

A women's thesaurus in Turkey

Addressing the challenge¹

Meral Alakus

In April 1994, as part of a Project on the Promotion of Education and Employment of Working People in Turkey, a Women's Documentation Center was established for the collection of material relating to women's studies, particularly in Turkey. This article explores issues related to the preparation of the bilingual thesaurus.

A Women's Documentation Center for Turkey

On 5 April 1994, the World Bank and the Government of Turkey signed an Agreement for a Project on the Promotion of Education and Employment of Working People. One section of the Project was concerned with the 'Improvement of women's employment' in all areas of work, including those known to be male dominant.² This was to involve collecting information, undertaking surveys, preparing reports, and finally forming policies on the education and employment of women. An important aspect was the establishment of a Women's Documentation Center to collect material relating to women's studies in general but with particular reference to Turkey; to compile bibliographies; to organize the material and to form a database to assist specialists working on the project, other members of the Directorate General, and anyone doing research on women in government departments and universities both in Turkey and the wider world. The Center would collect books on the subject, reference works and journals in Turkish and in foreign languages, single articles, theses, national and international seminar and conference reports, unpublished reports, individual country reports, newspaper clippings, pictures, postcards, and other visual materials. Conferences and meetings organized by the Directorate General would be recorded for the collection.

All of this, to be of any use, would need organizing. So a key element was to be the preparation of a bilingual thesaurus.

Preparing a multilingual thesaurus: methodology

The *IFLA Guidelines* (2005) describe three approaches to the development of multilingual thesauri:

1. Building a new thesaurus from the bottom up:
 - a) starting with one language and adding another language or languages to it or
 - b) starting with more than one language simultaneously.
2. Combining existing thesauri:
 - a) merging two or more existing thesauri into one new (multilingual) information retrieval language to be used in indexing and retrieval

- b) linking existing thesauri and subject heading languages to each other; using the existing thesauri and/or subject heading languages for both indexing and retrieval.

3. Translating an existing thesaurus into one or more other languages: In this case the languages involved are not treated equally and the language of the existing thesaurus becomes the dominant language.

Method 3 (translating existing thesauri) is the one most commonly adopted. But language differences often imply cultural and conceptual differences so method 1, starting with a pool of terms drawn from all the relevant languages, would almost certainly give better results as all the languages would have an equal chance to contribute concepts and conceptual relationships.

Conceptual differences in multilingual thesauri

ISO 5964:1985 (International Standard for Multilingual Thesauri) suggests that all the languages in a multilingual thesaurus should be regarded as having equal validity. It also takes the view 'that the majority of procedures and recommendations contained in ISO 2788 [for monolingual thesauri] are equally valid for a multilingual thesaurus. This applies particularly to general procedures concerning, for example, the forms of terms, the basic thesaural relationships, and management operations such as evaluation and maintenance.'

True enough perhaps, as far as it goes, but account has to be taken, as any indexer knows, of the lack of true equivalence between concepts in different languages (see, for example, Booth, 2006). A particular word in one language may well cover an area which will be slightly different in another language. (Hudon (2001)) explains the difficulties in translating concepts:

- The differences in conceptual meanings depend very much on the social, political, cultural, religious and philosophical circumstances existing in different societies.
- It is difficult to translate a term in one natural language and to be sure that it will have the same meaning in another. In a multilingual thesaurus, therefore, there might be more than one 'cross-lingual quasi-synonym' to

express a descriptor in another language. In such a thesaurus, what appears as a relation of equivalence (RT) may be closer to a relation of similarity.

- Concepts expressed by a single word in one language may be expressed by more than one term in another language.
- A term in one language may have no cross-lingual equivalence in another language.

Hudon (2001) describes the five degrees of cross-lingual equivalence as follows:

- Exact equivalence: *Sun = Soleil*
- Inexact equivalence: *Alcohol Education = Education Antialcoholism*
- Partial equivalence: *Nuts = Noix*
- Single-to-multiple equivalence: *Readability and Legibility = Lisibilité*
- Non-equivalence (with no corresponding synonymy) *Latchkey Children = ?*

There is also the problem of asymmetric equivalence. For example, the English 'education' in French may be 'éducation', 'enseignement' or 'formation'. 'Teaching' would always be 'enseignement' whereas 'instruction' in English in French would require both 'enseignement' and 'instruction'. It is difficult, to say the least, to maintain symmetry in a multilingual thesaurus. The *IFLA Guidelines* (2005) give some suggestions as useful to the back-of-the-book indexer as to someone constructing a thesaurus:

