-Creation

(7 November 2011, London, United Kingdom)

… in conjunction with LONDON INTERNATIONAL CONFERENCE on EDUCATION 2011

"At MCQ-Creation we discuss empirical studies of MCQ creation methodologies and then suggest improvements."

Table of Contents

Preface

Welcome to the proceedings from the first MCQ-Creation event!

By hosting MCQ-Creation within the LIC Education conference, the committee aimed to bring together educationalists from industry, governmental examining bodies, universities and schools to examine the merits and pitfalls in traditional processes for creating Multiple Choice Question (MCQ) test items. The output from the workshop was to be a variety of proposals for new (or adapted) MCQ creation methodologies appropriate to the domains defined by the presenters.

Presenters were to give an overview of their domain of discourse (ie the context of their assessments) and a description of how they evaluated an established formal method for MCQ test item creation or defined and evaluated a demonstrably **NEW** MCQ test item creation methodology within their domain of discourse.

Six abstracts were received. Each submission was reviewed by the Workshop Committee, who have experience in both academia and industry. Three invitations were made on the basis of these decisions but only one presenter was available to present a paper at the workshop (the other invited presenters indicated a willingness to attend next year's event).

These online proceedings (*http://www.testcentres.co.uk/MCQ-Creation/Proceedings2011.pdf*) include the program for the event, the abstracts that were accepted, and the paper and poster that were provided by the presenter at the workshop.

MCQ-Creation Committee

Programme

00:00	Welcome and introductions
00:02	The Aims of the MCQ-Creation workshop
00:04	Repeat Call for FULL papers (deadline 31/01/12)
00:05	EXERCISE 1: Your current MCQ-Creation skills
00:10	TALK 1: Defining domains
00:15	EXERCISE 2: Definition of a domain for which you wish to create MCQs
00:20	TALK 2: Defining Objectives
00:25	EXERCISE 3: Prepare a specific objective for your chosen domain
00:30	TALK 3: Thomas M. Haladyna's MCQ item creation taxonomy
00:40	EXERCISE 4: What problems do you anticipate or have you encountered with Thomas Haladyna Taxonomy?
00:45	CASE STUDY 1: Preparing effective questions for use during an audit of staff who are to become authorised in the use of a Cable Spiking Gun
01:00	CASE STUDY 2: A study of MCQs for refreshing safety knowledge of operators of Heavy Lifting Plant
01:15	CASE STUDY 3: The Construed Antonym Realization Exercise (CARE) generation methodology.

Abstracts

Paper ID: 003

Title: Preparing effective questions for use during an audit of staff who are to become authorised in the use of a Cable Spiking Gun

Abstract: The question creation process described in this paper is used regularly by the author who is an experienced auditor of safety procedures for a UK Electricity Distribution company. The specific context of the quoted example is a safety audit in which the auditor aims to ensure that correct Standard Techniques are applied when using an Electricity Cable Spiking Gun. Evaluation of the quality of the created question items is achieved through a combination of reflective statements by the auditor, feedback from candidates and a demonstration of the limitations of all reasonable alternative forms of the question.

Paper ID: 004

Title: A study of MCQs for refreshing safety knowledge of operators of Heavy Lifting Plant

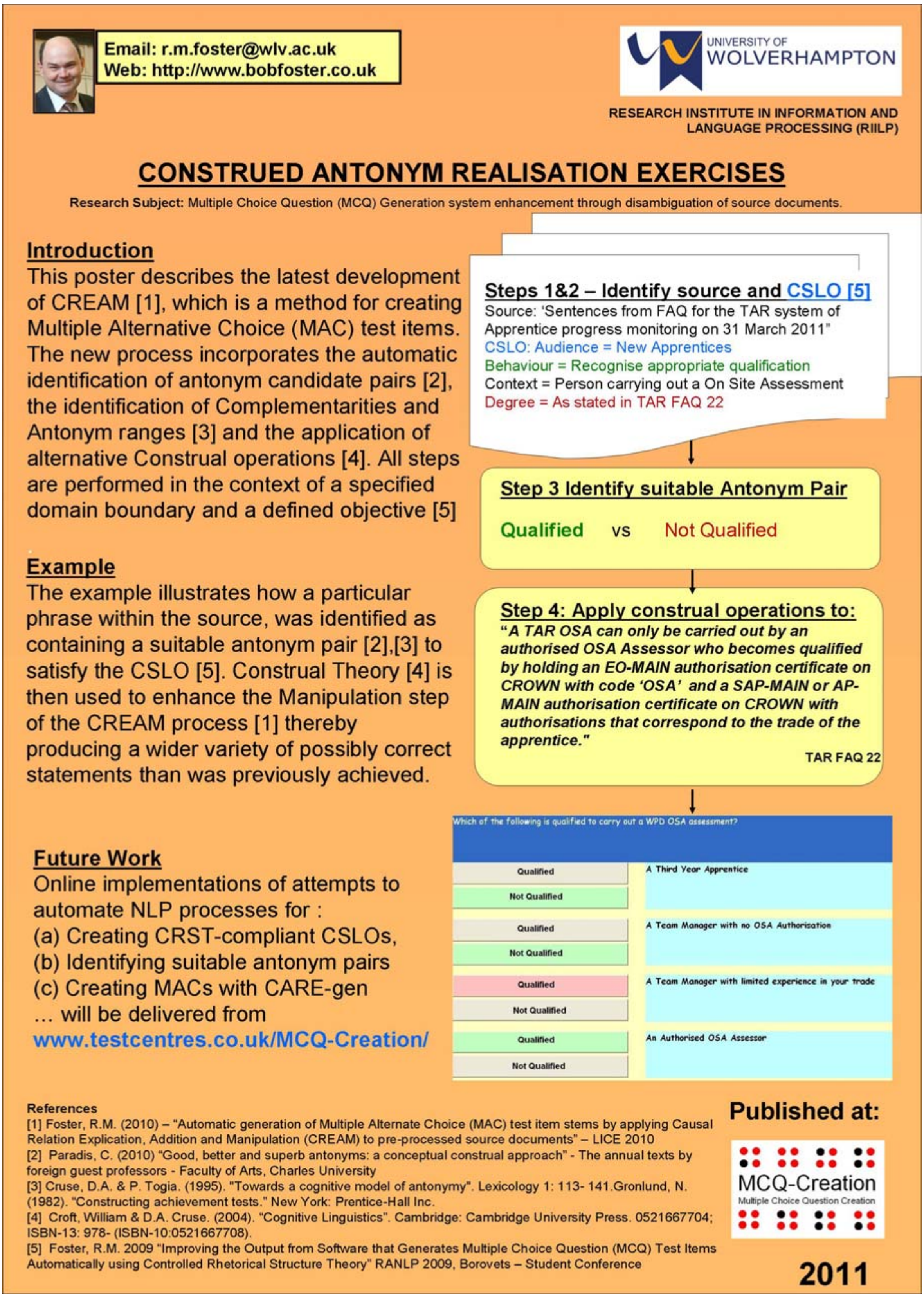
Abstract: This study examines the processes used to create the Multiple Choice Questions (MCQ)s that are used to confirm successful retention of safety messages by staff who operate heavy duty plant for a UK company. The scope of the study is the routines that refresh the knowledge of staff who operate lorry loaders, fork lift trucks, mobile elevated working platforms and gantry cranes. The process for creating questions is described and evaluated using analysis of response data. The outcome is a list of suggested changes to the content of the MCQs, and improvements to the MCQ creation process.

Paper ID: 005

Title: The Construed Antonym Realization Exercise (CARE) generation methodology

Abstract: We explore the process for the creation and maintenance of a MCQ question bank. MCQs from this question bank are used pre-test delegates who attend High Voltage systems operations training courses. Other questions from the same bank are also used to follow up and refresh the knowledge of the delegates after they have returned to their daily work. The scope of the study is the routines for High Voltage Switching operations and the issues surrounding the issue and receipt of safety documents, with a particular emphasis upon Permit to Work. Also included are MCQs that test knowledge of UK legislation and the distribution safety rules. The study includes analysis of response data which results in suggestions for changes to both the format and content of the MCQs.

