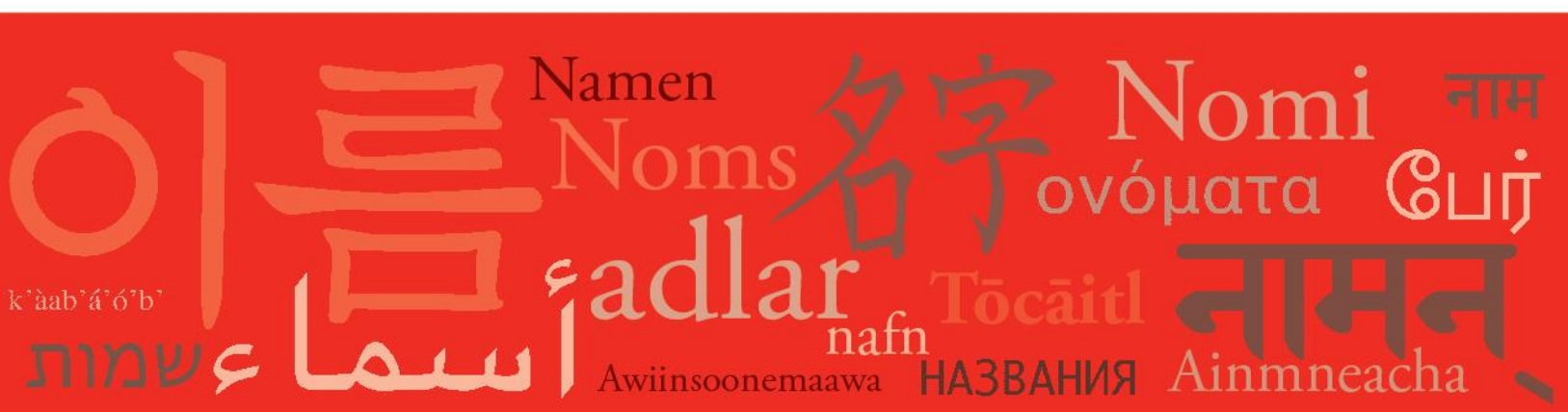


Names | A Journal of Onomastics



Developing the Gaois Linguistic Database of Irish-language Surnames

Brian Ó Raghallaigh

Dublin City University

Michal Boleslav Měchura

Dublin City University

Aengus Ó Fionnagáin

University of Limerick

Sophie Osborne

Dublin City University

ans-names.pitt.edu

ISSN: 0027-7738 (print) 1756-2279 (web)

Vol. 69, Issue 1, Winter 2021

DOI 10.5195/names.2021.2251



Articles in this journal are licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).



This journal is published by the [University Library System](https://www.library.pitt.edu/), University of Pittsburgh.

Abstract

It is now commonplace to see surnames written in the Irish language in Ireland, yet there is no online resource for checking the standard spelling and grammar of Irish-language surnames. We propose a data structure for handling Irish-language surnames which comprises bilingual (Irish–English) clusters of surname forms. We present the first open, data-driven linguistic database of common Irish-language surnames, containing 664 surname clusters, and a method for deriving Irish-language inflected forms. Unlike other Irish surname dictionaries, our aim is not to list variants or explain origins, but rather to provide standard Irish-language surname forms via the web for use in the educational, cultural, and public spheres, as well as in the library and information sciences. The database can be queried via a web application, and the dataset is available to download under an open licence. The web application uses a comprehensive list of surname forms for query expansion. We envisage the database being applied to name authority control in Irish libraries to provide for bilingual access points.

Keywords: anthroponymy, surnames, database, Irish, Celtic, Gaelic

Introduction

In this paper, we describe the development of the *Gaois* linguistic database of Irish-language surnames. Irish-language surnames are defined here as surnames written in Irish, whether of Gaelic origin or not. We refer to them as Irish-language surnames throughout to avoid confusion with surnames of Gaelic ethnic origin specifically, as our database contains surnames of both Gaelic and non-Gaelic origin, e.g. English, Norman. The database could be modified to include new Irish-language surnames of Polish, Chinese, Nigerian or other origins in the future, should these be gaelicised. The language is referred to as *Irish* and not *Irish Gaelic*, in keeping with ISO 639 and to distinguish it, crucially, from the Gaelic ethnicity, as well as from the (Scottish) *Gaelic* language. The research is based on surname data from the 26 counties of the Republic of Ireland, but could be expanded to include data from the 6 counties of Northern Ireland which is part of the UK.

