

Cormac Breathnach and Pádraig Ó Mianáin

MAKING A MOLEHILL OUT OF A MOUNTAIN: TECHNICAL AND EDITORIAL CONSIDERATIONS IN PRODUCING THE CONCISE ENGLISH-IRISH DICTIONARY (2020)

A Discussion of Technical and Editorial
Considerations in Producing a 1800-Page Hardback
Dictionary Containing 30,000 Headwords From
an Online Database of 48,000 Headwords

Abstract The *Concise English-Irish Dictionary* (CEID), published in 2020 and the first major English-Irish dictionary published in print form since the 1950s, is a 1800-page hardback dictionary containing 30,000 headwords and 80,000 senses, along with a substantial style and grammar section. Flying in the face of retro-digitisation, this printed dictionary was derived from the *New English-Irish Dictionary* (NEID), a much larger online dictionary published 2013-2017 and containing 48,000 headwords and 145,000 senses. Simply printing the entirety of the online content would have doubled the size of the printed dictionary, so this necessitated a number of measures to whittle the online content down to a single-volume dictionary. This paper outlines some of the challenges and measures involved, such as selection or deselection of lexicographical content, reformatting for print, and the technical process of outputting the same entry to both screen and paper.

Keywords Irish language; online to print conversion; bilingual

1. Introduction

The *Concise English-Irish Dictionary* (CEID), published in 2020, is the first major English-Irish dictionary published since Tomás de Bhaldraithe's *English-Irish Dictionary* in 1959 and the first major bilingual dictionary published in Ireland since Niall Ó Dónaill's *Foclóir Gaeilge-Béarla* [Irish-English Dictionary] in 1977.

It was produced by Foras na Gaeilge, the cross-border implementation body with responsibility for the promotion of the Irish language in Ireland, both North and South, including responsibility for the development of lexicography and terminology. CEID was the final product of the New English-Irish Dictionary project which had started in 2000. The online dictionary which is available free of charge at www.foclóir.ie had been published incrementally from 2013 until 2017, and an accompanying app had been available since 2015. The digital versions contain some 48,000 headwords and 145,000 senses.

The New English-Irish Dictionary project was initially intended to produce a printed dictionary and an accompanying CD-ROM, but by the time the writing phase of the project began in 2008, the rapidly developing digital revolution in lexicography had resulted in online dictionaries replacing printed dictionaries as the primary medium, and ultimately what began as a book and CD-ROM project ended up with the editorial team concentrating on a website and an app.

When the new online dictionary was completed in 2017, it was by no means a given that a printed version would follow. There was a strong argument for turning immediately to the urgent need for a major monolingual Irish dictionary, which would be the first ever of its kind, and a new Irish-English dictionary to succeed the previous one from 1977. However, there were also compelling arguments for producing a printed version: it was always felt that it would do well commercially as it would be the first such English-Irish dictionary in sixty years, and it was also recognised that there was a large demographic of likely Irish-language dictionary users to whom the digital platform was and would remain beyond reach.

A significant challenge resulting from the late decision to produce a printed version was that, as none of the editorial work on the online entries had been done with a print edition of the dictionary in mind, the fact editors had enjoyed zero space limitations up to that point proved to be a major hurdle for the printed dictionary. Another challenge was maintaining one single database for both the digital and printed versions of the dictionary as it was vital to keep both products in the same database to avoid having to amend or correct two versions of each entry continually. The book was finally sent to the printers three years later during the first Covid-19 lockdown in May 2020.¹ In this paper we discuss some of the challenges faced when moving in the unusual direction of online to print, as well as some of our solutions to those problems.