Poster



Email: r.m.foster@wlv.ac.uk
Web: http://www.bobfoster.co.uk

UNIVERSITY OF WOLVERHAMPTON

RESEARCH INSTITUTE IN INFORMATION AND LANGUAGE PROCESSING (RIILP)

CONSTRUED ANTONYM REALISATION EXERCISES

Research Subject: Multiple Choice Question (MCQ) Generation system enhancement through disambiguation of source documents.

Introduction

This poster describes the latest development of CREAM [1], which is a method for creating Multiple Alternative Choice (MAC) test items. The new process incorporates the automatic identification of antonym candidate pairs [2], the identification of Complementarities and Antonym ranges [3] and the application of alternative Construal operations [4]. All steps are performed in the context of a specified domain boundary and a defined objective [5]

Example

The example illustrates how a particular phrase within the source, was identified as containing a suitable antonym pair [2],[3] to satisfy the CSLO [5]. Construal Theory [4] is then used to enhance the Manipulation step of the CREAM process [1] thereby producing a wider variety of possibly correct statements than was previously achieved.

Steps 1&2 – Identify source and CSLO [5]

Source: 'Sentences from FAQ for the TAR system of Apprentice progress monitoring on 31 March 2011"
CSLO: Audience = New Apprentices
Behaviour = Recognise appropriate qualification
Context = Person carrying out a On Site Assessment
Degree = As stated in TAR FAQ 22

Step 3 Identify suitable Antonym Pair

Qualified vs Not Qualified

Step 4: Apply construal operations to:

"A TAR OSA can only be carried out by an authorised OSA Assessor who becomes qualified by holding an EO-MAIN authorisation certificate on CROWN with code 'OSA' and a SAP-MAIN or AP-MAIN authorisation certificate on CROWN with authorisations that correspond to the trade of the apprentice."

TAR FAQ 22

Which of the following is qualified to carry out a WPD OSA assessment?

Qualified Not Qualified	A Third Year Apprentice
Qualified Not Qualified	A Team Manager with no OSA Authorisation
Qualified Not Qualified	A Team Manager with limited experience in your trade
Qualified Not Qualified	An Authorised OSA Assessor

Future Work

Online implementations of attempts to automate NLP processes for :
(a) Creating CRST-compliant CSLOs,
(b) Identifying suitable antonym pairs
(c) Creating MACs with CARE-gen
… will be delivered from
www.testcentres.co.uk/MCQ-Creation/

References

[1] Foster, R.M. (2010) – "Automatic generation of Multiple Alternate Choice (MAC) test item stems by applying Causal Relation Explication, Addition and Manipulation (CREAM) to pre-processed source documents" – LICE 2010

[2] Paradis, C. (2010) "Good, better and superb antonyms: a conceptual construal approach" - The annual texts by foreign guest professors - Faculty of Arts, Charles University

[3] Cruse, D.A. & P. Togia. (1995). "Towards a cognitive model of antonymy". Lexicology 1: 113- 141.Gronlund, N. (1982). "Constructing achievement tests." New York: Prentice-Hall Inc.

[4] Croft, William & D.A. Cruse. (2004). "Cognitive Linguistics". Cambridge: Cambridge University Press. 0521667704; ISBN-13: 978- (ISBN-10:0521667708).

[5] Foster, R.M. 2009 "Improving the Output from Software that Generates Multiple Choice Question (MCQ) Test Items Automatically using Controlled Rhetorical Structure Theory" RANLP 2009, Borovets – Student Conference

Published at:

MCQ-Creation
Multiple Choice Question Creation

2011

Paper

The Construed Antonym Realisation Exercise (CARE) generation methodology

Author: Robert M Foster

Affiliations:

Research Institute for International Language Processing – Wolverhampton University

And

Western Power Distribution (plc)

Published: 20 March 2012

Abstract

This paper presents and tests a new process for automatically creating Multiple Alternative Choice (MAC) test items. Research has shown that summative assessment tests, comprising MAC-formatted Multiple Choice Question (MCQ) test items, improve the effectiveness of induction training through more precise identification of learner knowledge gaps.

The process begins by pre-processing the source documents in the context of a Controlled Specific Learning Objective (CSLO) within the domain defined by a collection of source documents. This is done by explicating, and where necessary adding, causal coherence relations to sentences within the source document that are relevant to the CSLO. Statistical Natural Language Processing (NLP) techniques are then used to identify pairs of lexical items that co-locate within complementary and antonymic syntactic patterns. Two lists of statements are then generated by applying each construal process from a recognised system for categorising construal operations, to the contexts in which the word pairs were identified. One of the lists illustrates ‘‘correct’ construal. of these word pairs in relation to the CSLO and the second list of statements illustrates erroneous’ construal A MAC template is then used to generate test items from the word pairs and the two lists of statements representing different instances of available construal operations.

The process is tested using a domain specific evaluation method. A. domain expert selects one ‘AC item set’ from each of 32 pairs of ‘AC item sets’ (a ‘AC item set’ consists of a stem question and two alternative responses). One AC item set from each pair (the ‘CARE’) was generated using the proposed process while the other was created using traditional manual methods. Combinations of the selected AC item sets from each pair were combined to form Multiple Alternative Choice (MAC) items for use in the test routine. The results table for this experiment indicates how many of the CARE AC item sets were used without applying any changes, how many were used following changes and how many were rejected in favour of the manually created AC item set. The results provide sufficiently positive evidence to support the application of future experiments described in the last section.

1) Introduction

"What if a software program could generate Multiple Choice Questions directly from my employer's policy documents of rules and procedures?"

WPD Computer Based Training Manager 2006

In his description of 'the great leap forward' in his popular science book 'The Ancestors Tale' (Dawkins 2004), Richard Dawkins provides a clear illustration of the importance of questions to the development both of individual human minds and of the collective human consciousness. The question he identified as so important is 'what if...?' as an expression of an exploratory imagination. The current study was motivated by the above question.

The WPD Computer Based Training Manager is responsible for producing and then maintaining a MCQ item bank from which test routines are compiled for a variety of purposes in accordance with Western Power Distribution Policy. This paper describes the context and the latest achievements of a study that began following the granting of permission by WPD, to use documents from their policy library as source documents for some experiments. The requirements from the study are as follows:

> *“WPD seek improvements to their current systems for the creation, delivery and maintenance of Multiple Choice Question (MCQ) test items that will*
>
> *(a) Provide evidence to all interested parties that WPD have met (or exceeded) their responsibilities under the Health and Safety at Work, etc Act 1974 for all relevant fields of knowledge.*
>
> *(b) Reduce likelihood of disputes over validity by drawing content directly from a trace-able source text for an identified target population covering a clearly defined field of knowledge.*
>
> *(c) Manage Change by avoiding expensive revisions of manually created MCQ test item banks following changes in fields of knowledge (eg changes to national legislation)*
>
> *(d) Increase efficiency of MCQ test item creation by saving the time of our item designers, validators and users (trainers and trainees),”*

Before examining the processes involved in generating MCQ test items, some important discoveries were made during a review of theories about learning theories and about MCQ format which support current manual methods for creating MCQ test items.

1.1 A Review of Learning Theories

A WPD internal review of relevant Learning theories CRI (Mager 1975), ACT-R (Anderson 1990) and the Taxonomy of Cognitive Learning (Bloom 1958) concluded that MCQ test items are appropriate for formative and summative assessment of knowledge in preparation for WPD skills assessments. However, a review of MCQ test item formats (Haladyna 2002) suggested a change from traditional 4-option Multi Choice format to the Multiple Alternative Choice (MAC) format. Experiments showed (Foster 2010d) that knowledge assessment tests comprising MACs improve the effectiveness of induction training into the featured domain, through more precise identification of knowledge gaps.

