

Towards a lexicon of Irish-language idioms

Katie Ní Loingsigh
Fiontar, Dublin City University, Ireland
katie.niloingsigh@dcu.ie

RESUME

Vers un lexique d'idiomes de la langue irlandaise

Le présent exposé fournit un éclairage sur un lexique d'idiomes de la langue irlandaise rassemblés par *Foclóir Gaeilge-Béarla* (Ó Dónaill, 1977) et *Foclóir Gaedhilge agus Béarla* (Dinneen, 1927), deux références en matière de dictionnaires de gaélique élaborés au cours du vingtième siècle. Cette lexique d'idiomes est le fruit de recherches effectuées dans le cadre d'études doctorales sur les idiomes de langue irlandaise fondées sur les publications de Peadar Ó Laoghaire, auteur gaélique emblématique du 20ème siècle, et en est un sous-produit utile. La présente compilation offrant une ressource exploitable pour des analyses et recherches à venir en phraséologie, linguistique informatique et lexicographie gaélique.

ABSTRACT

Towards a lexicon of Irish-language idioms

This paper presents the development of a lexicon of Irish-language idioms as collected from *Foclóir Gaeilge-Béarla* (Ó Dónaill, 1977) and *Foclóir Gaedhilge agus Béarla* (Dinneen, 1927), the two primary Irish-English dictionaries compiled during the twentieth century. This lexicon of idioms is a beneficial by-product of doctorate research undertaken on Irish-language idioms from the published work of Peadar Ó Laoghaire, one of the foremost Irish-language authors of the twentieth century. The lexicon of idioms presented in this paper can be used as a resource for future analysis and research in the areas of phraseology, computational linguistics and Irish language lexicography.

MOTS-CLÉS: phraséologie, idiome, langue irlandaise, ressources lexicales, lexique.

KEYWORDS: phraseology, idioms, Irish language, lexical resources, lexicon.

1 Introduction

If natural language had been designed by a logician, idioms would not exist. (Johnson-Laird, 1993 cited in Cacciari and Tabossi, 1993, vii)

There has been a steady increase in idiom-related research in the area of phraseology over the past three decades. Despite this development, there are still many terminological issues which cause difficulties in the classification and description of various linguistic units to the extent that “phraseology is ‘bedivilled’ (Cowie’s description) by the proliferation of different terms for the same category and by conflicting uses of the same terms” (Pawley, 2001). Idioms are classified as a subset of a more general linguistic unit in phraseology which has been described using numerous

terms by various authors, e.g. ‘multiword unit’ (Grant and Bauer, 2004; Wulff, 2008); ‘word-combination’ (Zgusta, 1972); ‘fixed expression’ (Carter, 1987; Alexander, 1984); ‘phraseme’ (Mel’čuk, 1995); ‘composite’ (Howarth, 1996; Cowie, 1981); ‘phrasal lexeme’ (Moon, 1998; Lipka, 1990); ‘phraseological unit’ (Gläser, 1986); ‘multiword expression’ (Fernando, 1996) and ‘conventional expression’ (Pawley, 2001), etc. In this paper an idiom is defined as a type of phraseme or multiword expression (MWE) 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. This description follows the definition of idioms as laid down in the literature (e.g. Abdou, 2012; Hanks, 2004; Howarth, 1998 and Fernando, 1996), for example:

- *Rud a chur ar an méar fhada*¹, ‘to put something off indefinitely’;
- *Muc i mála a cheannach*², ‘to accept an offer or deal foolishly without being examined first’;
- *Teacht aniar aduaidh ar dhuine*³, ‘to take someone unawares’.

This paper focuses specifically on the development of a linguistic resource for Irish, namely a lexicon of Irish-language idioms for reference and research. This lexicon can be used as a resource in the study of Irish-language idioms in phraseology but also in research related to phrasemes or multiword expressions in the area of natural language processing (NLP). A brief background to the creation of the lexicon is given in Section 2. Section 3 focuses on the areas of phraseology and computational linguistics, while the methodology and compilation of the lexicon is discussed in Section 4. Future work is discussed in Section 5 and the lexicon itself will be made available as an open data set available at <http://www.gaois.ie/> under a non-restrictive license.