2. Size Matters

The first consideration for the editorial team was the design and physical size of the printed dictionary which would include entries as well as a substantial style and grammar supplement. Early calculations indicated that simply printing the entirety of the online content – all 48,000 headwords, 145,000 senses, almost 130,000 example sentences and around 2 million words – would have run to 3000-3500 pages not including the supplement. That would require two volumes, which would obviously significantly increase costs but would also inevitably greatly reduce the predicted uptake and sales of the product. When the decision was made to go for a one-volume dictionary of around 1500-1800 pages, that meant a 50% reduction of the physical size initially calculated, necessitating editorial measures to select or deselect content from the online version and technical solutions to reformat the dictionary content as space-efficiently as possible.

¹ It was reviewed in the *International Journal of Lexicography* in March 2022 (Pödör, 2022).

3. Trimming by Editing

It was obvious from the beginning of the book project that, aside from a drastic measure such as removing all example sentences, most of the downsizing would have to be done editorially by a staggered selection process, from headword or entire entry level right down to register and domain level. Each level of selection posed its own challenges for the editorial team, as it would prove impossible to achieve a uniform calibration of in/out decision-making among a team of editors, many of whom were working remotely most of the time. While the initial selection at headword level could be based largely on frequency, the lexicographer's dilemma grew as we moved on to sense and example selection where the onus was largely on the individual editor to decide the fate of any given content.

3.1 Selection at Headword Level

The obvious first step was to determine which headwords would be included in the book, and which headwords would not, and to that end the entries in the general NEID dictionary database were divided into three categories for inclusion.

3.1.1 Category 1

Category 1 comprised c. 5,600 headwords selected on frequency basis as outlined in (a) and (b):

(a) The most common 1000 words in English

At the beginning of the translation phase of the NEID project the top 1000 lemmas in order of frequency were given special status and treatment (Ó Mianáin & Convery, 2014, p. 322), and as such they were automatically included in CEID.

(b) Comparison with other dictionaries

Another measure taken during the NEID project was to analyse the headword lists of three medium-sized monolingual English dictionaries and compare them with the headword list for NEID. A score between 1 and 3 was given to each NEID headword depending on its inclusion status in the three control dictionaries. For example, a common headword such as 'table' appeared in all three dictionaries and therefore was given a score of 3. The headword 'bronze' was only present in two dictionaries and was given a score of 2, and so on. Headwords present in all three dictionaries were also included in Category 1 and were automatically included in the headword list for CEID. This amounted to about 4,600 entries.

3.1.2 Category 2: Headwords With a Score of 2 or 1 in the Control Dictionaries

All headwords present in either two or one of the control dictionaries, that weren't in Category 1, were marked as Category 2 and added to the CEID headword list automatically. This amounted to about 6,400 entries.

3.1.3 Category 3: Corpus Frequency Scores and Selection by Editors

The remaining 19,000 or so headwords in CEID were selected using both (a) statistical analysis of the project corpus and (b) coverage checks carried out by editors on a continuous rolling basis.

(a) Corpus frequency scores

A 1.7-billion-word corpus was used as a basis for the English-language content in the NEID project (Ó Mianáin & Convery, 2014). The primary criterion in Category 3 was word frequency in the corpus for the 37,000 headwords not already been included through Categories 1 and 2 above. These headwords were ranked in order of frequency in the corpus and then grouped into 20 bands. There was no set frequency threshold for Category 3 as the final number of headwords in this category depended on how much of the available page-space was being taken up by other categories. Instead, we moved out through the bands of frequency, starting with the top 5%, until we could no longer include any more entries in the book. Approximately 13,000-14,000 headwords were selected using this measure.

(b) Headwords selected by editors

A number of checks were carried out by editors to ensure that CEID would include essential headwords that may not necessarily have been picked up based on frequency in the corpus:

- I. We endeavoured to ensure that specific sets of words such as numbers, directions, colours, musical instruments, etc., were adequately covered and that the main closed sets would be included in their entirety.
- II. Cultural entries relevant to Irish life, culture, history, music and sport, also Hiberno-English.
- III. Encyclopaedic and school-subject entries, of which about 4000 were added, as it was anticipated that the dictionary would probably be used in schools and we were keen to cover a broad range of topics relevant in an educational setting.
- IV. Using Google Analytics, we extracted lists of the most frequently searched items on www.focloir.ie both in the overall seven-year lifespan of the dictionary up to that point and also in the 12-month period leading up to our publication date in 2020, and used those lists to identify additional candidates for inclusion.
- V. Neologisms not in the original DANTE database on which NEID was based but that had been added to the online dictionary since going online in 2013. The online dictionary is regularly updated with new words and phrases that had come to prominence, such as ‘selfie’, ‘craft beer’,

‘emergency accommodation’ and ‘fake news’, and also with words and phrases identified by analysing the unsuccessful searches on the site. A selection of these entries was added to CEID, including a number of Covid-19 terms added in the very final weeks of the book project, such as ‘social distancing’, ‘coronavirus’ and ‘self-isolation’.

- VI. Gender balance, for instance where occupational entries such as ‘doorman’ were reciprocated with ‘doorwoman’ etc.

Approximately 5,000-6,000 headwords were selected using these measures.

The entire headword selection task involving Categories 1-3 resulted in the reduction of the number of headwords from the initial 48,135 entries in NEID to 30,635 in CEID; a 36% reduction.

3.2 Selection at Sense Level

The removal of over a third of the headwords, however, did not by any means equate to a similar reduction in size as the excluded headwords were by nature relatively small in terms of number of senses and examples. In effect, the 36% reduction at headword level translated into an 18% reduction in the number of overall senses and only 8% in the number of usage examples.

The next step in the reduction process was to select or deselect senses under the remaining headwords. Unlike the exclusion process at headword level, which was largely a mechanical process based on word frequency either at corpus level or in the comparison with other dictionaries, the selection of senses for exclusion was primarily down to the judgement of individual editors. Essentially, it was at each editor’s discretion to decide whether or not each sense of the online dictionary entry was sufficiently significant in terms of general language use to merit inclusion in the printed dictionary. In some cases, editors consulted other general bilingual dictionaries in order to determine the importance of word senses in entries. The senior editors would subsequently have an opportunity to promote excluded senses if required, or similarly to demote senses not excluded in the initial editorial pass.

Beginning with an online database of up to 145,000 senses contained in 120,000 sense containers, the final print dictionary was reduced to about 85,000 total senses. As mentioned above, the exclusion of 36% of the headwords accounted for an 18% reduction in sense containers (or about 23% of overall available senses). The exclusion of sense containers by editors in the remaining 30,635 entries led to a further 21% reduction in sense containers (or circa 26% of senses) from the final print product.

3.3 Selection at Micro-Level (Intra-Sense Decisions)

Other items within the remaining sense units were excluded in the CEID entries where they were not considered essential, ranging from examples and their accompanying translations to more granular elements such as domain, style and register labels. Some translations were also excluded, and the sense disambiguators were revised and trimmed where possible. The way usage examples and variants were dealt with is discussed below.

3.3.1 Usage Examples

A lot of the entries in the online dictionary are example-heavy as it is primarily a dictionary for encoding into Irish, but for the printed version the editors were deciding which example sentences would be included. Although there were editorial guidelines for this task, the decisions were largely down to the judgment of individual editors. We recognised from early on that considering every single example sentence would be a far too arduous and time-consuming task and so allowed editors to suppress example sentences on an ad hoc basis.

In the end, over 4000 usage examples in English were suppressed along with their corresponding 4500 translations to Irish. This amounted to a total saving of about 37,500 words, or almost 40 pages of text.

3.3.2 Variant Forms

As a rule, all variants of headwords, phrases and subforms/derived forms were excluded from the print edition of the dictionary. The online version of the dictionary contains almost 10,000 variants of words at various levels. For example, American English spelling of headwords is usually provided (colour/color, analyse/analyze, aesthetic/esthetic). It was felt, however, that the print edition wouldn't suffer for the lack of variants and that their removal would save a lot of space. Indeed, in some cases the variant of a headword is just the headword itself with or without a hyphen (vice captain/vice-captain). This would have looked clunky and unnecessary in the book. This was a relatively straightforward measure to implement but still accounted for a saving of about 21,000 words or over 20 pages of text.

