

Category 7: Lexicography for specialised languages – terminology and terminography

www.focal.ie TERMINOLOGY PROJECT

Úna Bhreathnach, Caoilfhionn Nic Pháidín, Donla Uí Bhraonáin
FIONTAR, Dublin City University

ABSTRACT

This paper presents an EU-funded project for Irish and Welsh which will be accessible from 2007 at www.focal.ie. This is a major initiative in two Celtic languages. The context and content of the Irish implementation strand are discussed here. An overview is presented of the development of terminology in Irish, and the project presented demonstrates the potential for innovation and renewal through the development of sophisticated linguistic tools for users of all levels. This paper explores some of the linguistic and editorial challenges faced by the project team. It demonstrates the need for transnational cooperation in sustaining European linguistic diversity, and a potentially pioneering role for lesser-used languages through modern communications technologies.

WWW.FOCAL.IE Terminology Project

1 Project description

This project was established by Fiontar, Dublin City University in partnership with the University of Wales Lampeter to develop high-level language tools for Irish and Welsh. It is valued at €1.5 million and it is funded by the EU through ERDF INTERREG IIIA (Ireland-Wales) for the period 2004-2007. The Irish government co-funds the Irish implementation strand through Foras na Gaeilge. The project provides access for EU citizens in the Ireland-Wales cross-border region to high-level ICT tools for Irish and Welsh, ranging from school children to sophisticated high-level users. The principal outputs are a bidirectional Welsh/English on-line dictionary and a database of Irish language terminology containing 200,000 items, also available on-line. Both of these are freely accessible through a trilingual portal in English, Irish and Welsh at www.focal.ie. This project brings two of Europe's older written languages from the perceived periphery to the forefront of technological innovation. EU funding has enabled the development and sharing of transnational expertise: innovation in language technology tools at this level cannot be resourced and developed at national level only. The international partnership forges an important link, both on symbolic and practical levels, between peripheral languages and modern technology. This paper explores the Irish strand of the project.

2. Overview of Terminology in Irish

The Irish language has absorbed and created terminology to deal with evolving social, political, ecclesiastical and technological developments for at least the last fifteen centuries. The brief overview below of the modern development of terminology in Irish will explain the starting point and context of our own project. It will show the reasons why the raw material for our project differs considerably from source data available in major languages with better resources.

As a threatened language emerging from colonialism, building a base of modern terminology in Irish was a daunting prospect. During the struggle for national independence and later after the foundation of the Irish State in 1922, its development was driven largely by the needs of the education system, and to a lesser degree to serve state requirements in law and public administration. The media was also a significant driver of development and dissemination.

The Irish government established an official Terminology Committee in 1927 to create and validate terms. This lapsed in 1939, and in 1968 the Permanent Terminology Committee was established under the Department of Education. Under the British-Irish Agreement of 1999 this committee passed to the all-Ireland body for language promotion, Foras na Gaeilge. One terminologist is employed by this agency and under-resourcing of this important function has been an issue for decades.

The system of term creation and validation in Irish works by a series of voluntary, Government-appointed, domain-specific committees, monitored in turn by a steering committee. Since 1968, 22 domain-specific dictionaries have been published by An Gúm, the state-run Irish-language publisher of dictionaries and textbooks in Irish. 7 dictionaries have been published by other bodies, including two by FIONTAR, DCU, in collaboration with the terminology committee, and over 20 further lists have been developed. Some of these have been issued in electronic format, but no integrated resource containing all available Irish-language terms existed before this project. The bulk of these 200,000 terms were developed in response to public or institutional demand.

There has recently been an increase in the demand for terminology in Irish, due to several factors. The Irish-language radio station, set up to provide radio programmes for people in Irish-speaking regions in 1972, is now broadcast countrywide and is on air 24 hours a day since 2001. An Irish-language television station, TnaG (now TG4), followed in 1996. FIONTAR, an Irish-language centre in Dublin City University established in 1993, offers 2 primary degree programmes and 2 postgraduate programmes in Business, Information Technology and Communications, all through the medium of Irish. The Official Languages Act 2003 requires all public bodies to provide a bilingual service to customers, both written and oral, and to publish annual reports and other documents bilingually. In 2005, Irish was given official working status by the EU Commission and provision will have to be made for the resulting demand for terminology.

