The Application of XML to the historical lexicography of Old, Middle, and Early-Modern Irish: a Lexicon based analysis

Julianne Nyhan

NATIONAL UNIVERSITY OF IRELAND, CORK

FACULTY OF ARTS

DEPARTMENT OF HISTORY

Degree of PhD

May 2006

Head of Department: Prof Dermot Keogh

Supervisor: Prof Donnchadh Ó Corráin
# Table of Contents

Acknowledgements ......................................................... viii
Abstract ........................................................................... xi

1 Introduction: an overview of the history of lexicography ........................................ 1  
1.1 Abstract ........................................................................ 1  
1.2 Proto-Dictionaries and Greek innovators ........................................... 2  
1.3 Development of lexicography in Greece, Rome and Byzantium  
    from *De verborum significatu* to the *Suda* ................................... 4  
1.4 Ordering systems .......................................................... 12  
    1.4.1 Thematic origins ......................................................... 12  
    1.4.1.1 Pliny the Elder’s *Historia naturalis* ......................... 14  
1.5 Evidence of Alphabetisation ............................................... 15  
    1.5.1 Alphabetic origins ....................................................... 15  
    1.5.2 Alphabetisation in Old-English Glossaries ...................... 18  
    1.5.3 Alphabetisation in Medieval-Latin Dictionaries ................ 19  
1.6 Introduction ................................................................... 22  
1.7 Old-Irish Glossaries ........................................................ 22  
    1.7.1 Metrical glossaries ....................................................... 23  
    1.7.2 Text-dependent glossaries ........................................... 23  
    1.7.3 Medical glossaries ....................................................... 23  
1.8 Scholarship on the compilation of glossaries ........................................ 25  
1.9 List of manuscripts containing major medieval Irish-language  
    glossaries .................................................................. 26  
1.10 O’Davoren’s Glossary ...................................................... 28  
1.11 Lecan Glossary ............................................................. 30
# The history and limitations of the *Dictionary of the Irish Language* 53

## 2.1 Abstract 53

## 2.2 Description of the *Dictionary of the Irish Language* 53

### 2.2.1 Dialectical information supplied in the *Dictionary of the Irish Language* 55

## 2.3 Early history of the *Dictionary of the Irish Language* 56

### 2.3.1 The dictionary in 1910 60

### 2.3.2 The appointment of Marstrander 61

### 2.3.3 Marstrander’s *D* 63

### 2.3.4 The correspondence between Meyer and Best 64

### 2.3.5 Controversy 67

### 2.3.6 Comparison of Marstrander’s *D* with Meyer’s *Contributions* 72

### 2.3.7 History of *DIL* from 1913 to 1975 75

### 2.3.8 Editors and assistants of the *Dictionary of the Irish Language* 77

## 2.4 Limitations of the *Dictionary of the Irish Language* 80

## 2.5 Navigating *DIL*: inadequate cross-referencing 80

### 2.5.1 air- 80

#### 2.5.1.1 Analysis of *airábind* to *airbaid* 81

### 2.5.2 ar-fo- 83
Table of Contents

2.5.2.1 Analysis of *ar-fócair* and *ar-fogni* .......... 84
2.5.3 Internal orthography .................................. 84
2.5.4 Lemmata of complex verbs ............................ 85
2.6 Inconsistent presentation and treatment of information in *DIL* . . 88
  2.6.1 Abbreviations ....................................... 88
  2.6.2 Structure ............................................. 88
  2.6.3 Words treated of in incorrect lexical entries .......... 89
  2.6.4 Subjective information labels in *DIL*: the terms *earlier* and *later* ........................................... 90
2.7 Categorisation of limitations ............................. 91

3 Mark-up and XML: preliminary theoretical and historical considerations 93
  3.1 Abstract ................................................. 93
  3.2 Views of electronic text ................................ 94
  3.3 What is mark-up? ...................................... 98
  3.4 Types of mark-up that can be applied to electronic documents . 105
    3.4.1 Mark-up applied to electronic texts ................. 105
    3.4.2 Presentational and Procedural mark-up ............... 105
    3.4.3 Descriptive mark-up .................................. 106
  3.5 Theories of textuality .................................. 107
  3.6 The History of XML .................................... 110
  3.7 What is XML? .......................................... 113
  3.8 The Text Encoding Initiative ............................ 115

4 XML: the foundation of the electronic Lexicon 119
  4.1 Abstract ................................................. 119
# Table of Contents

