The MOLOR Lemma Bank: A New LLOD Resource for Old Irish

Theodorus Fransen[†], **Cormac Anderson**[‡], **Sacha Beniamine**[‡], **Marco Passarotti**[†]
[†]Università Cattolica del Sacro Cuore, Milan, Italy, [‡]University of Surrey, Guildford, United Kingdom
{theodorus.fransen, marco.passarotti}@unicatt.it, {cormac.anderson, s.beniamine}@surrey.ac.uk

Abstract

This paper describes the first steps in creating a Lemma Bank for Old Irish (600-900CE) within the Linked Data paradigm, taking inspiration from a similar resource for Latin built as part of the LiLa project (2018–2023). The focus is on the extraction and RDF conversion of nouns from Goidelex, a novel and highly structured morphological resource for Old Irish. The aim is to strike a good balance between retaining a representative level of morphological granularity and at the same time keeping the number of lemma variants within workable limits, to facilitate straightforward resource interlinking for Old Irish, planned as future work.

Keywords: Old Irish, Lemma Bank, morphology, Linked Data

1. Introduction

While text-and-lexicon interlinking for Irish of the early medieval (and modern) period has been the subject of earlier studies and projects (Nyhan, 2008), these efforts have so far not resulted in a published resource. This is due to the highly variable nature of the language of this period in combination with the absence of sufficiently structured digital resources. The electronic Dictionary of the Irish Language, or eDIL (eDIL 2019), is the standard dictionary for Irish of the medieval period; however, it does not exhaustively list all possible inflections and spelling variants and does not use a consistent orthography for the spelling of headwords.

The lexical database Corpus PalaeoHibernicum (CorPH) "Corpus of Old Irish" (Stifter et al., 2021), covering the period ca. 6th–mid 10th century CE, is a more comprehensive and better-structured resource, but, like eDIL, does not provide word-level links to the source texts. Additionally, due to manual annotation practice, the spelling of headwords in the CorPH database is not entirely consistent, and the way it segments and stores complex morphological forms inhibits easy resource interoperability and interlinking.

The most linguistically principled and best structured lexical resource is Goidelex (Anderson et al., 2024), currently based on the 8th-century Old Irish Würzburg glosses (Kavanagh and Wodtko, 2001), which are not included in CorPH. Goidelex contains normalised lexemes with fine-grained morphophonological information. Like the aforementioned resources, however, it is not available in the Resource Description Framework (RDF) and published as Linked Data yet.[1]

To alleviate the issues around resource interlinking for medieval Irish, this paper puts forward a Lemma Bank: a collection of canonical forms for interlinking lexical and textual resources. The resource is developed as part of the MOLOR — Morphologically Linked Old Irish Resource — project, which aims to create a new lexicographic model and standard for Old Irish for linking inflected forms in a text with a full-form lexicon, with a focus on the Würzburg glosses. The project takes inspiration from the Linking Latin (LiLa) project,[2] as part of which a Lemma Bank has been developed (Moretti et al., 2023), which was conceived — and has proven to successfully act — as a hub for linking lexical resources and texts for Latin (Passarotti et al., 2020).

Admittedly, while early medieval Irish sources represent the largest corpus of pre-twelfth-century European vernacular material, other than Latin and Greek (Stacey, 1991; Eska, 2019), a mature Natural Language Processing (NLP) pipeline comparable to Latin does not yet exist for automatic processing of medieval Irish texts, which is an urgent desideratum in the field. Dereza et al. (2023) attribute poor performance of NLP models to the lack of a linguistic and editorial standard for historical Irish and prompt Celticists and historical linguists to engage in further discussion. Promising advancements have nonetheless been made over the last 10–15 years, including on NLP core tasks such as tokenisation, lemmatisation, part-of-speech (POS) tagging, and morphological analysis and generation (Lamb and Fransen, forthcoming).

The creation of digital text archives for medieval Irish, however, goes much further back with the establishment of CELT — Corpus of Electronic Texts — in 1997, Ireland's longest-running Humanities Computing Project (Ó Corráin et al., 1997); this resource contains 688 source texts in Irish (or Scottish Gaelic), albeit without linguistic annotation.

[1]For RDF see https://www.w3.org/RDF/. See also Tim Berners-Lee's Web Design Issues (Berners-Lee, 1996–present), particularly his Design Principles and his four rules about Linked Data

[2]https://lila-erc.eu/

LDL: Linked Data in Linguistics @ LREC-COLING 2024, pages 37–43
20-25 May, 2024.