The two component parts of a MAC are the question stems and the Alternative Choice pairs that provide the candidate with two options for each question stem. The WPD CBT Manager decided to explore existing methods for automatically generating these component parts of MAC test items, thereby increasing the availability of domain familiarising exercises for domain newcomers without expensive input from human item designers and testers. If no suitable method could be identified then the CBT Manager would devise a new system..

1.2 A Review of Existing Systems for generating MCQ test items

Many of the available systems for generating MCQ test items focus upon language familiarisation or vocabulary assessment. For example, Brown's study (Brown et al. 2005) focuses upon testing vocabulary knowledge. Although there might be a place in the WPD MCQ test item banks for a few items that test vocabulary within our corporate sub-language,, the WPD requirement is for a more flexible system that can produce test items addressing a range of learning objectives.

Systems requiring the construction (or pre-existence) of some form of knowledge base, eg (Tsumori and Kaijiri 2007) were discounted since the construction costs for such a system would be very hard to justify for our domain in the current economic climate. The maintenance burden of a knowledge base is also likely to present a barrier.

The most promising approach seemed to be a MCQ test item generator (Mitkov and Ha 2003, 2006) that generates MCQ test items directly from source documents. Mitkov and Ha report that generating MCQ items using the generator can speed up the process by 4 times without compromising the quality of the output. However preliminary experiments applying the system to a document from the WPD policy library delivered no usable MCQ test items. Changes to the internal methodology of the system and techniques for pre-processing the source documents are investigated in this paper.

1.3 Investigations into available theories for Source Document pre-processing

In 2010, the CREAM technique for generating MAC stems (Foster 2010e) was proposed. The CARE creation process is a development of CREAM, which introduces two theoretical frameworks from Cognitive Linguistics into the process..

Initial research into the field of generating the required Alternative Choice component of MAC items identified several possible systems for categorising opposites (Paradis 2010, Vogel 2003, Cruse and Togia 1995). The CARE creation process uses the categories of: Complementarities and Antonym ranges (Cruse and Togia 1995).

The process also includes a step for identifying instances of erroneous construal of the identified opposites that might arise in the minds of newcomers to the domain. The CARE creation process uses the construal operation categorisation system proposed in the book 'Cognitive Linguistics' (Croft and Cruse 2004) in preference to Imaging systems (Talmy 2000) and Focal adjustments (Langaker 1987). Using the Cognitive Linguistics categorisation system, each construal operation applied to the sentences from the WPD policy library can be labelled as either 'correct' or 'incorrect' in the context of the Controlled Specific Learning Objective (Foster 2009a) that defines the purpose of the test routine for which the MCQ test items are to be generated.

1.4 Overview

Section 2 of this article gives an overview of the Context for the experiments that seek to identify appropriate enhancements for the process for generating MAC test items. Section 3 summarises the relevant theories from the literature: Section 4 illustrates how the CARE creation process combines these theories to generate a domain specific MAC formatted MCQ test item using sentences from the source documents. Section 5 describes an experiment in which the technique was tested as part of on-going efforts to improve MAC creation processes within WPD.. Section 6 provides a statement of the conclusions and describes plans for future work..

2) - Context

2.1 Multiple Choice Questions chosen to facilitate Health and Safety Training

Western Power Distribution (WPD) is a UK company who provide a service to residents of the United Kingdom in the South West and South Wales regions. Their primary function is the distribution of electricity to homes and businesses in those regions. WPD undertakes extensive policy development, training & assessment activities in order to ensure a professional approach to the delivery of these services so as to meet (or exceed) their responsibilities under UK legislation. In recent years, the preferred method for conducting both formative and summative knowledge check assessments at WPD has been moved away from classroom quizzes and towards Computer Based Tests (CBTs).

‘ST:OS7D – Relating to Audits of Operational field staff’ states that

“3.1 All Senior Authorised and Authorised Persons who hold an authorisation for HV Operational Work (11SW, 33SW, 66SW, 132SW and restricted variations) shall complete an annual CBT test to the satisfaction of an Examining Officer qualified to examine for that authorisation.”

ST:OS7D is a ‘Standard Technique’ from the WPD Policy library

Support for this change is provided by experiments (Foster 2010a, 2010d) showing that MCQs can deliver more comprehensive feedback within formative assessments and more targeted identification of knowledge gaps during summative assessments, than is possible using more traditional assessment methods. In his book “A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment” (Haladyna 2002) has also highlighted the need for research to improve assessment of cognitive learning.

> *"Item analysis has been a stagnant field in the past, limited to the estimation of Item difficulty and discrimination using CTT or Item Response Analysis, and the counting of responses to each distractor. .... The future of item response validation will never be realized without significant progress in developing a workable theory of item writing."*

Thomas Haladyna 2002

Haladyna calls for the educational academic community to provide better methodologies for creating assessment test items that are suitable for use in the education domain. The most extensive response has been the CRESST REPORT 778 (Behrens, J.T. et al 2010), however this work has a limited relevance to the current study since they focus upon the assessment of school-children, and upon higher levels of cognitive learning, as measured by Bloom's taxonomy of cognitive learning - see section 3.

2.2 Initial training for new staff at Western Power Distribution

WPD take on between 30 and 40 new Apprentices each year and in recent years have taken on and trained an additional 200 staff in order to provide the capacity necessary to satisfy the requirements of recent legislation (Electricity Supply Quality and Continuity Regulations 2002).

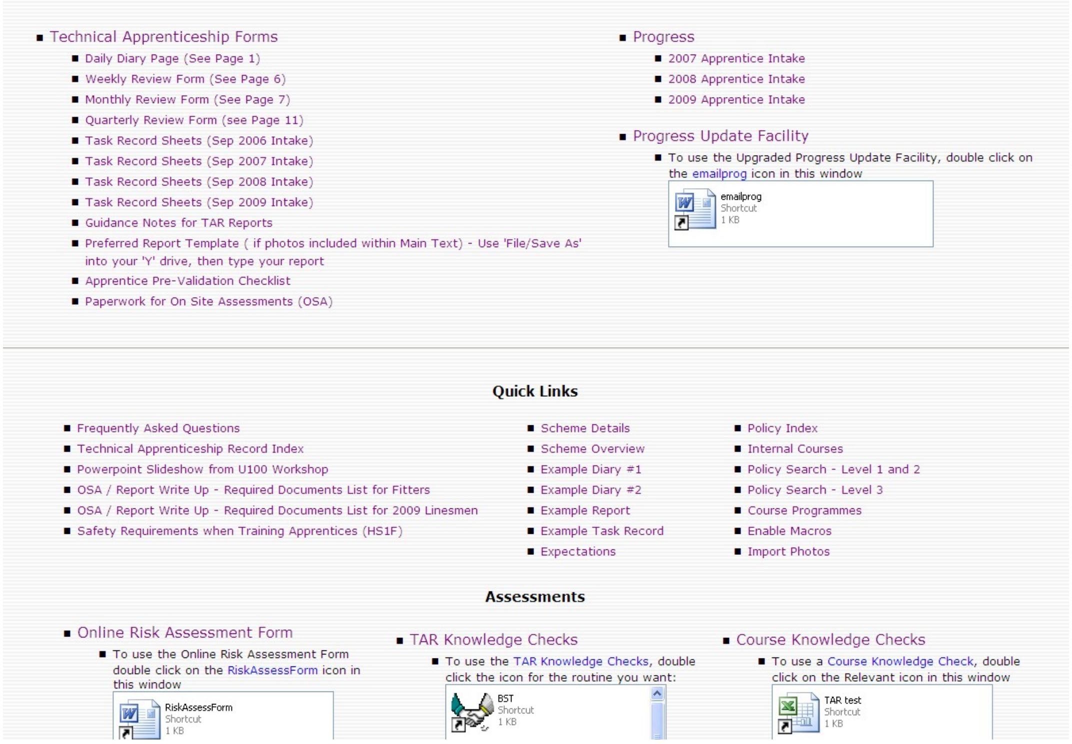
UK legislation requires WPD to provide training for new staff:

> *“13. Capabilities and training:*
>
> *(2) Every employer shall ensure that his employees are provided with adequate health and safety training*
>
> *(a) on their being recruited into the employer's undertaking; ….”*

Management of Health and Safety at Work Regulations 1989

The recent increase in the number of new employees to the company led to a review and development of the specification of how this initial training is conducted and assessed. This review highlighted the need for a formal system of progress monitoring and assessment on completion of the course of training.

