Creating a High Voltage Cable-Jointing Knowledge check using the CARE Generation Methodology

Robert Michael Foster
University of Wolverhampton,
Wulfruna Street, Wolverhampton, WV1 1LY,
Research Institute in Information and Language Processing
R.M.Foster@wlv.ac.uk

Abstract

This paper presents the latest application of a new process for creating Multiple Alternative Choice (MAC) test items. The method is named Construed Antonym Realization Exercise (CARE) generation. The first step is to pre-process the source documents in the context of a series of Controlled Specific Learning Objectives (CSLOs).

Statistical NLP is then used to identify significant content and complementary and antonymic syntactic patterns. A system of construal operations is then used to produce test items which allow identification of 'correct'/'incorrect' construal of these word pairs in relation to the CSLOs. A MAC template is then used to generate test items.

The process is tested using a domain specific evaluation method. The results provide sufficiently positive evidence to support future experiments.

1. Introduction

Initial research into methods for automatically generating the Alternative Choice component of MAC (Multiple Alternative Choice) test items, identified several possible ways for defining opposites [14], [17], [3]. The new process described in this paper makes use of the categories of Complementarities and Antonym ranges as defined in [3].

The new process also involves a step for identifying instances of erroneous construal of the identified opposites that might arise in the minds of domain newcomers. The description of construal operations proposed in the book 'Cognitive Linguistics' [2] has been chosen to address this requirement in preference to Imaging systems [16] and Focal adjustments. Selected sentences from the featured UK company's policy document library are re-written following the application of each of the construal operations defined by the Cognitive Linguistics framework. The resulting sentences can then be identified by a domain expert as being either 'correct' or 'incorrect' in the context of a CRST-compliant Controlled Specific Learning Objective [6].

The rest of this paper is organized as follows: section 2 describes the Context of this study and provides a description of component processes. Section 3 provides an example to illustrate how some of the AC-item-sets featured in the experiment were created and Section 4 describes how the experiment was conducted before presenting the table of results. Conclusions and descriptions of proposed applications can be found in Section 5.

2. Context

After describing the background to this study, this section describes the component theories for the proposed CARE methodology. Controlled Rhetorical Structure Theory [6], Causal Coherence Relations [15], Construal Operations [2] and Domain specific complementarities [3].

2.1. Background

A promising MCQ test item-generation system was identified from the literature [12],[13]. During initial experiments applying this system to a particular policy document from the featured company's policy library most of its clauses were filtered out. The result was that 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 [6] was successfully applied to the source documents giving some improvement, 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 [12],[13] are applicable to many educational contexts. However in the context of this study 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 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 [12],[13] 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 [7]. However the possibility of more benefits was identified when the following actions by item designers were considered:

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

ii. Item designers sought to identify the salient features of the source documents within potentially very complex Antonym ranges [2] and then rationalise them into relatively simple domain-specific and construal-specific Complementarities.

2.2. Controlled Rhetorical Structure Theory

Rhetorical Structure Theory [11] defines some widely used tools for Natural Language Discourse Processing. Controlled RST [6] adapts some of these tools to guide the controlled construction of discourse elements within a well specified domain. CRST unites standard Rhetorical Structure Theory [11] with the theory of Controlled Specific Learning Objectives (CSLO) [6].

The inherent restriction of CRST to a well defined domain does not present a problem in the context of this research since the domain is well defined by the company’s policy document library from which the source documents are taken. The application of the CRST standard enforces clarity of content and the communicative goals of the source document writer. This paper presents the results of an experiment when MAC test item stem templates were applied to the output from the pre-processing described in this and subsequent sections and the MAC test items produced were reviewed by a domain expert.

2.3. 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 [15]. 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 [15] but can be summarized as follows:

i. Basic operation (causal vs additive)

ii. Source of coherence. (semantic vs pragmatic)

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

iv. 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.

2.4. Construal Operation Categories

William Croft explains in chapter 3 of 'Cognitive Linguistics' [2] that the construal operation classification system proposed in his 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 …. 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 [16] and Focal adjustments. 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 [5] are missing and the more complex cognitive processes of Metaphor [9] and Image schemas [8],[10] 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' [2]

i. How do construal operations interact within this domain?

ii. How should we characterise the processes of Language vs Thought vs Experience?

2.5 Domain Specific complementarities

Research into possible methods for generating the required Alternative Choice component of MAC items identified several contradictory studies [14], [17], [3]. If we take a dynamic construal approach to sense relations then we can adopt the general categories of Complementarities and Antonyms [3] 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, but 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 system.

In the featured domain, when tackling an un-familiar 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. Assessment of a new concept 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 [3] the relationship changes from a straight-forward 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 [3], 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 system for identifying opposites [3] gives detailed descriptions of antonym pairs and readers of this paper are advised to read chapter 7 of the book 'Cognitive Linguistics' [2] if a deeper insight into this topic is required.

