

# Localization into Irish

*Michal Boleslav Měchura*

Irish is one of Europe's minority languages, spoken by approximately 300,000 people in Ireland. As such, it has only entered the digital age fairly recently. Even though some localized software has existed in Irish for a long time, mostly in the open-source camp, localized software only became a real option for most Irish-speaking computer users in 2005 when Microsoft launched its first-ever Irish version of Windows XP and Office 2003.

This, as well as other recent developments, provides a good test case for what happens when a lesser-used language is introduced to computing. In this article, I will take a look at some of the characteristics of Irish and other Celtic languages and consider the challenges they pose for software.

## The basics

Irish (ISO 639-1 code *ga*) is a language spoken by a minority of the population in the Republic of Ireland and in Northern Ireland. Both Irish and English are official languages in the Republic of Ireland, but, in reality, most of the country's public administration, business and the media function exclusively through the medium of English. Practically all Irish speakers also speak English natively and are often more literate in it than in Irish.

Irish is not a dialect of English, nor is it similar to it. Note that there exists a dialect of English called Irish English or Hiberno-English, and this is also completely distinct from the Irish language. Irish is a Celtic language, a family which also includes Welsh, Cornish,

Ór í an ĴaeilĴe an teanĴa náiriúnta  
ir í an phríomĴeanĴa oifigiúil í.

Ór í an ĴaeilĴe an teanĴa náiriúnta  
ir í an phríomĴeanĴa oifigiúil í.

Ós í an Ghaeilge an teanga náisiúnta  
is í an phríomhtheanga oifigiúil í.

Sample text in Irish, shown first in the traditional Gaelic script, then in a slightly modernized version of the same script and finally in the Latin alphabet used for writing Irish today (Gaelic fonts designed by Vincent Morley).

Breton, Manx Gaelic and Scottish Gaelic. Of these, Irish is most similar to Scottish Gaelic, the language spoken in the Highlands of Scotland. Irish and Scottish Gaelic are to some extent mutually comprehensible, although not without effort.

Sometimes, the word *Gaelic* is used to refer to the Irish language, especially in North America. However, the term is rarely used in Ireland itself. It is also slightly inaccurate because Gaelic is actually the name of a sub-family of the Celtic languages containing Irish, Scottish Gaelic and Manx Gaelic. In some versions of Microsoft Office, Irish appears listed as "Gaelic (Ireland)." This is an error, presumably resulting from a confusion as to

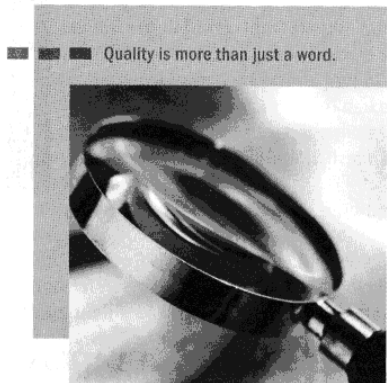


*Michal Boleslav Měchura works in Fiontar, an Irish-language teaching and research centre in Dublin City University, where he builds lexical databases. The author thanks Delyth Prys and Dewi Jones of Canolfan Bedwyr at the University of Wales in Bangor for checking his claims about Welsh.*

# Languages

the status of the three Gaelic languages as separate languages or as dialects.

In 2004 the Official Languages Act was passed in the Republic of Ireland that, like similar legislation in Canada, Wales and other countries, compels state bodies to produce large amounts of content bilingually. In 2007 Irish became an official language of the European



## ErrorSpy 4.0

The quality assurance software for translations

How can you ensure that the 5,000 technical terms of your company are used correctly and consistently in a 300 page manual?

### ErrorSpy checks:

- Proper use of terminology
- Accuracy of numbers
- Format of your texts
- Completeness of the translation
- Integrity of tags
- ... and much more

### ErrorSpy automatically generates:

- a list of errors and
- a customized evaluation of the translation quality.

### Formats supported:

- All Trados formats (RTF, TTX...)
- Translation Memories (TMX, Trados)
- Bilingual text files
- Transit files

Freelance version: 249 € / \$320

Full version: 490 € / \$630

Please add local taxes where appropriate

Server-based version also available.



D.O.G. Dokumentation ohne Grenzen GmbH

E-Mail: [info@dog-gmbh.de](mailto:info@dog-gmbh.de)

[www.dog-gmbh.de](http://www.dog-gmbh.de) · [www.multilingual-products.com](http://www.multilingual-products.com)

Union (EU). Along with the launch of localized Irish Windows and Office, this has somewhat increased the language's public profile and has also increased the demand for electronic resources such as dictionaries and terminology and for software tools that are either localized into Irish or at least capable of processing content in the language.

