

Entry Structures and Phonetics Policy Report

TASK: To design structures for every category of entry in the dictionary; to establish a policy for handling phonetics.

DELIVERABLE: (1) Entry structures: a set of policies, tested on real lexical items, for handling every category of entry, both lexical entries (all word classes, idioms and multiword phrases, etc.) and encyclopaedic entries (biographical, geographical, cultural, etc.) (2) Phonetics: a complete set of policies for handling phonetics.

Table of Contents

Plan of the Deliverable	2
Part 1 : NEID Entry Structures.....	3
1. NEID approach to Entry Structures	3
2. Major entry layout and content decisions outstanding	3
2.1 High-level structural division of lexical entries	4
2.2 Treatment of multiword expressions and derived forms	4
2.3 Policy on linguistic labels in the dictionary and database	5
2.4 Function word entries: content, style and layout.....	5
2.5 Irish regional variants in target-language items	6
2.6 Use of literary Irish as alternative translations	6
2.7 Strategies for handling lexical gaps in the target language	7
3. Entry Structures implicit in the DTD.....	7
3.1 Appendix 3: Introducing the DTD	8
3.2 Appendix 4 : DTD inter-element relationships	8
3.3 Appendix 5 : Anatomy of a complex entry – <i>box</i>	8
3.4 Appendix 6 : Valency menus for the STR field	8
4. Entry Structures: Conclusion	8
Part 2: NEID Phonetics Policy	9
1. Options and recommendations: the print version of the dictionary	9
1.1 Should the dictionary show phonetics at all?	9
1.2 Where should the phonetics appear?	9
1.3 In what form should phonological information be given?	9
1.4 How should dialectal variations be handled?.....	10
1.5 Recommendations on phonetics policy in the print dictionary	11
2. Options and recommendations: the electronic version of the dictionary.....	11
2.1 Should the electronic version show a written representation of Irish phonetics?	11
2.2 How should audio pronunciations be generated?.....	11

Plan of the Deliverable

This Deliverable consists of the following sections:

1. Report : Part 1 (Entry Structures) and Part 2 (Phonetics Policy)

This Report sets out the issues involved in defining entry structures for a bilingual dictionary, relating this deliverable to the work being done on the DTD (Document Type Definition: Task J) and on developing the sample entries (Task E). It also includes a proposed phonetics policy for the NEID.

2. Appendices

Appendix 1 : Six Entry Types

This appendix contains six disparate entries (*box*, *efficient*, *gravely*, *however*, *e.g.*, and *Cork*) chosen to illustrate points made in this Report.

Appendix 2: Major Structural Options

This appendix sets out two versions of the entry for *box*: Version A in a layout initially dependent upon the senses of the lemma, and Version B in the conventional layout on the basis of the parts of speech of the lemma. The objective of this is to clarify the discussion of these two approaches in the main Report.

Appendix 3 : Introducing the DTD

This appendix contains a commentary on some of the details of the DTD, together with a listing of all elements currently in the prototype DTD. Its objective is to give an overview of the extent and power of the DTD.

Appendix 4 : DTD Inter-element Relationships

This appendix takes the form of a graphic representation of the Sense Block (SenBlk) element from the DTD: the objective is to clarify relationships within the DTD.

Appendix 5 : Anatomy of a Complex Entry

The objective of this appendix is to demonstrate how the various DTD elements are instantiated in entries. The elements required to structure the complex entry for *box* are factored out and explained individually.

Appendix 6: The tagged entry

This appendix shows how the elements in the DTD are used to structure the print entry.

Appendix 7 : Menu Lists for Valency STR Fields

Although not strictly speaking part of entry structures, the content of this file demonstrates the amount of detail which closed-set fields will contain within the DWS.

