Abstract
This paper presents a database of Irish-language idioms along with a description of its internal schema. This is a dialect-specific synchronic study of idioms in the Irish of West Munster from the published work of Canon Peadar Ó Laoghaire (1839–1920). Firstly, a brief overview is given of previous research undertaken on Irish-language idioms in the area of phraseology. The database schema is outlined in the latter section of the paper. Each of the individual categories in which idioms are recorded, categorized, explained and cross-referenced in the database are described in detail. A selection of sample entries is provided in the Appendix to this paper. This research provides an innovative approach to the collection and categorization of Irish-language idioms and also a foundation for further analysis of idioms in other dialects and languages and for future studies in this area.

Keywords: lexicography; language technology; phraseology; collocation; database schema

1 Introduction
This paper focuses on the creation of an idiom database from the published work of the primary Irish-language author of the late nineteenth and early twentieth century, Canon Peadar Ó Laoghaire (1839–1920). Following the description of idioms as laid down in the literature (e.g. Abdou 2012; Hanks 2004; Howarth 1998; Fernando 1996; etc.), an idiom is defined in this paper as a type of phraseme which has a figurative meaning in terms of its whole, or a unitary meaning that cannot be derived from the meanings of its individual components and whose components can only be varied within restricted definable limits.

Firstly, a general overview is given on previous research undertaken on Irish-language idioms. A brief background of Peadar Ó Laoghaire will be provided along with a synopsis of his work. A corpus of 1,864,015 tokens was compiled as part of this research but an account of the compilation and the search methodology is beyond the scope of this paper. However, a detailed description of this aspect of the project is available in Ní Loingsigh and Ó Raghallaigh (2016). The primary focus of this paper is the creation of a database in which idioms are recorded, classified, explained and cross-referenced. The database schema is described in detail and a select number of sample entries from the database are available in the Appendix to this paper. The database will be made publicly available at the following link: http://www.gaois.ie/en/.

2 Background
The Irish language belongs to the Celtic Indo-European language group. There are three main dialect varieties in use today which correspond primarily with the most northerly, westerly and southerly geographic regions in Ireland in the coastal counties of Ulster, Connaught and Munster. Ó Corráin’s (1989) work on the Ulster dialect of Irish and idiomatic expressions identified in the work of the author Séamus Ó Grianna (1889-1969) is the only other academic study undertaken on Irish-language idioms in literary works to date. This concordance of idiomatic expressions was undertaken before the widespread availability of large-scale corpora and followed a different methodology to the corpus
approach followed in this project. Despite the limited amount of academic research on Irish-language phraseology, idioms have been collected and analysed by individual scholars and in various lexicographic works from the late nineteenth century onwards.

This paper builds on the foundations established in Ó Corráin’s (1989) initial work on Irish-language idioms and specifically focuses on idioms in the work of Peadar Ó Laoghaire and on the southern dialect of Irish, specifically the Múscraí region in County Cork. Ó Laoghaire is recognised as a seminal writer of the Irish-language revival period in Ireland at the turn of the twentieth century and his work left an indelible mark on generations of writers and students. Ó Laoghaire produced over five hundred individual pieces of work, including extensive prose compositions, translations, newspaper and journal articles (Ó Cuív 1954). Additionally, hundreds of unpublished letters, articles and notes relating to his work are currently held in various libraries throughout Ireland (Ó Dúshláíne 2015). It was this volume of work, but also the undisputed authority of his work, that gained him praise and respect both nationally and internationally as a writer (e.g. Ó Macháin 2015; Ó Cearúil 1999; Mac Mathúna 1987; Keane 1970; Maol Muire 1939; Zimmer 1911.) His self-proclaimed writing style became known as ‘caint na ndaoine’ (the people’s speech) which focused on the spoken language as a base for literature rather than building on the remnants of a more archaic classical literary tradition.

