Institution: University of Essex

| **Unit of Assessment:** 28 – Modern Languages and Linguistics |

| **Title of case study:** Developing prototypes for natural-language interfaces in collaboration with BAE Systems |

1. Summary of the impact

Essex research into the practical deployment of computational grammar theories, tools and techniques led to the expertise of Dr Doug Arnold being sought between 2009 and 2011 by BAE Systems, a leading UK manufacturer of advanced defence and security systems. Arnold advised the company on the design of two prototype natural-language interfaces for responding to emergency situations and sharing sensitive data across organisations. The projects' goals were met and his contribution enabled BAE Systems to develop feasibility-of-concept demonstration systems. His practical expertise in Natural Language Processing provided the company with an appreciation of the limits of particular tools and helped it to avoid undertaking over-ambitious projects.

2. Underpinning research

Doug Arnold's (Senior Lecturer at Essex) research has focused on the real-world deployment of computational grammar theories, tools and techniques, and the formal description of English in grammatical frameworks that can be used for computational purposes. In short, his work has concentrated on the problems, both theoretical and practical, of how the English language can be 'understood' by a computer. Arnold has not only produced a considerable volume of publications in this area, but working in this field has also allowed him to gain hands-on experience of developing practical applications and acquire a working knowledge of cutting-edge tools and techniques.

Arnold has published on a range of topics relevant to the practical application of Natural Language Processing (NLP) since the mid-1990s, including how systems should be evaluated (Arnold et al., 1993), the design of resources for testing systems (Balkan et al., 1995), and how systems for one purpose can be reused for other purposes (Arnold, 1994). Much of the crucial research underpinning the impact described in Section 4 has been Arnold's knowledge of, and expertise in, the application of the syntax and semantics of a wide range of grammatical constructions in English (see e.g. Arnold, 2007; Arnold and Sadler, 2010, which deal with the syntax and semantics of relative clauses). This work is combined with hands-on, practical knowledge of extant programming languages and techniques, as well as, most significantly, their limitations (see Arnold and Linardaki, 2007, which addresses the limitations of certain statistical approaches to NLP).

Arnold's work on the syntax-semantics interface is concerned with the effective automatic analysis of natural language with a degree of accuracy that makes it ideal for real-world applications. Many state-of-the-art techniques are probabilistic and while robust, they do not possess the accuracy required for sensitive applications where a high degree of precision is required. However, because Arnold's research and expertise is within the rule-based classical approach to NLP it is ideal for developing 'high-precision' techniques. It is this particular concern with 'high-precision' approaches, allied to practical expertise, which led BAE Systems to ask Arnold in 2009 to act as a consultant on projects that required detailed grammatical analysis of state-of-the-art techniques.

3. References to the research

Arnold, D., R. L. Humphreys and L. Sadler (1993) Evaluation: An assessment. *Machine Translation*, 8(1-2): 1-24. DOI: 10.1007/BF00981238

Arnold, D. (1994) Reusability — general considerations. In S. Markantonatou and L. Sadler (eds.), Grammatical formalisms: Issues in migration and expressivity, *Studies in Machine Translation and Natural Language Processing*, vol 4, pp. 11-34. Office for Official Publication of the European Communities, Luxembourg, ISSN 1017-6568. [Available from HEI on request.]

Balkan, L., D. Arnold and S. Meijer (1995) Test suites for Natural Language Processing. *ASLIB Proceedings*, 47(4): 95-98. DOI: 10.1108/eb051385

Arnold, D. (2007) Non-restrictive relatives are not orphans. *Journal of Linguistics*, 43(2): 272-309. DOI: 10.1017/S0022226707004586

Arnold, D. and E. Linardaki (2007) Linguistic constraints in LFG-DOP. In M. Butt and T. Holloway King (eds.), *Proceedings of the LFG07 Conference*, pp. 66-86, Stanford, CA: CSLI Publications. http://www.stanford.edu/group/cslipublications/cslipublications/LFG/12/lfg07.pdf

Arnold, D. and L. Sadler (2010) Pottsian LFG. In M. Butt and T. Holloway King (eds.) *Proceedings of the LFG10 Conference,* pp. 43-63, Stanford, CA: CSLI Publications. http://cslipublications.stanford.edu/LFG/15/papers/lfg10arnoldsadler.pdf