2 Background

The Irish language belongs to the Celtic branch of the Indo-European family of languages and is one of two official languages of Ireland, the other being English. The study of Irish-language idioms within the field of phraseology is a relatively new and underdeveloped area of research. Even though idioms have been collected and analysed as part of general lexicographic studies from the late nineteenth century onwards, there has been only one major academic study undertaken on Irish-language idioms, i.e. *A concordance of idiomatic expressions in the writings of Séamus Ó Grianna* (Ó Corráin, 1989). Additionally, it is acknowledged that in comparison to other official languages of the European Union, Irish-language technology is under-resourced (Lynn, 2014; Ó Raghallaigh and Měchura, 2014).

The idiom lexicon presented in this paper is a useful by-product of doctorate research which involved the creation of a database of Irish-language idioms from the published work of Peadar Ó Laoghaire (Ní Loingsigh 2016). The idiom database was created in *Leacsclann*, an online platform used for building dictionary writing systems and terminology management systems as well as other lexicographic and reference applications (Měchura 2012) and is used in various research projects developed in Fiontar, Dublin City University (Ó Raghallaigh and Měchura, 2014). To facilitate the search and extraction of idioms from Peadar Ó Laoghaire’s published work, a lexicon

¹ Literal meaning: ‘to put something on the long finger’.

² Literal meaning: ‘to buy a pig in a sack/bag’.

³ Literal meaning: ‘to come upon someone from the northwest’.

of idioms was manually compiled from the two primary Irish-language dictionaries of the twentieth century and it is this lexicon which is presented in this paper. The most common lemmas from this lexicon were categorized in order of frequency and were used to search a corpus of Ó Laoghaire’s published work, which was compiled using *Sketch Engine* tools (Kilgarriff et al., 2004) and a morphological analyser and a part-of-speech tagger (Ui Dhonnchadha, 2009). The search methodology, which used “idiom-prone words” (O’Keefe, McCarthy and Carter, 2007), or the most frequent lemmas from the lexicon of idioms presented here, to search the corpus, will not, however, be examined in this paper.

3 Phraseology and NLP

The central topics of research on idioms in Europe during the past three decades have focused on five main areas: (i) syntax of idioms, (ii) semantics of idioms, (iii) pragmatics of idioms (including text-related modifications), (iv) cognitive approaches to idioms, and (v) contrastive research on idioms (including cultural specifics and cross-cultural comparison of idioms) (Dobrovol’ski and Piirainen, 2005, p.30). Additionally, there has been ongoing research undertaken on idioms as a subtype of MWEs in the field of NLP. Multiword expressions are recognized (Sag et al., 2002) as a key problem for the development of NLP technology and are underappreciated in the field at large. Colson (2015) highlights two serious shortcomings of computational phraseology:

1. There is no universally accepted algorithm for the automatic extraction of phraseology, especially not for ngrams larger than bigrams.
2. There is no consensus as to the proportion of set phrases in relation with the rest of the vocabulary: according to Jackendoff (1995), there are about as many fixed expressions as there are single words in the dictionary, but others (such as Mel’čuk 1995) hold the view that fixed expressions far outnumber single words. (Colson, 2015, p.7)

In a similar vein, the development of tools to recognize and extract Irish-language idioms as a type of MWE from corpora are lacking. This is not a problem specific to Irish-language idioms or Irish-language MWEs but a common stumbling block faced in NLP research on MWEs.

4 Methodology

In this paper, idioms are categorized as a subclass of MWEs following the definition set out by Sag et al. (2002 cited in Baldwin and Kim (2010, p. 269)):

- (3) Multiword expressions (MWEs) are lexical items that: (a) can be decomposed into multiple lexemes; and (b) display lexical, syntactic, semantic, pragmatic and/or statistical idiomaticity.