Ó Laoghaire’s primary and most renowned work Séadna (Ua Laoghaire 1904) received widespread acclaim upon its release. For these reasons Peadar Ó Laoghaire’s published works were selected for this project, not only as a comprehensive source of native idiomatic expressions, as a basis of phraseological work and linguistic studies “but also towards the general understanding of language and of texts as cultural products” (Biber, Breiteneder and Dobrovolskij 2002). An annotated corpus was used to search for idioms following a mixed-methods approach, i.e. both corpus-based and corpus-driven methods were followed in the idiom search. Following the collection of idioms, a database schema was set out specifying which categories of information needed to be recorded. The classification and categorization of idioms in the database will be described in detail in the following section.

3 Methodology

The database was created using Léacslann (Měchura 2012), a web-based platform used for building dictionary writing systems and terminological management systems and is currently used in a number of Irish-language based projects in Fiontar, Dublin City University, e.g. the National Terminology Database for Irish (www.tearma.ie); the project to digitize the National Folklore Collection of Ireland (www.dúchas.ie) and others. Moon (1998: 45) lists several databases of fixed expressions and idioms in various languages and suggests that each database “records and prioritizes only certain kinds of information”. The database schema in this project was created to record, classify and cross-reference the collected Irish-language idioms. Léacslann holds various XML-based stocks which cater to the needs of several research projects in Fiontar (Měchura 2012) and a discreet stock with a specific schema was created to cater for the needs of this project. The idiom database is stored in this stock in Léacslann under the title An tAthair Peadar and each category within this stock is described in detail below.

There are seven main categories in the database, each of which are displayed by Léacslann alphabetically: aicme comhréire (syntactic class), aicme séimeantaice (semantic class), aicme úsáide (usage class), faisnéiseoir (informant), foinse bhreise (extra source), foinse chorpais (corpus source) and nath (idiom)1 (see Figure 1). Entries of each category, when created in Léacslann, are stored as XML and comprise of their root element and a set of precomposed child-elements. Entries can be

1 While labels in Léacslann are in Irish, the title of each label will be translated into English for the purposes of this paper. All translations are the authors’ own unless stated otherwise.
edited and saved (sábháil) or deleted (scrios). The user can revert to a previous version of the entry using the history (stair) option.

The data captured in each of the seven categories of the database are described in the following section with particular focus on the most important entry category of the database, i.e. nath (idiom).

### 3.1 Database categories

**Aicme comhréire** (syntactic class): Eleven primary syntactic classes of idioms were recorded as entries of type aicme comhréire following an analysis of the idioms collected. This is a broad syntactic classification as a detailed classification of the various syntactic structures of Irish-language idioms has yet to be undertaken, i.e. idioms containing a verbal noun are recorded as verbal noun idioms (e.g. ag imeacht le gaoth²), idioms containing an irreversible binominal are recorded as irreversible binominal idioms (e.g. ó mhaidin go hoíche³) and so forth.

**Aicme séimeantaice** (semantic class): Four main semantic classes were set out in the database as entries of type aicme séimeantaice. This follows the semantic classification as developed and reported in the literature (e.g. Howarth 1998; Fernando 1996; Cowie, Mackin and McCaig 1983; etc.) and as summarised by Mulhall (2008: 1358). Idioms are classified as pure idioms, figurative idioms, semi-idioms and non-word idioms.

**Aicme úsáide** (usage class): Several usage classes which related to the use of certain idioms were recorded as entries of type aicme úsáide, i.e. common intensifiers (e.g. numerics, colours, etc. in idioms (ó chúig cúigí na hÉireann⁴)); negative usage of idioms (e.g. idioms which are not used in their positive sense or meaning) (níl leathphingin rua aige⁵)); restrictions on lexical variation, etc.

**Faisnéiseoir** (informant): Due to the dialectal and synchronic nature of the idioms recorded in the database, certain idioms were discussed with native Irish-speakers from Múscraí to clarify their underlying meanings. This primarily concerned idioms of which there was little or no evidence in other sources regarding their meaning and description. The respondents name, address, profession and age were recorded as entries of type faisnéiseoir.