4.2 Examples and description of ten different types of encoded lexical entry .................................................. 120
   4.2.1 Brief description of *ecal* ................................................. 120
   4.2.2 Brief description of *didiu* ............................................. 122
   4.2.3 Brief description of *clár* ............................................. 123
   4.2.4 Brief description of *mo* ............................................. 124
   4.2.5 Brief description of *mé* ............................................. 124
   4.2.6 Brief description of *prim* ......................................... 125
   4.2.7 Brief description of *Gall* ......................................... 125
   4.2.8 Brief description of *cen* ......................................... 126
   4.2.9 Brief description of the definite article ......................... 126
   4.2.10 Brief description of *téit* ......................................... 126
   4.2.11 Description of elements and attributes ....................... 127

4.3 The Document Type Definition .................................... 134
   4.3.1 The Document Type Declaration ................................ 135

4.4 Commentary on and explanation of lex.dtd .......................... 137
   4.4.0.1 Element declaration ........................................ 137
   4.4.0.2 Content model ............................................... 138
   4.4.0.3 Declaring an attribute .................................... 139
   4.4.0.4 Internal parameter entity ................................ 141

4.5 Character references .................................................... 143

4.6 Interrelationship of elements and attributes in lex.dtd .......... 145

4.7 XSLT examples .......................................................... 150

4.8 Examples of lexical entries .......................................... 156

5 Advantages and disadvantages of XML ................................ 240
   5.1 Abstract .......................................................... 240
5.2 Advantages and disadvantages ..................................................... 241
  5.2.1 The problem of date and context ........................................... 241
    5.2.1.1 Inadequate lexicographical criteria .................................. 241
    5.2.1.2 Limitation of OHCO theory of textuality .............................. 243
    5.2.1.3 Multilex root element .................................................. 246
  5.2.2 Advantages ............................................................................ 248
    5.3.1 The Lexicon as a stand-alone research tool .............................. 248
    5.3.2 Lexicon and Text interlinked ............................................... 252
    5.3.3 Lexicon, Dictionary, and Text interlinked ............................... 257
  5.4 Partial automation of encoding of citations ................................. 260
    5.4.1 Patterns of citation in DIL .................................................. 262
    5.4.2 Limitations of early regular expressions ................................. 263
  5.5 Predictable and unpredictable data patterns ................................ 265
    5.5.1 Predictable exceptions type A ............................................. 265
    5.5.2 Predictable exceptions type AA ........................................... 265
    5.5.3 Predictable exceptions type B ............................................. 266
    5.5.4 Predictable exceptions type AB ........................................... 267
    5.5.5 Mark-up that cannot feasibly be automated ............................. 267
    5.5.6 Variations within types of citation ....................................... 268
  5.6 Methodology ............................................................................. 269
    5.6.1 Routine 1.0 .......................................................................... 269
      5.6.1.1 Routine 1.1 .................................................................. 271
      5.6.1.2 Routine 1.2 .................................................................. 273
      5.6.1.3 Routine 1.3 .................................................................. 274
      5.6.1.4 Routine 1.4 .................................................................. 276
    5.6.2 Routine 2 ............................................................................ 276
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.6.3</td>
<td>Routine 3 and <em>sed</em>'s pattern space</td>
<td>278</td>
</tr>
<tr>
<td>5.6.4</td>
<td>Routine 4</td>
<td>279</td>
</tr>
<tr>
<td>5.6.5</td>
<td>Accuracy rates</td>
<td>280</td>
</tr>
<tr>
<td>6</td>
<td>Conclusion</td>
<td>281</td>
</tr>
<tr>
<td>7</td>
<td>Bibliography</td>
<td>288</td>
</tr>
<tr>
<td>7.1</td>
<td>Bibliographies consulted</td>
<td>288</td>
</tr>
<tr>
<td>7.2</td>
<td>Guidelines, Technical Reports and Working Group notes</td>
<td>289</td>
</tr>
<tr>
<td>7.3</td>
<td>Manuscripts</td>
<td>290</td>
</tr>
<tr>
<td>7.4</td>
<td>Dictionaries &amp; printed lexicographical sources</td>
<td>291</td>
</tr>
<tr>
<td>7.5</td>
<td>Secondary works</td>
<td>295</td>
</tr>
</tbody>
</table>
List of Figures

4.1 Processing instruction and document type declaration . . . . . . . 137
4.2 List of symbols and terms used in the DTD . . . . . . . . . . . . . 138
4.3 Element declaration . . . . . . . . . . . . . . . . . . . . . . . . . . 138
4.4 Model group . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 139
4.5 Attribute declaration . . . . . . . . . . . . . . . . . . . . . . . . . . 140
4.6 Internal parameter entity . . . . . . . . . . . . . . . . . . . . . . . 142
4.7 Invoking an internal parameter entity . . . . . . . . . . . . . . . . 142
4.8 Character references . . . . . . . . . . . . . . . . . . . . . . . . . . 143
4.10 Output text: TEI conformant mark-up . . . . . . . . . . . . . . . 155
Acknowledgements

I am indebted to my supervisor, Professor Donnchadh Ó Corráin for his generosity with both his time and expertise. His expert knowledge of Irish History and Old, Middle, and Early Modern Irish was invaluable to me throughout the course of my postgraduate studies. Further, the vision he showed in setting up the Corpus of Electronic texts, one of the first scholarly encoded electronic text corpora in Ireland, gave me (and many others) the opportunity to take my first steps in SGML and XML and to develop my research in a broad, applied context.