While idioms are recognized as a central class of phrasemes or MWEs, other categories such as similes, proverbs, routine formulae and certain restricted collocations, which exhibit similar features to idioms, are not examined here.⁴ The majority of idioms collected in the lexicon presented in this paper are primarily lexical items that display semantic and lexical idiomaticity as set out in Baldwin and Kim (2010). However, it is often difficult to distinguish certain idioms from

⁴ See Dobrovol’ski and Piirainen (2005) for a more indepth overview of this topic.

other types of MWEs depending on factors such as a speakers age, cultural and linguistic background, etc. Phrasal verbs are not included as “they are such a large group... that they merit separate and thorough research of their own” (Grant 2003, p.19). Additionally, compound words and functional expressions such as proverbs, greetings, blessings, and terms of endearment are not included (Abdou 2012, Moon 1998). The lexicon of idioms presented in this paper provides a base for future research on MWEs and it is accepted that further analysis and scrutiny of the list is needed to ensure its completeness and to also verify the quality of the idioms collected.

According to the definition of idiom as set down in Section 1 of this paper, idioms were manually selected and recorded from the two primary Irish-language dictionaries of the 20th century, *Foclóir Gaeilge-Béarla* (Ó Dónaill, 1977) and *Foclóir Gaedhilge agus Béarla* (Dinneen, 1927). Each individual headword in both dictionaries was manually examined and any lexical items falling within the definition of an idiom were manually extracted and recorded. *Foclóir Gaedhilge agus Béarla* (Dinneen, 1927) was the first major Irish-English lexicographic work undertaken during the early twentieth century and is still a valuable resource for students and writers alike (Ó Murchú, 2005). The first edition of Dinneen’s dictionary, which was published in 1904, is available as part of the Corpus of Electronic Texts (CELT) project in University College Cork in searchable PDF format⁵ and the second more comprehensive edition, which was published in 1927 and used as a basis for analysis in this paper, is also available as a digitized and fully searchable online resource.⁶ The foremost Irish-English dictionary available at present is *Foclóir Gaeilge-Béarla* (Ó Dónaill, 1977). Ó Dónaill’s dictionary is still recognized as the principal orthographical source for the spelling of the language and “provides the most comprehensive coverage of the grammar and other aspects of words in Irish” (*Foclóir Gaeilge-Béarla* (Ó Dónaill), 2016). It is available as a searchable electronic resource as part of the *Leabharlann Teanga agus Foclóireachta*⁷ project. Due to the lack of NLP resources available and the limited research undertaken on Irish-language MWE’s, both dictionaries were examined manually and 5,437 idioms were collected. While this approach proved time-consuming, the resultant lexicon of idioms provides a comprehensive base for further research. This lexicon includes some duplication due to the inclusion of variant forms of certain idioms used as usage examples in various entries.

Bealtaine, *f.* (*gs.* ~, *pl.* -ní). May, Lá ~, May Day. Oíche Bhealtaine, eve of May Day. Mí na ~, month of May. Idir dhá thine Bhealtaine, in a dilemma. (Ó Dónaill, 1977, s.v. Bealtaine.)

FIGURE 1: Idiom: *idir dhá thine Bhealtaine*⁸ (‘in a dilemma’).

Bealtaine, *g.id., f.*, (oft. pron. Beallthaine), the Irish May Festival, the month of May... idir dhá theine (uisce) lae Bealtaine, in a dilemma, from the practise of driving cattle between two fires with a view to their preservation. (Dinneen, 1927, s.v. Bealtaine.)

FIGURE 2: Idiom: *idir dhá theine (uisce) lae Bealtaine*⁹ (‘in a dilemma’).

The canonical form of the idiom in Figure 1 and Figure 2 is recorded as *idir dhá thine Bhealtaine* (‘in a dilemma’) and the variant morphological form can be seen in Figure 2 which still retains the same idiomatic meaning, *idir dhá theine (uisce) lae Bealtaine* (‘in a dilemma’). The idiom in

⁵ Available: <http://www.ucc.ie/celt/Dinneen1sted.pdf>

⁶ Available: <http://glg.csisdmsz.ul.ie/index.php>

⁷ Available: <http://www.teanglann.ie/ga/fgb/>

⁸ Literal meaning: ‘between two May fires’. Emphasis added.