---

2 *Ag imeacht le gaoth*: literal meaning: ‘going with wind’; idiomatic meaning: ‘in decline’.
3 *Ó mhaidin go hoíche*: literal meaning: ‘from morning until night’; idiomatic meaning: ‘all day long, continually’.
4 *Ó chúig cúigí na hÉireann*: literal meaning: ‘from the five provinces of Ireland’ (which historically had five provinces); idiomatic meaning: ‘from the whole of Ireland, from everywhere’.
5 *Níl leathphingin rua aige*: literal meaning: ‘he does not have half a red penny’; idiomatic meaning: ‘he is penniless, he has little or no money’.
Foinse bhreise (extra source): A complete list of additional resources examined during the analysis of each idiom was recorded. Each source was stored as an entry of type *foinse bhreise*. The title, author and year of publication of each resource was recorded. These extra sources are primarily lexicographic and dialectal sources used as aids to explain the collected idioms.

Foinse chorpais (corpus source): A complete list of corpus sources, which recorded the title, author and year of publication of each source, was compiled and the items on the list were stored as entries of type *foinse chorpais*.

*Nath* (idiom): Each idiom was recorded in its baseform or canonical form. Idioms were stored as entries of type *nath*. The structure of this entry category is described in the following section of this paper. Example entries of this type are provided in the Appendix to this paper.

### 3.2 Internal structure and contents of entries of the category Nath in the database

*Buunfhoirr* (baseform): Each idiom consists of a baseform or canonical form, “the simplest morfo-syntactic form that an expression can take” to activate its specific idiomatic meaning and function (Barkema 1996: 141). Due to the dialectal nature of this study, each idiom was recorded in its baseform in standard orthography, as established by *An Caighdeán Oifigiúil* (Rannóg an Aistriúcháin 1958). However, a small number of entries contain some non-standard components. These exceptions to the general rule were included when strong evidence from the corpus suggested that the idioms collected contained certain fossilized components, e.g. *ag stracadh na bpíobán,*\(^6\) *do lá agus d’oíche,*\(^7\) etc. Priority is given to the non-standardized forms of idioms due to the prevalence of their use in the spoken language. Prioritizing and recording the standardized form of these idioms would be of minimal benefit to those who wish to use these idioms as they predominantly occur in non-standardized form. However, both a standardized and a non-standardized version of the idiom could be recorded in the database in the future as a basis for comparison. In instances such as these, the idioms concerned were discussed with language respondents in the *Múscraí* region and left in non-standardized form.

In cases where an idiom contained a verb and a noun as part of its baseform, these idioms were recorded in the third person, singular form with the representative verbal noun, e.g. *fuil a dhoirteadh; anáil a tharraingt; bob a bhualadh (ar); carbhat caol a chur (ar); sciot scot a dhéanmh (de);* etc. Additionally, when evidence from the corpus suggested that certain verbs and prepositions formed an integral part of the idiom concerned, these were included in the baseform. This reflects the importance of certain prepositions and verbs in Irish-language syntax (Ó Baoill and Ó Domhnailláin 1975). The relative preposition was recorded in brackets as shown in the previous examples.

*Ceannfhocal* (headword): In each idiom the primary nouns, adjectives and adverbs, in singular form, along with the infinitive form of each verb were recorded. This is a benefit of using a digital database over a printed dictionary of idioms as the researcher/user is not limited to selecting one headword. These headwords can be further utilized as a cross-referencing tool to search for idioms containing similar headwords.

*Sampla corrais* (corpus example): Up to three contextual examples of each idiom from the corpus were recorded as *sampla corrais* child elements within entries of type *nath* (idiom) in the database. Each contextual example was transcribed directly from the source text and recorded in non-standardized form.