I am also indebted to Peter Flynn, XML expert and manager of the Research and Academic Electronic Publishing Unit at University College Cork. Without Peter’s expert guidance it would not have been possible for me to devise the Lexicon or write this work. Not only did he mentor me in XML and XSLT, but also in sed, awk, xpath, regular expressions, the Linux operating system, the Emacs editor, and \LaTeX. The generosity of Professor Ó Corráin and Peter Flynn with their time and expertise was immense and I feel privileged to have worked with them both.

I am also very grateful to Beatrix Färber, manager of the CELT corpus, for the practical support and encouragement she gave me over the past five years.

I also wish to thank Dr Gregory Toner and Dr Maxim Fomin in the University of Ulster, Coleraine. I worked with them on the Linking Dictionaries and Text project, a cross-border initiative between University College Cork and the University of Ulster, Coleraine, and also on the
electronic Dictionary of the Irish language project. The research I carried out for these projects in XML, XSLT, and Linux utilities helped me to develop a clearer understanding of the content and structure of the Dictionary of the Irish language.

Special words of thanks also go to Professor Pádraig Ó Macháin and the Dublin Institute of Advanced Studies. In 2004 Professor Ó Macháin organised a one day colloquium in the Institute, entitled ‘Digital Text and Digital Image’. As well as giving me the opportunity to air some of my research findings, I appreciated the opportunity to meet scholars involved in many aspects of digitising Irish material. As a result of the colloquium I set up the Digiquest mailing list in 2005 and my thanks also goes to the members of that list.

Thank you, also, to the staff of the Department of Early and Medieval Irish Department at University College Cork, especially Dr Kevin Murray, Dr Caitriona Ó Dochartaigh and Dr Clodhagh Downey. I was privileged to attended a range of excellent classes and seminars in the department during the course of my postgraduate studies. I also wish to thank the staff of the Ionad na Gaeilge Labhartha, especially Seán Ó Laoi.

I would also like to thank the staff of the History Department, UCC, as well as the staff of the CELT corpus, especially Benjamin Hazard, Emer Purcell and Michelle O’Halloran. I am also grateful to Tomas Ó Riordán, manager of the Multitext project and Sinead Deignan, manager of the ARCH project.

I wish to thank the Higher Education Authority who provided funding for the five years of my postgraduate studies.
Acknowledgements

I would also like to thank the members of TEI-L mailing list, the Mulberry XSLT mailing list, and the members of the ACH, ALLC, and DRH. The mailing lists and conferences organised by these groups gave me the opportunity to present and develop my research with the assistance of internationally recognised experts in Humanities Computing.

On a personal note, I would also like to thank my friends, especially Jade and Anna Hogan, Joe Gilday and Barry O Flynn for their love and support. I would like to thank my parents John and Eileen and my siblings Pat, Daniel and Muireann for giving me an upbringing that made an undertaking such as this possible as well as desirable. I am also grateful for their love, and for all the different kinds of help and support, too numerous to list, that they gave me over the years. Finally, I wish to thank Joey and Reimar Faust for their love, support, patience and inspiration.
Abstract

This work seeks to apply the eXtensible mark-up language (hereafter XML) to the lexicography of Old, Middle, and Early Modern Irish by designing an electronic Lexicon of Medieval Irish, based on the Royal Irish Academy’s *Dictionary of the Irish language*. It further seeks to evaluate the technical and academic advantages and disadvantages of this application from the perspective of both the specialist and non-specialist user of electronic resources in Irish.