2.1 User Profile

The user profile of terminology in Irish differs from that of major languages. The native speaker base has continued to decline all through the 20th century. Developments in industry, modern media and tourism in the heartland, or Gaeltacht, have been achieved at a high linguistic price. Only a small proportion of modern terms in Irish have general currency among native speakers, usually those LGP terms which are aired by the broadcast media.

In marked contrast to this, and not typical of speaker profiles elsewhere, generations of Irish people have learned Irish since 1922, and a sizeable minority of these use the language daily, albeit many of them in school only. There are 339,541 daily speakers in the State. Of these 33,789 live in the Gaeltacht, and 153,700 or almost half the total, live in Leinster (CSO 2002).

These L2 children of the Revival are typically well-educated, urban, many living and working on the east coast. Some thousands of daily speakers use Irish in a variety of high-level domains, principally education, media, public administration and law. This is reflected by the predominance of certain domains in the demand for terminology. Activities like sport, leisure, cooking and gardening are therefore under-represented in our project database and, conversely, business, computing, pedagogical terms and popular science terms (rather than specialised medical terms, etc.) feature very prominently. Unlike Catalan, for example, Irish was never established as a language of industry or trade, or in the health sector, major areas where education and practice have been conducted almost exclusively through English.

3 The www.focal.ie project

3.1 Perspective

The www.focal.ie project provides a solution to the difficulties of public access to Irish terminology. Before this project, sources were dispersed, and the provision was utterly inadequate, bearing in mind that study of Irish is mandatory for all Irish children while in school, normally up to the age of 18. The latest authoritative English-Irish dictionary was published in 1959. The New English-Irish Dictionary project, commenced by Foras na Gaeilge in 2003, will take some years to complete, and is limited to 50,000 headwords. This puts the need for www.focal.ie and its potential user base into perspective. We are particularly encouraged by the exponential growth in the number of searches on the Welsh dictionary (from 209 searches per month when it was launched in 2002 to 200,000 per month this year). This resource has largely displaced sales of printed dictionaries.

3.2 Project Development and Content

As inheritors of old – sometimes very old – data from a huge range of fields, we have faced some editorial challenges which will be familiar to anyone working in the development of resources for lesser-used languages. There are two broad reasons: our terms are general and our term-concept relations are looser than a terminologist would like.

Firstly, our terminology is very general. Because we are including *all* the official terminology developed by State terminological committees over the last forty years, we are drawing from an extensive list of sources. Some lists and dictionaries deal with broad subjects like computing, philosophy and home economics; there are other miscellaneous lists containing the terminology committee's responses to queries by the public.

Some entries would not normally be considered strictly LSP: verbs like 'scheme' or 'schedule' are included. This is so because the bilingual English/Irish dictionaries currently available are limited and outdated (de Bhaldraithe 1959; Ó Dónaill 1978); general language words therefore have to be created or rediscovered and publicised, especially with the growth of the translation sector.

Secondly, we have decided not to do any data modelling. Modelling is a project best carried out at the term-creation stage anyway, and would be virtually impossible with the vast quantities of old data we have inherited. However, we have included a way to create links between concepts, a kind of high-tech 'see also'.

Thirdly, we do not have a strict correlation 1 concept-1 term, as a strictly LSP collection would. We do not even have this model for the concepts in our collection which are *definitely* LSP. This situation has evolved for several reasons. Due to a seriously under-resourced information storage and publishing system in the past, it was difficult for those responsible for coining new terms to find out easily if and how a term had previously been translated. Terms were published in lists of assorted terms or in dictionaries in which they sometimes did not logically 'fit'. For instance, 'nursing home', 'drunken driver' and 'paperweight' are all included in a list of telecommunications terminology. The search system improved with the use of computing, but was still slow, complicated and unreliable. As a result, a small but significant number of terms were translated several times, and either spelled differently or translated in an entirely different fashion. Furthermore, users presented with a range of non-integrated sources found it difficult (and, often, were not qualified) to investigate and differentiate between these versions. This problem had of course been noticed but only became a major issue with the compilation and integration of data in this project. At this stage some of these terms have entered the vocabulary; others have not and can safely be deleted by the terminology committee.

In our project, it was necessary for us to include all terms which have been published or are in use. In order to deal with the issue of multiple terms for each concept, we are using an LSP

concept-based rather than an LGP headword-based approach¹. This also means that the database is reversible. We do, however, label terms as ‘preferred’, ‘familiar’ etc, and show preferred terms at the top of the term list. This will be a resource for terminologists in future, as they work their way through the data and make decisions about its status.