Surnames first came into use in Ireland in the 10th century (Ó Cuív 1986, 34; Mac Mathúna 2006, 65) when Irish was both the literary and vernacular language of Ireland. Most Irish-language surnames are patronymic multiword expressions and take the form ‘grandson/descendant of X’, e.g. *Ó Briain* or ‘son of X’, e.g. *Mac Cárthaigh*. In Irish, the form of the name changes in different grammatical contexts, e.g. *Seán Mac Mathúna* ‘Seán McMahon’ (NOM), *feirm Mhuintir Mhic Mhathúna* ‘The McMahon family farm’ (GEN), *a Shíle Nic Mhathúna* ‘[Dear] Síle [daughter of] McMahon’ (VOC). Traditionally, the form of the name also changes with regard to sex and marital status, e.g. *Seán Ó Ceallaigh* ‘Seán descendant of Ceallach’, *Síle Uí Cheallaigh* ‘Síle wife of Ceallach’, *Síle Ní Cheallaigh* ‘Síle [unmarried] daughter of Ceallach.’

Following the collapse of the Gaelic order in the mid-seventeenth century, the linguistic hierarchy changed, and English became the predominant language in the administrative and legal spheres in Ireland (O’Rahilly 1976, 15; Ó hUiginn 2008, 8). From that point onwards Irish-language surnames were anglicised, usually phonetically (Woulfe 1923, 36; de hÓir 1973, 192). *Ó Briain* became *O’Brian*, *Mac Cárthaigh* became *McCarthy*, and so on. From the late nineteenth century, however, due to a revival of the Irish language throughout the island of Ireland in concurrence with the nationalist movement that led to Irish independence, use of the Irish forms of surnames has become commonplace again, and began to be supported by Government policy towards the end of the 20th century:

The [Irish language steering group of the Department of the Environment] recommends that the forms of personal names should not be subject to arbitrary alteration during the administrative process and, in particular, that the Irish forms of surnames and Christian names, when used by members of the public, should never be changed by Local Authority staff to the associated English forms; Seán Ó Conaill should not be altered to John O’Connell, nor vice versa. (Department of the Environment 1995, 12).

It is in this context of the acceptance and the normalisation of Irish-language surnames once again, that we approach this topic. Nowadays, the spelling of Irish-language surnames is largely standardised since the publication of official spelling rules for Irish in 1958 (Government of Ireland 1958) and the publication of the surname dictionary *An Sloinnteoir agus an tAinmneoir* in 1966 (Ó Droighneáin 1966). Our aim is to provide an online resource where users can check the spelling and grammatical forms of surnames in Irish, and our hope is that our dataset will be used to enhance search algorithms in library catalogues and citation databases.

Context

Since 2003, with the enactment of the *Official Languages Act* (Government of Ireland 2003), the Government of Ireland supports the use of both official languages of the state (i.e. Irish and English (Government of Ireland 1937)) for official purposes. And while Irish-language surnames are now generally accepted, actually using one can prove challenging. About 40% of the population of Ireland can speak Irish (Central Statistics Office 2016). This leaves about 60% of the population, however, that might have difficulty in writing an Irish-language surname. In addition, information systems (e.g. web forms, databases, digital displays, etc.) do not always allow for Irish-language surnames, which can include spaces and accented characters, either in place of English-language names or as alternative forms. While there is no longer a valid technical reason for this, organisations do not tend to prioritise fixing this issue. Classic examples of this are the Irish Rail ticketing/seating system (Irish Language Commissioner 2020, 20–21; The Irish Times View 2020) and the Central Statistics Office (CSO) vital statistics records. The CSO were criticised during Irish Parliamentary Questions (Ó Cuív 2018) for continuing until recently to anglicise all Irish names in their records by removing all accents (Central Statistics Office 2017).

Our primary goal is to provide an online resource that supports the use of Irish-language surnames in Ireland by giving users and application developers access to the standard Irish-language spellings as well as grammatical information via the web. Existing resources do not provide this. While *An Sloinnteoir Gaeilge agus an tAinmneoir* (Ó Droighneáin 2013) provides the standardised spellings, it does not provide grammatical information. Its structure as an English–Irish list also limits its usefulness. Other reference works, including Muhr and Ó hAisibéil’s *A Dictionary of Family Names of Ireland* (2021), the Hanks et al. *Oxford Dictionary of Family Names in Britain and Ireland* (2016), and MacLysaght’s *The Surnames of Ireland* (2013), are all focused on the anglicised surname forms primarily, and spellings in MacLysaght are sometimes out of date. None provide inflected Irish forms.

Our secondary goal is to provide users with an equivalence mapping between standard Irish-language and English-language forms of surnames in Ireland. It is generally accepted in onomastics that anthroponyms (i.e. people’s names) do not have synonyms, and are not translatable. This is because personal names, unlike words which are easily translatable, are essentially labels, without immediate lexical significance, that refer to an individual (Coates 2006, 312). In Ireland, however, names are regularly anglicised and de-anglicised. All Gaelic Irish surnames in Ireland have two forms, one in Irish and one in English, e.g. {*Ó Riain, Ryan*}. Some have more than two forms, e.g. {*Ó Dochartaigh, Doherty, O’Doherty*}, {*Mac Amhlaoibh, Cawley, McAuliffe, McCauley, McAuley*}. Given this situation, we use the terms *synonym* and *equivalent* in relation to surnames in this paper. Our use of these terms, however, is not at odds with the theory, as described in Coates (Ibid.), because we do not use them with regard to the meaning of the surname. We use them as a kind of shorthand to refer to the relationship between one label and another label when it can easily be claimed that the labels are forms of the same name as opposed to two different names. If the two labels are in the same language, we call them synonyms, and if the two labels are in different languages, we call the relationship between them an equivalence.