A new system was created to address this requirement. The system was named the WPD Technical Achievement Record (TAR) system and has been developing over the past six years. The TAR support systems are delivered over the WPD LAN from an intranet website and many of the rules for operating the system have been presented in the form of a FAQ (Frequently Asked Questions) document. The front menu webpage for the TAR system is shown below.



- Technical Apprenticeship Forms
  - Daily Diary Page (See Page 1)
  - Weekly Review Form (See Page 6)
  - Monthly Review Form (See Page 7)
  - Quarterly Review Form (see Page 11)
  - Task Record Sheets (Sep 2006 Intake)
  - Task Record Sheets (Sep 2007 Intake)
  - Task Record Sheets (Sep 2008 Intake)
  - Task Record Sheets (Sep 2009 Intake)
  - Guidance Notes for TAR Reports
  - Preferred Report Template (if photos included within Main Text) - Use 'File/Save As' into your 'Y' drive, then type your report
  - Apprentice Pre-Validation Checklist
  - Paperwork for On Site Assessments (OSA)
- Progress
  - 2007 Apprentice Intake
  - 2008 Apprentice Intake
  - 2009 Apprentice Intake
- Progress Update Facility
  - To use the Upgraded Progress Update Facility, double click on the emailprog icon in this window

Quick Links

- Frequently Asked Questions
- Technical Apprenticeship Record Index
- Powerpoint Slideshow from U100 Workshop
- OSA / Report Write Up - Required Documents List for Fitters
- OSA / Report Write Up - Required Documents List for 2009 Linesmen
- Safety Requirements when Training Apprentices (HS1F)

- Scheme Details
- Scheme Overview
- Example Diary #1
- Example Diary #2
- Example Report
- Example Task Record
- Expectations

- Policy Index
- Internal Courses
- Policy Search - Level 1 and 2
- Policy Search - Level 3
- Course Programmes
- Enable Macros
- Import Photos

Assessments

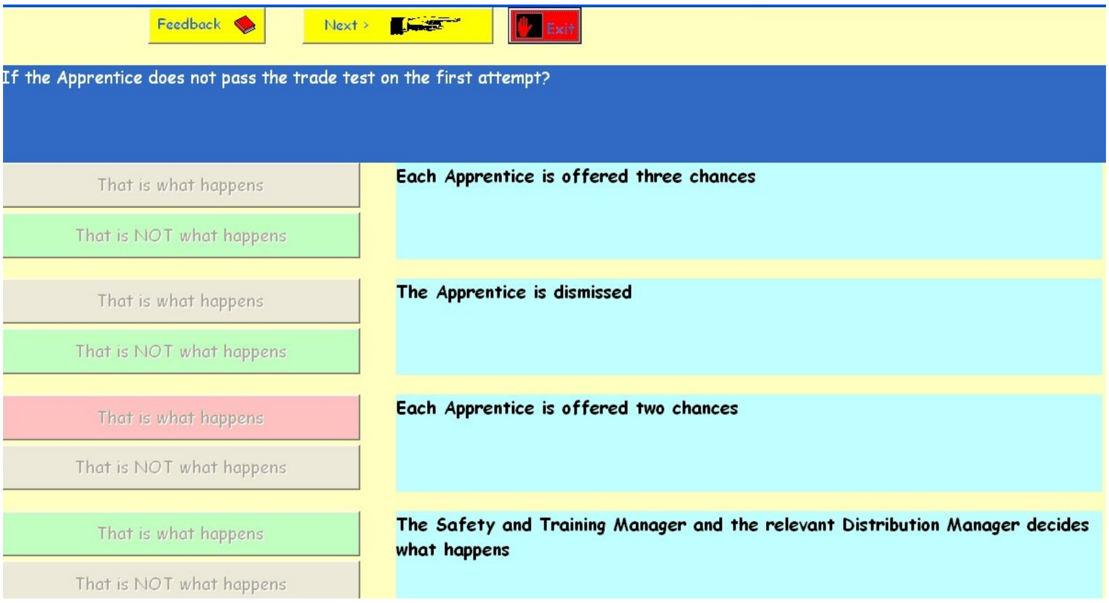
- Online Risk Assessment Form
  - To use the Online Risk Assessment Form double click on the RiskAssessForm icon in this window

- TAR Knowledge Checks
  - To use the TAR Knowledge Checks, double click the icon for the routine you want:

- Course Knowledge Checks
  - To use a Course Knowledge Check, double click on the Relevant icon in this window

Figure 1 – Front Screen for the WPD Technical Achievement Record (TAR) system

The implementation of this system includes the use of a series of formative quizzes’ that are used during the training of newcomers to the company as they familiarise themselves with the way the TAR system operates. These quizzes use MAC-formatted MCQ test items in an interactive mode. After receiving training and formative assessments, apprentices take a summative assessment of their knowledge, which uses a different presentation of the MAC-formatted test items. Some examples of these items are provided below. It is these formative and summative assessment MAC items that are used in the illustration of the MAC generation technique proposed in this paper.



If the Apprentice does not pass the trade test on the first attempt?

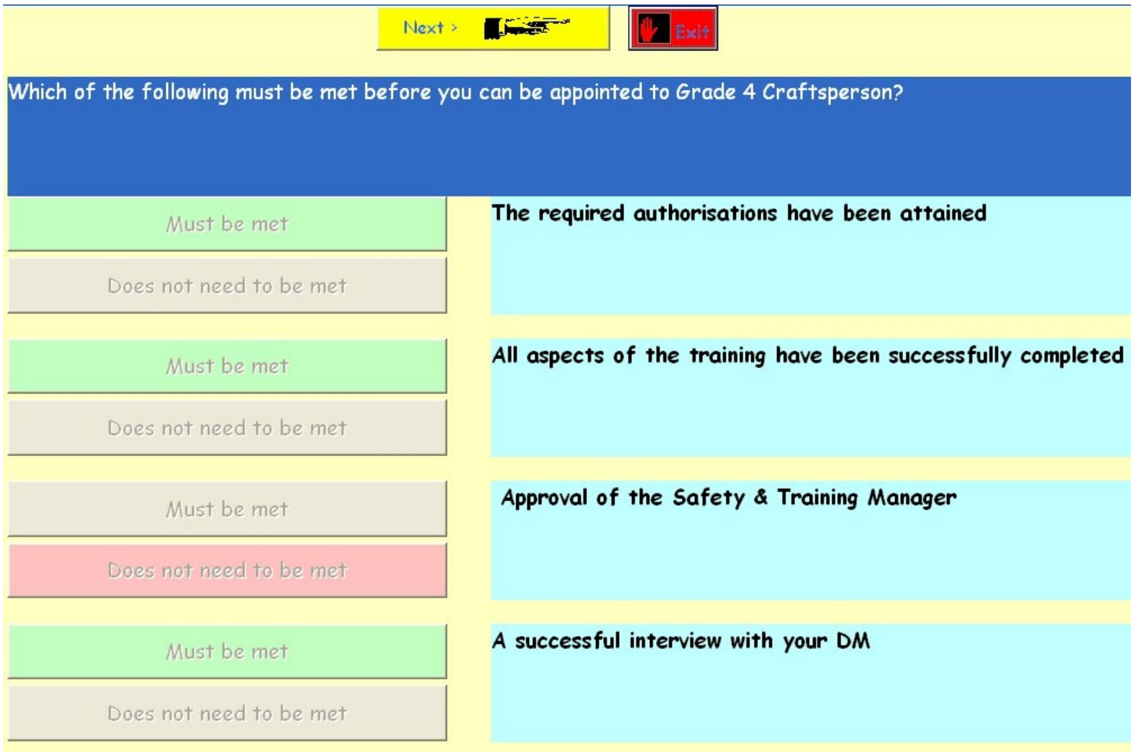
Each Apprentice is offered three chances

The Apprentice is dismissed

Each Apprentice is offered two chances

The Safety and Training Manager and the relevant Distribution Manager decides what happens

Figure 2 – MAC testing knowledge of events following a failure (referral?) at the first attempt at the trade specific 'trade test'



Which of the following must be met before you can be appointed to Grade 4 Craftsperson?