Variants of phrasal verbs were, however, included in the book as they represent a more important part of language usage, for example 'clock off' and 'clock out' are both included.

3.3.3 Variant Forms of Phrases and Idioms

The handling of variants in the case of phrases and idioms was a trickier task. Phrases are by nature quite a bit longer than headwords or phrasal verbs, and as such, the corresponding variant forms are also quite long. We were keen to include the variants deemed linguistically significant, however the inclusion

of ‘to call somebody to account’, ‘to bring somebody to account’ and ‘to hold somebody to account’ would constitute far too much text at the beginning of the phrase. As a solution, it was decided to display a merged version of the phrase, i.e., ‘to call/hold/bring sb to account’. To this end all phrases and their variants were suppressed in the print stylesheet and a new element was introduced that would house the merged version of the phrases.

to call sb to account (*also to bring sb to account, to hold sb to account*)

duine a thabhairt chun cuntais

a iarraidh ar dhuine míniú a thabhairt

she was held to account for the overspend bhí uirthi míniú a thabhairt ar an róchaitheamh

who'll be brought to account for the delay? cé a thabharfar chun cuntais as an moill?

• **to call/hold/bring sb to -** duine a thabhairt chun cuntais, a iarraidh ar dhuine míniú a thabhairt; **she was held to - for the overspend** bhí uirthi míniú a thabhairt ar an róchaitheamh; **who'll be brought to - for the delay?** cé a thabharfar chun cuntais as an moill?

Fig. 1: Idiom ‘to call sb to account’ on NEID website (left) and in CEID (right)

3.3.4 Cross-Referencing to Exemplar Entries

The online dictionary has a number of templated sets of entries, where entries such as countries, languages, letters, numbers and pronouns all contain a lot of identical examples to show various usage and grammar elements that can be complicated in the Irish-language equivalents. In the online dictionary, for instance, the entry for ‘you’ contains 14 senses and 156 examples, the entry for ‘Spanish’ contains 4 senses and 18 examples, and the entry for ‘B’ contains 6 senses and 13 examples. As a space-saving measure in the book, however, it was decided to reduce the majority of such entries to their minimum where possible and to cross-reference them to exemplar entries where users could see the usage patterns relevant to the type of word involved. Not all entries could be cross-referenced due to the complexity of the grammar rules in their Irish-language equivalent, as is the case with pronouns and numbers; in the print version, for instance, the entry for ‘you’ has 8 of its original 14 senses and 148 of its original 156 examples.

Ultimately, nearly 500 countries, languages, nationalities and letters of the alphabet were cross-referenced in this manner in the print version, which was a significant saving in terms of number of pages. The printed entry for ‘B’, for instance, was reduced from its original 6 senses and 13 examples to one sense and a cross-reference to the entry for ‘A’, which is a comprehensive entry. Although ‘Spanish’ kept all 4 senses in the print version, all of its 18 examples are suppressed and are replaced by a cross-reference to the entry for ‘German’, an actual reduction in terms of column centimetres from 12 cm to 2.5 cm.

Spanish

1 ADJECTIVE GEOG

Spáinneach *adj1* 🗣️ CMU

(de chuid) na Spáinne

he's Spanish is Spáinneach é

a Spanish soldier saighdiúir Spáinneach, saighdiúir de chuid na Spáinne

the Spanish government rialtas na Spáinne

2 ADJECTIVE LING

Spáinnise *gs as adj* 🗣️ CMU

a Spanish book leabhar Spáinnise

Spanish class rang Spáinnise

Spanish dictionary foclóir Spáinnise

Spanish exam scrúdú Spáinnise

Spanish teacher múinteoir Spáinnise

a Spanish word focal Spáinnise

the Spanish language an Spáinnis

3 NOUN LING

Spáinnis *fem2* 🗣️ CMU

she speaks Spanish tá Spáinnis aici

they're speaking Spanish tá siad ag labhairt Spáinnise, tá siad ag labhairt i Spáinnis

do you know any Spanish? an bhfuil aon Spáinnis agat?