To achieve our goals, we have devised a structure that comprises clusters of surname forms that are synonyms or equivalents of each other. We give the standard form or forms, and since it is not always possible to make a binary determination, our structure allows us to capture additional forms that are in use. These are labelled as either historical or modern alternative forms. We do not store forms that we determine to be non-standard or incorrect (i.e. where spelling rules are broken) because these would be too numerous. Unlike countries like Sweden, where there is legislation encouraging the use of standard forms of names (Brylla 2016), the situation in Ireland enables the establishment of non-standard or incorrect versions of names, particularly in Irish. Our web application helps users to find the standard version from a non-standard or misspelled version with query expansion, based on the Irish Surname Index (National Folklore Collection 2020), and using standard spelling distance algorithms (i.e. spell checking). To provide

grammatical information, we have developed an algorithm for generating all inflected forms of all Irish-language surnames. The output of the algorithm is included in the dataset and is available via the web application. This is a substantial result in itself in that we have created the first morphological database of Irish-language surnames. There already exists a morphological database of Irish placenames (Government of Ireland 2020) and a morphological database of Irish vocabulary (Foras na Gaeilge 2020; Méchura 2014), but no such resource exists for surnames to date.

We envisage this resource will be used in a number of different contexts. The first likely context is the educational sphere in Ireland, specifically within the primary school system. Up to 2013, all pupils were registered with the Department of Education and Skills using the Irish-language version of their name. Since 2013, they can be registered in either English or Irish (Department of Education and Skills 2013, 6). Although no longer obligatory, it is still common practice for schools, especially Irish-medium schools, to register pupils with an Irish-language version of their name. This process often involves an ad hoc and historically inaccurate association with an Irish-language equivalent. The second context is the cultural sphere, specifically within Gaelic Games, i.e. Gaelic football, hurling/camogie, handball and rounders. The Gaelic Athletic Association (GAA), the umbrella body for these sports, promotes the use of the Irish language by requiring that player registrations (Gaelic Athletic Association 2020a, 14) and team sheets given to referees (Gaelic Athletic Association 2020b, 8) contain only the Irish-language form of players' names, except where no Irish-language form exists. The third context is the public sphere, where members of the public may choose to change their name to Irish or use the Irish-language form of their names in certain contexts, or indeed to register the names of their children in Irish. Access to the standard Irish-language spelling and grammatical information will be useful in these contexts.

Another application of this resource will be in the library and information sciences, specifically within personal name authority control. Personal name authority files, such as that of the Virtual International Authority File (VIAF), are an important resource for libraries in terms of being able to identify individuals unambiguously (Shi and Jia 2018). Authors and other public figures are identifiable by librarians and archivists by their personal name authority identifier or heading, thus making their related works more discoverable. Kimura (2014, 743) highlights how the absence of properly cross-referenced local variants in name authority data can result in inconsistent search results. As the VIAF does not systematically include Irish-language forms of names in Irish personal name authority records, surname information made available via our resource could be used to build virtual bilingual access on top of the VIAF or other name authority files in Irish libraries.

Research by Byrne and O'Malley (2013) into the origins of the political party systems in Ireland (North and South) used surnames to test their hypothesis that the party system in the Republic of Ireland has some basis in family origin, as reflected in surnames (Gaelic Irish vs Old English) going back to the 12th century. The authors had to handle both anglicised and gaelicised names in their research data. Such research might also benefit from our resource.

Method

The database was created in *Lexonomy*, lexonomy.eu, a cloud-based, open-source platform for writing and publishing dictionaries (Méchura 2017). A custom XML schema was developed for this project. Each entry is a cluster of surname forms (i.e. surnames). Each cluster, in theory, pertains to a particular version of an Irish-language surname. In other words, all *Ó Murchú* surnames are clustered together in one entry, all *Ó Briain* surnames are clustered together in one entry, and so on. It is our proposition that Irish-language surnames can be and are best structured as clusters according to a particular form of an Irish-language surname. If our theory is correct, there will not be much overlap between different clusters. Overlap is possible, however, and is acceptable where there is any uncertainty as to which surname cluster a particular surname belongs. In these cases, the same surname will be found in two or more clusters.