The required authorisations have been attained

All aspects of the training have been successfully completed

Approval of the Safety & Training Manager

A successful interview with your DM

Figure 3 – MAC testing knowledge of expectations of Apprentice performance

3.0 Related Academic Work

This section begins by describing the learning theories that are significant for this study. There is a particular focus upon the Cognitive Domain within the Bloom Taxonomy of Learning. This is followed in section 3.2 by a summary presentation of the theory of Cognitive Primitives upon which the CREAM method for generation MAC-formatted MCQ test items is based. The chosen categorisation system for construal operations from the book 'Cognitive Linguistics' (Croft and Cruse 2004) is then summarised in section 3.3, Section 3.4 provides a description of the selected theory about opposites (Cruse and Togia 1995) which uses the concepts of Complementarities and Antonyms. Section 3.5 describes the intended meaning of the term 'Realisation' in the title of this document..

3.1 Learning Theories - Measuring progress in Human Cognitive Learning

The introduction made a brief reference to 'the great leap forward' (Dawkins 2004). Using this 18 thousand year old example from our human evolutionary history, Dawkins shows how the miracle of Human Cognitive Learning (HCL) resulted in an astonishing increase in the rate of progress made by human beings when people began to ask a significantly different question. In the case of the 'the great leap forward' for humanity the question was the expression of an exploratory imagination 'what if...?'. However leaps of this kind are made by billions of individuals every day in homes, schools and businesses throughout the world as the miracle of HCL gives them survival and social benefits. Each leap begins when an individual asks a question they have never asked before and then accepts an answer as their 'truth'.

A history of more recent theories about learning which are significant to this study is provided on the Wikipedia page for 'Designing Learning Programs' The theory which has most immediate relevance to the industrial training featured in this study is the theory of Criterion Referenced Instruction (Mager 1975). This theory provides training designers with a method for measuring an individual's progress in their learning in relation to specified criteria. Again the asking of a question is the start of the process, but instead of focusing upon the INTERNAL (formative) questions, Mager suggests the use of EXTERNAL (summative) questions. Mager proposes that analysis of the responses provided by learners to these summative assessment questions can determine whether or not their learning has progressed to a required standard. Criterion Referenced Instruction enables us to make precise measurements of HCL progress in relation to specific learning objectives by analysing learner responses to summative assessment questions.

Organisations have used summative assessments for many years to confirm that an individual has achieved the required prior learning before attending more advanced training courses or before they are allowed to take up more responsible posts within the organisation's reporting structure. UK Organisations are also increasingly required by law (eg MHSWR 1989) to demonstrate that the staff they have appointed to certain responsible positions, possess a minimum level of knowledge. Organisations therefore require accurate methods for identifying gaps in the knowledge of their staff. They also must identify gaps in staff knowledge that might arise due to the passage of time or when significant changes have been made to their job descriptions. When such gaps in staff members' knowledge are significant, the law requires them to be filled before staff can be (re-)authorised to carry out certain tasks or to fulfil certain duties and responsibilities within the organisation's reporting structure.

Management of Health and Safety at Work Regulations 1989 – Section 13

> *13. Capabilities and training:*
>
> *(2) Every employer shall ensure that his employees*
>
> *are provided with adequate health and safety training*
>
> *(a) on their being recruited into the employer's undertaking; and*
>
> *(b) on their being exposed to new or increased risks because of—*
>
> *(i) their being transferred or given a change of responsibilities within the employer's undertaking,*
>
> *(ii) the introduction of new work equipment into or a change respecting work*
>
> *(3) The training referred to in paragraph (2) shall—*
>
> *(a) be repeated periodically where appropriate;*
>
> *(b) be adapted to take account of any new or changed risks to the health and safety of the employees concerned; and*
>
> *(c) take place during working hours.*

Management of Health and Safety at Work Regulations 1989

In the cognitive domain of his Learning Taxonomy (Bloom 1958), Bloom identified a progressive series of cognitive learning categories. The theory states that satisfactory cognitive performance at the higher levels in the structure (analysis, creativity and judgment) is only possible when the lower levels (knowledge, understanding and skills) have been mastered. This theory of human learning whereby a foundation of Fact knowledge is required before other forms of learning can proceed is also supported by John Anderson's ACT-R theory. There have been challenges levelled at Anderson’s theory, based upon apparently uncertain distinctions between declarative memory and procedural memory. However, these seem once again to be focused upon concerns in the education domain and do not arise in the industrial safety training domain. Whether the learner’s internal representation of the fact or procedure is stored as a procedural fact or a declarative fact is not important. The important point is that they can recall the fact or procedure quickly and accurately when the need arises.

In light of this survey of well established learning theories, the cognitive taxonomy learning level of question/answer pairs to be generated, whether for formative or summative assessment purposes, has been deliberately restricted to testing fact and procedure knowledge at Bloom’s Taxonomy Cognitive learning level 1 only.

3.2 Causal coherence relation primitives

A useful method for causal coherence relations analysis is proposed in the literature which requires the assumption that all relations are cognitively basic (Sanders et al. 1993). The proposal is that only four cognitive primitives are required to express the primitive causal coherence relations necessary for communication. Combination of these four primitives by a writer can then present increasingly sophisticated types of causal coherence relation between a text's information units. The primitives are described in detail in the literature (Sanders et al. 1993), but can be summarized as follows:

1) Basic operation (causal vs additive)

2) Source of coherence. (semantic vs pragmatic)

3) Order of information units (basic vs complex).

4) Polarity (positive vs negative)

The utility of this theory in the context of MAC stem creation is illustrated in section 4 whereby a standard can be applied to policy documents insisting that source texts clearly define causal coherence relation clauses.

3.3 A system for categorising Construal Operations

William Croft explains in chapter 3 of 'Cognitive Linguistics (Croft and Cruse 2004) that the construal operation classification system proposed in their book combines the observations of previous systems from the Linguistic Semantics and Cognitive Psychology communities.

> *"*A basic premise of Cognitive Linguistics is that *'Language is an instance of general conceptual abilities'.* The classification of construal operations in table 3.1 is not intended to be a reduction of construal operations in just four processes. The various construal operations listed under the four headings are all distinct cognitive processes. The analysis we propose is that the various construal operations are manifestations of the four basic cognitive abilities in different aspects of experience. "
>
> 'Cognitive Linguistics' - Cruse and Croft 2004

The two other systems described in the literature for categorising construal operations which were considered for use within CARE are: Imaging systems (Talmy 2000) and Focal adjustments (Langaker 1987) . However, as William Croft explains there are several important construal operations which are inadequately represented in these systems. For example the fundamental considerations of Framing (Fillmore 1975, 1977a) are missing and the more complex cognitive processes of Metaphor (Lakoff and Johnson 1980) and Image schemas (Lakoff 1987, Johnson 1987) are also inadequately covered for our purpose.

This paper embodies a domain specific response to the following questions, which are posed in the conclusion (section 3.6) to Chapter 3 within 'Cognitive Linguistics' (Croft and Cruse 2004)

1) How do construal operations interact within this domain?

2) How should we characterise the processes of Language vs Thought vs Experience?