More recently created digital corpora include the text archive as part of CorPH (Stifter et al., 2021) as well as two Old Irish corpora with a combined total of 98 syntactically annotated glosses following the Universal Dependencies (UD) framework, including 42 glosses from the Würzburg corpus; 3,469 POS-tagged glosses containing 21,749 tokens from the St. Gall UD corpus have been the basis for machine-learning-based POS-tagging experiments on diplomatically edited Old Irish text (Doyle and McCrae, 2024).

Interconnecting medieval Irish corpora with lexical resources using the Linked Data paradigm would be a major boon to medieval Irish studies. Despite the current lack of a mature NLP pipeline, a Lemma Bank, functioning as a central hub and interface, is considered to be a vital component in the envisaged MOLOR Knowledge Base of interlinked textual and lexical resources.

The focus of this short paper is on the first steps in creating a Lemma Bank for Old Irish (600–900CE), with a focus on the Würzburg glosses: the extraction and conversion into RDF of nouns contained in Goidelex (Anderson et al., 2024). We report on the design choices in selecting canonical forms, striking a balance between, on the one hand, linguistic granularity and, on the other hand, a workable amount of canonical forms (i.e. lemmas), while adhering to standards and best practices that have emerged in the area of Linguistic Linked Open Data (LLOD), notably the LiLa project and OntoLex (McCrae et al., 2017).

This paper is structured as follows. Section 2 introduces the resources instrumental in creating an Old Irish Lemma Bank. The LiLa Lemma Bank proved useful as an example for design choices, while a subset of the Goidelex data was used for the Lemma Bank's content. The conversion of this content into RDF using existing ontologies for linguistic annotation is the topic of section 3. Some preliminary conclusions, as well as planned future research directions, are discussed in section 4.

2. Resource context

2.1. The LiLa Lemma Bank

The goal of the ERC-funded LiLa project (2018–2023) was to interconnect distributed (lexical and textual) resources and NLP tools for Latin by using the Linked Data paradigm, which is the basis of the so-called Semantic Web (Berners-Lee et al., 2001). As Passarotti et al. (2020, 187) have pointed out, "The core of the LiLa Knowledge Base consists of a large collection of Latin lemmas: interoperability is achieved by linking all those entries in lexical resources and tokens in corpora that point to the same lemma". The resulting Lemma Bank currently contains 215,102 Latin dictionary forms (Mambrini and Passarotti, 2023).

The design principles of the LiLa Lemma Bank are according to the specification of the lexicon model for ontologies (OntoLex-Lemon) as resulting from the work of the W3C Ontology Lexicon Community Group.[3] This specification has emerged as the *de facto* standard for describing the content of lexical resources in the Linked Data framework.

It should be stressed that the LiLa Lemma Bank is not a lexical resource, that is, consisting of individuals belonging to the OntoLex-Lemon Lexical Entry class (`ontolex:LexicalEntry`). Instead, it is merely a collection of entities subsumed under the OntoLex Form class (`ontolex:Form`), for which an in-house class was devised — `lila:Lemma`. Being an OntoLex Form, a LiLa lemma can be linked to a Lexical Entry in any lexical resource via the property `ontolex:canonicalForm`, a subproperty of `ontolex:lexicalForm` (see Figure 1), connecting all other lexical resources compiled using the OntoLex-Lemon formalism (Passarotti et al., 2020).

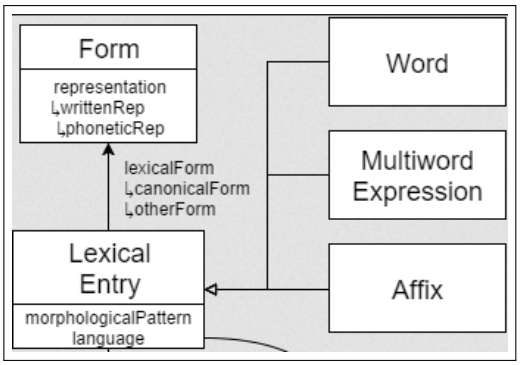

Figure 1: Part of the OntoLex-Lemon core model (Cimiano et al., 2016): the relationship between the classes `ontolex:LexicalEntry` and `ontolex:Form`