The entries themselves have a reasonably simple structure. Each entry basically contains a list of surnames. Each surname is classified by language, either Irish or English, and by whether it is primary, historical, or alternative. Each Irish-language surname contains one base form, e.g. *Ó Briain*, as well as all the inflected forms derived from the base form, e.g. *Uí Bhriain* and *Ní Bhriain*. All inflected forms are generated from the base form using an expansion algorithm developed for this project. The expansion algorithm recognises which inflectional pattern the surname belongs to. We have identified five patterns:

1. The *Ó* pattern.
2. The *Mac* pattern.
3. The *Mac Giolla/Mac Con* pattern.
4. The *-ch* adjectival pattern, e.g. *Breathnach*.
5. The *anything else* pattern, which does not inflect.

Since the inflectional pattern can be recognised from the base form of the surname, no more information is required. In theory, therefore, it would suffice to encode each surname in one XML element, as follows:

```
<surname-irish>Ó Briain</surname-irish>
```

However, to allow for cases yet to be discovered, where the inflectional pattern cannot be recognised from the base form only, the base form is placed in a redundant child element of `<surname-irish/>`. This is done so that additional child elements can be added to store extra information if necessary. Entries also contain a title, which is the etymological root or origin of the surname, encoded in one element, for example:

```
<title>Brian</title>
<surname-irish>
<form>Ó Briain</form>
</surname-irish>
```

The data can be viewed via the web application and is available to download in two formats, minimal and expanded. In the minimal format, only the base form is stored explicitly in the entry. In the expanded format, in addition to the base form, male/wife/daughter forms and inflected forms are generated, added, and marked up, for example:

```
<surname-irish>
<form gender="male" case="nom"><pre>Ó</pre> Briain</form>
<form gender="male" case="gen"><pre>Uí</pre>
B<mut>h</mut>riain</form>
<form gender="male" case="voc"><pre>Uí</pre>
B<mut>h</mut>riain</form>
<form gender="female" familyStatus="wife" case="nom"><pre>Uí</pre>
B<mut>h</mut>riain</form>
<form gender="female" familyStatus="wife" case="gen"><pre>Uí</pre>
B<mut>h</mut>riain</form>
<form gender="female" familyStatus="wife" case="voc"><pre>Uí</pre>
B<mut>h</mut>riain</form>
<form gender="female" familyStatus="daughter"
case="nom"><pre>Ní</pre> B<mut>h</mut>riain</form>
<form gender="female" familyStatus="daughter"
case="gen"><pre>Ní</pre> B<mut>h</mut>riain</form>
<form gender="female" familyStatus="daughter"
case="voc"><pre>Ní</pre> B<mut>h</mut>riain</form>
</surname-irish>
```

This method of providing minimal and expanded formats is inspired by a similar distinction made in the Irish National Morphology Database (Méchura 2014). The algorithm which converts from minimal to expanded format exists in two implementations, once as an XSL stylesheet (available publicly for download) and once as a function in the C# programming language (used internally by our web application). The database can support any of the alphabetic sorting norms commonly practiced (Ó Droighneáin 2013; MacLysaght 2013; Nic Cóil 2011; Plassard 1996), using information that can be retrieved from the dataset for every form of every Irish-language name currently stored in the dataset, as follows. If the user wants to sort female forms under male forms, the web application has access to information on what male form (e.g. *Ó Raghallaigh*) relates to the female forms (e.g. *Uí Raghallaigh*, *Ní Raghallaigh*). If the user wants

alphabetic sorting that ignores initial particles such as *Ó, Nic*, etc., and/or the initial mutations that they effect, those segments are tagged in the expanded format that is available to the web application for processing (e.g. `<form><pre>Ní</pre> M<mut>h</mut>athúna</form>`). If the user wants to sort in a way that does not distinguish between forms (i.e. synonyms and equivalents) of the same name, they can, because the synonyms and equivalents are in the same cluster.

Seeding of the database was data-driven and began with an ordered frequency list of surnames registered for babies born in Ireland in 2017. The babies' surnames list, which included surnames registered 3 times or more in 2017, was obtained directly from the Central Statistics Office and sourced from the General Register Office. The first editorial pass involved the following steps, for each surname on the list of babies' surnames, down to a frequency of 4. Resource constraints precluded us from processing surnames registered three times or less:

1. Check in *The Surnames of Ireland* (6th ed.) (MacLysaght 2013) to ascertain if the surname is considered to be of Gaelic origin, or to have been Gaelicised by the 17th century, before the collapse of the Gaelic hierarchy and the decline of the Irish language, e.g. *O'Brien*.
2. Determine primary modern Irish form by consulting *An Sloinnteoir Gaeilge agus an tAinmneoir* (12th ed.) (Ó Droighneáin 2013), as well as other sources, i.e. the *National Database of Irish-Language Biographies* (Cló Iar-Chonnacht 2020), and the *Irish Surname Index* (National Folklore Collection 2020), e.g. *Ó Briain*.
3. Establish cluster title, usually the nominative case of the primary Irish surname's main element, e.g. *Brian*. Consult *Irish Names* (2nd ed.) (Ó Corráin and Maguire 1990) to help determine etymology.
4. Determine if primary Irish form has any synonyms. Primary synonyms, where more than one modern form is in widespread use, e.g. *Breatnach* \approx *Breathnach*, are input under the `<surnames-irish-main/>` element. Historical synonyms that remain commonplace, are dialectal, are morphologically rich, or are significant in some other way, e.g. *Ó Murchadha* \approx *Ó Murchú*, *Ó Dubhghaill* \approx *Ó Dúill*, are input under the `<surnames-irish-historical/>` element. Alternative synonyms, where a less common form is known to be in use, e.g. *Ó Cuív* \approx *Ó Caoimh*, are input under the `<surnames-irish-alternative/>` element.
5. Input English equivalents of all primary Irish forms, using surnames found on babies' surnames list, e.g. *O'Brien*. Historical forms are not considered, and surnames known to be in use in Ireland but not present on the babies' list are input as alternatives on an ad hoc basis, e.g. *Connellan* \approx *Conlan*. Variants primarily found in the UK and North America are not included for the most part, e.g. *Murry* (*Murray*), as our resource is primarily aimed at users in Ireland.

While going through steps (4) and (5) above, two general rules were applied while populating the database. Firstly, as introduced in the previous section, all Irish forms in each entry (i.e. cluster) are synonyms of each other, i.e. that you can easily claim that they are the same surname, except for differences in spelling or grammar. For example, this is true of *Breatnach/Breathnach*, and of *Bhailís/de Bhailís*, but it is not true of *Breatnach/Bhailís*. Even if a surname is derived from another, it is not a synonym if it is significantly different in form, e.g. *Madagáin* > *Madáin*. Therefore, *Breatnach/Breathnach* can be in the same entry, and *Bhailís/de Bhailís* can be in the same entry, but *Breatnach/Bhailís* cannot be in the same entry, even though they may have the same English equivalent (i.e. *Walsh*). Secondly, all English forms in each entry are equivalents of all the Irish forms in the entry. For example, this is true for *Breatnach/Breathnach/Walsh*, *Breatnach/Breathnach/Walsh*, *Breatnach/Breathnach/Brannagh*. Therefore, those three English forms can be in the same entry as *Breatnach/Breathnach*. It does not matter if the English forms are not synonyms of each other. For example, *Walsh/Brannagh* are not synonyms, but this does not stop them from being in the same entry. With the application of these rules, what we end up

with in each entry is the Irish-language surname as a kind of dictionary headword (sometimes with multiple variants), and a list of English-language equivalents of that surname. In other words, what we have is a database of Irish-language surnames and how they are anglicised.

The database we have created is accessible to the public at www.gaois.ie/en/surnames. Users are able to browse an alphabetical index of surnames or clusters, or can search by text. The search results are presented in three sections: exact matches (i.e. links to clusters where one of their surnames match the user's query exactly), distant matches (i.e. links to clusters where one of their surnames match the user's query after it has been query-expanded by the Irish Surname Index (National Folklore Collection 2020)) and spelling suggestions (i.e. surnames that are similar to the user's query as computed by the Levenshtein (1966) edit-distance algorithm). The user can click a spelling suggestion to search again, or click one of the cluster links to display the cluster. Once the user has located the cluster they are looking for, they can see all surnames in it including the automatically generated inflected forms of Irish-language surnames (i.e. the cluster in its expanded format).

Results and discussion

Work on populating the database to date has resulted in 664 surname clusters, containing in total 932 Irish-language surnames and 1,070 English surnames. In 602 clusters, only one primary Irish-language surname is given. In 62 clusters, where a binary choice was not possible, more than one primary Irish-language surname is given. 130 clusters contain historical Irish forms, and 59 clusters contain alternative Irish forms. In 469 clusters, one primary English surname is given, and in 195 clusters, more than one primary English surname is given. 106 clusters contain alternative English forms.

The data structure allows for the promotion of standardised forms, but also permits the establishment of acceptable alternative forms. Establishing the standard Irish-language form presented some challenges. For the most part, the form given in Ó Droighneáin (2013), which adheres to modern standard Irish spelling rules, was taken as the standardised Irish-language form. In some cases, however, more common forms were selected or minor corrections were made, e.g. *Ó Doibhilin* > *Ó Doibhlin*. This was done on the basis of ancillary surname searches in the National Database of Irish-Language Biographies (Cló Iar-Chonnacht 2020) and in Google. Google searches were used as a guide to the prevalence of one form over another, as no population-wide surname frequency data was available at the time of writing.

