Institution: University of Wolverhampton

Unit of Assessment: 28 Modern Languages and Linguistics

a. Context

The overall aim of the submitted unit of assessment (UoA) is to produce high-quality linguistic research and to apply it in order to achieve positive impacts outside academia. The unit comprises two groups, RGCL (Research Group in Computational Linguistics) and CDCS (Centre for Discourse and Cultural Studies), both investigating linguistic topics of strategic relevance to healthcare, translation and education.

The research activities conducted by RGCL lie within the area of Computational Linguistics and more specifically Natural Language Processing (NLP), which is concerned with the development of computational techniques for analysing and representing naturally occurring texts for the purpose of achieving human-like language processing in a range of applications. Being focused on human communication and knowledge, NLP technologies have applications in a variety of fields for different types of users. Large organisations and companies as well as Small-Medium Enterprises (SMEs) who hold significant quantities of natural language data can exploit them more efficiently by using NLP technologies and applications, such as information retrieval, information extraction, semantic search, question answering, terminology processing and automated scoring technologies. Web users accessing pages written in languages they do not know, or people with reading comprehension difficulties attempting to access information in textual materials, will benefit from technologies such as machine translation and automatic text simplification as they help reduce the effect of language barriers and other obstacles (e.g. text complexity) when accessing textual data. Users involved in the generation of natural language resources, such as educators (producing educational and assessment resources), will benefit from computer-aided Multiple-Choice Question (MCQ) generation and contextual disambiguation.

The Centre for Discourse and Cultural Studies carries out research in the area of discourse analysis, focusing on topics such as illness experience, gender experience and the experience of the process of organ donation and tissue transplantation, bringing relevant contributions to medicine and psychology. In the medical domain, health authorities can rely on discourse analytical research to improve quality of care and support through a better understanding of the experience of organ and tissue transplantation. The discourse analysis of the linguistic forms used by mentally ill patients assists clinicians in the provision of psychiatric and psychological health care.

Table 1 lists the aforementioned groups of users, together with examples of actual users of the research of RGCL and CDCS, and specific technologies and methodologies exploited. Non-academic partners (further detailed below) such as *HERMES*, *NLP Technologies Inc., and NHSBT (National Health Service Blood and Transplant)* feature in Case Studies 1 and 2. In addition to the research described in the case studies, the *US National Board of Medical Examiners* (*NBME*) are currently using RGCL applications which assist them in their everyday activities. People with language difficulties and their carers are currently evaluating RGCL's textual data accessibility technologies within the framework of FIRST (an EC FP7-funded project).

User groups	**Examples of actual users benefiting from the unit's research**	**Methodologies, technologies and applications employed**
Large organisations and companies, SMEs who hold significant amount of natural language data	*NBME*	Semantic Processing, Information Retrieval, Information Extraction, Question Answering
	NBME	Automated Short Note Scoring
	NHSBT	Discourse Analysis
Large organisations and companies, SMEs, Carers, Individual users who need to generate and translate natural language texts	*HERMES,* *NLP Technologies Inc.,* *Translution* *DTIC*	Computer-Aided Translation
	Carers of patients with language difficulties	Text Simplification, Textual Data Accessibility
Individual users who need to access natural language texts	Web users	Machine Translation, Textual Data Accessibility
	Patients with language difficulties	Text Simplification, Textual Data Accessibility
	Clinicians	Patient Discourse Analysis

Table 1: User groups, specific users, and the methodologies, technologies and applications exploited by them

b. Approach to impact

The overriding strategic imperative of the UoA remains as it was at its foundation: to produce cutting-edge research in identified areas of NLP, discourse and cultural studies that, through their practical application, have the potential to benefit users in academia and beyond. The development of the two participating research groups, their outputs, their engagement with funders and users of research, their supportive environment for both staff and research students, and their readiness to provide consultancy, are elements that have helped both RGCL and CDCS achieve the impact of their research.

c. Strategy and plans

In order to maximise the impact of its research, the UoA adopts a three-point plan, which is in line with the overall strategy outlined in REF5:

1) focused applied research and development to address users' needs,
2) motivated scientific and social networking, and
3) dedicated consultative training.

Point 1) enables development of the methodologies that can be brought to users by activities undertaken in Points 2) and 3). Point 2) ensures dissemination of our research to as many potential users as possible and helps identify their needs, whereas carrying out Point 3) involves working with users to help them exploit the outputs of our research, maximising their benefits and the positive impact on their work. All these branches of the strategy are driven by the study of the needs of different users in society and the identification of linguistically informed solutions which can successfully address these needs.

1) The first part of the three-point plan is focused applied research and development with an emphasis placed on those areas that can bring about societal benefits. This involves identifying linguistically informed solutions which can successfully address the needs of different users in society, where the development of high-quality research, tools, and resources can have a considerable impact on non-academic users and can sustain the quality and standards of the research. As described in REF5, of the societal needs studied by RGCL and CDCS, the UoA has sought to apply natural language technologies and discourse analysis in three priority areas: medicine/healthcare, translation and education. Research topics contributing to these applications thus constitute the primary focus of the UoA. Case Study 1 demonstrates how it is possible to maximise the efficiency of translation using a tool developed by the unit. Case Study 2 shows the importance of discourse analytical research for educating clinicians and understanding the experience of friends and relatives of organ transplantation.