- The descriptor (or term) should preferably consist of a noun or noun phrase.
- Verbal nouns (i.e. gerunds or infinitives) are acceptable if their use is customary in the language under consideration.
- The use of the plural or singular form of nouns should follow the conventions of different languages where these are standardized: otherwise the same rules should be used for all languages.
- Adjectives without a noun, verbs and initial articles should be avoided except where a given language has a different convention.
- Terms should wherever possible be given in the script of the language concerned.
- Capitalization, punctuation, diacritics and special characters should follow the convention of the language in question.
- For spelling and alphabetization, the practice of the language in question should be followed.
- For proper names [institutional] there are various possibilities:
 - to use the proper name in one of the languages for all languages
 - to use the proper name in each language that has the name, and the proper name in the original language for those languages that lack it
 - to use translations as far as is sensible and possible. An example of this would be 'International Committee of the Blue Shield (ICBS)', 'Comité International du Bouclier Bleu (CIBB)', 'Internationaal Comité van het Blauwe Schild'.

Cross-language retrieval

For cross-language retrieval the user should be able to retrieve documents in any relevant language by using a term in his/her own language. Peters and Picchi (1997) define and analyze the advantages and disadvantages of different cross-language information retrieval (CLIR) techniques as outlined below.

Machine-translation techniques

Full machine translation (MT) systems are still far from achieving high-quality results and so are not a realistic answer to the problem of matching documents and queries across languages. The translation of entire collections of documents into another language is also very expensive. Research has therefore concentrated on ways of translating the *query* into the language(s) of the documents.

Knowledge-based techniques

- *Dictionaries*: dictionaries would seem an obvious tool for matching query to document. However automatic machine-readable dictionary (MRD) query translation has been found to be of limited effectiveness for three main reasons: the absence in general purpose dictionaries of specialized vocabulary, the presence of false translations, and a failure to handle compound terms.
- *Multilingual thesauri*: by contrast, the use of multilingual thesauri has been shown to give results as good as the use of thesauri for monolingual systems. The user is normally able to use a term in his/her language to find the same concept in a document in another language.
- *Ontologies*: a general-purpose multilingual ontology is being developed in the EuroWordNet Euro project. WordNet is a multilingual database which represents basic semantic relations between words for several European languages (Dutch, Italian, Spanish and English). For each of the languages involved, monolingual word-nets are being constructed maintaining language-specific cultural and linguistic differences.

So thesauri and ontologies are promising but:

they are expensive to build, costly to maintain and difficult to update. Language differences and cultural factors mean that it is difficult to achieve an effective mapping between lexical or conceptual equivalences in two languages; this problem is greatly exacerbated when several languages are involved. It is necessary to build some kind of interlingua to permit transfer over all languages; it is to be expected that the trade-off for multilinguality will be the loss of some monolingual specificity (Peters and Picchi, 1997).

Corpus-based techniques

Another approach to solving the problem of translations would be the corpus-based techniques in which information about the relationship between terms is obtained from observed statistics of term usage. Corpus-based approaches analyze large collections of texts and automatically extract the

information needed to construct application-specific translation techniques. These texts consist of 'parallel' or 'translation equivalent' sets, or they could be 'comparable' or 'domain-specific' texts, trying to bring together the terms with equivalent meanings in languages. The problem with using parallel translated texts is that they are very much domain specific and costly to produce. A new thesaurus would have to be built to perform retrieval on every new topic; it is also unclear how well this method can adapt to searching a large heterogeneous collection. Instead of 'translation equivalent' sets, there has been a growing interest in 'comparable corpora' or 'comparable document' collections, in which documents are aligned on the basis of the similarity between the topics rather than translation equivalence.

Constructing the Turkish Women's Thesaurus

All of the issues discussed above were very much to the fore when it came to constructing the Turkish Women's Thesaurus. It quickly became clear that, rather than using a general and universal classification scheme, it would be easier and more appropriate to design a special classification system and a thesaurus for specific and/or marginal subjects. For the purpose of indexing and retrieval of the documents in the Center, two women's thesauri in the English language were analyzed.