---

\(^6\) Non-standardized form: *ag stracadh na bpíobán*; standardized from: *ag sracadh na bpíobán.*

\(^7\) Non-standardized form: *do lá agus d’oíche*; standardized from: *de lá agus d’oíche.*
Cur síos (description): Each idiom was paraphrased and explained in clear and concise language. These explanations were based on information found in various lexicographic works, in specific dialectal resources and from fieldwork with Irish-language informants in Múscraí.

Sample breise (additional example): Any additional information obtained regarding the meaning of idioms was recorded under this element. This included information from both published and unpublished lexicographic works, specific dialectal resources, lexicographic manuscripts, private notes and correspondence from Peadar Ó Laoghaire, etc. The examples recorded were directly transcribed from the relevant sources. The source of each additional example, as well as the headword of each example, were stored as part of these elements.

Aicme úsáide (usage class): One or more usage classes were selected (if necessary) which related to the correct use of the idiom concerned. These elements contained a reference to an entry of type aicme úsáide (see 3.1).

Aicme séimeantaice (semantic classification): One semantic class was selected for each idiom. These elements contained a reference to an entry of type aicme séimeantaice (see 3.1).

Aicme comhréire (syntactic classification): One or more syntactic classes were selected for each idiom. These elements contained a reference to an entry of type aicme comhréire (see 3.1).

Crostagairt (cross-reference): Zero or more cross-references to mnemonic idioms were selected for each idiom. This option allows the user to connect idioms with a similar meaning together. These elements contained a reference to an entry of type nath (idiom).

Eolas breise (additional information): Any additional information received from the dialectal respondents in Múscraí was recorded under this element.

4 Results and Conclusion

A total number of 420 idioms in their baseform and in standard orthography were recorded in the database. Additionally, over 1,000 corpus examples of the idioms were recorded in which the variation and use of each idiom are available in contexts. Each idiom is classified both semantically and syntactically and cross-referenced with similar idioms in several instances through a specific cross-reference and also through similar headwords.

This research presents an overview of an indepth dialectal analysis of Irish-language idioms found in the published work of Peadar Ó Laoghaire. The collection of idioms is recorded, classified and cross-referenced in a digital database. This research builds on the foundations laid down by Ó Corráin (1989) and presents native dialectal idioms from the early twentieth century in a new light. The idioms collected provide an insight into the interrelationship between the Irish language and culture, which could be contrasted and compared with other dialects and also with other languages in the future research. The database schema can be used and adapted to cater for other languages and idiom-related projects and can also be expanded to cater for future phraseological-based research in the area.

5 References


**Acknowledgements**

This research was undertaken with support from Fiontar, Dublin City University.
### Appendix