Chapter 1 discusses the history of lexicography from Antiquity to Byzantium. In this context, an introduction to the major glossaries of medieval Ireland is presented, and their importance to later scholarship established. Chapter 2 sets out the history and limitations of the *Dictionary of the Irish language*, in order to establish the value of the application of XML to Irish lexicography. Preliminary theoretical considerations about the concept of mark-up, and the meta-mark-up language XML are presented in Chapter 3. Chapter 4 provides a detailed description of the XML mark-up used in the Lexicon. Examples of ten different types of XML-encoded lexical items are presented, namely, a possessive pronoun, personal pronoun, adjective, conjunction, verb, prefix, definite article, preposition and two nouns. The advantages and disadvantages of using XML to encode an electronic Lexicon of Old, Middle and Early Modern Irish are set out in Chapter 5. Chapter 6 concludes that the application of XML to the lexicography of Old, Middle and Early Modern Irish is a fruitful one. It has not only enabled the creation of an electronic
Abstract

Lexicon of Old, Middle and Early Modern Irish, but some important steps towards creating a first-class electronic infrastructure for Irish historical lexicography have been taken.
Chapter 1

Introduction: an overview of the history of lexicography

1.1 Abstract

Chapter one has two main aims. The first is to discuss the ways lexicographical information was presented and made accessible in pre-electronic contexts. The second is to provide a survey of the content and structure of the canon of lexicographical material that existed before the publication of the Dictionary of the Irish Language. The dependence of the Dictionary of the Irish Language for obscure and archaic words on the major glossaries of medieval Ireland will be pointed out. Further, in order to set out all that can have influenced indigenous Irish medieval glossatorial activity, directly or indirectly, an overview of lexicography in Greece, Rome, Byzantium and England will be presented.

Section one examines the advances made in lexicography in Western Europe from the earliest times until about the tenth century AD in order
1. Introduction: an overview of the history of lexicography

to establish a framework in which to assess the Irish lexicographical tradition. In addition to providing an introduction to the history of lexicography in Western Europe, techniques employed by lexicographers such as alphabetic and thematic organisation are explored.

The second section of this chapter discusses the structure and content of the major glossaries of medieval Irish and establishes the ordering systems used in such works. This is followed by a list of some printed Irish language dictionaries before 1913. Finally, the extent to which the lexicographers who worked on the *Dictionary of the Irish Language*—the subject of Chapter two and the work upon which the electronic Lexicon is based—were dependent upon the major glossaries of medieval Irish for some of the more rare and obsolete words will also be illustrated.

1.2 Proto-Dictionaries and Greek innovators

Before discussing the content and ordering structures of the major glossaries of medieval Irish further, sections § 1.2 to § 1.4 set out a summary of all that can have influenced indigenous Irish medieval glossatorial activity, directly or indirectly.

The word lists compiled by the Akkadians c. 2340 BC to aid the interpretation of the non-Semitic, agglutinative language spoken by their Sumerian predecessors\(^1\) have been described by some meta-lexicographers as proto-dictionaries.\(^2\) In terms of Western lexicography, however, the

---


\(^2\) Jonathon Green, *Chasing the sun: dictionary makers and the dictionaries they made*
1. Introduction: an overview of the history of lexicography

Greeks were the first to develop dictionaries proper, which they identified by the term *glossai* from the fifth century BC onwards.\(^3\)

The development of the Classical Greek term for a difficult word: *glossa* or *glotta*, has been discussed by MacArthur. He plots the semantic development of this word from the concrete meaning of ‘tongue’ to the more figurative sense of ‘language in general’. From this figurative sense the word begins to denote ‘a language that is not one’s own’, followed by a ‘foreign language’, and then develops the meaning of a complex or difficult word in a foreign language.\(^4\)

The earliest Greek glossaries were created to provide definitions of difficult words used by a single author.\(^5\) The works of Homer, which played a central role in classical education,\(^6\) were already beginning to be misunderstood by the time of Hesiod,\(^7\) and glossaries to Homer were compiled by scholars such as Protagoras of Abdera in the fifth century and Praxiphanes in the fourth century BC.\(^8\)

The scholarship of the pre-Christian period was not confined to Homer; Philetas of Cos produced the *Miscellaneous Glosses*,\(^9\) a work that treated

---


\(^4\) McArthur, op. cit. 76.


\(^8\) Collison, *Foreign-language dictionaries*, 26.
1. Introduction: an overview of the history of lexicography

of technical and dialectical vocabulary, in addition to that of Homer.

Specialised glossaries on other topics were also produced, such as the
glossary of medical terms compiled by Xenocritus of Cos and Philinus
of Cos.\textsuperscript{10} The poet Callimachus of Cyrene (c. 310 - c. 235 BC) compiled
glossaries of proper nouns such as the names of rivers and birds and
important lexicographical work was also undertaken by Aristophanes of
Byzantium (c. 257 - 180 BC.) who compiled a thematic dictionary based
on Greek dialects. The work of Aristophanes prompted the creation
of a school of lexicographers where reference works were produced by
scholars such as Nicander of Colophon.\textsuperscript{11}

1.3 Development of lexicography in Greece, Rome and

Byzantium from \textit{De verborum significatu} to the Suda

The Augustinian period witnessed the first Latin lexicon, compiled
by Verrius Flaccus (fl. 10 BC) and entitled \textit{De verborum significatu}\.\textsuperscript{12}
Flaccus, a famous tutor and teacher of the grand-children of the emperor
Augustus was the most distinguished scholar of Rome since Varro.\textsuperscript{13} His
alphabetised lexicon is considered by some as the earliest example of an

\textsuperscript{9} Frederick John Williams, ‘Philitas’, S Hornblower & A. Spawforth (ed), \textit{The Oxford

\textsuperscript{10} Shaw, \textit{Contributions}, §1.1.