- *Canadian Feminist Thesaurus Féminist du Canada*. Compiled by the Canadian Women's Indexing Group (CWIG)/compilé par de Groupe d'analyse de documents féministes canadiens (GADFEC). The Ontario Institute for Studies in Education/LInstitut d'études pédagogiques de l'Ontario. 1990. This is an English and French thesaurus using a standardized vocabulary designed to help in finding or organizing written materials by, for and about Canadian women, the Canadian Women's movement and Canadian women's studies.
- *A women's thesaurus: an index of language used to describe and locate information by and about women*. National Council for Research on Women (US) (1987.) Mary Ellen Capek's thesaurus is a listing of over 5,000 terms and remains an important work in this field.

Although there were a lot of discrepancies in both of these thesauri in relation to the Turkish society, history, education, traditions, religion, women's rights and the like, they made a useful starting point. Other examples of women's thesauri are:

- *Thesaurus on women and the family*, published by the National Women's Education Centre in Japan around the end of 1980s
- *Women in development thesaurus*, a joint project of the Indonesian Institute of Science and UNICEF, published in 1991
- *Vrouwen thesaurus*, The International Archives of the Women's Movement (IIAV), Amsterdam, 1993
- *Feministischer Thesaurus*, FrauenMediaTurm. Germany, 1994.

There are also a number thesaurus projects on women in progress in, for example, Austria, Spain, Italy, and South Africa.

Selecting the terms

- The *Canadian Feminist Thesaurus* and *A Women's Thesaurus* were used for the initial selection of descriptors.
- Additional descriptors were selected through cataloguing the documents available at the Center. Reports written in the directorate and journals in the fields of social studies, history, and education relating to women were searched in the National Library. More subject headings were compiled from the titles and abstracts or by scanning these documents and articles.
- A bibliography of books on women was prepared. Some indexing terms were also picked up from titles of these materials.
- Each term was written on a 7.5 × 12.5 cm index card as it was selected, and filed in alphabetical order. The source of each term was recorded on the cards.
- The compiled list of headings was reviewed by subject specialists working in the directorate.
- The descriptors derived from the original thesauri and from reading and cataloguing the materials, and those suggested by the specialists, were then grouped under 17 broad headings. Two class numbers were left free for possible new subject groups, bringing the total to 19 classes.
- The descriptors under each group were again fine-grained into smaller groups forming ontologies. Again some subclass numbers were left free for future additions.
- Under each group the descriptors and their broader/narrower/related terms were identified. Each entry was written on a separate card giving its BT/RT/NT relationships as assigned.

A separate entry was made for each BT/RT/NT and *see* reference, citing the relationships among these terms.

- They were then put into a single alphabetical order, including broader, related and narrower terms as well as the *see* references for non-preferred terms. Having this single list helped in the checking of new terms to prevent duplicate entries and to provide consistency in the syntax of the terms.
- Facets (that is, characteristics for classification purposes) and the terms under each facet were coded with class numbers so that the thesaurus could function as a classification system for the Center's collection. So far as possible, a hierarchical system was adopted for use in organizing the different types of materials on the shelves, in vertical files and digitally.

A copy of the faceted structure, which resulted, can be obtained from the author (meral_alakus@yahoo.ca).

Handling the cross-language problems:

- Terms from the two languages were put into a pool. English terms were mainly taken from the existing thesauri and translated into Turkish; Turkish descriptors derived, as described above, from Turkish materials, were translated into English. Each term was a 'one-to-one'

translation. When necessary, phrases were used to translate a single word.

- An important problem in a Turkish thesaurus is related to the changes that took place and are still taking place in the Turkish language following the establishment of the Turkish Republic. Some of the Arabic and Persian words used by the educated elite during the Ottoman Empire or in the early years of the Republic have been replaced by newly created words in Turkish.
- Old words have often been replaced by terms borrowed from French, English and other western languages. So there were *see* references for some descriptors for words borrowed from Western languages. For example:
 - *Haberleşme* (old Turkish still being used) *Komünikasyon* (communication) *İletişim* (new Turkish)
 - *Sosyal Bilimler* (social sciences) *Toplum Bilimleri* (new Turkish)
 - *Müracaat Kitapları* (old Turkish) *Referans Kitapları* (reference books) *Başvuru Kitapları* (new Turkish)
 - *Edebiyat* (literature, in old Turkish but still widely used) *Yazın* (new Turkish)
 - *Telefon* (in use) *Bilgisayar* (new Turkish) for computer, used as 'kompüter' only for a short while.

In recent years there has been a strong tendency to provide words in Turkish instead of borrowing from other languages. This creates more synonyms, thus more *see* references.

Old versions once they have clearly dropped out of use must be deleted.