#### Example A

<table>
<thead>
<tr>
<th>ID</th>
<th>4588205</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BUNFHIRM</strong></td>
<td><strong>ag imeacht le craobhacha</strong></td>
</tr>
<tr>
<td><strong>CEANNFHOCL</strong></td>
<td>creobh</td>
</tr>
<tr>
<td><strong>CEANNFHOCL</strong></td>
<td>imigh</td>
</tr>
<tr>
<td><strong>SAMPLA CORPAIS</strong></td>
<td>&quot;Is é crich an sghéil é,&quot; ar saisean, &quot;go bhfuil eagal orm go n-eirgheóchaigh a chroidhe ar Dhiarmuid mura bhfuil eirighthe cheana aige air, agus go n-imtheóchaigh Sadhbh bhocht le craobhachaibh.&quot;</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Seadna</td>
</tr>
<tr>
<td><strong>LEATHANACH</strong></td>
<td>145</td>
</tr>
<tr>
<td><strong>CUR SÍOS</strong></td>
<td><strong>Ag dul le báiní; bheith ar buile.</strong></td>
</tr>
<tr>
<td><strong>SAMPLA BREISE</strong></td>
<td>1. (e) Dúil, imeacht, le craobhacha, to go wild, mad. Tá sé ar, le, craobhacha, he is raging mad.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Foclóir Gaeilge–Béarla (Ó Dónaill)</td>
</tr>
<tr>
<td><strong>CEANNFHOCL</strong></td>
<td>craobh</td>
</tr>
<tr>
<td><strong>SAMPLA BREISE</strong></td>
<td>To rampage (about), bheith ag imeacht le craobhacha.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>English–Irish Dictionary (de Bhaldraithe)</td>
</tr>
<tr>
<td><strong>CEANNFHOCL</strong></td>
<td>rampage</td>
</tr>
<tr>
<td><strong>SAMPLA BREISE</strong></td>
<td>Dúil le craobhachaibh, to go mad; lit. to go on the bushes, i.e. to go wildly about, catching the clothes upon the bushes.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Foclóir do Shíasaíochta (Borthwick)</td>
</tr>
<tr>
<td><strong>CEANNFHOCL</strong></td>
<td>craobhachaibh</td>
</tr>
<tr>
<td><strong>AICME SÉIMEANTAICE</strong></td>
<td>Fionnath</td>
</tr>
<tr>
<td><strong>AICME COMHRÉIRE</strong></td>
<td>Nath an im braithartha</td>
</tr>
</tbody>
</table>
### Example B

<table>
<thead>
<tr>
<th>ID</th>
<th>4631149</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BUNFHÓIRM</strong></td>
<td><strong>cos ar bolg</strong></td>
</tr>
<tr>
<td><strong>CEANNFHÓCAL</strong></td>
<td><strong>bolg</strong></td>
</tr>
<tr>
<td><strong>CEANNFHÓCAL</strong></td>
<td><strong>cos</strong></td>
</tr>
<tr>
<td><strong>SAMPLA CORPAIS</strong></td>
<td>Ansin do tháining cealt aigne agus éitheach agus marbhughadh daoine, agus goid agus fudach agus éagóir, agus éigean agus lárnálaidir agus cos ar bolg.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Don Clochóta (Cuid II)</td>
</tr>
<tr>
<td><strong>LEATHANACH</strong></td>
<td>93</td>
</tr>
<tr>
<td><strong>SAMPLA CORPAIS</strong></td>
<td>Dhein na Sasanaigh cos ar bolg ar na daoine bochta, agus i n-inead aon chusaint a dhéanamh ar na daoine is amhladh a thog an dlighe páirt leis an eugcóir agus leis an gcos ar bolg.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Comhairle ar Leasa</td>
</tr>
<tr>
<td><strong>LEATHANACH</strong></td>
<td>155</td>
</tr>
<tr>
<td><strong>SAMPLA CORPAIS</strong></td>
<td>Ni bheas liom-sa, ar seisean, ná le righ Sasana, ná le righ Brittain, éagóir agus ansmacht agus cos ar bolg a dhíheisgint a dhéanamh aon treabhdhas daoine, suas ná iséal, gan gach ní bheadh ar mo chumas do dhéanamh chun cosg do chur le héagóir, chun láidir do leagadh agus lag do thógáil.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Lughaidh Mac Con</td>
</tr>
<tr>
<td><strong>LEATHANACH</strong></td>
<td>46</td>
</tr>
<tr>
<td><strong>CURSÍOS</strong></td>
<td>Ansmacht.</td>
</tr>
<tr>
<td><strong>SAMPLA BREISE</strong></td>
<td>Cos-ar-bolg, a trampling of people under foot.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Notes on Irish Words and Usages (Ua Laoghaire 1926)</td>
</tr>
<tr>
<td><strong>LEATHANACH</strong></td>
<td>79</td>
</tr>
<tr>
<td><strong>SAMPLA BREISE</strong></td>
<td>Is maith é an smacht gan amhras. Ni smacht atá agat-sa d’a dhéanamh anmhairch, ach cos ar bolg. Control is a good thing, certainly; but it is not control you are exercising, but downright oppression.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Mion-chaint: Cuid III (Ua Laoghaire 1902)</td>
</tr>
<tr>
<td><strong>LEATHANACH</strong></td>
<td>26</td>
</tr>
<tr>
<td><strong>SAMPLA BREISE</strong></td>
<td>Cos ar bolg, oppression.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Fodhór Gaeilge-Bearla (Ó Dónaill 1977)</td>
</tr>
<tr>
<td><strong>CEANNFHÓCAL</strong></td>
<td>cos</td>
</tr>
<tr>
<td><strong>SAMPLA BREISE</strong></td>
<td>D'éinleadh cos ar bolg ar na Gaedhil tarléis cath Cionnta'sile. D'éinleadh leathramh ortha.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>Cúrsa Caimse na Gaedhilge (Mac Séighin 1940)</td>
</tr>
<tr>
<td><strong>LEATHANACH</strong></td>
<td>19</td>
</tr>
<tr>
<td><strong>SAMPLA BREISE</strong></td>
<td>To tyrannize over someone, cos ar bolg, ansmacht, a imirt ar dhúine.</td>
</tr>
<tr>
<td><strong>FOINSE</strong></td>
<td>English-Irish Dictionary (de Bhaldráith 1959)</td>
</tr>
<tr>
<td><strong>CEANNFHÓCAL</strong></td>
<td>tyrannize</td>
</tr>
<tr>
<td><strong>AICME SEIMEANTAICE</strong></td>
<td>Fíonacht</td>
</tr>
<tr>
<td><strong>AICME COMHRÉIREACHTA</strong></td>
<td>Nath ainmhatadach</td>
</tr>
<tr>
<td><strong>CROSTAGAIRT</strong></td>
<td>gabhail de choisa</td>
</tr>
</tbody>
</table>
**Example C**