Modern Irish today is written in the Latin alphabet which includes the accented vowels *á, é, í, ó, ú*. Until approximately the 1950s, Irish was written in a variant of the Latin alphabet called the Gaelic script. This can be recognized by the distinctively round and wide shape of its letters. The Gaelic script was phased out with the onset of typewriters and is rarely used these days except for ceremonial or ornamental purposes.

The modern Latin alphabet of Irish is included completely in ISO 8859-1 (Latin-1), in Windows-1252 and in similar Latin-based encodings, so encoding issues have never been much of a problem for the language – unlike some of its Celtic cousins that possess more exotic accents such as the circumflex (ˆ) and the diaeresis (¨).

### Do I need to localize into Irish?

As practically all Irish speakers also understand and speak English natively, it is almost never necessary to localize software and content into Irish. Most Irish speakers are quite prepared to accept content in English and in some cases might even prefer it due to underdeveloped literacy skills in Irish.

This situation is common to minority languages in Europe. There is often another “major” language in the neighborhood, and this is the language that people mostly use for literacy-related activities such as using the computer, even though they might speak a minority language at home and with friends.

Even if localizing is, strictly speaking, unnecessary, presenting content or software in a minority language is bound to produce an effect of novelty and to make a product attractive in ways that could not be achieved through the majority language. When websites and software in Irish are launched, the event is almost guaranteed to be mentioned in the Irish-language media. Furthermore, public and state bodies in Ireland and institutions of the EU may be required by law to produce content in Irish.

Computer-related terms such as *mouse*, *screen*, *network* and *database* exist and are well-known among Irish speakers (they are *luch*, *scáileán*, *líonra* and *bunachar sonraí*, by the way). More specialized terms also exist, largely as a result of the work of the state-funded Irish Terminology Committee, but are not so well-known among the Irish-speaking public. When confronted with user interfaces localized into Irish or with texts of technical nature, some Irish speakers will inevitably be slightly lost at sea because of unfamiliar terminology.

An additional difficulty is that the syntax of Irish does not lend itself easily to constructing long noun phrases where chains of nouns modify each other, such as *database efficiency assessment methods*. When such terms are translated word for word, the result is usually very difficult for Irish speakers to parse mentally. Some careful rewording may be necessary, such as rendering the above as *methods to measure the efficiency of databases*. Unavoidably, Irish translations of English user-interface captions are going to be longer than the original, as is indeed common in many other languages.

When localizing into Irish, it is important to remember that localizers will, to a large extent, be setting new standards because there are no existing ones to follow and, like it or not, they will be either promoting or demoting the language in the users' minds. Too many incomprehensible translations on screen can turn the user away not only from the software but also from the language altogether. Vice versa, an easy-to-understand user interface is likely to encourage users to use more of the language as well as more of the software.

### Words for yes and no

One quirk of Celtic languages, including Irish, is that they have no words corresponding to *yes* and *no*. When giving answers, Irish speakers recycle the main verb from the question. For example, the question *Will you travel to Dublin?* produces the answer *Will travel* or *Will not travel* meaning *yes* and *no* respectively.

This poses difficulty for user interfaces where ready-made dialog boxes with *Yes* and *No* buttons are re-used again and again for all sorts of questions. One common strategy is to translate *yes* and *no* as verbs and

then make sure that all questions are phrased with the same verb. Exactly which verb is used differs among software packages, but it is always some generic verb that fits a number of situations. In Windows XP, all *yes/no* questions begin with the phrase *an bhfuil fonn ort...?* (literally *is there an inclination on you to...?*), and the pre-captioned answer buttons are *tá* for *yes* (literally *there is*) and *níl* for *no* (literally *there isn't*). In Microsoft Office 2003, a different phrasing with the same effect is used: questions begin with *an mian leat...?* (literally *is it a desire with you to...?*) and the answer buttons are *is mian* (literally *is a desire*) and *ní mian* (*is not a desire*). This works fine for most types of *yes/no* questions as long as localizers consistently phrase them in the prescribed way.