- Another problem was words that did not carry exactly corresponding meanings in Turkish or in English. For example:
 - 'Common law marriage' was translated as 'living together' (*Nikahsız yaşamak*). However this just means quite literally living together, as there are no laws in Turkey pertaining to the rights of partners living together as is the case in North America under common law.

İmam nikahi, on the other hand, meaning 'religious marriage', is not a legal or a required ceremony in Turkey. It can only be performed privately after the official marriage has taken place, and only if the family requests it, by an *imam*, a religious leader. Legal official marriage can be executed only by an assigned civil servant. Religious marriage gives no rights to either partner, so women usually are the disadvantaged party of such a union. There are still many rural families who despite the legal requirement for a civil marriage are living under *religious marriage* alone

Thesauri and indexing: the state of the art in Turkey

Turkish thesauri are virtually non-existent with only a few thesauri constructed for special in-house purposes. ULAK-BİM (Turkish Academic Network and Information Center) uses the following thesauri: MESH (medicine), JICST (engineering) and CABY (veterinary sciences). Descriptors are directly translated from English into Turkish, and additional terms in Turkish are added on the recommendation of subject specialists and are then translated into English. These added

entries are compiled and stored separately under another field called 'Free Index Terms'.

The treatment of non-Turkish names obviously raises problems. Some names are kept in their original form but others may appear in a completely different version. For example, in Turkish, Germany is *Almanya*; India is *Hindistan*; London is *Londra*. *Paris* and *New York* stay the same. UN (United Nations) becomes BM (*Birleşmiş Milletler*), and EU (European Union) is AB (*Avrupa Birliği*); NATO remains the same. Jean-Jacques Rousseau might be found under *Jan-Jak Russo*, Karl Marx as *Karl Marks* since there is no 'X' in Turkish. In recent years the tendency has been to follow the original form of an author's name.

Personal or other names in Turkish including letters such as ğ / ö / ü / ş / ç, as in *Ağaoğlu* or *Güçlü* or *Özgür* cause problems in both writing and filing together with other languages, databases and the Internet. How to resolve this problem is currently under discussion in Turkey.

General rules concerning semantics and syntax were set for descriptors to be used in the *Women's thesaurus* for the Turkish section. However these rules are not officially recognized and there is no official standardization for thesaurus construction or indexing. Problems regarding nouns, noun and adjective phrases, verbal nouns, singular or plural terms, no adjectives standing alone, capitalization, punctuation, spelling and so on have still to be standardized. Indexing as such has never been addressed academically or professionally. Some journals in social sciences and humanities are being indexed, but no thesauri exist for these, so there is no consistency of descriptors used. The *Women's thesaurus* was conceived as work in progress, to be developed and refined through use, with new descriptors being added and some changed or even deleted. It was essentially an in-house tool for classifying and cataloguing the various types of materials in the collection. There were discussions with subject specialists, but they had no background in indexing or thesaurus construction. Obviously subject specialists have a part to play, but thesaurus construction is a team project which must also involve librarians, lexicographers, indexers and the user.

For Turkey, the immediate challenge is to ensure that standards are set as soon as possible to cope with lexical and conceptual problems for thesauri construction and indexing. This would be a major step towards the classification of material relating to Turkey and its retrieval both in Turkey and in the wider world.

Machine versus the human mind

For those who think that the day has come when we can leave it all to the computer, I would urge them to think again. My brief overview of different types of retrieval techniques in multilingual mediums demonstrated just how limited are the uses of machine-based techniques at the moment. One should never overlook the fact of human endeavour and contribution not only in the retrieval of information but in every aspect of the translation process in multilingual and monolingual thesauri construction and any indexing activity. Machines and technology are there to facilitate, to help ensure consistency and to make things faster. Nothing has emerged as yet to replace the human mind.

The role of the thesaurus

For indexing of periodicals and other collective works such as monographs written by several authors or conference proceedings, in which contributors may have used variant terminology when dealing with the same concepts, the use of a thesaurus is virtually a necessity and the only effective means to achieve consistency in the choice of terms and the display of relationships in an organized and helpful manner. (Wellisch (1996))

Work on the Turkish Women's Thesaurus has demonstrated this to the full.

Notes

1. This article is based on a paper presented at the ASI-IASC/SCAD Joint Conference in Toronto, 15–17 June 2006.
2. See Appendix for a list of milestones in the history of women in Turkey.