According to the OntoLex specification (Cimiano et al., 2016), "A Lexical Entry [...] needs to be associated with at least one form, and has at most one canonical form".[4] In order to allow for the use of different canonical forms used in lexical resources and lemmatised corpora, and not impose a lemmatisation criterion, the LiLa project created the symmetric property `lila:lemmaVariant`, whose range is the LiLa lemma class, "making it possible to retrieve from the textual [and lexical] resources connected to LiLa all the tokens that belong to the same lex-

[3] https://www.w3.org/community/ontolex/

[4] For example, the Lexical Entry for Old Irish 'man' would be linked to the canonical `nom.sg` form *fer*, while inflected forms such as *fir*, *feraib* etc. would have the subproperty `ontolex:otherForm` (not part of the MOLOR Lemma Bank)

ical item, regardless of the lemmatization criteria followed in individual corpora" (Passarotti et al., 2020, 190).

Lemma variants were devised for cases where morphological properties differ as part of the same lexical item, ignoring orthographic and phonetic variation which has no inflectional implications. For example, citation forms such as *claudo*, *claudor*, *claudeo* and *claudeor* 'to limp' constitute four different (LiLa) lemmas (i.e. OntoLex Forms) since each belongs to a different conjugation pattern and may be used as a lemma in lemmatised corpora or lexical resources. The lemma (Form) *claudo*, however, subsumes the graphical variant *cludo* (alongside *claudo*), encoded as written representation belonging to the same lemma (Passarotti et al., 2020). It is important to note that lemma variants, each represented by an OntoLex Form, receive a separate Uniform Resource Identifier (URI),[5] while written representations (`ontolex:writtenRep`) are encoded as strings of the data property type with range `rdf:langString`, as such being merged as part of the same URI (see Figure 2).

2.2. Goidelex

Goidelex (Anderson et al., 2024) consists of a relational database (currently) containing 574 Old Irish nouns. It provides a normalised representation of lexemes (Fransen et al., 2023) and structured groupings into lexemes and flexemes (Fradin and Kerleroux, 2003; Thornton, 2018; Pellegrini, 2023). Lexemes are defined on the basis of shared meaning and POS type, while inflectional variants that belong to the same lexeme are analysed as separate flexemes. Each lexeme is associated with the corresponding identifiers in eDIL and CorPH (section 1), making Goidelex interoperable with existing resources. Flexemes are accompanied by phonemic transcriptions following Anderson (2016) as well as morphological and phonological properties. This resource also contains information about etymology and derivational morphology. Furthermore, it is designed to be compatible with the Paralex (Beniamine et al., 2023) and the Cross-Linguistic Data Format (Forkel et al., 2018) standards.

3. Conversion Principles

3.1. Motivations

Since Goidelex contains normalised forms and uses a principled approach to dealing with inflectional variation, it was found to be the most suitable resource among the ones mentioned in section 1 for starting to populate the Lemma Bank. However, Goidelex is too linguistically granular for the purposes of the Lemma Bank, necessitating a conversion process that often involves more-to-one mappings based on flexemes (see section 3.3).

While Goidelex is intended as a fine-grained morphological resource, the Lemma Bank's function is rather to offer standardised entities identified by URIs to which other resources can link. Passing through the Lemma Bank, resources referring to these URIs will be made interoperable with each other. Linking to Goidelex is built-in, as its lexemes correspond to `ontolex:LexicalEntry` linked to lemmas in the Lemma Bank. As such, in conjunction with the Lemma Bank, the contents of Goidelex will provide rich morphological and phonological information in a Linked Data-based infrastructure of texts, lexical resources, and tools.

3.2. Lemma Properties and Ontologies

The code snippet in Figure 2 illustrates part of the triples, serialised using Turtle,[6] that describe the resource : 94459, which is a `lila:Lemma`, a subclass of `ontolex:Form`.

```
<data/id/lemma/94459>
  a      lila:Lemma ;
  rdfs:label     "claudo" ;
  lila:hasInflectionType
      lila:v3r ;
  lila:hasPOS    lila:verb ;
  lila:lemmaVariant
      <data/id/lemma/94457> ,
      <data/id/lemma/94458> ,
      <data/id/lemma/94456> ;
  dcterms:isPartOf
      <data/id/lemma/LemmaBank> ;
  ontolex:writtenRep     "claudo"
      , "cludo" .
```

Figure 2: Part of the triples as part of the LiLa lemma *claudo*, serialised using Turtle. Strictly speaking, a language tag for Latin (e.g. ISO 639-3 code `lat`) is required with `rdfs:label` "claudo"

Apart from `ontolex:writtenRep`, LiLa uses its own ontology and namespace (`http://lila-erc.eu/`) for linguistic annotations, which are aligned with OLiA (Chiarcos and Sukhareva, 2015). This modelling decision has been emulated in MOLOR (see Figure 3).

