

Úna Bhreathnach, *Fiontar*, Dublin City University

This article gives a short overview of the www.focal.ie terminology project, including the contents of the database, the use of the resource since it became available online, some of the editorial work which has been done, and some of the challenges encountered.

Presentation

The website [focal.ie](http://www.focal.ie) is a new terminology database for the Irish language. It contains over a quarter of a million (270,500²) English and Irish words and terms (equivalent to 137,800 dictionary-style entries), from all areas of life. This is one of the largest databases of its kind in the world, and one of the most useful in terms of ease of use and richness of results.

The project was established by *Fiontar*, Dublin City University (in partnership with the University of Wales Lampeter) to develop language tools for Irish and Welsh. It is funded by the EU through ERDF INTERREG IIIA (Ireland-Wales). Foras na Gaeilge co-funds the Irish side. The principal outputs are [focal.ie](http://www.focal.ie) and a bilingual Welsh/English on-line dictionary (www.geiriadur.net).

Fiontar has been working on [focal.ie](http://www.focal.ie) since early 2005, with a team of up to eight. On the editorial side, most of the work involved inputting and checking 54 different dictionaries and term lists supplied by the Terminology Committee of Foras na Gaeilge, who are responsible for term creation and management for the Irish language. There was (and still is) also a considerable amount of editing work, as will be discussed below. On the technical side, the challenge was to create an accessible but highly sophisticated search mechanism, to allow the user to find the term he or she is looking for, or at least something close enough to be useful, if the exact term is not available.

¹ This article was originally published in *Translation Ireland* Vol. 17, No. 2 (2007), pp 11-18 and is reproduced by kind permission of the editor.

² All figures quoted are correct at the time of writing, but are liable to vary slightly as editorial work progresses. Statistics quoted are for the period March-October 2006.

At the end of the project, in March 2007, focal.ie will be handed over to the Terminology Committee, who will then be able to use it both as an internal term management tool and as a fast, interactive and flexible way of making terminology available to the public.

Contents

The database includes almost all of the term lists and dictionaries produced by An Coiste Téarmaíochta over the years. These cover a very wide range of subjects; the most comprehensive lists are in the areas of Trades and Technology (11,896 entries), Geology and Geophysics (13,220), Business (23,510), Geography, Planning and Archaeology (11,406), Science (17,209), Parliamentary Terms (8,929), Computing (10,864) and Religion (9,667).

Shorter lists cover areas such as Biology; Science; Geography; Religion; Education; History; Home Economics; Mathematics; Physical Education; Art and Crafts; Technical Drawing; Woodwork; Agriculture and Dairying; Astronomy; Countries, Peoples, Languages; Environment; Cooking, Food and Drink; Librarianship; Metalwork; Modular Work; Music; Plants and Animals; Physiology and Health; Shop Signs; Sport; Statistics; Telecommunications; and Television and Radio, as well as miscellaneous term lists.

School subjects predominate, reflecting the sociolinguistic context of the language revival policy. From 1968 until recently, when it became part of Foras na Gaeilge, the Terminology Committee functioned under the auspices of the Department of Education.

Only a relatively small number of terms belong to lists which largely comprise existing, traditional, terminology in such areas as Literature and Criticism (3,812 entries); Old Irish Grammar (347); Philosophy (2,214); Proper Names (650); Plant and Animal Names (2755); Sailing (3524). This is reflective of the domain loss suffered by the language over the years and, perhaps, of the fact that most speakers of Irish are not native speakers.

Use of the resource

Since the site was made available on a trial basis in March 2006, and especially since its well-publicised launch in September, the number of users and searches has climbed steadily, reaching 196,000 individual searches in October. Clearly such a service addresses a real need. Although we cannot tell who the users are, we have gathered some information through feedback and looking at the kinds of searches being made. This discussion is based on more than 500,000 searches made between March and October 2006.

It is clear both from the types of searches that are being done – often lists of complex terms to do with a specialised theme – and from feedback we have received that a considerable number of translators use focal.ie. They are probably the biggest single category of user.

Adequacy

There is a facility on focal.ie to request terms, and this gives an indication of the types of unsuccessful search. Many of the terms requested are new terms or reflect current interests or current affairs:

Google whack, burqa, whitewater rafting, landing page, winery, moolvee, blog

New terms such as these, once created by the Terminology Committee, can of course be added to the database.

