

# The SALT Cymru Feasibility Report and the resulting Special Interest Group

**Gruffudd Prys**

Bangor University  
Bangor, Gwynedd, Wales  
g.prys@bangor.ac.uk

**Abstract:** This paper examines the findings of a 2008 project examining the feasibility of establishing a Special Interest Group (SIG) focusing on Speech and Language Technology (SALT) research in Wales. It will also report on the current progress of the SALT Cymru Special Interest Group, which was established as a direct result of the feasibility study's recommendations.

**Keywords:** SALT, Wales, Welsh, Minority Language, Speech and Language Technology

## *SALT Cymru Feasibility Project*

### **Background**

The SALT Cymru Feasibility Project was the result of a successful application from Bangor University to the Welsh Assembly Government's EU-funded Knowledge Exchange Programme (KEF). KEF's aim was to stimulate Wales's economy by promoting knowledge transfer between universities and businesses. The study would explore the feasibility of establishing a special interest group that would bring together representatives of academia and industry in Wales who share an interest in speech and language technology. The final report would form the basis of an application for Welsh Assembly administrated European funding for the proposed special interest group.

The project was undertaken by Bangor University's Language Technologies Unit (LTU), which specializes in the development of multilingual language technology, in particular those that serve the needs of the Welsh language as a modern, living language.

Despite the unit's experience with developing Welsh speech and language technology, the SALT Cymru feasibility project and the proposed special interest group were not conceived as a language-specific project nor as a minority language project. Rather, the project's main aim, aligned with the Welsh Assembly's priorities, was to stimulate the Welsh economy by encouraging the participation of Welsh industry in the growing worldwide multilingual SALT market, whatever the language. However, the statutory

requirement for Welsh-language provision from public bodies in Wales was also regarded by the LTU as an economic opportunity for business in Wales.

Due to the multidisciplinary nature of SALT, the project team featured a variety of specialists. Delyth Prys, a linguist specializing in terminology, served as the project director and Rhys James Jones, a speech technology specialist, managed the project together with Gruffudd Prys, a linguist and web developer who served as the north Wales Co-ordinator. Additional expertise was provided by a team of software engineers including Dewi Bryn Jones, Ambrose Choy and David Chan, whose experiences include a number of natural language processing technologies.

### **Defining SALT**

A report examining issues to do with SALT naturally requires that SALT be precisely defined. However, as SALT is considered in academia to be a multidisciplinary area, its precise definition is not universally defined. A survey of the definitions of SALT used by various research institutions (see the *SALT Feasibility Report, Appendix A*) failed to identify a common description. As the SALT Cymru Feasibility Report (the project's ultimate output) required the creation of a standard definition for SALT for the sake of clarity and consistency, it was decided to follow the definition found in the *Survey of the State of the Art in Human Language Technology* edited by Varile and Zampolli (1997) and use the following:

*For the purposes of the project, SALT (speech and language technology) is defined as the inclusion of human language in software for processing text, speech and knowledge. It includes, but is not limited to the following fields: speech technology; written language input; language analysis, understanding and translation; automatic document processing; machine translation; multimodality; electronic language resources and SALT evaluation.*

The categories found in Survey of the State of the Art were also emulated, although for ease of a definition the thirteen categories were reduced in number to the following eight:

1. Speech technology (speech recognition, speaker recognition, text to speech techniques, speech coding and enhancement, multilingual speech processing)
2. Written language input (optical character recognition, handwriting recognition)
3. Language analysis, understanding and generation (grammar, semantics, parsing, discourse and dialogue)
4. Document processing (text and term extraction, interpretation, summarization)
5. Machine translation (including computer aided translation, multilingual information retrieval)
6. Multimodality (gesture and facial movement recognition, visualisation of text data)
7. Language resources (written and spoken corpora, lexica, terminology)
8. Evaluation (of all of the above)

## **Project Overview**

In order to ascertain the feasibility of establishing a Special Interest Group for SALT in Wales, the SALT Cymru Feasibility Project proposed to investigate the:

- Worldwide economic impact of SALT

- International Best Practice and State of the Art in SALT
- Academic SALT research base in Wales
- Enterprise research base in Wales
- SALT needs of industry in Wales

These are discussed in more detail below.

The economic importance of the SALT Sector worldwide is often talked about but the establishment of a funded SALT SIG required proof of the global economic significance of SALT. Several presentations and papers at the LangTech conference attended by SALT Cymru representatives provided figures demonstrating the significance of the SALT sector to the global economy. These included:

- The worldwide market in speech technology deployments alone is estimated to be worth \$3.2 billion by 2010 (*Understanding the Market Movements in Network Speech*)
- In 2005 the translation market, particularly in multilingual websites and software localization, generated \$8.8 billion in worldwide revenues. (*Venture Capital and Language Technologies*)
- 4 of the top 20 international companies site their HQs in London, the North of England and Ireland employing over 2300 people. (*Venture Capital and Language Technologies*)

This provided a strong economic argument for ensuring that academia and industry in Wales are involved in SALT research and development.