3.3. Mappings

Mapping flexemes in Goidelex (section 2.2) indiscriminately onto lemmas would lead to a multitude

[5] A string that uniquely and persistently identifies a resource or concept, most commonly on the web

[6] https://www.w3.org/TR/turtle/

	LiLa		Goidelex		MOLOR	
cluster cardinality	# lemmas	percentage	# flexemes	percentage	# lemmas	percentage
1	156,323	94.81%	467	67.19%	549	91.65%
2	7,344	4.45%	188	27.05%	44	7.35%
3	999	0.61%	36	5.18%	6	1%
4	164	0.10%	4	0.58%		
5	30	0.02%				
6	18	0.01%				
	164,878	100%	695	100%	599	100%

Table 1: Statistics relating to LiLa lemma variant clusters, Goidelex (nominal) flexeme clusters, and MOLOR (nominal) lemma variant clusters

of lemma variants. Performing lemmatisation of corpora at this very fine-grained level would be challenging, in turn impeding straightforward linking between resources. By adopting a coarser granularity, we intend to facilitate the creation of accurate lemmatisers. Following the approach taken in LiLa, it was decided to create separate lemmas (a subclass of `ontolex:Form`) only where flexemes differ in inflectional properties (i.e. inflectional class, gender).

Some statistics may lend support for this decision.[7] Table 1 shows how many LiLa lemmas are in each lemma variant cluster with a cardinality of 1 (no lemma variants) to 6 (six lemmas for a lexical item).[8] The same statistics are calculated for Goidelex, but with flexemes rather than lemmas (there are no lexemes with more than 4 flexemes). Contrasting the given percentages for both resources, it becomes clear that a flexeme-to-lemma mapping would translate into about one-third of the lemmas having more than one lemma variant, as opposed to only about 5% in the LiLa Lemma Bank, negatively impacting lemmatisation and straightforward resource interlinking.

The variation seen in the Old Irish data can be categorised according to a four-way typology:

i phonologically same, morphosyntactically same; e.g. *fer* 'man', masculine o-stem — realised by one `ontolex:Form`

ii phonologically different, morphosyntactically same; e.g. *muinter*, *muntar* 'community', feminine ā-stem (see also flexeme 74.1 and 74.2 in Figure 3) — realised by one `ontolex:Form` and two spellings (`ontolex:writtenRep`)

iii phonologically same, morphosyntactically different; e.g. *fius* 'knowledge', neuter or masculine u-stem, alternatively neuter o-stem — realised by three `ontolex:Form`s

iv phonologically different, morphosyntactically different; e.g. *brith* 'carrying', feminine i-stem, alternatively *breith*, feminine ā-stem — realised by two `ontolex:Form`s

The upshot of this decision is that flexemes with the same inflectional properties but a different phonological representation (and hence a different range of possible spellings)[9] are merged as part of a single `ontolex:Form`. Figure 3 illustrates a case with three flexemes mapped to two Forms, i.e. MOLOR lemmas, which are lemma variants of each other.

Consideration has also been given to lexemes that have different or additional stem classes for singular and plural, leading to inflectional micro-classes, e.g. *duine* 'person' (`gen_masc;stem_io;num_sg`) with suppletive plural *doíne* (`gen_masc;stem_i;num_pl`), and *demun*/*demon* 'demon, devil' (`gen_masc;stem_o;num_all`), additionally *demon* with both a different gender and inflectional pattern in the plural (`gen_neut;stem_i;num_pl`). In the case of defective nouns, Goidelex uses a combined class as part of one flexeme, e.g. *aipgitir* 'alphabet' (`indecl/i;num_all`; indeclinable in the singular, i-stem inflection in the plural).