Feedback from users has concentrated on two shortcomings of the database: the lack of general language terms, discussed below, (many users have suggested incorporating *Foclóir Uí Dhónaill* into the site) and the lack of placenames.

Does the terminology which *is* provided meet the needs of users? As discussed below, most searches are returning a result, which is a very positive outcome. However, we have found that 70% of English and 80% of Irish terms have never been requested, even once (and over half a million searches have been made at the time of writing). This is not surprising – there is a large amount of very specialised terminology (names of minerals and of stars, for example) in the database, which is called for only

occasionally. It is to be hoped that as new users and new requests come forward, this reserve of terms will prove useful.

The terms which are most often searched for tend to be general language words rather than domain-specific terms³. In a way, this is to be expected since such terms occur in many different contexts. It could also mean that the site is being used instead of an (online or paper) general dictionary. There is no contemporary bilingual dictionary: the latest authoritative English-Irish dictionary (De Bhaldraithe) was published in 1959, and the Irish English dictionary (Ó Dónaill) in 1978. Focal.ie therefore fills a need. Though better than nothing, this creates problems, especially for the inexpert user: the database is *not* a general dictionary and entries are not comprehensive or hierarchically ordered.

78% of all searches, including misspellings and other user errors such as entering whole sentences, returned either a precise result or a related term. Of the 823 terms which were searched for fifty times or more, only 35 yielded no direct results. Some of these were not really valid searches ('I love you', 'Happy Birthday', 'how are you'), some were misspellings ('slan', 'contibutor'), but others were general language words which would normally be found in a dictionary rather than in a terminology database ('sister', 'enjoy', 'for'). In nearly all cases, these could be found in the related terms section, though this may not always give the user the result they were looking for. 'Sister' returns

Association of Primary Teaching Sisters, Presentation Sisters, sister blocks, sister churches, sister colleges, sister hooks, teaching sister and the Seven Sisters

but nothing involving siblings. Clearly, most popular searches are returning *some* sort of answer to the user.

Some 'successful' searches, of course, are not returning the results that users want or need. A search on 'fixed', for example, gives the meanings 'secured' or 'permanent' ('fosaithe', 'buan', 'seasta', 'do-bhogtha', 'suite', 'feistithe', 'daingnithe',

³ These are the fifty most searched-for terms up to 19 October 2006: hello, performance, project, access, potential, objective, control, section, condition, area, design, contributor, initiative, report, consider, process, balance, commitment, experience, destruction, issue, consultation, review, procurement, application, website, relevant, conduct, specific, contribution, development, action, consideration, record, capacity, address, outline, sustainable, love, charge, quality, body, decision, benefit, return, request, maintain, ratio, response, model.

‘fosaitheach’, ‘ceangailte’) but not the more common meaning ‘repaired’. A search for ‘bail’ gives several meanings from agriculture and animal husbandry (‘maide’, ‘maide bactha’, ‘ceanglachán’, ‘ceanglachán cinn’) and from typewriting (‘clampóir’) but not the meaning of ‘bail bond’, the most usual one.

Part of the reason for these absences is that focal.ie is a terminology database, consisting of a large number of subject-specific dictionaries. It cannot be expected to have the scope and comprehensiveness of a general dictionary compiled from a carefully selected list of words and meanings, representative of the whole of the language. It is not based on a corpus of either language.

On the other hand, there are many entries in the database which are not, strictly speaking, terms (e.g. ‘pyjamas’, ‘mobile phone’, ‘punk’, ‘Christmas spirit’). In most languages they would not be found in a terminology database but in a dictionary. In the absence of general bilingual English-Irish dictionaries, requests for current phrases have been routed through the Terminology Committee in recent years.

There are several dangers attaching to this partial confusion of terminology and lexicography. Firstly, users are often disappointed or, worse, misled by the results achieved. As in the example of ‘sister’, above, the most common word, the one the user is most likely to be looking for (‘deirfiúr’, in this case), is omitted. Secondly, because *some* GPL (general purpose language) is included in the database, and because no comparable GPL dictionary database exists, users may start to depend on it to the exclusion of real GPL sources. This means that the user misses out on the advantages of a good bilingual dictionary: a thorough treatment of common words and idioms; a compilation of uses and usages; a corpus-based record of the language as it is spoken and used.

Editorial work

Most terms (78%) being searched for are in English, which suggests that users are usually translating into rather than out of Irish. It also suggests that users are seeking to express themselves in Irish rather than to aid comprehension of Irish. This means that it is very important to disambiguate meanings when there are two or more possible Irish equivalents of an English term.