3.4 The search for domain specific Complementarities

Research into possible methods for generating the required Alternative Choice component of MAC items identified several contradictory studies (Paradis 2010, Vogel 2003, Cruse and Togia 1995). If we take a dynamic construal approach to sense relations then we can adopt the general categories of Complementarities and Antonyms (Cruse and Togia 1995) to describe the different forms of semantic opposites in a domain. However, successful communications using these categorisations require both readers and writers of domain defining documents to share a common ground construal of the intended meanings for each set of opposites.

In many other application domains, this might present problems (eg general communications between governments and the wide variety of learning preferences and cognitive development stages of their citizens). However, the principal objective of the training process is the achievement of a common ground construal of certain key facts and concepts between writers and readers of the company's policy documents. Therefore this requirement for a shared common ground is an acceptable, and perhaps even a desirable, feature of this categorisation system.

> *"The notion of the construal of oppositeness is the negation of the original aspect."*
>
> (Cognitive Linguistics - Croft and Cruse 2004 :p164 examples 1 to 5).

In the featured domain, when tackling an unfamiliar topic, the cognitive response by both learner and trainer involves identifying good examples. Each example is judged according to how 'pure' and 'symmetrical' the opposition is between the two extremes. The judgement involves identifying properties that are either present or absent and identifying features of the construal that are not relevant to the opposition. When the reader begins to imagine the effects of more or less of the property upon the degree of oppositeness, then under Cruse and Togia's system, the relationship changes from relatively straightforward complementarity, into a significantly more complex, Antonymic relationship which is comparatively difficult to learn,. Similarly, when readers pre-suppose the presence of one property over it's opposite, then the complementarity's purity decays towards more complex networks of antonymic ranges, scales and oppositions.

Under Cruse and Togia's system, Complementarities must also be construed as both mutually exclusive and mutually exhausting. Therefore if a reader begins to imagine a third state which is not included in one or other of the complementarities then the relationship must again be re-categorised as an antonym and will therefore demand considerably greater effort from both teacher and learner before a secure assessment result can be obtained.

The chosen method for categorising Opposites (Cruse and Togia 1995) gives detailed descriptions of antonym pairs and readers of this paper are advised to read chapter 7 of the book 'Cognitive Linguistics (Croft and Cruse 2004) if a deeper insight into this topic is required.

3.5 The intended meaning of Realisation

The description of the CARE creation process in this paper includes a step for extracting word pair opposites from source documents followed by a description of how extracted word pairs can be ‘Realised’ to produce effective MAC-formatted test items. This section presents the intended definition for the word ‘Realisation’ as used in this paper:

The closest offering from Princeton’s WordNet is sense 5:

> *“5. making real or giving the appearance of reality [syn: realization]”*

WordNet® 3.0, © 2006 by Princeton University.

A better, although still incomplete, presentation of the meaning intended in this paper is provided in sense [6b] of the Collins English Dictionary

> *“6b. to reconstruct (a composition) from an incomplete set of parts “*

(Collins English Dictionary 2009)

The precise meaning can be described as follows:

> *“In the CARE creation process, each ‘AC item set’ is generated by re-uniting one of a pair of word pair opposites with a construal aware re-write of its original context sentence, and then offering test takers the other word from the word pair as an alternative. This has been labelled as ‘realisation’ in the sense that the word pair has been put into a ‘real’ context... it has been ‘realised’.”*

Readers might also choose to construe another meaning in this context. A person who selects the ‘incorrect’ response to a AC item set will be given an ‘incorrect’ indication and will perhaps then ‘realise’ that the other response was the correct response. This is a fortuitous alternative meaning, but was not the original intention of the authors of this paper.

The use of the term ‘Antonym’ as opposed to ‘Complimentary’ in the name of the method also requires explanation. The stated aim of the CARE creation process is to demonstrate that learning has taken place at level 1 of the cognitive domain of Bloom’s taxonomy. AC item sets can achieve this when the content to be learned is the ‘correct’ selection between two complimentary words. However, most learning requires ‘correct’ selections within complex antonymic ranges and scales. That is why this paper has proposed Multiple AC item sets, in accordance with recommendations by Thomas Haladyna (Haladyna 2002). When well designed, a MAC has the potential to demonstrate satisfactory learning of naturally complex antonymic ranges within the defined context because the important construal operations are adequately ‘realised’ by the combination of choices between complementarities for each of the important construal operations represented by the collection of question stems.

4.0 Creating Construed Antonym Realisation Exercises - Illustrative example

During initial experiments applying the most promising candidate MCQ test item-generation system from the literature to a particular policy document from the WPD policy library, most of its clauses were filtered out and so the number of usable MCQ test items produced was very small. In order to improve upon this performance, a new source document pre-processing technique (Foster 2009b) was applied to source documents. The results gave some improvement in the output and this paved the way for subsequent research.

The question generation patterns applied to source texts when 'transforming filtered clauses into questions' in the original method (Mitkov and Ha 2003, 2006) are applicable to many educational contexts. However in the WPD training context, a greater emphasis needs to be placed upon testing factual knowledge. This can be seen from a comparison of MCQ test items that have been created manually by WPD industrial trainers and MCQ test items that were generated by the system during initial experiments:

Source Sentence: "Make sure you complete all sections of the diary page. In the 'Work Carried Out' section you must give comprehensive details of your day's achievements."

Manually created question: "A brief description is all that is required in the Work Carried out' section - True or False?" (Correct: False)

Generated question: "What kind of details of your day's achievements must you give in the 'Work Carried out' section" (Correct: Comprehensive)

An analysis of existing MCQ creation techniques and a comparison with the steps within the original process (Mitkov and Ha 2003, 2006) identified the importance of coherence relations in source sentences during the item creation process. The response to this discovery was the development of the CREAM technique (Foster2010e). However the possibility of more benefits was identified when the following actions by item designers were considered::

1) Item designers sought to anticipate erroneous reader construal operations to identify instances of mis-construal following course attendance.

2) Item designers sought to identify the salient features of the source documents within potentially very complex Antonym ranges (Cruse and Croft 2004) and then rationalise them into relatively simple domain-specific and construal-specific Complementarities.

This section describes and then illustrates how skilful narrowing of construal operations available to those answering MCQ test items can reduce the level of Bloom Taxonomy Cognitive learning required to select a response. More specifically, the CARE creation process seeks to identify the salient features within potentially very complex Antonym ranges (Cruse and Croft 2004) and rationalise them into relatively simple domain-specific and construal-specific Complementarities extracted from source documents. The system then prompts designers to identify alternative construal operations that might be applied by readers, and then prompts categorisation decisions about whether each of the construal operations is either erroneous or correct in the context of the specified CSLO (Foster 2009a). A clever designer of MAC items should therefore aim to ensure that all 'correct' construal operations that a learner might apply to the stem within the context are associated with the 'correct' response option and all incorrect construal operations are associated within the incorrect response option."

The classification system for construal operations contains many familiar sets of antonymic ranges; Quantitative vs Qualitative (within the scalar adjustment category of the attention processes); Subjectivity vs Objectivity within the Perspective processes and the slightly more complex Figure vs Ground range (within the Comparison processes). However, the intention in applying the technique is that a careful designer of summative assessment items can create question stems and option word pairs that allow demonstration of the presence or absence in a learner's cognitive processes of a specifically identified item of knowledge.

The 23 construal operations identified within this construal operation classification system (Croft and Cruse 2004) are included within a development document that has been designed for use as a checklist by item designers during manual application of the technique. The original numbering system is retained to allow manual MAC item designers to facilitate references within the textbook 'Cognitive Linguistics' (Croft and Cruse 2004)..

The context chosen for illustrating and evaluating the proposed method provides a rich variety of illustrative contexts for entities and their boundaries. The participants use a full range of learning styles as they interpret the company's written rules. The example quoted in this paper demonstrates how mis-construal can occur when readers of written instructions mis- construe the intended meanings of written instructions. The example also shows the power of MAC formatted MCQ test items when they are used to identify different construal operations that might be applied by different readers to the same source sentence.