how do you say 'Ireland' in Spanish? cén chaoi a ndeirtear 'Éire' i Spáinnis?

what's the Spanish for 'food'? cad é an Spáinnis ar 'bia'?

say that in Spanish abair é sin i Spáinnis

a film in Spanish scannán Spáinnise

a degree in Spanish céim sa Spáinnis

the Spanish *NOUN GEOG*

na Spáinnigh *mpl1* 🗣️ CMU

muintir na Spáinne

Fig. 2: The entry 'Spanish' on the NEID website

Spanish **ADJ** ① GEOG Spáinneach
a1, (de chuid) na Spáinne ② LING
 Spáinnise *nmod*
N LING Spáinnis *f2*
the Spanish **N** GEOG na Spáinnigh
mpl1, muintir na Spáinne ▶ *for*
usage see German

Fig. 3: The entry 'Spanish' in CEID

3.3.5 Domain Labels and Sense Disambiguators

On the NEID website, the domain labels and sense disambiguators are presented on the first line of each sense, as in the first two senses of the entry 'access rights' below. In the case of the printed dictionary, however, it was deemed superfluous to include both the domain label and the sense disambiguator unless absolutely necessary.

access rights

- 1 *NOUN LAW* to see child
cearta rochtana CMU
cead cuairte familiar CMU
he wants access rights to his children teastaíonn cead cuairte lena pháistí uaidh
- 2 *NOUN COMP* to view file etc
cearta rochtana CMU
- 3 *NOUN* to enter location
cearta dul isteach

Fig. 4: Entry ‘access rights’ on NEID website

Running a job in DPS to *suppress all domain labels* or to *suppress all disambiguators* in the print style sheet was too risky considering the amount of data involved and would have necessitated a large checking task afterwards. Instead, we created a spreadsheet of all the senses in the book that contained both domain labels and sense disambiguators. The ideal scenario for us was to use the domain labels as often as possible as they are much shorter in length than the disambiguators. We went through the spreadsheet – some 8,500 lines – and decided on whether to suppress the domain label or the disambiguator. In a small number of cases, we kept both. Using the spreadsheet, a job was run on the data in DPS and the @hide=print attribute (discussed in more detail below) was added to the selected elements. The example from the spreadsheet below shows the decisions in column four on which element – domain label or disambiguator – would be suppressed in the book.

Headword	Domain label	Sense disambiguator	HIDE?
A	mus	note, key	sense dis.
A	med	blood group	domain
abbreviation	ling	of word, phrase	sense dis.
abdicate	pol	renounce throne	sense dis.
abdication	pol	renouncing of throne	sense dis.
aberrant	biol	of gene, cell, etc	sense dis.
abnormal	stat	of curve etc	domain
abolitionist	law, pol	supporting abolition of sth	sense dis.
abolitionist	pol, hist	supporting abolition of slavery	sense dis.
abort	comp, aero	stop prematurely	sense dis.
abrasion	med	injury	sense dis.
abridge	law	limit	sense dis.
abrogate	gov, law	repeal	sense dis.
abscess	med	infected swelling	sense dis.
abscond	police	escape	domain
absence	bus-adm, ed	being away from somewhere	sense dis.
absent	social	not part of household	domain
absentee	hist	not living in certain place	domain
absolution	rel	forgiveness of sin	sense dis.
abstract	art	of artwork, artist	sense dis.
abstract	art	artwork	sense dis.
accent	ling	long accent in Irish	domain
accent	ling	diacritic	domain
accent	ling	stress	domain
accent	ling, mus	stress	sense dis.
access	social	for disabled	domain
access	telecom	connection	domain
access	law	contact with child	domain
access	telecom	connect with	domain

Fig. 5: Example from domain and disambiguator spreadsheet

4. Technical Measures

All of Foras na Gaeilge's dictionary projects are hosted in IDM's Dictionary Production System (DPS), and all technical solutions involve customisations within DPS.