We are currently looking for satisfying ways to model this micro-variation, keeping in mind that a lemma is modelled as a form (`ontolex:Form`) rather than a lexeme, and ideally should not predicate features that only apply to parts of the paradigm, perhaps not even to the (lemmatic) form itself — the `nom.sg` in the case of nouns. However, it was decided to provisionally take over the inflectional micro-classes as they are encoded in Goidelex, with the exception of flexemes with different plural inflection, i.e. those encoded with `num_pl`, which were ignored (only 4 cases). This

[7] This comparison should not be understood as solely reflecting the difference in the range of variation found in these languages; different design decisions in the resources concerned undoubtedly play a role as well

[8] The lemma as represented here does not include multiple written representations (e.g. *claudo*, *cludo*), which would result in a higher number

[9] One spelling in Goidelex since this resource uses normalised spellings, i.e. a one-to-one mapping between phonological form and written representation

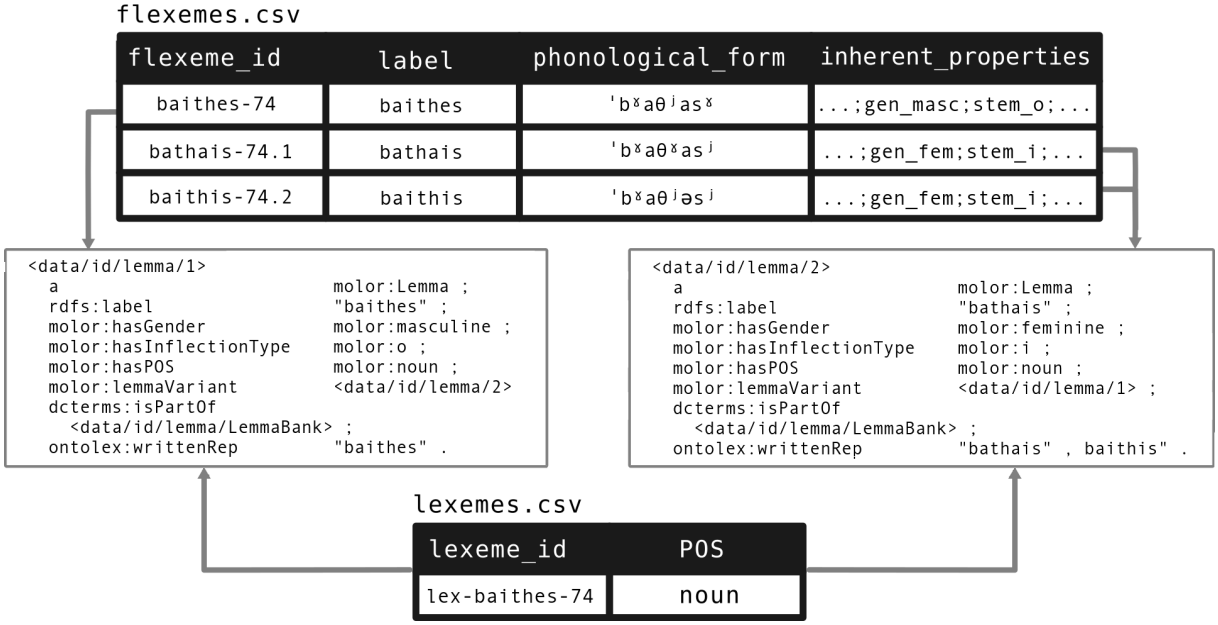

Figure 3: Example entry in Goidelex mapped to RDF-encoded MOLOR lemmas (`ontolex:Form`s). The required Old Irish language tag (e.g. ISO 639-3 code `sga`) with `rdfs:label` is not shown in this example

leads to the statistics given for MOLOR in the two rightmost columns in Table 1, which are now more similar to those of the LiLa Lemma Bank.

Lastly, we currently merge the Goidelex POS types `compound_noun`, `numeral_noun`, `prefixed_noun`, `proper_noun`, and `verbal_noun` into just `noun`; this information could be used at a future stage to establish derivational relationships (also encoded as part of Goidelex), likely to be modelled as an external resource similar to Word Formation Latin (Litta et al., 2020).

4. Conclusion and Future Work

This paper has described the first steps in converting the content of Goidelex (Anderson et al., 2024), a novel and highly structured lexical resource for Old Irish, into a Linked Data Lemma Bank, currently focussing on nouns. For design choices we have relied on the Lemma Bank developed as part of the LiLa Knowledge Base (Passarotti et al., 2020).