4.1 STEP 1 - Define the domain by specifying the corpus boundary condition

The first step in the application of the CARE creation process is to define the boundary to the corpus of documents that form the target domain. In the case of the illustrative example this boundary was identified as 'sentences contained within a date-specific specification document for the TAR system of progress monitoring described in 2.0

4.2 STEP 2 - Explication and Addition of Coherence Relations

The second step is to apply the Explication and Addition components of the CREAM method (Foster 2010e). Words and phrases within significant sentences from the source documents are identified that indicate the presence of coherence relations linking one or more significant information units. This is done by identifying all information units within the source document that relate in some way to the CSLO (Foster 2009a) and then ensuring that all causal coherence relations between these significant information units are explicit (Explication). If the original form of the source document contains implicit causal coherence relations between significant information units then explicit statements of the causal coherence relations are inserted (Addition). An example is provided below in section 4.4

4.3 STEP 3 - Extraction of candidate word pairs using Statistical NLP

The third step is to identify lexical items within the modified corpus that co-locate within syntactic patterns that have been previously identified as likely to co-locate with 'opposite' word pairs.

Complementarities (Cruze and Togia 1995) are awarded the highest scores, while increasingly complex and ambiguous antonyms are awarded progressively lower scores. Complementarities and Antonyms are relations between construal operations, not between lexical items (Cruse and Croft 2004 and Cruse and Togia 1995). Even though this categorisation system has been subsequently challenged (Paradis 2008), the process of generating MAC-formatted MCQ test items from source documents in relation to a stated CSLO turns this restriction into an advantage. The requirement for a shared construal of the source documents by reader and writer, reinforces the need for a shared Semantic communicative intention and a consistent Syntactic presentation if a correct response is to be selected

The illustrative example will be applied to the candidate word pair 'Qualified vs Not Qualified'

4.4 STEP 4 - Generate MAC stems by applying construal theory to word pair contexts

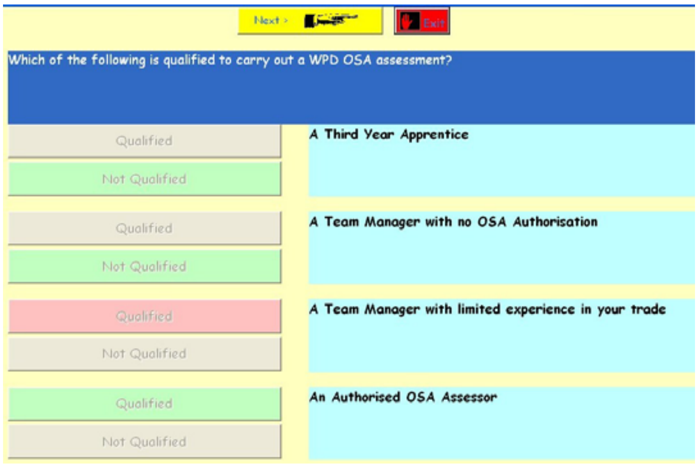
The fourth step of the CARE creation process gives more detailed guidance about the Manipulation step than is specified in CREAM (Foster 2010e). An attempt is made to perceive each of the identified word pairs through each of the available kinds of construal operations applied to the sentence in the source document from which the word pair was originally identified. This is done by re-writing each of the context sentences for the target word pair for as many of the categorised construal operations as are meaningful. The example will do this for the word pair: 'Qualified vs Not Qualified' and the Explicated Source sentence:

> *"A TAR OSA can only be carried out by an authorised OSA Assessor who becomes qualified by holding an EO-MAIN authorisation certificate on CROWN with code 'OSA' and a SAP-MAIN or AP-MAIN authorisation certificate on CROWN with authorisations that correspond to the trade of the apprentice."*

TAR FAQ 22:

4.5 STEP 5 - Present MAC stems and word pairs using MAC template

The CARE creation process concludes by inserting stems and word pairs into the MAC template. The MAC test item that resulted from the illustrative example is provided below:



Which of the following is qualified to carry out a WPD OSA assessment?

A Third Year Apprentice

A Team Manager with no OSA Authorisation

A Team Manager with limited experience in your trade

An Authorised OSA Assessor

Figure 4 – MAC testing knowledge of level qualification required to conduct an On-Site Assessment

5. Experiment

5.1 Hypothesis

The hypothesis is that test items created using the CARE creation process are indistinguishable from manually created MCQ test items.. This will have been proved if the domain expert selects as many CARE items as manually created items for inclusion in a test routine designed to confirm the knowledge of attendees the WPD TAR Induction training course.

5.2 Method

The application of the CARE creation process was achieved within a simulation as opposed to a reprogramming of the question generator in order to ensure careful and thorough application of the steps as described in section 4 above. The source document used in the experiment was taken from the policy library of the UK Company referred to in the introduction. The output sentences from the pre-processing were used as source documents during the manual simulation of the modified test item generator (Mitkov and Ha 2003, 2006) processes which included application of the CARE creation process. The simulated run of the MCQ test item generator produced 32 CARE item sets which were paired up with 32 manually created item sets covering equivalent content.

5.3 Evaluation

The final selection by the domain expert was to consist of 32 AC item sets, which addressed the Controlled Specific Learning Objective:

> *"TAR Induction course attendees must be able to recognise and accurately recall facts covered during the TAR training sessions".*

The domain expert had no involvement in the creation of either the manually or automatically generated items and had no prior knowledge of which were generated AC item sets, therefore these factors could not have any bearing upon his decision about which item sets to include in the test routine. The following usability scores were used to record the domain expert's assessments of the items:

A= Use the item unchanged

B= Make minor changes and then use the item

C= Do not use the item

5.4 Results

On the day of the experiment the 32 pairs of AC item sets was presented to the domain expert. Once the usability categories had been assigned for each of the 32 item sets, the following comparison table was produced:

Usability Score categories	***Generated AC item set was preferred***	***Manually Created AC item set was preferred***
A=Use the AC item set unchanged	28% (9 sets)	25% (8 sets)
B=make minor changes then use this AC item set	38% (12 sets)	9% (3 sets)
C=Do not use this AC item set	34% (11 sets)	66% (21 sets)

Table 1 – Usability categorization decisions for Generated vs Manually created AC item sets.

There were several cases in which AC item sets generated using the CARE creation process were mixed with manually created AC item sets in order to produce the final MAC test item. The number of AC item sets that were changed to make them usable varied considerably as each MAC was constructed, and in one case (content set U151B) one of the four CARE generated items was the ‘inspiration’ for the three new manually created items. These were counted as three ‘changed’, manually created AC item sets (Category B) and three of the original Manually created AC item sets for content set U151B were discarded.

<u>6) Conclusions and Future Work</u>

The illustration of the CARE creation process in this paper has shown how the wide variety of Cognitive processes identified by psychological and linguistic theorists over past decades, can be harnessed into a single method for generating MAC-formatted MCQ test items. The example was chosen to illustrate some of the ‘mistakes’ (erroneous construal operations in relation the stated rules and a Controlled Specific Learning Objective (Foster 2009a)) that can arise when readers of written instructions mis-understand the meaning originally intended by document writers.

This has been a fascinating and stimulating project to be involved with, and every new discovery has brought me closer to an answer to my original ‘what if…’ question. My conclusion so far is that the goal as originally intended and as widely interpreted:

> *“An automated system that can generate usable questions within an un-bounded field of knowledge, when provided simply with ‘a source text’ without any qualifications as to its structure and content”*

… is a goal without meaning or direction.