⁹ Literal meaning: ‘between two May Day (water) fires’. Emphasis added.

Figure 2 is displayed in non-standardized orthography. As *Foclóir Gaedhilge agus Béarla* (Dinneen, 1927) was compiled prior to the publication of an official standard of Irish (*An Caighdeán Oifigiúil* (Rannóg an Aistriúcháin, 1958)), these idioms are recorded in non-standardized orthography. However, these idioms can be standardized using *An Caighdeánaitheoir* (Scannell, 2009), an application which annotates pre-standard words with standardized forms. A more detailed description of *An Caighdeánaitheoir* can be found in Uí Dhonnchadha et al. (2014).

The idioms in the lexicon presented in this paper are listed according to the headword under which they are recorded in the dictionary and have not been classified or analysed. As the idioms are recorded as usage examples in both dictionaries, they are often recorded as part of a longer sentence and not systematically by canonical form. However, certain syntactic structures are more common than others. For example, a number of idioms in the lexicon contain a verbal noun at the beginning of the idiom. Figure 3 and Figure 4 below show an example of this structure through the following example, *ag imeacht le haer an tsaoil* ('pleasure-seeking, leading a gay life'). Although part of this idiom, *aer an tsaoil* ('the pleasures of the world'), is given under the headword *aer* in *Foclóir Gaeilge-Béarla* (Ó Dónaill 1977) in Figure 3, the canonical form of the idiom containing a verbal noun is recorded as an individual example under the same headword. Additionally, in Figure 4, a non-standardized canonical form of the idiom, *ag imtheacht le haer an tsaoghail* ('leading a purposeless, improvident life'), is given as a usage example under the headword *aer* in *Foclóir Gaedhilge agus Béarla* (Dinneen 1927).

aer, *m.* (*gs.* *air*)... 4. Gaiety, pleasure. ~ *an tsaoil*, the pleasures of the world. *Ag imeacht le h~ an tsaoil*, pleasure-seeking, leading a gay life. *Chaith sé a chuid airgid le h~ an tsaoil*, he spent his money on pleasure. (Ó Dónaill, 1977, s.v. *aer*.)

FIGURE 3: Idiom: *ag imeacht le haer an tsaoil*¹⁰ ('pleasure-seeking, leading a gay life').

aer, *g.* *air*, *m.*, the air, the sky, climate... *ag imtheacht le haer an tsaoghail*, wandering aimlessly about, leading a purposeless, improvident life. (Dinneen, 1927, s.v. *aer*.)

FIGURE 4: Idiom: *ag imeacht le haer an tsaoil*¹¹ ('wandering aimlessly about, leading a purposeless, improvident life').

Mulhall (2010, p.1358) refers to a limited number of idioms that contain non-words or idioms which are also referred to as 'unique sublexical items' (Gouws, 1991) and gives an example of the use of 'amok' in the idiom 'to run amok' in English. A number of idioms that contain non-words are recorded in the lexicon presented in this paper and can be seen in Figure 5 and Figure 6 below. The idiom *húm ná hám* ('a sound, a move') is recorded as a usage example in both dictionaries under the headword *húm* ('a jot, a word').

húm, *s.* (In phrase) *Ní raibh ~ ná hám* as, there wasn't a sound, a move, out of him. (Ó Dónaill, 1977, s.v. *húm*.)

FIGURE 5: Idiom: *húm ná hám* ('a sound, a move').

¹⁰ Literal meaning: 'Going with the pleasure of the world/of life'. Emphasis added.

¹¹ Literal meaning: 'Going with the pleasure of the world/of life'. Emphasis added.

húm, *m.*, a jot, a word; *ní* fhéadfadh sé húm ná hám do bhaint as an gcloich, he could not get a move out of the stone; *ní* dubhairt sé húm ná hám, he remained neutral. (Dinneen, 1927, s.v. húm.)

FIGURE 6: Idiom: *húm ná hám* ('a sound, a move').