Part 1 : NEID Entry Structures

1. NEID approach to Entry Structures

In a computer-assisted dictionary project, the emergence of the entry structures from the design process is inextricably linked with the development of the DTD for the Dictionary Writing System (Tasks J and L deliverables respectively in this project). Other tasks which impact on entry structures are the decisions made on the template entries (Task D deliverable); the policy on phonetics (Task C deliverable) and on linguistic labels (Task F deliverable); and editorial policy on other issues worked out in detail through the channel of the 100 sample entries (Task E deliverable). It is therefore premature to propose a set of finalized entry structures at this point in the workflow of Phase 1. Fortunately, the DWS software we are using allows for considerable flexibility in configuring the DTD. Provided that final decisions on the key entry-structure issues outlined below in Section 2 can be made within a reasonable timeframe, delivery of the final DTD by 31st August 2004 will not be jeopardised.

Consequently, the objective of this report is not to list and display a watertight finalized set of entry layouts; it is to demonstrate that the structures required for writing the NEID entries are already in place within the DTD and open to adaptation as necessary for at least another three months. Included as Appendix 1 Entry Types is a group of six disparate entries, all of which are catered for in the prototype DTD. Each entry is commented from the point of view of the lexicographic properties of the headword which must be handled by the entry structure. These exemplar entries are:

- **box** polysemous lemma including noun, modifier and verb uses, and associated idioms and phrasal verbs
- **efficient** polysemous adjective also forming productive suffix used in compounds
- **gravely** polysemous adverb (degree, manner)
- **e.g.** abbreviation in common use
- **Cork** encyclopedic entry
- **however** function word entailing predictably anisomorphic source & target language items

We shall discuss entry structures within the DTD (and its potential for developing further entry structures) in Section 3, after clarifying in Section 2 the seven principal decisions still to be made before any entry structures can be finalized.

2. Major entry layout and content decisions outstanding

These decisions, still to be made by Foras na Gaeilge on receipt of discussion documents as part of Task E (100 Sample Entries), relate to the following issues:

1. the major, high-level structural division within the entry: based on meaning or grammar?
2. multiword expressions (idioms, compounds, and phrasal verbs) and derived forms: their placement in the text
3. linguistic labels - policy, for both source-language (SL) and target-language (TL) items
4. function word entries: content, style and layout
5. Irish regional variants - what to include in the print dictionary, and how to show them
6. target-language level: use of 'literary Irish' as alternative translations
7. strategies for handling lexical gaps in the target language.

Each decision must be carefully considered, on the basis of evidence of how each option affects the dictionary text. The decision-makers need to be able to gauge the overall effect of each option on the dictionary as a whole, taking account of factors such as usability and the impact on overall

extent, and this means setting out the options in the form of alternative versions of the same entries.

In a separate communication we shall propose an approach to these issues which advances the due date of a proportion of the sample entries (Task E). In essence, each of the seven policy decisions listed above will form the focus of a brief discussion outlining the decision to be made, and illustrating the options by means of alternative displays of the same information in a carefully selected batch of entries. The rescheduling of the Task E deliverables into batches of entries each with its own topic and its own target date for LMC delivery and FnG response will be proposed by LMC during the week beginning April 5th. In the remainder of this section, we shall briefly set out the issues to be resolved. The numbering of the subsections reflects the initial list above.

2.1 High-level structural division of lexical entries

An illustration of the fundamental effect which this decision has on the way the material is presented is given in Appendix 2 (Major Structural Option), but no decision may be made in this regard without careful consideration of other entries which have different lexicographic properties from those of the lemma *box*.

This particular issue impacts on the whole of the DTD (as explained in Section 3), which is why it has priority status. It must be resolved as soon as possible, and will form the focus of the first 'policy-making' batch of sample entries. It concerns the fundamental decision on whether to 'navigate' the microstructure of the lexical entry on the basis of meaning or of wordclass.¹ Is it easier for English-speaking users to find their way around the entries by using their native-speaker instinct about what the headword can mean, or by using their knowledge of nouns, verbs, adjectives, adverbs and so on? Can we rely on the people who use the dictionary in secondary schools, in homes and offices, to know enough about grammar? In other words, should the first major division of entry text be made, as has traditionally been the case, according to parts of speech (treating noun uses, verb uses etc. separately even when they are closely linked semantically), or should the NEID entries be structured according to the various broad senses which the word has? In Appendix 2, version A of the *box* entry relies on the users' acquaintance with the various meanings of the word, while version B shows the traditional grammar-based approach.