**ID**

4744542

**BUNFHÓIRM**

**CEANNFHÓCAL**

Conchubhar

**CEANNFHÓCAL**

cúinne

**CEANNFHÓCAL**

mór

**SAMPLA CORPAIS**


**FOINSE**

Mo Scéal Féin

**LEATHNACH**

19

**OUR SÍOS**

**Easpa mhór; ganntanas bia.**

**SAMPLA BREISE**

Tá Conchubhar Mór li Mhathúna (Mathghathanna) sa chúnne aige = Tá laaraí sa drom aige (gothb) = An dromhaointeas.

**FOINSE**

Láimhscríbhinn (Acadamh Ríoga na hÉireann): M055 Cnuasach Focal é Mhúscraí (Ó Buachalla) Conchubhar

**SAMPLA BREISE**

bheadh "Conchubhar Mór sa chúnne" ag a lán: bheadh easpa mhór ar a lán. Conchubhar Mór = an gorta, an t-eoras.

**FOINSE**

Fodhóir agus nótaí do "Mo Scéal Féin" a scríobh an tAthair Peadar Ó Laoghaire (Ó hAodha 1970)

**LEATHNACH**

12

**SAMPLA BREISE**

Tá Conchubhar Mór sa chúnne aige. Tá easpa gus gádhtar air.

**FOINSE**

Cortha Caimnti na Gaedhilge (Mac Sibhgh 1940)

**LEATHNACH**

72

**SAMPLA BREISE**


**FOINSE**

Sil an eolas agus eagna an Ghaeil (Ó Cadhlaigh 1925)

**LEATHNACH**

190

**SAMPLA BREISE**

Conchubhar Mór sa chúnne: gátar nó gannntanas a bheith ann mar bhí ag daoine bochta aimsir an ghorta.

**FOINSE**

Lán-chúrsa na Gaeilge (Ó Riain 1952)

**LEATHNACH**

43

**AICME SÉIMEANTAICE**

Fiormh

**AICME COMHRÉIRE**

Nath abairte