However, if

(a) the field of knowledge DOES have a testable boundary and

(b) the source documents DO contain the facts that are to be tested, and

(c) the testable facts are described using a controlled lexicon and a controlled specification of all significant antonyms that can be applied to the terms in that controlled lexicon and

(d) the structure of the documents has been arranged so that these facts are clearly identifiable

.. then I believe a software system can be created to draw these components together in response to a well defined assessment objective and thereby satisfy the WPD requirements as originally stated:

> *"WPD seek improvements to their current systems for the creation, delivery and maintenance of Multiple Choice Question (MCQ) test items that will*
>
> *(e) Provide evidence to all interested parties that WPD have met (or exceeded) their responsibilities under the Health and Safety at Work, etc Act 1974 for all relevant fields of knowledge.*
>
> *(f) Reduce likelihood of disputes over validity by drawing content directly from a trace-able source text for an identified target population covering a clearly defined field of knowledge.*
>
> *(g) Manage Change by avoiding expensive revisions of manually created MCQ test item banks following changes in fields of knowledge (eg changes to national legislation)*
>
> *(h) Increase efficiency of MCQ test item creation by saving the time of our item designers, validators and users (trainers and trainees),"*

WPD have already made significant changes to their internal guidance rules concerning the preferred choices of MCQ test item format following earlier experiments (Foster 2010a). In addition to the possibilities offered by the CARE creation process, other linguistic devices are also being considered to combat mis-construal in the minds of newcomers to the company:

(A) a controlled process for the creation of new lexical items (Foster2010b)

(B) a controlled structure for written instructions (Foster 2009b)

Attempts will be made to automate the new techniques within software implementations as they become sufficiently rigorously defined. The intention is to offer the most successful automated implementations on a public domain website.

References

Ames, C. & Ames, R. (1989). Research in Motivation in Education, Vol 3. San Diego: Academic Press.

Behrens, J.T., Mislevy, R.J., DiCerbo K.E., Levy R. (2010) "CRESST REPORT 778 - An Evidence Centred Design for Learning and Assessment in the Digital World" - The National Center for Research on Evaluation, Standards, and Student Testing Graduate School of Education & Information Sciences

Bloom, B. (1956), Taxonomy of Educational Objectives: Book 1 Cognitive Domain, Longman, 1956.

Brown J.C., Frishkoff G.A. Eskenazi M., (2005) "Automatic Question Generation for Vocabulary Assessment" Processing (HLT/EMNLP), pages 819–826, Vancouver, October 2005. © 2005 Association for Computational Linguistics

Croft, William & D.A. Cruse. (2004). Cognitive Linguistics. Cambridge: Cambridge University Press. (ISBN-10: 0521667704; ISBN-13: 978- 0521667708)

Cruse, D.A. (1986) “Lexical Semantics” Cambridge: Cambridge University Press

Cruse, D.A. & P. Togia. (1995). "Towards a cognitive model of antonymy". Lexicology 1: 113-141.

Dawkins R. (2004) The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution: (ISBN-10: 0753819961; ISBN-13: 978-0753819968)

Dictionary.com, "realisation," in WordNet® 3.0. Source location: Princeton University. http://dictionary.reference.com/browse/realisation. Available: http://dictionary.reference.com. Accessed: January 15, 2011.

Dictionary.com, "realisation," in Collins English Dictionary - Complete & Unabridged 10th Edition. Source location: HarperCollins Publishers. http://dictionary.reference.com/browse/realisation. Available: http://dictionary.reference.com. Accessed: January 15, 2011

Fillmore, Charles J. (1985). "Frames and the semantics of understanding". Quaderni di Semantica 6: 222-254.

Foster, R.M. (2009) "Controlled Specific Learning Objectives" (Aston Graduate Corpus Conference 2009) http://acorn.aston.ac.uk/conf_speakers09/confRF.html

Foster, R.M. (2009) “Improving the Output from Software that Generates Multiple Choice Question (MCQ) Test Items Automatically using Controlled Rhetorical Structure Theory” RANLP 2009, Borovets – Student Conference

Foster, R.M. (2010) “Adapting multiple-choice item-writing guidelines to an industrial context.” In proceedings from ICEIS 2010, Funchal, Madeira.

Foster, R.M. (2010) “Cataphoric Lexicalisation” In proceedings for the ICEIS 2010 Conference, Funchal, Madeira

Foster, R.M. (2010) “Improve the output from a MCQ test item generator using Statistical NLP” ICALT 2010, Tunisia

Foster, R.M. (2010) – ‘Multiple Alternative Choice test items (MACs) deliver more comprehensive assessment information than traditional 4-option MC test items ‘ – LICE 2010

Foster, R.M. (2010) – ‘Automatic generation of Multiple Alternate Choice (MAC) test item stems by applying Causal Relation Explication, Addition and Manipulation (CREAM) to pre-processed source documents’ – LICE 2010

Gronlund, N. 1982. “Constructing achievement tests.” New York: Prentice-Hall Inc.

Haladyna, T.M., Downing, S.M., Rodriguez, M.C., (2002) “A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment” Applied measurement in education 15(3), 309–334

Hovy, E. (1993). Automated discourse generation using discourse structure relations. Artificial Intelligence: Special Issue on Natural Language Processing, 63(1-2):341–386.

Johnson, M. (1987) “The body in the mind” Chicago: University of Chicago Press

Lakoff, G. (1987) Women, fire and dangerous things: what cateories reveal about the mind. Chicago: University of Chicago Press

Lakoff G., and Johnson, M. (1980) “Metaphors we live by” Chicago: University of Chicago Press

Langacker, Ronald W. (1987). Foundations of Cognitive Grammar, vol. 1. Theoretical prerequisites Stanford, California: Stanford University Press.

Mager, R. (1975). Preparing Instructional Objectives (2nd Edition). Belmont, CA: Lake Publishing Co.

Mann, William C. and Sandra A. Thompson (1988). "Rhetorical Structure Theory: Toward a functional theory of text organization."

Manning, C.D, and Schutze, H., (2007) “Foundations of Statistical Natural Language Processing” The MIT Press

Marie Tarrant M., Ware J.,Mohammed A.M. (2009) “An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis” BMC Med Educ. 2009; 9: 40. (10.1186/1472-6920-9-40. PMCID: PMC2713226)

Mitkov, R., (2004). “The Oxford Handbook of Computational Linguistics “ Oxford University press

Mitkov, R., and L. A. Ha. (2003). “Computer-Aided Generation of Multiple-Choice Tests.” In Proceedings of HLT-NAACL 2003 Workshop on Building Educational Applications Using Natural Language Processing, pp. 17-22. Edmonton, Canada.

Mitkov, R., L. A. Ha, and N. Karamanis. (2006). “A computer-aided environment for generating multiple-choice test items.” Natural Language Engineering 12(2): 177-194.

Paradis, C. (2010) "Good, better and superb antonyms: a conceptual construal approach"- "The annual texts by foreign guest professors" - Faculty of Arts, Charles University

Sanders, T.J.M., Spooren, W.P.M., & Noordman, L.G.M. (1993) “Coherence relations in a cognitive theory of discourse representation.” Cognitive Linguistics, 8, 93-133.

Swanson, D. B., Holtzman, K. Z., Allbee, K., & Clauser, B. E. (2006) “Psychometric Characteristics and Response Times for Content-Parallel Extended-Matching and One-Best-Answer Items in Relation to Number of Options”: Academic Medicine Vol 81(10,Suppl) Oct 2006, S52-S55.

Talmy, L. (2000) “Toward a cognitive Semantics” vol 1 – Concept Structuring Systems

Tsumori S., Kaijiri K., (2007) "System Design for Automatic Generation of Multiple-Choice Questions Adapted to Students' Understanding" 8th International Conference on Information Technology Based Higher Education and Training, 10th to 13th July 2007, Kumamoto, JAPAN

UK Legislation Health and Safety at work, etc Act (1974)
http://www.hse.gov.uk/legislation/hswa.pdf

Vogel, Anna (2009) A cognitive approach to opposites: The case of Swedish levande 'alive' and död 'dead' - "Approaches to Language" - Research Unit for Variation, Contacts and Change in English (VARIENG), University of Helsinki http://www.helsinki.fi/varieng/journal/volumes/03/vogel/