For the moment, on the advice of the Chief Irish Editor, the prototype DTD is predicated on an initial division based on the broad senses of the headword, but more work needs to be done before this is set in stone. Indeed, the editorial team is not yet wholly convinced that such an approach is even viable for all types of lexical entries. Applying it to other very different headwords will make it clear how feasible it is to give this sense-based vs. grammar-based option.

2.2 Treatment of multiword expressions and derived forms

The question here is where to locate such *definienda* as

- **compounds** (e.g. *box player*, *boxing glove*, *chocolate-box adj*, *in spite of*, *all right*);
- **phrasal verbs** (e.g. *box in*, *box up*, *pass out*, *come up with*, *take aback*, *set up*);
- **idioms and other phrases**. (e.g. *box clever*, *think out of the box*, *kick the bucket*, *rain cats and dogs*);
- **derived forms** (e.g. *boxed adj*, *worrying*).

The two *box* entries in Appendix 2 show the same approach to this question. In both cases, the treatment is as follows:

¹ Function word entries and encyclopedic entries are not germane to this discussion.

- **compounds** are listed as headwords in their own alphabetical place in the dictionary text, and therefore not in the *box* entry at all;
- **phrasal verbs** are listed below the main *box* entry as secondary headwords with their own mini-entries, each with sense divisions if appropriate, but without phonetics;
- **idioms and other phrases** are treated in the secondary headword listing below the main entry;
- **derived forms** are entered within the entry as secondary headword, with part of speech indication and translation (**boxed** *adj* i mbosca/mboscai).

There are however other options for the treatment of these four distinct types of items, and they require to be explored further, and decisions made on the basis of ample evidence in the form of the second batch of sample entries, which will be accompanied by a summary of the options and of the decisions to be made, and submitted to Foras na Gaeilge as soon as is practicable. These decisions also impact, of course, on the final form of the DTD.

2.3 Policy on linguistic labels in the dictionary and database

There are a number of questions to resolve regarding linguistic labels in a database-dictionary (whether the dictionary is to be published in print or electronic form, or both):

Qu. 1 Which labels should be used?

Qu. 2 How deep should the label hierarchy be? i.e. should we have BIOCHEMISTRY, CHEMISTRY and SCIENCE, or only CHEMISTRY and SCIENCE, or only SCIENCE?

Qu. 3 When should labels be used in the database? i.e. should every item which might carry a label be given one? This would mean that for instance *violin* would be marked MUSIC, as well as *arpeggio* and one of the senses of *note* and *chord*, and so on.

Qu. 4 When should they appear in the dictionary text (print and electronic)?

Qu. 5 In what form should they be shown there?

Qu. 6 Where should they be located with regard to source- and target-language items?

Question 1 (list of labels to be used in the project) has already been broached, and the first part of the Task F deliverable submits a maximal list of labels to choose from. Batch 3 of the sample entries will focus on this issue, and illustrate the various options involved. Decisions on all of these questions will result in a linguistics labels policy which will then be reflected in the final version of the Task F Deliverable, and translated into elements in the DTD, thus finalizing the aspect of entry structures which has to do with labels. The labels menu listings will also be incorporated into the dictionary writing system (DWS).

2.4 Function word entries: content, style and layout

Included under the designation of 'function words' (often limited to 'grammatical words') are lexical items which are commonly used in order to structure discourse in some way. This is true of our example, *however*, which is shown in Appendix 1.

Design decisions relating to function word entries are of course quite different from those having to do with lexical entries (encyclopedic entries pose few problems, and will not be further discussed in this document). Current lexicographical approaches to function words in bilingual and other pedagogical dictionaries take advantage of this situation. In many of their uses, function words have no 'direct' translation, and their function in the source language must be implemented in the target language in a variety of different – and often unexpected – ways. This requires considerable research into the behaviour of each of the English function words during the analysis stage of the lexicography (which produces the frameworks).