The next steps involve adding more lemmas with different POS categories from lexical resources, with the verb being the first in line. There are undoubtedly new challenges to overcome; the Old Irish verbal system (McCone, 1997) is much more complicated than the noun, whose inflectional patterns, as shown in this paper, already show an intricate interplay between morphology and phonology. Since verbs have not yet been systematically incorporated into Goidelex, and in the absence of a resource similar to Goidelex, we will have to resort mostly to other resources following a somewhat less granular approach. Thanks to its comprehensiveness and tabular format, CorPH (Stifter et al., 2021), in conjunction with the Würzburg dictionary (Kavanagh and Wodtko, 2001), is considered to be the most suitable starting point.

It is hoped and indeed expected that an Old Irish Lemma Bank will be an important hub in an interlinked resource framework, making medieval Irish texts and the language's grammar more accessible to scholars with various backgrounds. In the meantime, the authors gladly receive feedback from the Linked Data community on best practices for modelling under-resourced historical languages, especially in relation to variability and uncertainty.

5. Acknowledgements

Theodorus Fransen has received funding from the European Union's Horizon Europe scientific research initiative under the Marie Skłodowska-Curie Actions (MSCA), grant agreement No 101106220 (MOLOR — Morphologically Linked Old Irish Resource). Cormac Anderson is funded by a British Academy Grant (GP GP300169) while Sacha Beniamine is funded by a Leverhulme Early Career Fellowship (ECF-2022-286). We would also like to thank the reviewers, especially reviewers 2 and 3, for their feedback and helpful comments.

6. Ethical Considerations and Limitations

To the best of our knowledge, there are no ethical concerns pertaining to this resource.

7. Bibliographical References

Cormac Anderson. 2016. *Consonant colour and vocalism in the history of Irish*. Ph.D. thesis, Adam Mickiewicz University, Poznań.

Cormac Anderson, Sacha Beniamine, and Theodorus Fransen. 2024. Goidelex: A lexical resource for Old Irish. Paper accepted to *Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA)* at LREC-Coling 2024.

Sacha Beniamine, Cormac Anderson, Mae Carroll, Matías Guzmán Naranjo, Borja Herce, Matteo Pellegrini, Erich Round, Helen Sims-Williams, and Tiago Tresoldi. 2023. Paralex: a dear standard for rich lexicons of inflected forms. In *Presentation at International Symposium of Morphology*. https://www.paralex-standard.org.

Tim Berners-Lee. 1996–present. Design issues: Architectural and philosophical points. https://www.w3.org/DesignIssues/. Accessed: March 31, 2024.

Tim Berners-Lee, James Hendler, and Ora Lassila. 2001. The Semantic Web. *Scientific American*, 284(5):34–43.

Christian Chiarcos and Maria Sukhareva. 2015. OLiA - ontologies of linguistic annotation. *Semantic Web*, 6:379–386.

Philipp Cimiano, John P. McCrae, and Paul Buitelaar. 2016. Lexicon Model for Ontologies: Community report. W3C community group final report, World Wide Web Consortium. https://www.w3.org/2016/05/ontolex/. Accessed: March 31, 2024.

Oksana Dereza, Theodorus Fransen, and John P. Mccrae. 2023. Do not trust the experts - how the lack of standard complicates NLP for historical Irish. In *Proceedings of the Fourth Workshop on Insights from Negative Results in NLP*, pages 82–87, Dubrovnik, Croatia. Association for Computational Linguistics.

Adrian Doyle and John P. McCrae. 2024. Developing a part-of-speech tagger for diplomatically edited Old Irish text. Paper accepted to *Third Workshop on Language Technologies for Historical and Ancient Languages (LT4HALA)* at LREC-Coling 2024.

Charlene M. Eska. 2019. *A Raven's Battle-cry: The limits of Judgment in the medieval Irish Legal Tract Anfuigell*, volume 27 of *Medieval Law and Its Practice*. Brill, Leiden and Boston.

Robert Forkel, Johann-Mattis List, Simon J. Greenhill, Christoph Rzymski, Sebastian Bank, Michael Cysouw, Harald Hammarström, Martin Haspelmath, Gereon A. Kaiping, and Russell D. Gray. 2018. Cross-linguistic data formats, advancing data sharing and reuse in comparative linguistics. *Scientific Data*, 5(180205).

Bernard Fradin and Françoise Kerleroux. 2003. Troubles with lexemes. In Geert Booij, Janet DeCesaris, Angela Ralli, and Sergio Scalise, editors, *Selected papers from the third Mediterranean Morphology Meeting*, pages 177–196. IULA – Universitat Pompeu Fabra.