Where Ó Droighneáin (2013) observed minor differences in spelling when mapping English surnames to Irish-language surnames, we usually did not. For example, Ó Droighneáin maps the English surname *Wolfe* to the Irish-language surname *Ó Mactíre*, whereas he maps the English surname *Woulfe* to the Irish-language surname *de Bhulbh*. This may be a product of the one-to-one mapping to which he constrained himself to. We have no such constraint and our structure allows us to give both *Wolfe* and *Woulfe* as equivalents of *de Bhulbh*. Further examples include *Haugh/Hough*, *Heary/Heery*, and *Henaghan/Henehan*. In another example of this, we see Ó Droighneáin map *Donegan* to *Ó Donnagáin* and *Donnegan* to *Ó Dúnagáin*. We treat *Donnegan* as a North American variant, omit it (having constrained ourselves to surnames commonly found within Ireland), and direct users to *Donegan*, currently the most frequently found form in Ireland. Other examples of variants more commonly used in North America (according to Google) and thus omitted include *Londrigan* (*Lonergan*), *Mannelly* (*Manley*), *Murry* (*Murray*), *Touhy* (*Tuohy*). Google and Wikipedia searches helped to identify UK and North American variants. We opted not to include such diaspora spelling variants to constrain the scope of our research. Diaspora variants could be included in future versions of the database, but for now, users will be directed to related clusters by the search expansion algorithm.

Query expansion (e.g. *Ó Rachtagáin* > *Ó Reachtgáin*), using variants found in the Irish Surname Index (National Folklore Collection 2020), and spelling suggestion using the Levenshtein (1966) edit-distance algorithm (e.g. *Ratigan* > *Rattigan*), are used to help bring the user to the correct cluster. Once at the correct cluster, the user is only provided with standard Irish-language surnames (e.g. *de Priondragás*) and their standard synonyms and equivalents (e.g. *Prendergast*, *Pender*). Where an English equivalent of one surname, e.g. *Minogue* (*Ó Muineog*), is commonly attracted to another surname, i.e. *Ó Mainchín/Mannix*, it is included as an alternative form in the latter. Where non-Gaelic English surnames, e.g. *Manning*, are commonly substituted for English versions of Irish-language surnames in Ireland, i.e. *Mannion* (*Ó Mainnín*), they are included as alternative forms. Non-Gaelic English surnames commonly used by way of semi-translation are included either as primary or alternative forms, e.g. *Reid* (*Ó Maoildeirg*), *Silke* (*Ó Síoda*), *Thornton* (*Ó Droighneáin*). This approach allows us to accurately capture how Irish-language surnames are conventionally associated with English forms in Ireland.

Regarding the data structure, we show that it is effective to structure a database of Irish-language surnames using clusters of surname variants, synonyms, and English-language equivalents. Our theory is that each cluster can effectively represent a particular version of an Irish-language surname. If our theory is correct, there will not be much overlap between different clusters. In our database of common Irish-language surnames, 900 out of 932 Irish-language surname forms (96.5%) occur in only one cluster. This supports our theory. Overlap between clusters, where the same Irish-language surname form is in two clusters indicating uncertainty as to which cluster it belongs to, occurs with only 3.5% of the Irish-language surname forms. None occur in more than two clusters.

The demarcation of clusters can be challenging. Where the surnames have the same etymological (not necessarily eponymical) origin in Irish, e.g. *Donncha*, i.e. {*Donncha: Mac Donncha, Ó Donncha, Ó Donnchaidh, Mac Donnchaidh, McDonagh, O'Donoghue, Donoghue, Dunphy*}, the cluster is not split. Where the surnames have different etymological origins, e.g. *Raghallach* {*Raghallach: Ó Raghallaigh, O'Reilly, Reilly*} vs *Raithileach* {*Raithileach: Ó Raithile, O'Rahilly*}, the surnames are separated into different clusters. With regard to providing the user with grammatical information about Irish-language surnames, we identified 5 patterns of inflection. Based on these 5 patterns, the inflectional paradigm can be correctly predicted from the base form, without any extra information, in the case of all Irish-language surnames in our database. This satisfactorily confirms the completeness of our analysis of the grammar of Irish-language surnames. If extra information is required in future cases, our extensible data model allows it.

Conclusion

Our first goal was to provide a resource that supports the use of Irish-language surnames in Ireland by giving human users and application developers access to the standard Irish-language spellings as well as grammatical information via the web. Our second goal was to provide users with an equivalence mapping between standard Irish-language and English-language forms of surnames in Ireland. The *Gaois* linguistic database of Irish-language surnames satisfies these two goals. Our ambition now is to expand the scope of the database to include all surnames, with an Irish-language form and a 4+ frequency, of children born in Ireland between 2015 and 2020, those yet to start school. We also hope to include data from Northern Ireland if available. In its current state, however, the database is reasonably comprehensive, and certainly covers the most common Irish-language surnames currently in use. We hope that easy online access to user-friendly and well-structured machine-readable data of this kind will support the use of Irish-language surnames in schools, libraries, government departments, and elsewhere.

Funding

This work was supported by the Faculty of Humanities and Social Sciences, Dublin City University. The sponsor had no other role in the study.