\textsuperscript{11} ibid.

\textsuperscript{12} John Edwin Sandys, \textit{A history of Classical scholarship from the sixth century BC to the
end of the middle ages} (Cambridge 1906) 200.

\textsuperscript{13} Robert K. Kaster, ‘Marcus Verrius Flaccus’, S. Hornblower & A. Spawforth (ed),
\textit{The Oxford Classical dictionary}, 1589.
encyclopaedic dictionary\textsuperscript{14} because of the scope of information contained in it.

In addition to lexicographical material, \textit{De verborum significatu} supplied information pertaining to grammar, history, law, and antiquities.\textsuperscript{15} The work survives only in part: just seventy of the 128 parchment leaves of the surviving copy are extant, and the letters after \textit{M} are lost.\textsuperscript{16} Indeed, that any of the work has survived is due to an abridgement made in the second century by Pompeius Festus, and a later abridgement of that work by Paulus Diaconus in the eighth century.\textsuperscript{17}

\textit{De verborum significatu} was to influence writers and lexicographers for many generations. Sandys notes that its influence can be traced in writers such as Quintilian, Gellius, Nonius and Macrobius,\textsuperscript{18} while Green holds that \textit{De verborum significatu} was a vital source not only for Calepino’s famous polyglot \textit{Dictionarium} of the sixteenth century, but also for many later English lexicographers.\textsuperscript{19}

Lexicographical research in the first and second centuries was carried out predominately by Greek scholars from Alexandrian circles, who were motivated by the need to explain Homer’s language, the ‘difficult words used by poets and dramatists, and … literary usage’.\textsuperscript{20} Thus,

\begin{thebibliography}{9}
\bibitem{14} Collison, \textit{Foreign-language dictionaries}, 26.
\bibitem{15} Sandys, \textit{A history of classical scholarship}, 200.
\bibitem{16} Green, \textit{Chasing the sun}, 46.
\bibitem{17} Sandys, \textit{A history of classical scholarship}, 200; Collison, \textit{Foreign-language dictionaries}, 26 and Shaw, \textit{Contributions}, §1.2.
\bibitem{18} Sandys, \textit{A history of classical scholarship}, 200.
\bibitem{19} Green, \textit{Chasing the sun}, 46.
\bibitem{20} Collison, \textit{Foreign-language dictionaries}, 33.
\end{thebibliography}
the lexicographers of the first and second centuries AD had much in
common with their pre-Christian predecessors, and the preoccupations
of many of the reference works of this era mirrored those of earlier times.

Pamphilus of Alexandria created an extensive dictionary of obscure
or complex words; Apion, the Alexandrian Greek, produced a glossary
of Homer’s language; and Theon of Alexandria is supposed to have
compiled a lexicon of comedy and tragedy based on materials collected
by Didymus. The lexicon compiled by Valerius Harpocration can be
considered somewhat unusual for its time in that it was based on the
language of the Attic orators and thus reflected the contemporary spoken
word.

Lexicography received a fresh impetus in the second century, when a
new genre of literature, based on the ‘prevailing fancy for imitating the
great Attic models of the past’, was developed by the new Sophists. In
order to facilitate the interpretation of difficult words in this literature,
a number of reference works, ‘probably the first puristic dictionaries in
the world’, were developed by lexicographers known as ‘Atticists’.

Lexicons of Attic words were compiled by Aelius Dionysius, Pausanias
the Atticist, Galen of Pergamon and Aelius Moeris. A selection of
words used by orators and writers such as Thucydides and Demosthenes

22 Collison, op. cit. 33.
23 Sandys, A history of classical scholarship, 200.
24 Shaw, Contributions quoting L. Zgusta and D. Georgacas, Lexicography of ancient
Greek, 1694-1704.
25 Collison, Foreign-language dictionaries, 35.
1. Introduction: an overview of the history of lexicography

was made by Julius Vestinus while his contemporary and townsman Valerius Pollio performed the same task for the language of the poets. In turn, Pollio’s son Diodorus compiled a lexicon based on the difficult terms used by the Attic orators.26