This solution is nothing more than a workaround. An ideal solution would be for localizers to have control of the button captions in each individual case and to translate them appropriately for each question. Unfortunately, this is almost never possible. Re-usable dialog boxes with pre-captioned buttons have been part of software architectures for decades. Until Celtic languages came along, apparently nobody considered the possibility that there may be a language in the world with no words for *yes* and *no*. Now, it is too late to go back to the drawing board, and it is the language that must accommodate the technology instead of the opposite. This situation is quite possibly the ultimate example of insufficient software globalization.

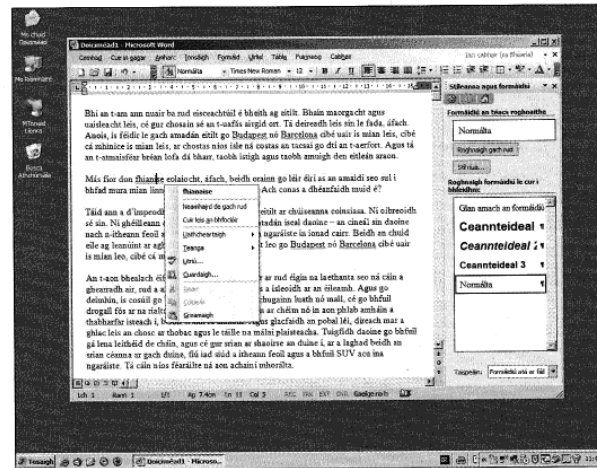
Interestingly, not all Celtic languages are affected. In Welsh, even though it doesn't have a word for *yes* or *no* either, it is acceptable to answer *yes/no* questions with the words *iawn* and *na*. These do not mean *yes* and *no* literally – *iawn* is an adjective/adverb meaning *good/well* and *na* is a particle that conveys negativity – but they are functionally equivalent to *yes* and *no*. These words are indeed the translations for *yes* and *no* that usually appear as button captions in localized software, including Welsh versions of Windows XP and Office 2003.

### Initial mutations

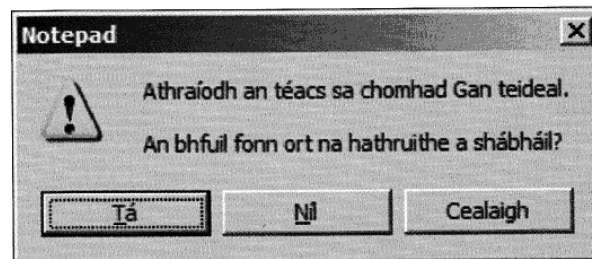
Like other Celtic languages, Irish displays a phenomenon called initial mutations. In Irish, words change not only at the end as the result of inflection, but also at the beginning on account of prepositions and other words preceding them. These changes at the beginning of words are called *initial mutations*. For example, the word for *woman* is *bean* but when preceded by the definite article *an*, it mutates to *an bhean* (*the woman*). The change from *b* to *bh* is an example of an initial mutation called *lenition*.

The existence of initial mutations produces some capitalization oddities which may look like typographical errors to non-Irish speakers. Some mutations work by prefixing a particular character to another character, such as the character *b* that under some circumstances mutates into *mb*. When the first character of the word is to be capitalized, it is the first character after the prefixed mutation that is capitalized. For example, in the phrase *ón mbean* (meaning *from the woman*), the noun must be capitalized as *ón mBean*.

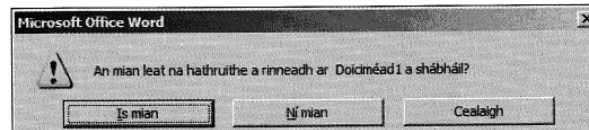
Capitalizing all the characters of a text into uppercase is problematic. Prefixed mutations should be left in lowercase, for example *ón mbean* should be capitalized as *ÓN mBEAN*. Capitalizing it as *ÓN MBEAN* would be viewed as a typographical error by most Irish speakers who cultivate literacy. Apart from



Microsoft Windows XP and Word 2003 localized into Irish, with the Irish-language spellchecker making a suggestion.



Example of a *yes/no* dialog box from Windows XP: The text in the Untitled file has changed. Do you want to save the changes? Yes – No – Cancel.



Example of a *yes/no* dialog box from Microsoft Office 2003: Do you want to save changes to Document1? Yes – No – Cancel.

instances of text where no prefixed mutations occur, this means that Irish text can never be capitalized completely.