While Figure 5 includes the prompt 'In phrase' which signifies the phrasal or idiomatic use of the headword, this prompt is not used systematically throughout the dictionary and cannot be relied on to identify all idioms recorded in the dictionary. In addition, it can be seen from Figure 5 and Figure 6 that the idiom *húm ná hám* is used in its negative sense only with the negative verbal particle, *ní*, prefixing any use of the idiom.

As the lexicon of idioms presented in this paper was primarily compiled as part of doctorate research to identify idiom-prone words which were used to facilitate a corpus search, it has not yet been subject to indepth linguistic analysis. This paper does not provide information in relation to the primary linguistic features of idioms listed in the lexicon and further analysis would ensure its quality and its potential use as a gold standard lexicon of Irish-language idioms.

5 Future work

This lexicon is the only current comprehensive representation of Irish-language idioms collected from both written and oral sources during the twentieth century. It provides a general representation of Irish-language idioms and can be used as a foundation for any future development of a comprehensive dictionary of Irish-language idioms. Additionally, it can be used as a base for analysis of the syntactic structure of idioms as a subset of MWEs. This will further the research and analysis on Irish-language syntax generally and Irish-language idiom syntax specifically. For example, the area of syntactic parsing benefits greatly from research in multiword expressions. As Baldwin et al. (2004) note, "a lack of MWE lexical items in a precision grammar is a significant source of parse errors". It follows that statistical parsers, which are trained on syntactically annotated treebanks, perform better if a treebank has multiword expressions identified and annotated. To date, the Irish Dependency Treebank (Lynn, 2012; 2016) does not contain these annotations, primarily due to the lack of sufficient linguistic resources available such as MWE corpora, MWE linguistic analysis or sufficient identification and categorization of MWEs in digital format. This work therefore constitutes a step towards bridging that gap in knowledge between Irish linguistics and NLP tools.

The lexicon of idioms can also be used as a starting point for research on semantics and pragmatics of Irish-language idioms, modification in Irish-language idioms, frequency of various idioms, language change over time and cognitive approaches to idioms. While the lexicon presented in this paper focuses on idioms as a subset of MWEs, it can be used as a stepping stone towards further analysis of various MWEs, e.g. collocations, similes, proverbs, etc. Multiword expressions present obstacles to the development of NLP tools and this lexicon provides a resource which can be developed and utilised to advance the development of other resources for Irish in this area.

Acknowledgements

This research was undertaken with support from Fiontar, Dublin City University. I would also like to thank the useful comments from the three anonymous reviewers.

References

- ABDOU A. (2012). *Arabic idioms: a corpus based study*. London: Routledge.
- ALEXANDER R.J. (1984). Fixed expressions in English: reference books and the teacher. *English Language Teaching Journal* 38(2): 127-132.
- BALDWIN T. and KIM S.N. (2010). Multiword Expressions. N. INDURKHYA and F.J. DAMERAU (eds.) *Handbook of Natural Language Processing*. 2nd ed. Chapman & Hall: United States of America, 267-293.
- BALDWIN T., BENDER E.M., FLICKINGER D., KIM, A. and OEPEN, S. (2004). Road-testing the English resource grammar over the British National Corpus. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, Portugal, 2047-2050.
- CACCIARI C. and TABOSSI P. (1993). *Idioms: processing, structure and interpretation*. Hillsdale: Erlbaum.
- CARTER R. (1987). *Vocabulary: applied linguistic perspectives*. London: Allen & Unwin.
- COLSON J-P. (2015). The contribution of corpus-based phraseology to translation studies: from experiments to theory. G.C PASTOR (eds.) *Computerised and corpus-based approaches to phraseology: monolingual and multilingual perspectives*, 6-9.
Online at: http://www.euophras2015.eu/euophras2015_bookoffullpapers/! [Retrieved 9 March 2016].
- COWIE A.P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics* 2(3): 223-235.
- DINNEEN P.S. (1927). *Foclóir Gaedhilge agus Béarla, being a thesaurus of the words, phrases and idioms of the modern language*. 2nd ed. Dublin: Irish Texts Society.
- DOBROVOL'SKIJ D.O and PIIRAINEN E. (2005). *Figurative language: cross-cultural and cross-linguistic perspectives*. Amsterdam: Elsevier.
- FERNANDO C. (1996). *Idioms and idiomaticity*. Oxford: Oxford University Press.
- FOCLÓIR GAELIGE-BÉARLA (Ó DÓNAILL). (2016). Online at: <http://www.teanglann.ie/ga/fgb/> [Retrieved 2 April 2016].
- GRANT L. and BAUER L. (2004). Criteria for re-defining idioms: are we barking up the wrong tree? *Applied Linguistics* 25(1): 38-61.
- GRANT L.E. (2003). *A corpus-based investigation of idiomatic multiword units*. PhD Thesis. Victoria: University of Wellington.