References

- Brylla, Eva. 2016. "A New Personal Names Act in Sweden?" In *Proceedings of the 25th International Congress of Onomastic Sciences* (Glasgow, August 25–29, 2014), edited by Carole Hough and Daria Izdebska, vol. 3, 27–33. Glasgow: University of Glasgow.
- Byrne, Kevin and Eoin O'Malley. 2013. "What's in a Name? Using Surnames as Data for Party Research." *Party Politics* 19, no. 6: 985–997.
- Central Statistics Office. 2016. "Irish Language and the Gaeltacht 2016." *Central Statistics Office Publications Website*. Accessed December 15, 2020. <https://www.cso.ie/en/releasesandpublications/ep/p-cp10esil/p10esil/ilg/>
- Central Statistics Office. 2017. "Irish Babies' Names 2017." *Central Statistics Office Publications Website*. Accessed December 15, 2020. <https://www.cso.ie/en/releasesandpublications/ep/p-ibn/irishbabies/names2017/bgn/>
- Cló Iar-Chonnacht. 2020. "National Database of Irish-Language Biographies." *Ainm Website*. Accessed December 15, 2020. <https://www.ainm.ie/>
- Coates, Richard. 2006. "Names." In *A History of the English Language*, edited by Richard Hogg and David Denison, 312–351. Cambridge: Cambridge University Press.

- de hÓir, Éamonn. 1973. "The Anglicisation of Irish Placenames." *Onoma* 27: 192–204.
- Department of Education and Skills. 2013. "Simplified Arrangements for the Maintenance of the Register, Roll Book and Daily Attendance Book. Department Circular 0028/2013." *Department of Education and Skills Website*. Accessed December 15, 2020. https://www.education.ie/en/Circulars-and-Forms/Active-Circulars/cloo28_2013.pdf
- Department of the Environment. 1995. *A Guide for Local Authorities on the Use of Irish*. Dublin: Government Publications Office.
- Foras na Gaeilge. 2020. "Irish National Morphology Database." *Teanglann Website*. Accessed December 15, 2020. <https://www.teanglann.ie/en/gram/>
- Gaelic Athletic Association. 2020a. "Official Guide – Part 1." *Gaelic Athletic Association Website*. Accessed December 15, 2020. <https://www.gaa.ie/api/pdfs/image/upload/ljywtjss059j50msvd3b.pdf>
- Gaelic Athletic Association. 2020b. "Official Guide – Part 2." *Gaelic Athletic Association Website*. Accessed December 15, 2020. <https://www.gaa.ie/api/pdfs/image/upload/vybckve2iipp56xmjhqd.pdf>
- Government of Ireland. 1937. "Article 8 (on the Languages of the State)." In *Irish Constitution*. Dublin: Government Publications Office.
- Government of Ireland. 1958. *An Caighdeán Oifigiúil*. Dublin: Government Publications Office.
- Government of Ireland. 2003. *Official Languages Act 2003, Number 32 of 2003*. Dublin: Government Publications Office.
- Government of Ireland. 2020. "Placenames Database of Ireland." *Logainm Website*. Accessed December 15, 2020. <https://www.logainm.ie/en/>
- Hanks, Patrick, Richard Coates, and Peter McClure. 2016. *The Oxford Dictionary of Family Names in Britain and Ireland*. Oxford: Oxford University Press.
- Irish Language Commissioner. 2020. "Monitoring Report 2019." *Irish Language Commissioner Website*. Accessed December 15, 2020. https://www.coimisineir.ie/userfiles/files/CT_Tuarascail_Faireachain_Monitoring_Report_2019_web.pdf
- Kimura, Maiko. 2014. "Differences in Representations of Japanese Name Authority Data among CJK Countries and the Library of Congress." *Information Processing & Management* 50, no. 5: 733–751.
- Levenshtein, Vladimir Iosifovich. 1966. "Binary Codes Capable of Correcting Deletions, Insertions, and Reversals." *Soviet Physics Doklady* 10, no. 8: 707–710.
- MacLysaght, Edward. 2013. *The Surnames of Ireland*. 6th ed. Dublin: Irish Academic Press.
- Mac Mathúna, Liam. 2006. "What's in an Irish Name? A Study of the Personal Naming Systems of Irish and Irish English." In *The Celtic Englishes IV*, edited by Hildegard Luise Charlotte Tristram, 64–87. Potsdam: Potsdam University Press.
- Měchura, Michal Boleslav. 2014. "Irish National Morphology Database: a High-Accuracy Open-Source Dataset of Irish Words." In *Proceedings of the First Celtic Language Technology Workshop* (Dublin, August 23, 2014), edited by John Judge, Teresa Lynn, Monica Ward and Brian Ó Raghallaigh, 50–54. Dublin: Association for Computational Linguistics and Dublin City University.
- Měchura, Michal Boleslav. 2017. "Introducing Lexonomy: an Open-Source Dictionary Writing and Publishing System." In *Electronic Lexicography in the 21st Century: Lexicography from Scratch—Proceedings of the eLex 2017 Conference* (Leiden, September 19–21, 2017), edited by Iztok Kosem, Carole Tiberius, Miloš Jakubiček, Jelena Kallas, Simon Krek and Vít Baisa, 662–679. Brno: Lexical Computing.
- Muhr, Kay and Liam Ó hAisibéil. 2021. *A Dictionary of Family Names of Ireland*. Oxford: Oxford University Press.
- National Folklore Collection. 2020. "Irish Surname Index." *Dúchas Website*. Accessed December 15, 2020. <https://www.duchas.ie/en/nom>
- Nic Cóil, Róisín. 2011. "Irish Prefixes and the Alphabetization of Personal Names." *The Indexer* 29, no. 2: C1–C6.
- Ó Corráin, Donnchadh and Fidelma Maguire. 1990. *Irish Names*. 2nd ed. Dublin: The Lilliput Press.
- Ó Cuív, Brian. 1986. *Aspects of Irish Personal Names*. Dublin: Dublin Institute for Advanced Studies.