Although it was clear that SALT formed a significant sector internationally, the focus of the SALT Cymru feasibility report project was its potential for the Welsh economy. As a project to be funded through the Welsh Assembly Government's Academic Expertise for Business (A4B) programme the special interest group's purpose was to facilitate the transfer of expertise from Welsh Universities to businesses in Wales for their economic benefit.

However, before the special interest group could be funded, the feasibility study had to establish that SALT research was to be found

within academia in Wales, and that the provision of funds for establishing a special interest group would be of benefit to both industry and academia in Wales.

As a result, the project was required to locate and quantify the existing SALT research to be found within academia and industry in Wales. Therefore, an audit of the SALT research in Wales was undertaken by members of the SALT Cymru team.

The relevant departments in all Higher Education institutions in Wales were identified and contacted to establish what activities, if any, they were undertaking in the field of SALT. It was discovered that the SALT development knowledge base in Wales was relatively small. In Welsh higher education institutions, it was estimated that the equivalent of fewer than ten full-time academics worked directly on SALT. Of these, only about half were permanently contracted to do so.

However, despite this small size a wide variety of SALT were found to be under development. They encompassed both speech and language technology, and seven of the eight research areas defined by the project as part of SALT were found being researched within Wales, namely:

- Speech technology (speaker recognition in Swansea, text-to-speech techniques and a speech recognition project in Bangor)
- Language analysis, understanding and generation (spelling and grammar checking, and a library of language tools in Bangor)
- Document processing (text and term extraction, in Bangor)
- Machine translation (a small pilot project in Bangor)
- Multimodality (SALT-associated research in Swansea)
- Language resources (developed in Bangor, Swansea and Lampeter)

- Evaluation (standardization work in Bangor and by Geolang, undertaken with ISO)

It was also noted that there was very little duplication of subject areas within Welsh SALT research. Only rarely were multiple companies or academic institutions found to be active in the same subfield of SALT, with each group seeming to have found their own individual niche. This finding was a positive one for the envisaged special interest group, as a lack of conflicts of interest within the group would likely lead to a more receptive environment for sharing ideas, and working together in a collaborative environment. Based on the growth of SALT internationally, it was felt that this knowledge base could thrive with appropriate funding and the establishment of greater ties with industry, and would be well placed in a multilingual marketplace due to the bilingual nature of Wales.

In addition to the pure SALT research and development that was identified, a significant amount of SALT-associated development work was found in Wales, both in academia and within industry. A number of companies were found to be exploiting SALT development together with other technologies, especially in the field of web design and in the multimedia and creative industries. Enabling these industries to benefit from the research underway at Welsh universities was seen as a priority for the intended Special Interest Group. This prompted an audit of the present state of play of industry in Wales and an assessment of its needs.

In examining the needs and current state of industry in Wales, special attention was paid to Welsh SMEs and their markets/potential markets.

These SMEs fell into three categories:

- (a) developers of software products and services e.g. software developers, web developers
- (b) developers of language-based assistive technology products e.g. disability aids
- (c) users and potential users of SALT-enhanced applications e.g. translation companies

Outside academia, many of those questioned about SALT were unfamiliar with the SALT label or grouping, even though they were daily users of SALT technologies. A number of SALT Cymru survey respondents reported that they did not use SALT despite mentioning in conversation that they used technologies such as spellcheckers and OCR relatively frequently.

One of the challenges when encouraging knowledge transfer between academia and business is therefore that of overcoming the language barrier formed by the specialized use of terminology in academia, which is unfamiliar to businesses and end users.

In addition to an audit by the project's researchers, information was also collected by means of a bilingual online survey. Participation was by invitation, and responses were used to identify potential focus group participants. A total of 48 complete forms were received, which was an encouraging number considering its narrow scope. Respondents included users, developers and potential developers of SALT, and the sample included a wide range of interest in various SALT subfields.

## Survey Results

The survey results provided a valuable insight into the SALT developed in Wales, with most types of SALT being represented, and no specific SALT type dominating research. Accessibility issues and the linguistic needs of a bilingual country were highlighted, as were the improvement of interfaces between users and objects or information, and of the identification of what constitutes relevant information to the user, and the meaningful categorization of information. Pure SALT development work in Wales appeared to be relatively fragmentary, geographically dispersed, and was in general accomplished by individuals or small teams. Associated SALT development within Wales appears to have a somewhat stronger research base.