- GLÄSER R. (1986). *Phraseologie der englischen Sprache*. Tübingen: Walter de Gruyter.
- GOUWS R.H. (1991). Toward a lexicon-based lexicography. *Dictionaries: Journal of the Dictionary Society of North America* 13: 75-90.
- HANKS P. (2004). The syntagmatics of metaphor and idiom. *International Journal of Lexicography* 17(3): 245-274.
- HOWARTH P. (1998). Phraseology and second language proficiency. *Applied Linguistics* 19(1): 24-44.
- HOWARTH P. (1996). *Phraseology in English academic writing: some implications for language learning and dictionary making*. Tübingen: Niemeyer.
- KILGARRIFF A., RYCHLY R., SMRZ, P. and TUGWELL D. (2004). The Sketch Engine. G. WILLIAMS and S. VESSIER (eds.) *In Proceedings of the 11th Euralex International Congress*, Lorient, France, 6-10 July 2004, 105-115. Online at: http://www.euralex.org/elx_proceedings/Euralex2004/011_2004_V1_Adam_KILGARRIFF_Pavel_RYCHLY_Pavel_SMRZ_David_TUGWELL_The_Sketch_Engine.pdf [Retrieved 3 November 2012].
- LIPKA L. (1990). *An outline of English lexicology: lexical structure, word semantics, and word-formation*. Tübingen: Niemeyer.
- LYNN T. (2016). Irish Dependency Treebanking and Parsing. PhD Thesis. Dublin City University.
- LYNN T., FOSTER J., DRAS M. and TOUNSI L. (2014). Cross-lingual transfer parsing for low-resourced languages: an Irish case study. *In Proceedings of the First Celtic Language Technology Workshop (CLTW '14)*, Dublin, Ireland, 41-49. Online at: <http://www.aclweb.org/anthology/W/W14/W14-4606.pdf> [Retrieved 8 April 2016].
- LYNN T., ÇETİNOĞLU O., FOSTER J., UÍ DHONNCHADHA E., DRAS M. and VAN GENABITH J. (2012). Irish treebanking and parsing: a preliminary evaluation. *In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC '12)*, Istanbul, Turkey, 1939-1946. Online at: http://www.lrec-conf.org/proceedings/lrec2012/pdf/378_Paper.pdf [Retrieved 20 May 2016].
- MĚCHURA M.B. (2012). LéacsIann: A platform for building dictionary writing systems. R.V. FJELD and J.M. TORJUSEN (eds.) *In Proceedings of the 15th Euralex International Congress*, Oslo, Norway, 7-11 August 2012, 855-861. Online at: http://www.euralex.org/elx_proceedings/Euralex2012/pp855-861_Mechura.pdf [Retrieved 23 November 2012].
- MEL'ČUK I.A. (1995). Phrasemes in language and phraseology in linguistics. M. EVERAERT, E-J. VAN DER LINDEN, A. SCHENK, AND R. SCHREUDER (eds.) *In Idioms: structural and psychological perspective*. Hillsdale: Erlbaum, 167-232.