The unit (particularly RGCL) also dedicates effort to developing working prototypes and demos which run on real-world data rather than expensive model datasets that are unrealistic in practical settings. This allows users to quickly assess the feasibility of the research in their application domain by using the prototypes to process their own data. Members of the unit also actively provide consultancy services for non-academic partners.

Other projects that have recently started and which will have significant impact in the coming years, include:

a) FIRST, which develops tools to simplify natural language texts for people diagnosed with Autism Spectrum Disorders (ASD),
b) EXPERT, which is a Europe-wide postgraduate training network in the field of Machine Translation and Translation Technology in general, with active participation from the industrial sector,
c) TELL-ME, which develops tools and materials to support language learning for professionals working in healthcare, and
d) DVC, which applies a new, corpus-driven theory of word usage and meaning and is discovering previously unrecorded systematic links between phraseology and meaning in text. It was the basis for an invited talk by Hanks on the role of phraseology in dictionaries at a symposium on the future of the Oxford English Dictionary and will shape future lexicographical endeavour.

2) The second branch of the strategy, motivated scientific and social networking, is followed by the UoA to establish and maintain links with industrial partners and end users, including joint research ventures, which have led to the development of user-orientated linguistically informed solutions. These activities contribute to research findings pertaining to the needs of different users in society.

Members of RGCL and CDCS disseminate their research at international conferences, attended both by academics and representatives from companies and governmental institutions, and in journals. *NBME* first encountered RGCL via the unit's publication and presentation of research in MCQ generation at the NAACL conference in Edmonton, Canada in 2003. Case Study 1 was partially built on collaborations established during the ASLIB 'Translating and the Computer' conference and other conferences with industrial participation. Case study 2 has developed as a result of presentations delivered at several workshops. Academic and non-academic professionals have been invited to visit the UoA, and members of the unit have visited non-academic partners. Examples of this include visits from the CEO of *QuestionMark*, the *NBME* management team, the NHSBT management team, and staff from the *NHS* involved in mental healthcare provision. Members of the unit have visited end users, studied their needs, discussed collaboration and, on occasions, given talks at organisations including *Research Research*, *Translution*, *New Cross Hospital, Fundación CIEN* and *NBME*, among others. These exchanges foster networking and collaboration between the UoA and non-academic partners, facilitating joint research and publications, joint projects, joint PhD supervisions, and funding. Researchers from the UoA were explicitly mentioned in the publicity material of several non-academic partners (e.g. *QuestionMark*). RGCL takes part in and coordinates EU and UK research councils funded research projects (QALL-ME, FIRST, MESSAGE, DVC), involving both academic and non-academic partners. This partnership demonstrates that the second branch of the strategy has helped to secure the impact of our research outside academia.

3) The third branch is dedicated consultative training, with a view to transferring available knowledge from academia to professional end users and stakeholders. Members of the UoA are involved in the provision of training for professionals on a regular basis and include dissemination of research results and consultancy activities. For example, in the EC-funded MESSAGE project, RGCL delivered training for professionals working in the emergency services on how to write more effective messages, alerts and protocols for use in response to terrorism and other security-related events. These training activities were based on the research into controlled languages carried out by RGCL. In the FIRST project that started in October 2011, RGCL acts with other project partners to deliver training for professionals working to support people with ASD, showing how to use software developed in the project in order to improve users' access to written materials. The TELL-ME project is developing resources for language learning for medical professionals. Research conducted by members of CDCS has been incorporated into a specialised training programme for psychology clinicians in the Lower Silesia region of Poland. With support from the University of Wolverhampton, the UoA actively engages with small and medium-sized businesses to build links and provide academic support. For example, RGCL was involved in a voucher-facilitated collaboration with *Translution* which enabled academic support to investigate the integration of dictionaries within online translation engines.

d. Relationship to case studies

Both case studies presented below result/benefit from Points 1), 2) and 3) of the strategy. The two case studies serve as detailed examples of how the unit's research has impacted different types of users. Case Study 1 describes how RGCL's post-editing tool has been used by *Hermes*, *NLP Technologies Inc.* and *the Department of Translation, Interpreting and Communication of Ghent University* to better understand the process of post-editing machine translation. This has resulted in better selection of machine translation software, and saves time in the translation process. In addition, the tool was successfully employed in teaching, a purpose which was not initially foreseen. Case Study 2 describes how analytical discourse research has had an immediate impact on communicating information in the context of mental health care and in transplantation medicine. Through a better understanding of communication activities in these clinical settings, significant changes have taken place in the educational and professional practices of clinical psychologists, as well as in the process of obtaining consent for organ and tissue donation in the UK.