Of those who responded:

- 40% stated that they were SALT developers
- 31% stated that they were prospective SALT developers, who might be interested in developing such technologies in the future

- 29% stated that they were users or prospective users of SALT: they did not develop SALT and had no intention to do so in the future, but were interested in the technology from a user perspective

A wide variety of SALT areas were mentioned by respondents. However these were the areas which appeared to be of consistent interest to developers, potential developers and users.

- Text-proofing tools
- Speech-enabled technologies (whether speech synthesis as a module, or speech-enabled communication aids as finished products)
- Speech recognition
- Intelligent web searching
- Keyword and trend spotting from text
- Machine translation

It would therefore appear that current SALT development in Wales is aimed at:

- the accessibility needs of disabled users
- the linguistic needs of a bilingual country
- the improvement of interfaces between users and objects or information
- the improvement of the identification of what constitutes relevant information to the user
- the meaningful categorization of information

## Interviews, investigations and findings

In addition to the research undertaken by the SALT Cymru project's own researchers, the advice and opinions of experts working in the field of SALT were sought through the medium of face-to-face interviews. A sample of

individuals actively involved in SALT R&D from across Wales was selected, and a list of key players was compiled, together with an overview of their interest(s) in SALT.

Interviews were carried out with individuals in internationally regarded laboratories researching the state of the art in the discipline. These interviews reflect the wide-ranging nature of SALT. Key findings included the importance of multimodality to future developments, and the need to nurture a knowledge base from the earliest opportunity, i.e. in schools in addition to universities. A summary of these interviews can be found in the SALT Cymru Feasibility Report (see <http://www.saltcymru.org>).

To further investigate the state of the art, the annual LangTech international conference was attended. The conference provided a broad overview of a wide range of SALT technologies in research and commercial contexts, and addressed future directions in international SALT development. A focus group with professional translators, held in Wales provided further opportunities to examine the needs of an important group of SALT users and their specific needs and priorities.

The project discovered a significant appetite for speech and language technologies (SALT) within Wales, both by end users and amongst SMEs that are currently developing such technologies or which have an interest in doing so in future. Investment, a building up of the knowledge base, and an increase in awareness activities were seen to be required in order that this sector of the economy may capitalize on Wales's privileged position as a bilingual nation and grow to fulfil its full potential in exploiting worldwide markets.

An important part of the feasibility report was the evaluation of relevant open source software and standards. Particular attention was paid to the potential of resources to be enhanced for use in Wales in a pre-competitive research stage, including their suitability for further development by industry in a non state-aid environment.

Five key components of SALT were examined, based on the main areas of interest highlighted in the survey results. These areas were:

- Speech synthesis
- Speech recognition

- Intelligent web searching; keyword and trend spotting from text
- Optical character recognition

Open source examples of software resources corresponding to the above were investigated, and a series of papers were published as appendices to the SALT Cymru Feasibility Report.

The software investigated were as follows:

- Moses (machine translation)
- Festival (speech synthesis)
- Sphinx (speech recognition)
- UIMA (intelligent web searching; keyword and trend spotting from text)
- Tesseract (optical character recognition)

It is important to note that a sensitive approach to licensing requirements is needed to cater for the complex and differing needs of private sector companies. Flexibility and choice of licences for individual circumstances is welcome by developers working in a commercial environment. In selecting the above resources the project found that the more permissive BSD or MIT-style licensed software were generally preferred by developers. As a project aimed to promote commercial SALT development, it was important that the licensing terms for the software should be permissive enough to be attractive to as wide a range of commercial enterprises as possible.

The LTU sees facilitating commercial developers to create commercial minority language end products as being a significant step in normalizing such languages in relation to their more resourced counterparts. Currently, the LTU is involved with commercial developers working on a number of minority language products and the catalyst for many of these collaborations has been the availability of the software with a permissive open source license.

## **Report recommendations and terms of reference**

In light of the research undertaken during the course of the project, the SALT Cymru Feasibility Report recommended that:

1. A SALT Cymru specialist interest group (SIG) be established.
2. The SIG shall have an international watching brief for SALT.
3. The SIG shall work to strengthen the research base in Wales.
4. The SIG shall draw up a prioritized programme to develop a toolkit of basic language resources.
5. The SIG shall maintain and further develop the SALT Cymru website to include a resource portal.
6. The SIG shall guide a Welsh SALT education and training programme.
7. The SIG shall address the training and communication needs of its own members.
8. The SIG shall address evaluation and quality control issues for SALT developers.
9. The SIG shall seek adequate funding to enable it to fulfil its terms of reference.

### ***The Establishment of the SALT Cymru Special Interest Group***

The above recommendations are now being put into action, as, following the report, funding was secured from the Welsh Assembly Government to finance the envisaged Special Interest Group.