There are many different ways of showing function word information in a bilingual dictionary: many dictionaries still opt for the now rather outdated approach of treating these words as though

in all their uses they contained little else but lexical meaning. This is the approach enshrined in de Bhaldráithe's *English-Irish Dictionary*, as well as those catering to a different readership such as the *Oxford-Duden English-German Dictionary*. However, thanks to the development of pedagogical dictionaries over the past two decades and the search for user-friendly methods of presenting facts about a foreign language, modern scholarship now leans towards a less formal approach to function words, as exemplified in the *Oxford-Hachette English-French Dictionary* and the *Macmillan English Dictionary for Advanced Learners*. In these works, many function word entries are preceded by a piece of discursive text, a kind of usage note, in which the pragmatic functions of the headword are described in jargon-free terms for the inexperienced dictionary user. It is this latter approach that is illustrated in the *however* entry in Appendix 1. The hypothesis on which this rests is that the information contained in the box which precedes the entry proper will not be of interest to experienced linguists such as language teachers and translators, but will be very helpful for the school student (for instance). This could be called the 'horses for courses' approach, and is essential if NEID is to supply the needs of the full range of its intended users. In such an approach, some information meets the needs of advanced users, and that will be couched in text using technical linguistic terms, while some information is targeted at inexperienced users and that will be expressed in non-technical terms, even if this means blurring the edges of the precise statements which linguists would expect to find.

The NEID approach to function words will be based on decisions made in the light of Batch 4 of the sample entries, where a number of function word entries will be offered, showing the same material presented in various ways. For the moment, suffice it to say that most of the elements of the Appendix 1 *however* entry already exist in the prototype DTD: only the boxed introductory note is still to be documented. Decisions here will depend on reactions to Batch 4.

2.5 Irish regional variants in target-language items

This is a topic which has given rise to considerable discussion both in the Dublin and Galway Translators' Workshops and at the Introductory Day for the Linguistic Advisory Board (LAB). There is no unanimity on whether or not to include regional variants, let alone on how they should be treated in the dictionary. Consequently, a discussion document entitled *Regional variants in Irish language items: proposed policy* was posted on the LAB intranet site together with the first entry to be commented. It produced a lively debate, most of which criticised some aspect of the policy proposals, but which offered no consensus on this topic. Questions to be resolved here include (for both print and electronic dictionaries):

Qu. 1 whether simply to show 'standard' forms, or ...

Qu. 2 whether to show variants when they are 'direct' translations of the headword and/or when they occur in translations of examples within an entry

and, if it is decided to include regional variants, then:

Qu. 3 whether to mark their region of origin

Qu. 4 how to order them

Qu. 5 how to deal with them in translating example phrases

All of these issues will be raised and exemplified in Batch 5 of the sample entries, which will focus on the treatment of regional variants in Irish-language items and which will propose an innovative compromise approach to the knotty problem of standard forms vs. regional variants.

2.6 Use of literary Irish as alternative translations

This arose as a topic for consideration because of the anxiety expressed at every linguistic meeting, whether Translators' Workshop or LAB introductory day, that the richness of the Irish language would inevitably be very badly served by the new 21st century NEID dictionary. The argument goes that the source-language text in this dictionary is (rightly, in view of its user profile) not literary English, and in the normal course of events would not be translated by literary Irish. Therefore, the richness of the Irish language will be betrayed.

The vast proportion of the language of the dictionary is of course unmarked, but where this is not the case the policy of most bilingual dictionaries is remarkably uniform: editors make a conscious effort to line up the language level of the source-language item with that of its target-language equivalent, i.e. formal SL items are if possible translated by formal TL items and so on. All of the language types flagged by labels listed in the current Task F file (covering vocabulary marked for domain, region, evaluation, register, style, sublanguage and time-period) are treated in this way.