- Ó Cuív, Éamon. 2018. "Cúrsaí Gaeilge. Dáil Éireann Debate, April 18 2018, Question 19." *The Oireachtas Website*. Accessed December 15, 2020. <https://www.oireachtas.ie/en/debates/question/2018-04-18/19/>
- Ó Droighneáin, Muiris. 1966. *An Sloinnteoir Gaeilge agus an tAinmneoir*. 1st ed. Dublin: C. J. Fallon.
- Ó Droighneáin, Muiris. 2013. *An Sloinnteoir Gaeilge agus an tAinmneoir*. 12th ed. Dublin: Coiscéim.
- Ó hUiginn, Ruairí. 2008. "The Irish Language." In *A New View of the Irish Language*, edited by Caoilfhionn Nic Pháidín and Seán Ó Cearnaigh, 1–10. Dublin: Cois Life.
- O’Rahilly, Thomas Francis. 1976. *Irish Dialects Past and Present, with Chapters on Scottish and Manx*. Dublin: Dublin Institute of Advanced Studies.
- Plassard, Marie-France. 1996. *Names of Persons: National Usages for Entry in Catalogues*. 4th ed. Frankfurt: K. G. Saur.
- Shi, Yanqing and Junzhi Jia. 2018. "Use and Presentation of Personal Name Components in Chinese Authority Files." *The Journal of Academic Librarianship* 44, no. 4: 477–485.
- The Irish Times View. 2020. "The Irish Times View on State Bodies and the Irish Language: A Litany of Failures. August 14, 2020." *The Irish Times Website*. Accessed December 15, 2020. <https://www.irishtimes.com/opinion/editorial/the-irish-times-view-on-state-bodies-and-the-irish-language-a-litany-of-failures-1.4329599>
- Woulfe, Patrick. 1923. *Sloinnte Gaedheal is Gall: Irish Names and Surnames*. Dublin: M. H. Gill & Son.

Note on Contributor(s)

Brian Ó Raghallaigh is an Assistant Professor at Dublin City University (DCU) in Fiontar & Scoil na Gaeilge and is responsible for the technical management and development of the Ainm, Dúchas, Lex, Logainm and Téarma projects as part of the Gaois research group. He is the Principal Investigator on the Logainm placenames project. He has a background in computer science, linguistics and Irish. His PhD thesis was on multidialect phonetisation for Irish at Trinity College Dublin (TCD). He is the author of *Fuaimeanna na Gaeilge* (Cois Life, 2014). He is on the Management and Editorial Committees of *Foclóir Stairiúil na Gaeilge* in the Royal Irish Academy.

Michal Boleslav Měchura is a software engineer who specializes in language technology and digital humanities. He has a master’s degree in computational linguistics from Trinity College Dublin (TCD) and had previously worked as a software localization engineer for Microsoft. He is currently doing a PhD in e-lexicography at Masaryk University.

Aengus Ó Fionnagáin is an Assistant Professor in Irish at the University of Limerick (UL). He teaches courses on Irish literature and language from 1600 to the present, including: modern Irish grammar, onomastics, early-modern prose literature in Irish, dialectology, Gaeltacht literature, the Schools’ Folklore Collection 1937-8, and Celtic Civilisation. His research focuses primarily on Irish onomastics, particularly the historical development of Irish surnames and placenames, but he has also published work on historical language shift and bilingualism in Ireland.

Sophie Osborne is an Irish and linguistics graduate from University College Dublin (UCD). She has a master’s degree in Irish from University College Dublin as well. She is currently working as an awareness campaign executive with *Conradh na Gaeilge*. Previously she worked as a research assistant at Dublin City University.

Correspondence to: Dr. Brian Ó Raghallaigh, Dublin City University, brian.oraghallaigh@dcu.ie