3. Methodology

This section describes and then illustrates how skilful narrowing of the range of construal operations offered to those answering MAC test items can reduce the level of Bloom Taxonomy [1] Cognitive learning required to arrive at the correct response. More specifically, the Construed Antonym Realisation Exercise (CARE) creation process allows designers to identify the salient features within potentially very complex Antonym ranges [2] and rationalise them into relatively simple domain-specific and construal-specific Complementarities which can be traced back to exact phrases within the source documents.

The item design processes can then identify alternative construal operations that might be applied by readers, and prompt decisions about whether each of the construal operations is either erroneous or correct in the context of the specified CSLO [6]. All 'correct' construal operations that a learner might apply to the stem are associated with the 'correct' response option and all incorrect construal operations are associated with the 'incorrect' response option.

The context that has been chosen to illustrate and evaluate the proposed method in this paper provides a rich variety of illustrative work situations and domain specific entities, relationships and boundaries. The participants use a full range of learning styles as they interpret the documented rules. The quoted examples demonstrate how reader mis-construal of written instructions can be anticipated. The examples also show the power of MAC formatted MCQ test items when they are used to explicate different construal operations that might be applied by different readers to the same source sentence.

3.1 Step 1 - Define Objective and Corpus boundary

The first step in the application of the CARE creation process is to define the objective of the test routine and to describe the boundary of the corpus of documents that will form the target domain. For example, the manual application of the method that is described in the next section was provided with this specification:

> The aim of the test routine is to:
> *"Identify candidates who can correctly recognize LV general jointing procedures for live and dead working (ST:CA1C)"*
>
> The boundary of the corpus is:
> *"Sentences contained within a date-specific version of ST:CA1C 'Relating to General Requirements Low Voltage Jointing"*

3.2 Step 2 - Explicate (and if necessary Add) Coherence Relations then produce CRST-compliant CSLOs

The second step is to apply the Explication and Addition components of the CREAM method [7]. Textual patterns are identified which indicate implicit or incomplete coherence for information units within the source document that relate in some way to the CSLO [6]. Incomplete coherence relations between these significant information units are then Explicated or if the original form contains implicit causal coherence relations between these significant information units, then explicit statements of the coherence relations are Added. For example, one of the original sentences identified as containing significant information was:

> *"If the flame is accidentally extinguished do not attempt to relight the gas and do not allow any naked flame near, until the accumulated gas has been dispersed by opening ventilators, tent flaps or doors"*

ST:CA1C – General Requirement 1

There was a need for the Coherence Relation to be Explicated before this sentence could be put forward to the next step because the sentence did not make sense without including a reference to the title of the section from which it was taken. The result of the 'Explication' action taken in step 2 of the CARE creation process was:

"If the flame **of a gas torch** is accidentally extinguished do not attempt to relight the gas and do not allow any naked flame near, until the accumulated gas has been dispersed by opening ventilators, tent flaps or doors"
ST:CA1C – General Requirement 1 (modified)

3.3 Step 3 – Identify useful concordance target and then extract antonym pairs

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 [3] 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 [2] and [3].

The illustrative example provided with this paper has been generated following the identification of the frequent occurrences of the lexeme 'if' in the corpus and subsequent extraction of all sentences containing this indicator of a likely statement of LV jointing procedure that could be tested within the context of the antonym pair: 'mentioned in policy ' vs 'NOT mentioned in policy'.

3.4 Step 4 – Apply construal operations in the context of identified antonym pairs

The fourth step of the CARE creation process gives more detailed guidance about the Manipulation step than is specified in CREAM [7]. An attempt is made to perceive each of the sentences from the source document through as many of the construal operations as are sensible. Each of the identified antonym pairs is then applied to form an AC item set. One of the created MAC test items includes the result of applying a Spacio-temporal mis-construal to the source sentence:

"The use of the disposable plastic gloves provided should eliminate any contact, but if it occurs, the affected area should be washed immediately with plenty of soap and water."

ST:CA1C – General Requirement 37

.. in the context of the antonym pair: "Correct action vs NOT the correct action"

3.5 Step 5 – Generate AC item sets Method

The CARE creation process concludes by using the 'correct' and 'incorrect' statements and associated antonym pairs to generate AC item sets. AC item sets which use the same antonym pair are grouped together in order to produce MAC test items ready for evaluation. The MAC test item that resulted from the illustrative example process application in this section is provided below:

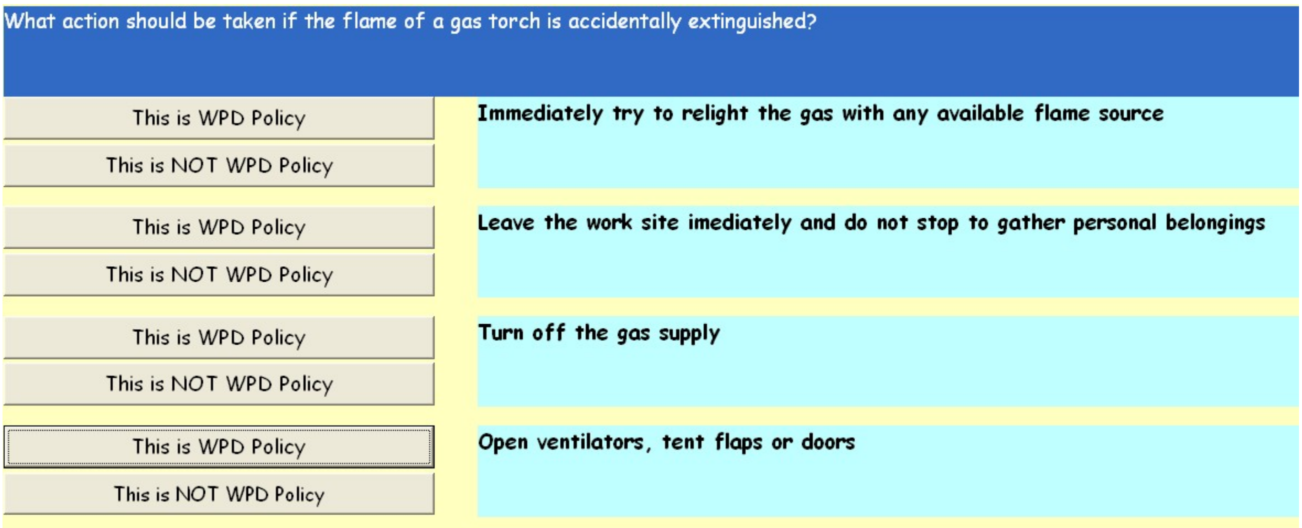

Figure 1 – Example generated MAC test item

4. Experiment

This section describes the latest experiment that has tested the CARE generation process in action. The Corpus boundary was defined as:

"Sentences contained within a date-specific version of ST:CA1C 'Relating to General Requirements Low Voltage Jointing"

4.1 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 an equal or greater number of AC item sets that were generated using the CARE creation process CARE as manually created items for inclusion in a test routine designed to:

"Identify candidates who can correctly recognize LV general jointing procedures for live and dead working (ST:CA1C)"

4.2 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 the Methodology section. The output sentences from the pre-processing were used as source documents during the manual simulation of the modified test item generator [12],[13] processes which included application of the CARE creation process. The simulated run of the MCQ test item generator produced 68 AC item sets which were paired up with 68 manually created item sets covering equivalent content before being presented for evaluation.

4.3. Evaluation

The final selection by the domain expert was to consist of 68 AC item sets, which addressed the Controlled Specific Learning Objective stated in section 4.1. 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 AC item set unchanged
B= Change the AC item set Antonym pair
C= Change the AC item set Statement
D= Do not use the AC items set

4.4 Results

On the day of the experiment the 136 AC item sets were presented to the domain expert who then compiled a routine, often using a combination of generated and manually created AC item sets 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 several cases one of the four CARE generated items was the 'inspiration' for the new manually created items. These were counted as 'changed', manually created AC item sets (Category B) and a corresponding number of the original Manually created AC item sets for this content set were discarded.

Once the usability categories were assigned for each of the 68 generated AC item sets, the following comparison table was produced:

Table 1 - Domain expert decisions for Generated vs Manually created AC item sets.

	Generated AC Item sets	**Manually created AC item sets**
A=Use unchanged	**28% (19 sets)**	**48% (32 sets)**
B=Change Antonym Pair	**13% (9 sets)**	**0% (0 sets)**
C= Change Stem	**12% (8 sets)**	**0% (0 sets)**
D= Do not use	**47% (32 sets)**	**52% (36 sets)**

5. Conclusions and Future Work

Applying the decision about category according to a clearly observable action allows the same process to be repeated consistently in other MCQ test item generation experiment evaluations. The most encouraging outcome from this experiment is that the number of manually created AC item sets that were excluded from the final routine was slightly higher than the number of excluded generated AC item sets.

Another encouraging feature of this experiment when compared to previous experiments is that the domain expert's decision was made entirely upon the merit or otherwise of the generated item as opposed to the inapplicable content, which has been a problem with previous experiments.

The fact that a significant number of the generated AC item sets required changes before they could be used is unfortunate. However, these item sets will provide useful guidance in the construction of an automated implementation, which is the planned next step for this project.

This paper has prepared the ground for future experiments seeking improvement in performance of the featured software [12], [13] by pre-processing source documents. The CARE creation process has been applied and a pragmatic, domain specific evaluation method of output from the system has been used.

Automation of the process is feasible thanks to the well defined boundary for the policy document corpus which is protected by a well organized change management system, and this is primary focus for future work.

Acknowledgements

I wish to thank Steve Loveridge and Steve Davies from my employing company Western Power Distribution and my Phd supervisors Dr Le An Ha and Professor Ruslan Mitkov from Wolverhampton University for their continued guidance and support.

References

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

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

[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] Cruse, D.A. (1986) "Lexical Semantics" Cambridge: Cambridge University Press

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

[6] 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

[7] 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

[8] Johnson, M. (1987) "The body in the mind" Chicago: University of Chicago Press

[9] Lakoff G., and Johnson, M. (1980) "Metaphors we live by" Chicago: University of Chicago Press

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

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

[12] 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.

[13] 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.

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

[15] 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.

[16] Talmy, L. (2000) "Toward a cognitive Semantics" vol 1 – Concept Structuring Systems

[17] 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/