An example of the type of problem which this topic is designed to deal with is taken from the current NEID formulation of the *command* entry, where the literary Irish words *foláireamh* and *foráil* have been added to other unmarked items such as *ordaigh*, *fógair*, and *ordú*:

<p>command</p> <p>v 1 (<i>order</i>) <i>ordaigh</i>, <i>fógair</i>, <i>foráil</i> (ar) <i>liter</i>; to command sb to do sth <i>ordú do dhuine rud a dhéanamh</i>; ‘open the door!’ he commanded ‘<i>oscail an doras!</i>’ a d’<i>ordaigh</i> sé;</p> <p>[...]</p> <p>n 1 <i>gen, Mil, IT (order, instruction) ordú m (gen sg ordaithe, pl orduithe), foláireamh m1 liter, foráil f3 liter</i>; to give a command <i>ordú a thabhairt</i>; he gave the command to retreat <i>thug sé an t-ordú dul ar gcúl</i>; they halted at his command <i>stadadar ar ordú uaidh</i>;</p>
--

Batch 6 of the sample entries will provide a brief discussion document on this topic, together with rich exemplification of the options set out in a number of different dictionary entries.

2.7 Strategies for handling lexical gaps in the target language

This is a problem which occurs on every page of every bilingual dictionary, no matter what language constitutes the source and what the target. One example of this problem appeared in the first Translators’ Workshop in Dublin, when it came to light that the various lexical units of the verb *forge* in the English SL were not paralleled by the lexical units of the verb *gaibhnigh*. As a result, there was no immediately obvious single-word Irish equivalent for the English verb in such contexts as:

- | |
|---|
| <ol style="list-style-type: none">1. youth of the world should be mobilized to forge a global partnership2. serve as a vehicle for forging a bipartisan coalition in the House3. Jan is hoping to forge a career in broadcasting4. Joan Curran forged a close relationship with the technical college5. New Delhi forged a close alliance with the Soviet Union |
|---|

Opinions varied about strategies to employ in such cases: some suggested creating a new sense of the existing Irish (blacksmith) verb; others preferred the more traditional approach of paraphrasing the SL sense in the TL; and so on. This problem will form the focus of the seventh batch of ‘policy-making’ sample entries.

3. Entry Structures implicit in the DTD

The Document Type Definition is the program which supports the computationally tractable lexical database, and which is developed gradually during the course of Phase 1, reflecting the needs of the lexicography and specifically the decisions made on lexicographic issues such as those outlined in Section 2. All the options on entry structures touched upon in Section 2 are currently

implicit in the prototype DTD which is in the process of being refined. The final version of the DTD itself constitutes Deliverable J, due at end August 2004. At that point, the entry structures which have been cumulatively discussed with and agreed by the Chief Irish Editor (with the assistance of the Linguistic Advisory Board), and accepted by Foras na Gaeilge, will themselves be finalised.

Four of the appendices to this Report have been compiled with the objective of explaining the structure of the DTD, and clarifying its role as holder of entry structure potential. A brief overview of these is given in the Sections 3.1 – 3.4 inclusive.

3.1 Appendix 3: Introducing the DTD

In this appendix we set out part of the structure of the DTD, and explain the relationship of the elements within it. All the elements in the current prototype DTD are listed in tabular form, together with an explanation of what each is. The elements which are used in the mark-up of the *box* entry in Appendix 5 are highlighted in this table.

3.2 Appendix 4 : DTD inter-element relationships

This organization chart illustrates in graphic format some of the subordinate elements of the Sense Block element, which are shown and explained in Appendix 3.

3.3 Appendix 5 : Anatomy of a complex entry – *box*

In this appendix, the entry for the lemma *box* is shown in a print dictionary format, followed by a deconstruction of the entry in tabular form, indicating how the DTD elements are realized in the entry.

3.4 Appendix 6 : Valency menus for the STR field

This appendix contains the provisional listings of grammatical (complementation) constructions for English verbs, nouns, and adjectives to be recorded in the frameworks. These lists will be embedded in the DWS in the form of menus for the relevant fields in the DTD. Menus will be provided for all datatypes which constitute a closed set. As such, they perform a useful function in reminding the lexicographers of relevant constructions to be noted in the corpus data and recorded in the framework, as well as controlling the input of data to these fields and preventing lexicographer errors and inconsistencies.