Meral Alakus, a freelance information consultant, is a graduate of University of Illinois Graduate School of Library and Information Science. She has worked in university, newspaper and private libraries in Turkey as well as in Canada at Dalhousie University Libraries. For the last ten years she has been principally interested in bilingual applications of indexing and thesaurus construction, with particular reference to Turkish. Her aim is to raise awareness of the need for form standards and regulations in Turkish. She can be contacted at meral_alakus@yahoo.ca.

Bibliography

- Booth, Pat F. (2006) Translating and indexing: some thoughts on their relationship. *The Indexer* 25(2), 89–92
- Hudon, Michèle (2001) *Relationships in a multilingual thesauri*, in Carol A. Bean and Rebecca Green (eds.) *Relationships in the organization of knowledge*, pp. 67–81. London: Kluwer Academic.
- IFLA. (Apr 2005) *Guidelines for multilingual thesauri*. Working Group on Guidelines for Multilingual Thesauri Classification and Indexing Section. Gerhard J. A. Riesthuis et al. <http://www.ifla.org/VII/s29/pubs/draft-multilingualthesauri.pdf>, accessed on 21 March 2006.
- ISO 2788: 1986. *Documentation – Guidelines for the establishment and development of monolingual thesauri*. www.collections.canada.ca/iso/tc46sc9/standard/2788e.htm and www.collectionscanada.ca/iso/tc46sc9/standard/5964e.htm, accessed on 21 March 2006.
- ISO 5964: 1985 *Documentation – Guidelines for the establishment and development of multilingual thesauri*. www.collections.canada.ca/iso/tc46sc9/standard/5964e.htm#1, accessed on 21 March 2006.
- Olson, A. Hope. Linking a women's thesaurus and DDC: the variables. www.ualberta.ca/~holson/femclass/variables.htm, accessed on 11 March 2006.
- Olson, A. Hope and Ward, Dennis B. (1999) Finding spaces for feminism in traditional library classification: Dewey Decimal Classification, Windows and the World Wide Web. www.ualberta.ca/~holson/femclass/, accessed on 8 March 2006.
- Peters, Carol and Picchi, Eugenio. (May 1997) Across languages, across cultures: issues in multilinguality and digital libraries. *D-Lib Magazine*. www.dlib.org/dlib/may97/peters/05peters.html, accessed on 29 February 2006.
- Wellisch, Hans H. (1996) *Indexing from A to Z*. Bronx, N.Y.: H.W. Wilson.

Appendix I: Some important milestones for women in Turkey

Ottoman Empire (1299–1923):

- 1843 Women accepted for the first time in the medical school to become licensed midwives.
- 1847 Equal inheritance rights for men and women accepted by law.
- 1856 Buying and selling of women as slaves prohibited.
- 1869 First women's weekly journal (*Terakk-i Muhadderat*) published.
- 1869 Law on general education imposing compulsory education for girls published for the first time.
- 1870 First teachers' training college (*Dar-ul Muallimat*) for women opened.
- 1871 Ottoman Civil Law (*Mecelle*) published requiring civil marriage to be performed by a state officer (rather than by men of religion), with a legal age of marriage of 18 for men, and for 17 for women. The law also banned marriages where force was applied to either party.
- 1876 First Constitution (*Kanun-i Esasi*) imposed elementary school education for girls.
- 1897 Working women given the rights of paid workers.
- 1913 Women admitted to work as state employees for the first time.
- 1914 Women allowed to work in trade and to own small business.
- 1914 A higher educational institution (*Inas Darülfünunu*) opened for girls.
- 1921 University became coeducational.
- 1922 Seven women students registered for medical education for the first time at the Faculty of Medicine.

Republican Era – 1923–

- 1923 29 October. After the founding of the Republic of Turkey, reforms were realized and many laws were passed to encourage women to take part in public life and for the general betterment of the social status of women.
- 1924 3 March. National uniform education law, which aimed to secularize education, to ban all religious schools and courses, and to place elementary and secondary education under the responsibility and administration of the National Ministry of Education. Education became coeducational and on equal terms.
- 1930 Maternity leave for mothers accepted.
- 1934 10 June. Directorate of Technical Training for Girls established to provide vocational education for girls.
- 1934 5 December. Women given the right to vote and to be able to stand for election.
- 1935 8 February. 17 women elected to the Grand National Assembly.
- 1965 The sale and distribution of birth control devices permitted and abortion legalized.
- 1966 22 December. ILO treaty of 1955 signed and the principle of equal pay for equal work accepted for male and female workers.