3.3 Users and how best to serve them

As discussed above, the www.focal.ie project encompasses all the terminology which has been produced by the Irish State terminology committee, and this covers a very broad range. Our potential users are also very varied. As there is no other accessible source for terms, we expect to have users from primary school age up to professional level searching our system.

Definitions are not included because it is commonly agreed that they are of little benefit to the general user, and that usage samples and context notes are better for demonstrating use. We have included some usage samples where possible, as is the norm in bilingual dictionaries, each phrase accompanied by a translation. From a resource perspective, this feature is an additional aspirational deliverable which will not be fully implemented within the lifetime of the current project. It is, however, of considerable interest to the editors. As in other minority languages, Irish has lost many idioms in favour of direct translation of English phrases. Usage samples transmit some of the richness and flavour of the language to the user. We feel that we can do this precisely because our terminology is not aimed at the specialist but at the general user.

Unusually for a terminology database, we decided to provide grammatical information – labels and grammatical forms. As we noted above, many of our users will not be specialists, either in the subject area or in Irish grammar. For most users, Irish is a second language, and grammatical labels and forms will assist them in using terms correctly. Therefore each term (English or Irish) has at least one grammatical note attached to it; genitive and plural forms are given for Irish nouns, and verb forms are given.

We feel that this flexible approach to the boundary between LSP and LGP, between terminology and lexicography, works well in our situation, mediating the territory between our data and our users.

3.4 Television and Radio Terminology

¹ Mechura (2006, forthcoming), also presented at this conference, deals with this issue in more detail.

An important output of this project is an expanded and updated version of the *Dictionary of Television and Radio Terminology*. This was originally published in 1996, prior to the establishment of TG4, the Irish-language TV station mentioned previously. Numerous independent production companies, most working bilingually to some degree, are also flourishing. Due to the developments in broadcasting technology since 1996, mainly from analogue to digital, and the fact that a staff of mostly L1 speakers of Irish had been working in the sector for 10 years, an expanded and updated dictionary was overdue.

Our methodology was more bottom-up than that which applied heretofore in term-creation in Irish. We acquired lists of digital-broadcasting terminology in English from the national broadcasting station, RTE, and more general broadcasting terminology from other professionals and academics. With the cooperation of TG4 in the Conamara Gaeltacht, we established a working-group of 11 professional radio and television broadcasters and film-makers who work through the medium of Irish. We collaborated in compiling a list of the Irish-language terminology in everyday use and a list of English terms for which there was no Irish equivalent. Each member of the working group systematically examined the long list of English terms – about 1,200 terms in all – and suggested Irish equivalents.

This approach was adopted to maximise acceptance of terms by professionals and their use in the sector. It is important to note that the members of the working group undertook this work with great enthusiasm and interest and that a further list of suggested terms was later submitted by the independent film company Macalla, again validating our decision to involve the end-user in ownership of the process. All of these terms, both those already in use and those newly created (approximately 1,500 terms) were submitted to the state terminology committee and most were approved, some with minor changes. This methodology was possible due to the existence of a thriving Irish-language broadcasting and film sector with a critical mass of L1 Irish speakers. It proved to be an efficient, enjoyable and cost-effective working model for term creation, which may be of relevance to future projects. It also reflects the emphasis on consultation and public mandate favoured by the UNESCO Guidelines for Terminology Policies, 2005.

4 Conclusion

Our approach to the management of a large volume of terminology for a diverse public may well be of interest to others concerned with language corpus planning in lesser used languages. We are

indebted to our international colleagues working with terminology with whom we have discussed our methodologies, and we look forward to launching www.focal.ie as the main resource for terminology in Irish. Our project aims to create state-of-the art linguistic tools for Irish and Welsh, providing our users with attractive interactive ways of using these languages effectively.

References

A. Dictionaries

De Bhaldraithe, T. (1959) *English-Irish Dictionary*, Dublin, An Gúm

Ó Dónaill, N. (1978) *Foclóir Gaeilge-Béarla*, Dublin, An Gúm

B. Other Literature

Breathnach, C. (1997) 'Gnó an Choiste Téarmaíochta agus Obair an Iriseora' in Nic Eoin, M., and Mac Mathúna, L. (eds.) *Ar Thóir an Fhocail Chruinn*, Dublin, Coiscéim

UNESCO (prepared by Infoterm) (2005) *Guidelines for Terminology Policies. Formulating and implementing Terminology Policy in Language Communities*, Paris, UNESCO