4. Entry Structures: Conclusion

This part of the Task C Deliverable Report set out to show how the current version of the DTD holds the potential for any number of options on dictionary entry layout, and to give a flavour of what some of these options are.

It is important to stress that the *linguistic content* of the entries shown in these Appendices is not relevant to the present discussion, which focuses on broad issues of editorial policy in relation to the structure of the entries. Comments on entry content and quality are therefore not appropriate at this stage, but there will of course be plenty of opportunity to consider these aspects and comment on them when the batches of sample entries making up Task E are delivered.

Part 2: NEID Phonetics Policy

Deliverable

A complete set of policies for handling phonetics in the dictionary

1. Options and recommendations: the print version of the dictionary

1.1 Should the dictionary show phonetics at all?

It is not usual lexicographic practice to give phonological information for target-language items. With the NEID, however, a strong case can be made for providing users – many of whom will be learners of Irish – with enough information to enable them to pronounce the words they find in the dictionary. In this document, therefore, we are assuming that Irish phonetics *should* be accounted for, and we recommend policies for handling this information. Conversely, we take it as a given that English pronunciations will not be shown, and do not need to be shown.

1.2 Where should the phonetics appear?

This question addresses the location of phonetics in the *print dictionary only*; the electronic version of the dictionary offers quite different opportunities, and these are discussed below.

The options are:

1. show phonetics adjacent to each Irish translation, for example:

box 1 *n* bosca /boskə/ *m4*, bocsa /boksə/ *m4*

2. show phonetics in an index at the back, where each Irish lemma that appears as a translation in the A-Z text is listed alphabetically and given a transcription

Option 1 is impractical for a number of reasons:

- it is extremely space-consuming: the same Irish word is likely to appear as a translation for many different source-language meanings, so the same pronunciation will be repeated – sometimes dozens or even hundreds of times – throughout the dictionary
- it will make the entries look cluttered and dense, detracting from the dictionary's appeal and user-friendliness.

Option 2 – or at least a version of it – is therefore preferable, and is further discussed under **Coverage** in 1.3 below.

1.3 In what form should phonological information be given?

Showing pronunciations

The options here are:

1. to show phonetics using IPA (the International Phonetic Alphabet)
2. to use a 'respelling' system based on Irish pronunciations of English

IPA is the system of choice in most quality bilingual dictionaries published in Europe, and in all the monolingual learners' dictionaries of English, and there is an increasing trend towards its use in 'trade' dictionaries for the native-speaker market; the *New Oxford Dictionary of English*, for example, uses IPA in those cases where a pronunciation is shown. The big advantage of IPA is that the system is delicate enough to enable dictionaries to record pronunciations with a high degree of precision.

Nevertheless, there are compelling counter-arguments in the case of the NEID: while some users will be linguists for whom IPA poses no problems, a majority (whether school students, 'family' users, or journalists) are likely to be unfamiliar with IPA, and this group of users – the people who most need the phonetic information – will not get the help they require. On the other hand, since all users of the dictionary are assumed to be fluent speakers of English, a carefully-devised 'respelling' system draws on what they already know, and is therefore likely to cater for a high percentage of the user-group. Respelling systems are not without their problems, of course: there are some distinctions (as between broad and slender consonants) for which English has no simple equivalent. But in general the disadvantages (in terms of descriptive adequacy) are outweighed by the benefits to the majority of users.

Coverage

If phonetics are not to be inserted into the target-language material (see the discussion in 1.2 above), and Option 2 is selected, a decision needs to be made between

- listing in the index every Irish lemma that appears in the A-Z text, and
- showing irregular items only, and including a brief discursive section on Irish phonology.

The latter approach seems preferable, partly on space grounds (a full listing – even in compressed format – might run to well over 100 pages), but principally on the grounds that much of the phonological system is regular. It is therefore proposed to include in the dictionary a reference section describing the regularities of the phonetic system of Irish (and of its major dialects), with information about frequently-occurring words that exhibit irregular phonological behaviour. Here the NEID can benefit from work already done by Dónall Ó Baoill and Seán Briain for the *Foclóir Scoile English-Irish / Irish-English Dictionary* (1986 Dublin: An Gúm).