4.1 Suppressing not Deleting: the @hide=print Attribute

As mentioned in the introduction, both the digital and printed dictionaries were being produced from the one dictionary database as we didn't want to have two separate databases to maintain – one for the book and one for the website. Consequently, it was of vital importance that dictionary content within the database would not be deleted in order to save space in the book. There is a huge amount of content on display on the website that could not be accommodated in the book. It was decided, therefore, that content that was to be excluded from the print dictionary would be suppressed or 'hidden' in the stylesheet rather than deleted from the database entirely. The @hide attribute was added to most elements in the database which could be set to any of the following values: never, always, print, or online.

The @hide=print attribute gave editors the option of hiding a sense, a string of text, a label or some other element in the database from the print stylesheet, without having to delete the element altogether and thus losing it from the online version of the website. As well as a WYSIWYG stylesheet to see the entry in the print layout, a previewer was also developed within DPS which would 'grey out' material that was set to @hide=print. In the example below, the fourth sense is fully hidden or suppressed as well as some example sentences in sense one and the sense disambiguator in sense three.

addition
1 *NOUN* sb or sth added
 it's a great addition to the village **cuireann sé go mór leis an sráidbhaile**
 she was a last-minute addition to the list **cuireadh ar an liosta ag an nóiméad deiridh í**
 Seán is the latest addition to the family **is é Seán an páiste is déanaí a rugadh sa chlann**
 an index would be a useful addition **bheadh sé áisiúil dá mbeadh innéacs leis**
 any amendments or additions? **aon rud le leasú nó le cur leis?**
 to make an addition to sth **breis a chur le rud, cur le rud**
 they made some additions to the text **chuir siad roinnt rudaí leis an téacs**

2 *NOUN* process of adding
 since the addition of law to the programme **ó cuireadh an dlí leis an gclár**
 it's sweetened by the addition of fruit juice **milsítear é trí shú torthaí a chur leis**

3 *NOUN MATH* of numbers
 suim iú *m asc*
 suim iúchán *m asc 1*
 repeated addition **s uimíú leantach**

3 *NOUN (US) CONSTR* extension
 píos a breise
 síneadh *m asc 1*
 to build an addition to a house **síneadh a chur le teach**

Fig. 6: Entry 'addition' in the hybrid CEID stylesheet displaying content excluded from the print dictionary

4.2 Trimming by Formatting

Once the lexicographical content of the printed dictionary had been decided upon, the next challenge was to find a workable way of presenting that content in as concise a manner as possible while ensuring that the dictionary was user-friendly. Quite a lot of time went into typographic design features (fonts, colours, spacing, leading, etc) and other aesthetic elements. We used a sans-serif font for the English-language content (Atlas Grotesk Bold) in 7.8pt and a serif font for the Irish-language content (Lyon Text Regular No. 2) in 8.5pt. This font is larger than that which would usually be expected in a dictionary of this size, however as this was the first major bilingual Irish dictionary in over 40 years, we were keen to produce a premium quality book that was both readable and functional.

The layout of entries in the online dictionary is quite vertical in style, with a line break before every sense, example and base translation, for example:

gentleman

1 *NOUN* man

fear *masc1* 🗣️ **C M U**

ladies and gentlemen! *a dhaoine uaisle!*

an older gentleman *seanfhear* *masc1*, *seanduine* *masc4*

he's a gentleman of leisure *tá saol an mhadra bháin aige, níl cúram sa saol air*

2 *NOUN* polite, considerate man

fear uasal 🗣️ **C M U**

he's a true gentleman *fear uasal cruthanta is ea é, corp an duine uasail atá ann*

Fig. 7: Entry 'gentleman' on NEID website showing use of line breaks

For the printed version, we needed to remove as much white space as possible and to fit the entries into as small a space as possible without detriment to the readability or usability of the dictionary.