- 1985 Turkey signed CEDAW (Convention for the Elimination of All Forms of Discrimination Against Women).
- 1990 Directorate General on the Status and Problems of Women established.
- 1995 Purple Roof Shelters Foundation for Women, which gives consultancy services to abused women, opened its first shelter home.
- 23 June 1998 The Court of Constitution annulled the law on adultery for women on the grounds that it was in conflict with the constitutional principle of equality.
- 1998 Women entitled to fill in their tax forms independently of their husbands, who were considered as the head of the household.
- 2001 New civil law approved in the Grand National Assembly.
- 2002 Women given the right to ownership of property acquired during marriage.

An/the/an index of confusion

Phillip O'Leary, Boston College

For those working with the Celtic languages, the definite article looms large. The problems it creates are in no way balanced by the fact that the Celtic indefinite article creates no problems at all because it doesn't exist. I recently finished coediting the two-volume *Cambridge history of Irish literature*. After five years of work, I and my colleague Margaret Kelleher saw the end in sight, particularly when Cambridge announced they would provide the index. That index, imposing in its length and detail, duly arrived. With its arrival, the end of the project quickly receded from our field of vision, to be replaced with another six months of work – almost all of it the result of the Irish definite article (or articles, as will soon become clear).

That definite article in the singular is *An*, and when we looked at the proofs of the index for the first time and saw that the letter 'A' went on for several pages we knew at once what had happened — entry after entry began with *An*. All would need to be rerouted to their proper places under the initial letter of the following noun.

An vs. an, and worse

Nor was the job to be that simple. First of all, there were also a few entries misplaced under the English definite article 'An' and under the Irish interrogative particle *An* (as well as titles in the interrogative misplaced under the initial letter of the word following this particle in various titles).

Moreover, Irish has a vocative particle *A*, which created further confusion for the indexers, who understandably enough saw it as an indefinite article and accordingly ignored it as they alphabetized those entries. Needless to say, the plural form of the definite article, *Na*, created its own chaos down the road in the 'N' section of the index. And, to add to the confusion, since this was a history of Irish literature from its beginnings, it, and its indexers, had to deal with earlier forms of the language (and of the definite article, which can appear as – and this list is by no means complete – *in*, *int*, *ind*, *a* in the singular and *ind*, *in*, *int*, *inna*, *na* in the plural).

All bad enough, but it gets worse.

Like all the Celtic languages, Irish is characterized by mutations of the initial letters of words caused by the words

or particles that precede them. Thus the feminine definite article changes *b* to *bh*, *c* to *ch*, *d* to *dh*, *f* to *fh*, *g* to *gh*, *m* to *mh*, *p* to *ph*, and *t* to *th*. What then does one do with an entry on 'the cow' – *an bhó*? Should it be under *b*? *bh*? *b(h)*? The masculine article is less problematic, although it does prefix *t* to words beginning with a vowel, leaving the problem of where to put 'the father' – *an t-athair*.

Incidentally, that hyphen is a useful device here as other Irish verbal particles can also create troublesome mutations. *Ar nathair* can mean either 'our father' or 'our snake,' perhaps explaining St. Patrick's eagerness to get those confusing serpents off the island before they created major theological problems for his mission. The hyphen solves this problem, but can leave one wondering whether *an t-athair* should be filed under 'T.' The proper listing would be '*athair*, *an t*'.

Feminine/masculine awkwardness

The feminine singular article has no effect on a following vowel, but does, in certain cases, prefix *t* to nouns beginning with an *s*. (Unfortunately, the Irish word for 'ant,' *seangán*, while beginning with an appropriate *s*, is masculine, denying us the convenient mnemonic *an t-seangán*.) The plural article in both genders prefixes *h* to nouns beginning with a vowel, creating a new problem since Irish really doesn't use *h* except to indicate this mutation (and for a handful of often exotic loan words like *héileacaptar* (helicopter), *hipirglycémie* (hyperglycaemia), *holograf* (holograph), or *homaisiogót* (homozygote), none often heard in pubs. Thus, for all practical purposes, the *h* listing in a careless Irish index could include nothing but misplaced plural nouns.

Still, for all these potential pitfalls, Irish indexers can take comfort that they don't have the problem faced by their counterparts in Wales: the Welsh definite articles *y* and *yr* could create a situation whereby readers could leaf through and index of the first 24 letters of the alphabet before meeting up with a noun at all.

This article first appeared in *Annals of Improbable Research*, July–August 2006.