Phrynichus (fl. 180) has been described as one of the most interesting lexicographers among the Atticists.27 He is likely to have taught rhetoric in Bithynia under Marcus Aurelius and Commodus and he compiled a lexicon called the Rhetorical magazine in 37 books.28 The work drew, inter alia, on the lexicon of Aelius Dionysius, orators such as Aeschines and Critias, and the writings of dramatists such as Euripides and Aeschylus.29

The Rhetorical magazine has been lost apart from a section published in Bekker’s Anecdota i pp. 1-74 and a summary of it made by Photius.30 The Lexica of Dionysius, Moeris and Phrynichus, among others, were further complemented by an anonymous publication, the Anti-Atticist, which quoted authorities such as Orus31 on words that were proscribed in Attic usage.32

Another important lexicon produced in this century is the Hermeneumata

---

26 Sandys, A history of classical scholarship, 323.
27 ibid.
28 ibid.
31 Sandys, A history of classical scholarship, 325.
32 Collison, Foreign-language dictionaries, 35.
1. Introduction: an overview of the history of lexicography

Glossary. It consisted of a collection of verbal paradigms followed by a number of colloquia. Foremost among the lexica of this period was Julius Pollux’s Onomasticon, a work that was to influence later dictionaries such as Withals’s shorte dictionarie for yonge beginners (1553).\(^{33}\) The Onomasticon is thematic in arrangement and dedicated to Commodus.\(^ {34}\) It is of interest in regard to the specialised language\(^ {35}\) that it covers: Book IV treats of music, dancing and the Greek theatre; Book VII the language of law and bureaucracy; while Book IX pertains to numismatics.\(^ {36}\) The work also lists a large number of synonyms and antonyms and has 33 different terms of opprobrium for tax collectors.\(^ {37}\)

Latin lexicography in the second century was represented by Aulus Gellius in his Noctes atticae.\(^ {38}\) This compendium contains chapters based on notes or excerpts he made while reading on a wide variety of topics such as grammar, philosophy and law and approximately a quarter of the work treats of lexicography.\(^ {39}\) The Noctes atticae was widely read by later lexicographers such as Nonius Marcellus, Ammianus and Macrobius.\(^ {40}\)

Little information remains in regard to lexicography in the third

\(^{33}\) Green, Chasing the sun, 47.
\(^{34}\) ibid.
\(^{35}\) Collison, Foreign-language dictionaries, 34.
\(^{36}\) Sandys, A history of classical scholarship, 327.
\(^{39}\) Shaw, Contributions, §1.2.
1. Introduction: an overview of the history of lexicography

century. However, Collison has drawn attention to the *Learned banquet* of Athenaeus of Naucratis (fl. 200) which mentions thirty-five compilers of glossaries.\(^{41}\)

In the fourth century the encyclopaedic work *De compendiosa doctrina* was compiled by Nonius Marcellus of Thubursicum in Numidia. It contains three sections: antiquarian, lexicographical and grammatical, and the latter section draws heavily on the work of Verrius Flaccus, among others.\(^{42}\) The work has been described as ‘careless’\(^{43}\) and its prime value lies in its excerpts from early Latin literature.\(^{44}\) The etymological lexicon created by Orion, consort of Theodosius II at Constantinople, was to become one of the most important reference works of this century since it was used as a source of many later etymological compilations of the Byzantine epoch. A lexicon of synonyms was also prepared by Ammonius, though Sandys notes that it appears to be a revised edition of the lexicon of Herennius Philon.\(^{45}\)

The lexicon of Hesychius of Miletus is known to us only through the many citations that are included in the *Suda* and it is thought to have been completed in the fifth century.\(^{46}\) Hesychius was an associate of Theodosius the Grammarian,\(^{47}\) and his sources included Aelius Dionysius

---

41 Collison, *Foreign-language dictionaries*, 35.
43 ibid.
44 Collison, *Foreign-language dictionaries*, 35.
46 Collison, *Foreign-language dictionaries*, 36.
47 Green, *Chasing the sun*, 47.
and Herennius Philon. Other notable works produced during this century included the geographical lexicon of Stephanus of Byzantium and the glossary of St Eucherius, bishop of Lyons.

In the sixth century the *Expositio sermonum antiquorum* was prepared by the North African Fabius Planciades Fulgentius (c. 468 - 533). Placidus also prepared a glossary that contained lexicographical information in the form of brief notes on archaic words as far as the letter *P*. His work was also to become well known and influential among medieval glossators and lexicographers.

The Christian cleric Photius (c. 810 to c. 891), who became the patriarch of Constantinople on two occasions, is one of the most important figures in Byzantine classical studies. In addition to his work in areas such as theology, he wrote a lexicon that drew on the works of earlier lexicographers such as Aelius Dionysius, Pausanias, and Diogenianus.