4. Details of the impact

In working with BAE Systems, a leading UK manufacturer of advanced defence and security systems, Arnold has so far contributed to two projects: the first, 'Free Text Processing', took place from September to December, 2009; the second, 'Advanced Data Sharing Agreements', ran from March to November, 2011. Given their potential uses in emergency situations and high-security operations, both projects required systems for NLP that utilised 'high-precision' techniques. In his consultancy role Arnold was able to draw upon his published research in the formal description of English and computational grammar, as well as his expertise in the syntax-semantics interface. It was not the case that BAE Systems simply made use of Arnold's research and ideas. Instead, he was actively involved in supplying practical guidance and was thus a key factor in realising applications of his research and expertise. BAE Systems' University and Collaborative Relationships Manager has corroborated the claims made in this section and all quotations below come from a statement that he has provided.

The theme common to the two projects is the aim of allowing ordinary users to communicate with various automatic systems using a subset of English ('controlled English'). In the case of the first project, the aim was to allow Emergency Service personnel (e.g. police or ambulance crews) to communicate precise narrative detail about a developing emergency situation. This communication would take place through a discourse like "A car and a bicycle have collided on Coval Road. The cyclist has injured her leg. She has been taken to Chelmsford A and E.", which might generate a pictorial display on a command and control room map. This type of discourse poses serious analytic challenges at the limit of current understanding and technology in NLP. BAE Systems have stated that "the project successfully demonstrated that 'free text' messages, in the abbreviated and specialised language used by emergency workers, could provide a valuable enrichment of the operational picture".

The second project was similar in that it involved automatically transforming text into a pictorial display. This project involved developing a prototype interface for facilitating secure sharing of sensitive data across organisations. Data-sharing agreements between organisations determine

which documents can be read by whom and in what circumstances. The aim was to produce a piece of software that allows data-sharing agreements to be visualised in order to verify that the text is consistent (i.e. not self-contradictory). This means that an agreement can be generated automatically from typescript in English, but no special skills are required to either write it, or to read and understand it. The software recognises contradictions, such as someone being allowed to view documents under certain circumstances but forbidding them from being in such circumstances.

As BAE Systems have confirmed, both projects achieved their goals as a feasibility-of-concept demonstration system was produced for each. Drawing on Arnold's research and expertise, the BAE Systems project teams were able to greatly improve their understanding of the problems and the techniques and technologies available. BAE Systems have gone on record as saying:

> "Dr. Arnold's role in both projects was critical, in two main ways. First, his understanding and command of the fundamental theoretical concepts meant that the projects were able to focus on the fundamentally important issues, this was important both in the design and execution of the projects. Second, his awareness of the capacities and limitations of advanced research tools and toolsets for Natural Language Processing meant that the projects were able to make more informed and appropriate choices among existing tools, and were able to focus resources on areas where existing tools are inadequate"

He was thus able to provide BAE Systems with tools and models for conceptualising what they were trying to build. In so doing, he allowed staff at BAE Systems to better understand the parameters of the types of system they were attempting to construct.

His guidance to BAE Systems included supplying information about which systems were research-only and providing an appreciation of the limits of different technologies. The latter proved crucial as his contribution involved not only helping to develop certain systems but also giving advice about what should be abandoned and what was not possible. Indeed, BAE Systems have stated that Arnold's contribution enabled the teams "to contemplate a range of novel projects, but also avoiding projects whose goals are over-ambitious given the current state of the art." BAE Systems added that avoiding over-ambitious projects is an "aspect of technology transfer [that] is often neglected in university-industry collaboration and Dr Arnold's contribution has been greatly appreciated".

The success of the two projects is demonstrated by BAE Systems inviting Arnold to work on a further initiative in March and April 2013. This involved collaboration on a proposal for a future project on the extraction of information from dialogue. Arnold prepared a report on the current state of the art, which again relied on his research expertise on rule-based machine translation.

5. Sources to corroborate the impact

All documents are available from HEI on request.

1. University and Collaborative Relationships Manager, BAE Systems.

2. Contract confirming Arnold's work with BAE Systems in March and April 2013.

3. Email from member of staff at BAE Systems outlining project on the extraction of information from dialogue.