- MOON R. (1998). *Fixed expressions and idioms in English: a corpus-based approach*. Oxford: Oxford University Press.
- MULHALL C. (2010). A semantic and lexical-based approach to the lemmatisation of idioms in bilingual Italian-English dictionaries. A. DYKSTRA, and T. SCHOONHEIM, (eds.) *In Proceedings of the 14th Euralex International Congress*, Leeuwarden/Ljouwert, The Netherlands, 6-10 July 2010, 1355-1377. Online at: http://www.euralex.org/elx_proceedings/Euralex2010/129_Euralex_2010_9_MULHALL_A_Semantic_and_Lexical-Based_Approach_to_the_Lemmatisation_of_Idioms_in_Bilingual_Italian-Eng.pdf [Retrieved 18 October 2015]
- NÍ LOINGSIGH K. (Forthcoming 2016). *Tiomsú agus rangú i mbunachar sonraí ar chnuasach nathanna Gaeilge as saothar Pheadair Uí Laoghaire*. PhD Thesis. Dublin City University.
- O'KEEFE A., MCCARTHY M. and CARTER R. (2007) *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- Ó CORRÁIN A. (1989). *A concordance of idiomatic expressions in the writings of Séamus Ó Grianna*. Belfast: The Institute of Irish Studies, the Queen's University Belfast.
- Ó DÓNAILL N. (1977). *Foclóir Gaeilge-Béarla*. Baile Átha Cliath: An Gúm.
- Ó MURCHÚ M. (2005). Dineen and Ó Dónaill. P. RIGGS (ed.) *Dinneen and the Dictionary 1904-2004*. Irish Texts Society: Dublin, 78-101.
- Ó RAGHALLAIGH B. and MĚCHURA M. B. (2014). Developing high-end reusable tools and resources for Irish-language terminology, lexicography, onomastics (toponymy), folkloristics, and more, using modern web and database technologies. *In Proceedings of the First Celtic Language Technology Workshop*, Dublin, Ireland, August 2014, 66-70. Online at: <http://www.aclweb.org/anthology/W/W14/W14-4610.pdf> [Retrieved 24 March 2016].
- PAWLEY A. (2001). Review of *Phraseology, linguistics and the dictionary*, by Anthony Paul Cowie. *International Journal of Lexicography* 14(2): 122-134.
- RANNÓG AN AISTRIÚCHÁIN (1958). *Gramadach na Gaeilge agus litriú na Gaeilge: an caighdeán oifigiúil*. Baile Átha Cliath: Oifig an tSoláthair.
- SAG I.A., BALDWIN T., BOND F., COPESTAKE A, and FLICKINGER D. (2002). Multiword expressions: a pain in the neck for NLP. *In Computational linguistics and intelligent text processing. Third International Conference, (CICLing 2002)*, Mexico City, Mexico, 17-23 February 2002. Berlin: Springer, 1-15.
- SCANNELL K. (2009). *Standardization of corpus texts for the NEID*. Presentation given in Saint Louis University, 22 May 2009. Online at: <http://borel.slu.edu/pub/naaclt09.pdf> [Retrieved 3 December 2012].

UÍ DHONNCHADHA E. (2009). *Part-of-speech tagging and partial parsing for Irish using finite-state transducers and constraint grammar*. PhD Thesis. Dublin City University.

UÍ DHONNCHADHA E., SCANNELL K., Ó HUIGINN R., NÍ MHEARRAÍ E., NIC MHAOLÁIN M., Ó RAGHALLAIGH B., TONER G., MAC MATHÚNA S., D’AURIA D., NÍ GHALLCHOBHAIR E, and O’LEARY N. (2014). *Corpas na Gaeilge (1882–1926): integrating historical and modern Irish texts*. N. CALZOLARI, K. CHOUKRI, T. DECLERCK, H. LOFTSSON, B. MAEGAARD, J. MARIANI, A. MORENO, J. ODIJK AND S. PIPERIDIS (eds.) *In Proceedings of the Workshop: Language resources and technologies for processing and linking historical documents and archives, (LREC 2014)*, Reykjavik, Iceland, 26 May 2014. Online at: <http://borel.slu.edu/pub/ria.pdf> [Retrieved 9 June 2014].

ZGUSTA L. (1972). *Manual of lexicography*. The Hague: Mouton.

WULFF S. (2008). *Rethinking idiomaticity: a usage-based approach*. London: Continuum.