The title of his Lexicon has been translated as the *Collection of words arranged alphabetically by which works of historians and orators are especially beautified*. It was known only through the *Codex Galeanus* at Cambridge, and was thought to have been lost to posterity until it was rediscovered.

---

50 Shaw, *Contributions*, §1.2.
51 Kaster, op. cit. 1175.
52 Paul Lemerle, *Byzantine humanism: the first phase; notes and remarks on education and culture in Byzantine from its origins to the 10th century*, tr. Helen Lindsay and Ann Moffatt (Canberra 1986) 216.
in a monastery in Macedonia in 1959.\footnote{Green, \textit{Chasing the sun}, 49-50.}

The final work of lexicography noticed in this section is the Byzantine \textit{Suda},\footnote{David Whitehead (ed), \textit{Suda on-line: Byzantine lexicography} http://www.stoa.org/sol/} a work that appeared either in the reign of the Emperor John Tzimisces (969 - 76) or Basil II (976 - 1025). The word \textit{Suda} was assumed by commentators—beginning in the later Byzantine era and continuing until the mid nineteenth century—to have been the name of the person who compiled the dictionary. However, later research indicates that the work is more likely to have been compiled by a group of scholars.\footnote{Green, op. cit. 47-9; Lemerle, op. cit. 343-45.}

The \textit{Suda} is a mixture of both lexicographical and encyclopaedic materials and has been described as: ‘a succession of some thousands of items, their extent varying from a single word to a page or more. In it we find the explanation of a difficult form ... or of a rare word, a grammatical point, elucidations of words with several meanings, as well as notes on people, places, institutions, and on concepts. It is essentially an historical and literary encyclopaedia, but it is also a collection of proverbs and a kind of dictionary of quotations’.\footnote{Lemerle, \textit{Humanism}, 345.} The sources of the \textit{Suda} included Hesychius of Miletus,\footnote{Sandys, \textit{A history of classical scholarship}, 387.} the \textit{Palatine anthology}, dramatists such as Aristophanes and, of course, Homer.
1.4 Ordering systems

Chapters four and five will discuss in detail the application of aspects of twenty-first century information science to the historical lexicography of Old, Middle and Early Modern Irish. The ways that XML can be used to develop sophisticated lexicographical resources that offer superior search and interrogation facilities will also be illustrated. This section aims to set out an introduction to the historical background of the use of information science in pre-electronic Irish lexicographical resources. In this introduction, special emphasis will be placed on the ways that lexicographical material was organised and made accessible in the pre-electronic textual environments of the major glossaries of medieval Ireland.

Two types of ordering systems, alphabetic and thematic, have been employed by compilers of dictionaries, lexica, and glossaries throughout the ages. Both systems have been used in varying degrees since the earliest reference works were compiled, and the course of their evolution has not always run smooth. Alphabetic ordering may be ubiquitous in present-day lexicography, but its struggle for dominance over thematic ordering is an interesting and lively one, influenced by factors ranging from philosophical to economic.

1.4.1 Thematic origins

MacArthur states that from Antiquity to medieval times, thematic organisation was employed more frequently than alphabetic organisation
1. Introduction: an overview of the history of lexicography

because it reflected the predominant ideological and philosophical theo-
rries. Indeed,

‘Alphabetisation may also have been offensive to the global scholastic
view of things. It must have seemed a perverse, disjointed and
ultimately meaningless way of ordering material to men who were
interested in neat frames for the containing of all knowledge’.59

The process of defining taxonomies of knowledge began with Socrates
who advocated learned inquiry over unquestioning acceptance, and
Plato continued the process with his theory of forms. In turn, Aristotle
steered the argument away from Plato’s idealism towards realism. He
wrote thematic treatises60 on the subjects of logic, ethics, biology and
metaphysics, among others, thus establishing and delineating such groups
of thoughts as discrete subjects.61

Alexander the Great, was educated by Aristotle and according to
Plutarch:

regarded the Iliad as a handbook of the art of war and took with
him on his campaigns a text annotated by Aristotle, which became
known as ‘the casket copy’ and which he always kept under his
pillow together with his dagger.62

59 McArthur, Worlds of reference, 76-77. It is interesting to note that, in present day
computing, the alphabetic ordering of every word in a document is one method
of encrypting its meaning.

60 Nissen’s assertion in Rheinisches Museum, 47 (1892), 187-88 that within each
section of the Constitutions followed the names of the states followed alphabetic
sequence is convincingly disproved by Lloyd D. Daly, ‘Contributions to a history of
alphabetisation in Antiquity and the middle ages’, Collection Latomus 90 (Brussels
1967) 17.