An additional advantage of the proposed index is that it could also include *morphologically* irregular Irish lemmas, giving pointers to fuller information about this aspect of Irish as well.

Hearing the sounds of Irish

Many users of the NEID would also welcome the opportunity to hear the sounds of Irish on a CD or cassette bundled with the print dictionary, and we include this in our recommendations in section 1.5 below.

1.4 How should dialectal variations be handled?

To some extent, this question forms part of the larger policy issue regarding the degree to which the dictionary should account for Irish dialectal variation in all its forms (whether lexical, syntactic, or phonological): section 2.5 of Part 1 above addresses this issue and proposes a methodology for resolving it. The view of the Chief Irish Editor is that most dialect-related variation in the *sounds* of Irish words is rule-governed, and the rules can also be explained in the proposed discursive section on Irish pronunciation described in 1.3 above. (A possible future Irish–English dictionary, of course, could address dialectal variation much more extensively.)

1.5 Recommendations on phonetics policy in the print dictionary

LMC and the Chief Irish Editor therefore recommend the following phonetics policy for the print NEID:

1. that the back matter should contain an index of all the Irish words with irregular pronunciations which appear as translations in the dictionary text, together with their phonetic 'respelling'; words appearing in this list can carry a symbol to that effect within the entry, without extending the text length unduly.
2. that there should be a discursive section in the dictionary in which the regularities of the phonetic system of Irish (and of its major dialects) are described, along with information about frequently-occurring words that exhibit irregular phonological behaviour.
3. that an audio CD or cassette should be bundled with the print dictionary, giving an introduction to the sounds of Irish (this is distinct from the electronic version of the dictionary, to be discussed later.)

2. Options and recommendations: the electronic version of the dictionary

The electronic medium imposes fewer space constraints, and offers significant opportunities to provide users with high quality information in a user-friendly form. The philosophical arguments against extensive coverage of dialectal variation remain valid, but aside from this the dictionary can provide very detailed information.

2.1 Should the electronic version show a written representation of Irish phonetics?

There is probably no need to provide this information in written form if the electronic version allows users to *hear* the sounds.

2.2 How should audio pronunciations be generated?

It is usual for dictionaries on CD-ROM (or other electronic platforms) to allow users to select a word and hear its pronunciation. In some dictionaries of this type, a degree of dialectal variation is also catered for: in the *Macmillan English Dictionary for Advanced Learners*, for example, users can elect to hear a British or American pronunciation of any headword.

There are two main approaches for generating sounds, and these will need to be considered in detail when decisions are made regarding the functionality of the electronic version of the dictionary. These are:

1. using human input: this is currently the most common approach. Native-speaking actors read out all the headwords in the dictionary (or in this case, all the words used as translation equivalents in the target-language side of the dictionary), and these recordings are saved as sound files on the CD-ROM and activated by clicking an icon. Consideration should be given here to ensuring a good representation of the different dialects.
2. using an automatic 'text-to-speech' (TTS) system which will generate the pronunciation of any item in the text. This is a very ambitious option, but definitely worth exploring. In

essence, it starts from a set of audio files produced by a speaker who reads texts that are carefully selected to cover all the phonological possibilities of a language. This is marked up in such a way as to produce a full database of sounds for every combination and sequence, and for every prosodic situation. These sounds are then recombined (using statistical models of which segments fit together in which situations) to produce an artificial but human-like voice. Using a system of this type, the dictionary user could select any word or string of words in the dictionary (i.e. not only the canonical forms of Irish words used as translation equivalents) and hear a realistic pronunciation. Progressing this option would entail a major research programme which would need to be initiated fairly soon if its output is to be available at the point when the dictionary is published. However: (a) some of the groundwork has already been done (by scholars such as Ailbhe Ní Chasaide from the Phonetics and Speech Lab in TCD) and (b) there may well be interest from industrial partners such as Eircom and BT.