The editorial work on focal.ie involves integration of these terms according to meaning. Sometimes this is very straightforward. The English term ‘inconsequent drainage’ can refer either to a drainage system (‘sruthchóras neamhréireach’) or to a process (‘draenáil neamhréireach’). It is sufficient to label the two translations as such: ‘inconsequent drainage (system)’ and ‘inconsequent drainage (process)’.

Other examples are more complicated. The source dictionaries, although largely consistent, do not always concur on Irish translations of a given meaning of an English term. This can occur for several reasons: the dictionaries were written over many years and by different groups of experts with different priorities; the lack of a central database made checking for consistency difficult; and one dictionary may amend an earlier source. The main reason, however, is that some terms in the database are so general that they should really be treated as lexicographic and not terminological items.

A typical example is the term (or word) ‘incomplete’. Two dictionaries (*Dictionary of Science*, *Fiontar Dictionary of Terminology*) translate this as ‘neamhiomlán’. It is translated as ‘easnamhach’ in the dictionary *Library Terms*, and in a miscellaneous list with the domain note ‘librarianship’. The dictionaries *Technical Drawing* and of *Trades and Technology* give four translations, ‘neamhiomlán’; ‘easpach’; ‘easnamhach’; ‘uireasach’. What is to be concluded? It seems clear that, in librarianship, the only acceptable term is ‘easnamhach’. But can only ‘neamhiomlán’ be used for science? And what about the many other areas in which the term ‘incomplete’ is to be used?

The question arises as to whether, for a language such as Irish, whose terminology collection is very general and very comprehensive, it is more useful to have a ‘one-stop-shop’ for all language resources, or whether a distinction must be made between terminological and lexicographical resources.

Additional tasks which would greatly improve focal.ie as a resource, but which are beyond the scope of Fiontar’s project, include the addition of usage samples. It is relatively difficult to find examples of Irish language terms in use, due to their

underrepresentation on the Internet and the lack of a large, comprehensive corpus (and also because of the fact that many terms are very seldom used outside of textbooks). Usage samples, particularly for verbs and for cases – such as ‘incomplete’, above - where there are several Irish equivalents for an English term, would be particularly useful for the learner. Although it has proved beyond the scope of the project to include more than a few thousand usage samples, a partial remedy is the inclusion of an auxiliary glossary containing terms and phrases from the database of Rannóg an Aistriúcháin (Translation Section), Office of the Houses of the Oireachtas. As well as containing additional terms, this glossary sometimes gives sentences and phrases which illustrate usage, albeit in a limited, legal context.

Conclusion

The availability of focal.ie has made a substantial contribution by showcasing the enormous collection of terminology – over a quarter of a million terms – compiled and created over the years by members of the Terminology Committee. This terminology largely meets the demands placed on it, although use of the database as a substitute dictionary inevitably leads to errors and confusion. Until the outdated English-Irish dictionary (1959) is replaced in a decade or so, the reality is that focal.ie will function by default as a lexicographic resource.

The amount of traffic on focal.ie (well in excess of expectations) shows the demand that exists for high-quality language tools for Irish. It also demonstrates the willingness of the Irish public to engage interactively with new technologies. This is very encouraging for projects like the planned new English-Irish dictionary, which, as a general language resource, can expect to have far more users than a terminology database. It is obvious that such a dictionary must be available online as well as in paper format. The need for such a dictionary is clear and its availability (perhaps in the same place) would complement focal.ie by providing fuller, detailed descriptions of general language use.

References

- De Bhaldraithe, T. (1959) *English-Irish Dictionary*, Dublin, An Gúm
Ó Dónaill, N. (1978) *Foclóir Gaeilge-Béarla*, Dublin, An Gúm
Uí Bhraonáin, D. (Ed.) (2004), *Foclóir Fiontar Gaeilge-Béarla/English-Irish Dictionary of Terminology*, Dublin, Fiontar, DCU

Department of Education (1994) *Foclóir Eolaíochta/Dictionary of Science*, Dublin, An Gúm

Department of Education (1992) *Foclóir Ceirdeanna agus Teicneolaíochta: Innealtóireacht, Foirgníocht, Líniúocht, etc./A Dictionary of Trades & Technology: Engineering, Construction, Drawing, etc.*, Dublin, An Gúm