Unfortunately, no major software package I am aware of supports the intricacies of Irish capitalization, not even the localized version of Microsoft Office, although recognizing mutation in text is relatively straightforward. Certain sequences of characters at the beginning of a word, such as *bh* and *mb*, are practically always the result of a mutation. Interestingly, this is not the case in some other Celtic languages. For example, both Irish and Welsh possess the mutation of the */t/* sound into the */d/* sound at the beginning of a word. In Irish, this is written as *dt* and is easily recognizable as a mutation. In Welsh, it is written as *d* (the *t* disappears), and there is no easy way of telling whether a word beginning with a *d* is mutated or not.

### Alphabetical sorting

The Irish alphabetical order is the same as it is in English and many other Latin-based orthographies. As is common in many

## Languages

languages that use accented vowels, accented vowels in Irish are listed after their unaccented base vowels (*á* comes immediately after *a*, for example).

Initial mutations present a problem for alphabetical sorting. In theory, mutations should be ignored when sorting. For example, *bhean* should be sorted alphabetically as if it were unmutated (as *bean*). In fact, no software application I am aware of supports this, even though recognizing mutations in text is quite straightforward. Essentially, character sequences that result from mutations should be treated as digraphs for sorting purposes, similar to the Dutch *ij* or the Czech *ch* – except that in Irish this only applies when they appear at the beginning of a word.

Initial mutations are not always a problem for sorting. For example, noun phrases in the nominative case rarely begin with a mutation. However, it becomes a large issue when alphabetizing lists of place names or other proper names that begin with the definite

article *an* and its plural equivalent *na*. The article often induces a mutation on the word that follows it. When the article is moved to the end for alphabetization (for example, *An Bhó Dhearg*, a place in Dublin, becomes *Bhó Dhearg, An*), the resulting item begins with a mutation and the mutation should be ignored by the software in order to sort the item correctly.

Sorting Irish-language names is also rather complicated. Surnames in the Irish language usually start with a preposition which indicates the gender and family status of the person. For example, in the surname *Ó Briain* the preposition *Ó* indicates the person is male. Its female equivalent is *Uí Bhriain*, where the preposition *Uí* and the mutation it causes indicate the person as female and married. By the way, these surname patterns should not be confused with their anglicized equivalents, such as O'Brian. When handling Irish-language names, it is important to respect the specific typographical conventions that pertain

to them so as not to anglicize them accidentally.

The status of these prepositions is not clear. In many other languages where this phenomenon occurs, such as *von Lemberg* or *d'Estang*, it is largely accepted that these should be ignored for alphabetical sorting purposes. In Irish, such a consensus does not exist. Some people consider them integral parts of the surname, some do not. Some even prefer to spell them uncapitalized to emphasize the point.

A further issue is that male and female versions are slightly different from each other, as we have seen in the example of *Ó Briain* and *Uí Bhriain*. Even if the preposition is ignored for sorting, the male and female forms of the same surname will not end up near each other in an alphabetically sorted list due to the mutation that most female names carry – unless, of course, the sorting algorithm is mutation-aware.

Unfortunately, no established conventions exist for dealing with these problems. Compilers of phone books and similar directories in Ireland usually “solve” the problem by avoiding it altogether and by applying a purely alphabetical sorting algorithm that does not do justice to the conceptual structure underlying Irish-language surnames.

One existing strategy for making sure that male and female versions of the same surname appear next to each other is to add a sorting key. The male form of the surname is typically chosen as the sorting key. So for example, a woman of the surname *Uí Bhriain* would be sorted as “*Ó Briain: Uí Bhriain*,” thus appearing next to a man of the same surname. This strategy, systematic as it is, has the disadvantage of being quite wordy, and, when the sort key is included in the listing, the resulting impression is confusing. The bottom line is that anybody wishing to implement a smart sorting algorithm for Irish should consult widely with literate Irish speakers before making any decisions, as no consensus currently exists.

This, after all, is a useful strategy for any aspect of computing with minority languages. Localizers and content producers are well advised to negotiate with the target language community early and often in order to avoid terminological misunderstandings and unanticipated problems such as the above-mentioned lack of translation equivalents for *yes* and *no*. **M**

## Lost in the translation shuffle?

It's time you dealt yourself a winning hand. NCS gives you a dedicated project manager, team members hand-picked for your specific project, and knowledgeable DTP specialists. All backed by customer focus and quality service. Think of it as a five-of-a-kind for your business.

Are you in? Call us to learn more. We'd love to get you in the game.



[www.ncs-pubs.com](http://www.ncs-pubs.com)

Tel.: 412.278.4590