Theodorus Fransen, Cormac Anderson, and Sacha Beniamine. 2023. Towards a normalised orthography for Old Irish. Paper at *36th Irish Congress of Medievalists*, Dublin, 22–23 June 2023.

Séamus Kavanagh and Dagmar S. Wodtko. 2001. *A lexicon of the Old Irish glosses in the Würzburg manuscript of the epistles of St. Paul*. Verlag der Österreichischen Akademie der Wissenschaften, Vienna.

William Lamb and Theodorus Fransen. Forthcoming. Irish and Scottish Gaelic language technology. In Joe Eska, Paul Russell, Peadar Ó Muircheartaigh, and Silva Nurmio, editors, *Palgrave Handbook of Celtic Languages and Linguistics*. Palgrave Macmillan, London.

Eleonora Litta, Marco Passarotti, and Francesco Mambrini. 2020. Derivations and connections: Word formation in the LiLa Knowledge Base of linguistic resources for Latin. *The Prague Bulletin of Mathematical Linguistics*, 115:163–186.

Francesco Mambrini and Marco Carlo Passarotti. 2023. The LiLa Lemma Bank: A Knowledge Base of Latin canonical forms. *Journal of Open Humanities Data*, 9(1):28.

Kim McCone. 1997. *The Early Irish verb*, 2nd edition. An Sagart, Maynooth. Revised edition with *index verborum*.

John P. McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. The OntoLex-Lemon model: Development and applications. In *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference*, pages 587–597, Brno, Czech Republic. Lexical Computing CZ s.r.o.

Julianne Nyhan. 2008. Developing integrated editions of minority language dictionaries: the Irish example. *Literary and Linguistic Computing*, 23(1):3–12.

Marco Passarotti, Francesco Mambrini, Greta Franzini, Flavio Massimiliano Cecchini, Eleonora Litta, Giovanni Moretti, Paolo Ruffolo, and Rachele Sprugnoli. 2020. Interlinking through lemmas. The lexical collection of the LiLa Knowledge Base of linguistic resources for Latin. *Studi e Saggi Linguistici*, 58(1):177–212.

Matteo Pellegrini. 2023. Flexemes in theory and in practice. *Morphology*, 33:361–395.

Robin Chapman Stacey. 1991. Law and order in the very old West: England and Ireland in the early Middle Ages. In Benjamin T. Hudson and Vicki Ziegler, editors, *Crossed paths: methodological approaches to the Celtic aspect of the European Middle Ages*, pages 39–60. University Press of America, Lanham and London.

Anna M. Thornton. 2018. Troubles with flexemes. In Oliver Bonami, Gilles Boyé, Hélène Firaudo, and Fiammetta Namer, editors, *The lexeme in descriptive and theoretical morphology*, pages 202–321. Language Science Press, Berlin.

8. Language Resource References

Cormac Anderson, Sacha Beniamine, and Theodorus Fransen. 2024. *Goidelex: A Lexical Resource for Old Irish*. Zenodo. PID https://doi.org/10.5281/zenodo.10898228.

David Stifter, Bernhard Bauer, Elliott Lash, Fangzhe Qiu, Nora White, Siobhán Barrett, Aaron Griffith, Romanas Bulatovas, Francesco Felici, Ellen Ganly, Truc Ha Nguyen, and Lars Nooij. 2021. *Corpus PalaeoHibernicum (CorPH) v1.0*. Maynooth University. PID http://chronhib.maynoothuniversity.ie.

Donnchadh Ó Corráin, Hiram Morgan, Beatrix Färber, Benjamin Hazard, Emer Purcell, Caoimhín Ó Dónaill, Hilary Lavelle, Julianne Nyhan, and Emma McCarthy. 1997. *CELT: Corpus of Electronic Texts*. University College Cork. PID http://www.ucc.ie/celt.

eDIL 2019. *An Electronic Dictionary of the Irish Language,* based on the Contributions to a Dictionary of the Irish Language (Dublin: Royal Irish Academy, 1913–1976) (www.dil.ie 2019). PID https://www.dil.ie.

Giovanni Moretti, Marco Passarotti, Rachele Sprugnoli, Paolo Ruffolo, and Francesco Mambrini. 2023. *CIRCSE LiLa Lemma Bank (V1.2)*. Zenodo. PID https://doi.org/10.5281/zenodo.8300851.