4.2.1 The Point of no RETURNS

Keeping line breaks to a minimum was the first priority in the print edition. Each new entry would, of course, have to start with a line break but all other options were on the table. Having consulted a range of similar bilingual dictionaries and considered all the options in InDesign – the publishing software used to design the book – it was decided to insert line breaks based on the following criteria, something we felt was a happy medium between readability and space-efficiency:

- before each new entry;
- before each part of speech, except the first one in the entry;
- before each phrasal verb;
- before each multi-word entry block (i.e., containing phrases and/or phrasal verbs);
- before each phrase and phrasal verb within those blocks, except the first one;
- before each subform of entries.

Apart from the instances mentioned above, content would run on from line to line. The most discussed item here was the decision on whether to put a line break between part of speech blocks, something that works reasonably well in mid-sized and large entries but looks somewhat clunky or untidy in small entries with no example sentences. In ‘abseil’ below, for example, you can see from the surrounding entries that there would have been space on the first line for the POS marker and for the first word of the next sense. On the other hand, in the mid-sized entry ‘acid’ the POS marker starting on a new line makes the entry clearer and more readable. This becomes even more apparent in large entries.

abscond v (*escape*) v₁ éalaigh, teith
abseil v v_T & v₁ rópáil anuas
N rópáil anuas
abseiling N rópadóireacht anuas
absence N ① (*non-existence*) easpa
*f*₄, uireasa *f*₄, ceal *m*₄; **in the ~ of**

Fig. 8: Entry ‘abseil’ in CEID

acid N ① CHEM, MED aigéad *m*₁;
folic ~ aigéad fólach; **lactic** ~ aigéad
 lachtach; **stomach** ~ aigéad goile ②
 (*drug*) aigéad *m*₁; **an ~ trip** siabhrán
 aigéid
ADJ ① CHEM aigéadach *a*₁, aigéid
nmod; ~ **soil** ithir aigéadach ② (*crit-*
ical) binbeach *a*₁, nimhneach *a*₁; **he**
gave me an ~ look thug sé féach-
 aint bhinbeach orm, thug sé amharc
 nimhe orm ③ (*sour*) searbh *adj*, géar
*a*₁; **an ~ flavour** blas géar

Fig. 9: Entry ‘acid’ in CEID

4.2.2 Swung Dashes

It is a common practice in printed dictionaries to use swung dashes or tildes to denote the headword within that headword’s entry (Bringhurst, 2012, p. 318), however it wasn’t as simple as replacing every instance of the headword with a swung dash and moving on.

It was a simple decision to use swung dashes in example sentences. We felt users could easily understand the convention and also be able to handle ‘~s’ or ‘~ed’ for plurals or verbs.

adjust v ① (*alter slightly*) vT socraigh, deisigh; **he ~ed his tie** shocraigh sé a charbhat, dheisigh sé a charbhat; **you can ~ the acidity of the soil** is féidir aigéadacht na cré a shocrú ② (*for accuracy, effect, etc*) vT

Fig. 10: Part of entry 'adjust' in CEID

We decided to also use the swung dash in the headings of phrases and idioms. They were generally in close proximity to the headwords and, again, we felt users wouldn't have a problem dealing with them. See example below from 'addition':

PHRASES ▪ **in ~** chomh maith leis sin, freisin, fosta, ina theannta sin, lena chois sin; **in ~, it's available online** chomh maith leis sin, tá sé le fáil ar líne; tá sé le fáil ar líne chomh maith
▪ **in ~ to** PREP chomh maith le, i dteannta, le chois; **in ~ to music, they do art** chomh maith le ceol, déanann siad ealaín