Coordinated by Bangor University's Language Technologies Unit, the SIG commenced in February of 2009 and was officially launched by the Deputy First Minister of Wales on the 20<sup>th</sup> of March. Funding will run for two years, by which time it is hoped the SIG will be sufficiently mature to be sustainable in the long term.

The international watching brief has been maintained through visits to UNESCO's Infoterm offices, Austria, and the NAACLT (North American Association of Celtic Language Teachers) conference and Welsh Office in New York. A surprising amount of international interest in Welsh SALT has been noted since the SIG's inception, and interest from amongst the traditionally TV based media sector in Wales is growing as the boundaries between old and new media become blurred.

A foundation for strengthening the research base has been established by the setting up of a SALT Cymru steering group comprising of representatives from many Welsh Universities. Work is continuing on the development of the SALT Cymru website into a SALT resource portal, which will hopefully serve as a catalyst for SALT research and development in Wales. The website is to be found at [www.saltcymru.com](http://www.saltcymru.com).

A series of quarterly seminars, networking events and newsletters is addressing the communication and knowledge transfer needs of the members of the SALT Cymru special interest group, and an annual conference is currently in the planning stages.

The SALT Cymru Special Interest group also offers SALT consultation to business which is paid for by the Welsh Assembly Government. These consultation sessions are tailored to the needs of the individual enterprises, and have benefitted from the research into open source resources undertaken as part of the SALT Cymru Feasibility Report.

So far, 6 enterprises (from a target of 48) have been assisted by the SALT Cymru SIG project team, and the project is currently on track.

These include:

- An enterprise specialising in education wishing to target its products at the multilingual market
- An enterprise specializing in subtitling and translation wishing to develop machine translation capabilities
- A translation company wishing to improve proofing tools for software translation
- A software developer specialising in development for mobile devices with an interest in SALT
- A software developer with an interest in proofing tools and translation memory systems
- A SALT developer interested in the SALT capabilities of mobile devices

Interestingly from a minority language perspective, the majority of these enterprises wished to apply SALT to Welsh-language business opportunities, with a number using the language as a step towards multilingualism in their products that would facilitate entry into international markets in the future.

Some of the consultations have resulted in Knowledge Transfer Partnerships (KTPs) between Bangor University and Welsh Enterprises. These are partnerships between an enterprise and a university who team together to accomplish a project that will be of benefit to the company, but would not have been feasible without the KTP. The enterprise receives 60% funding towards the employment costs of an 'associate', a suitable graduate identified and supervised throughout the course of the project by the university whilst working on location at the enterprise. The university also receives funding from the government for its part in the KTP, which contributes to the ongoing survival of the language Technologies.

## **Conclusion**

The SALT Cymru Feasibility Report successfully led to the securing of funding to establish a SALT special interest group for Wales by highlighting the increasing economic significance of SALT globally, and identifying the potential of the existing knowledge base in Wales whilst drawing attention to its precarious future. The feasibility study also highlighted the fact that SALT is not a label that is recognised within businesses that do not have strong links to academia. This has led to refocusing discussions between SALT team member and enterprises on individual relevant technologies rather than on SALT as a whole, as enterprise are understandably more interested in technologies useful to themselves than on a complete overview of SALT as discipline.

This paper has not focused on a specifically minority-language project as neither the feasibility report nor the special interest group project are specifically minority language projects. However it is relevant to a minority language discussion as it illustrates that when enterprises are able make a business case for minority language services and products, economic stimulus projects such as these may validly address the issues facing a minority language such as Welsh. This is seen by the LTU as a step towards normalising the Welsh

language by funding aspects of its development through funding streams not specifically set aside for minority language or 'cultural' purposes.

In fact, most of the companies so far assisted that have been interested in SALT technologies for economic reasons have been interested in its application for the Welsh language. This interest has primarily been driven by the demands of the public sector where it is legislated that, 'where practicable', Welsh and English should be treated on the basis of equality. By increasingly making Welsh-language support 'practicable' within the information and communication systems of today, SALT Cymru is playing an important role in the continuing relevance of Welsh as a modern language.

For more information about SALT Cymru, including the full report and appendices, see [www.saltcymru.org](http://www.saltcymru.org).

## **Bibliography**

- Varile, Giovanni Battista, and Antonio Zampolli. 1997. *Survey of the state of the art in human language technology*. *Linguistica computazionale*, v. 12-13. Cambridge [England]: Cambridge University Press.
- Jones, Rhys, and Prys, Gruffudd et al. 2008. *Project Closure Report for SALT Cymru*. Bangor University. On-line version available at: <http://saltcymru.org/english/adroddiadsalt.html> (accessed 24.08.09)