Fig. 11: Phrases and idioms in entry 'addition' in CEID

One complication that is evident in the examples under 'addition,' however, is that the swung dash in the example sentences only stands for the headword of the entry rather than the sub-headword or the phrase in question. This is also an issue in the case of variant forms of phrasal verbs:

act on (*also ~ upon*) ① (*take action*) vT
gníomhaigh de réir, déan de réir;
to ~ on the recommendations of sb gníomhú de réir moltaí duine;

Fig. 12: Phrasal verb 'act on' in CEID

In the case above, an unskilled user could take it that the swung dash stands for the entire phrasal verb and as a result could think that the variant form of 'act on' in this case is 'act on upon' or even that the example sentence could be read as 'to act on on the recommendations of sb'. On balance, however, it was felt that in general users of the dictionary would understand what the swung dashes stand for in such cases. Secondly, it was not guaranteed that a swung dash in place of the full phrasal

verb would be any clearer than the approach we employed. And finally, the value of the amount of space saved with the swung dashes was too high to forego in order to include full headwords in the phrasal verb sections.

4.2.3 Abbreviations

In the online version of NEID, the parts of speech are spelled out in full for the English headwords and partially for the Irish translations, however they were all abbreviated further in the print edition and explained in the front matter of the dictionary in English and in Irish. The resultant space-saving may be observed in the entry ‘affray’ below, where *noun* became *n*, and *masc1* [masculine noun, first declension] became *m1*.

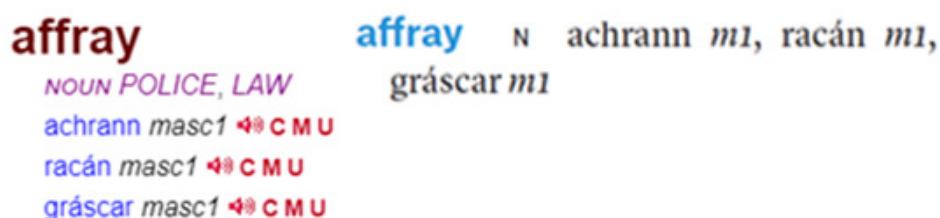


Fig. 13: Entry ‘affray’ on NEID website (left) and in CEID (right)

5. Conclusion

The measures outlined above reduced the dictionary content of CEID from an initial estimate of between 3,500 and 4,000 pages to 1,700 pages of dictionary content plus 100 pages of front matter and supplementary material. The dictionary was printed in 2020 and reprinted in 2021.

References

- Bringhurst, R. (2012). *The Elements of Typographic Style* (4th ed.). Hartley & Marks Publishers.
- de Bhaldraithe, T. (Ed.). (1959) *English-Irish Dictionary*. An Gúm.
- Ó Dónaill, N. (Ed.). (1977). *Foclóir Gaeilge-Béarla [Irish-English Dictionary]*. An Gúm.
- Ó Mianáin, P. (Ed.). (2020). *Concise English-Irish Dictionary*. Foras na Gaeilge.
- Ó Mianáin, P. (Ed.). Foras na Gaeilge. (2024, January). Retrieved from New English-Irish Dictionary: <https://www.focloir.ie>
- Ó Mianáin, P., & Convery, C. (2014). From DANTE to Dictionary: The New English-Irish Dictionary. In A. Abel, Ch. Vettori, & N. Ralli (Eds.), *Proceedings of the 16th EURALEX International Congress* (pp. 807–817). Institute for Specialised Communication and Multilingualism.

Ó Mianáin, P., & Convery, C. (2014). New English-Irish Dictionary. *Journal of the Dictionary Society of North America*, 35, 318–333.

Pödör, D. (2022). Pádraig Ó Mianáin (chief editor). 2020. Concise English-Irish Dictionary. Foclóir Béarla-Gaeilge. *International Journal of Lexicography*, 35(1), 129–135.

Contact Information

Cormac Breathnach

Foras na Gaeilge
cbreathnach@forasnagaeilge.ie

Pádraig Ó Mianáin

Foras na Gaeilge
pomianain@forasnagaeilge.ie