62 Plutarch, The age of Alexander, tr. & ed Ian Scott-Kilvern (Harmondsworth 1973)
260.
1. Introduction: an overview of the history of lexicography

As he conquered lands across West Asia and as far as Egypt, he disseminated the ideas of the philosophers, and set up the great Library at Alexandria. The system of Greek schooling, was also transmitted by Alexander, including the *enkýklios paideía*, the system of education based on literature and mathematics that boys were introduced to at the age of fourteen.\(^63\)

Green has discussed the roles played by later scholars such as Flavius Aurelius Cassiodorus (c. 490–585), St Isidore of Seville,\(^64\) and above all, St Thomas Aquinas (c. 1225—74) in sustaining and developing the thematic world view.\(^65\)

1.4.1.1 *Pliny the Elder’s Historia naturalis*

Pliny the Elder’s *Historia naturalis* is an early example of a work that is broadly\(^66\) encyclopaedic or thematic in structure: Book IX covers fish, Book XIII exotic trees and unguents, and Book XVIII covers grain. Furthermore, the *Historia naturalis* clearly cites the sources of some information, and attempts to present the information in an objective manner. Indeed the medieval Latin word that passed into English as encyclopaedia was the result of a scribal error when the phrase *enkýkl ios*

---


\(^{64}\) It should, nonetheless, be noted that while the broad structure of Isidore’s *Etymologiae* is thematic, Book 10 shows first letter alphabetisation.

\(^{65}\) Green, *Chasing the sun*, 56-57.

\(^{66}\) Daly, op. cit. 35-6 notes that portions of the *Historia naturalis* are alphabetised, such as the 110 species of plants that are listed at the end of Book 26. However, he states that Pliny ‘generally resorts to alphabetic arrangement only when he has exhausted all other principles of organisation, and even then is, at times, apologetic for what he apparently felt to be a rather unimaginative procedure‘.
paideía used in the Historia naturalis was copied as the compound noun enkýkliopaideía.67

1.5 Evidence of Alphabetisation

The earliest physical evidence of an ordered system of the letters of the alphabet is found on an Ugaritic tablet from the 15th century BC. Inscribed upon it are thirty letters of the Ugaritic script including the twenty-two North Semitic letters in the same sequence as they appear in the modern Hebrew alphabet. This order of letters was transmitted by Hebrew tradition and by the early Greeks and Etruscans.68

References to alphabetisation are to be found in literature in antiquity, for example, in the humorous epigram of Ammianus who tells his brother, the physician Lucius:

If you have decided to bury only those members of the council whose names begin with alpha, then there is still your brother, but if, as one may well suppose, you are proceeding letter by letter, I want to tell you that my name is Origen.69

1.5.1 Alphabetic origins

In considering the conditions that gave rise to the alphabetic classification of dictionaries, most meta-lexicographers trace the origins of alphabetisation to a physical environment where either the letters of the alphabet

69 ibid. 10
or indeed, the separate lexical entries of the reference works in question, existed as tangible objects.

Absolute alphabetisation is an exact process and one that becomes more complex as reference works become larger. Daly suggests that in Antiquity and the middle ages such a system was possible to implement only if the lexicographical material were available as ‘slips’, though he is not able to present contemporary evidence from the periods in question to support his assertion.\(^70\)

In an application of Daly’s theory to Old-English glossaries, Healey searched the Dictionary of Old English Corpus in Electronic form for linguistic evidence to support such a practice in Old English. Her findings did not lend support to Daly’s theory. Her research concluded that the word for a ‘slip’ was not current during the Old-English period, but entered the language at a later stage: ‘The noun slip in the sense of [a] piece of paper or parchment, esp[ecially] one which is narrow in proportion to its length is not attested until 1687’.\(^71\)

MacArthur has presented another theory in regard to alphabetisation. He argues that the printing press is the major factor in its development and subsequent dominance. With the advent of movable type, the letters of the alphabet existed for the first time as tangible, individual, hard-metal objects. This prompted people involved in the printing press to think about the alphabet and alphabetisation in a new and more practical

---

\(^70\) Lloyd D. Daly, Contributions to a history of alphabetisation, 86-89.

1. Introduction: an overview of the history of lexicography

way. As they began to touch and re-order the letters, the advantages of the alphabetical system was impressed upon them, and gradually an awareness of this system spread from people involved in making fonts to people who thought and theorised about letters and words:

Where scholars and copyists had previously been unaccustomed to thinking of words and even parts of words alphabetically, printers were now spending a great part of their time doing nothing else. Sheer familiarity with hard physical objects in a very practical craft appears, therefore, to have promoted interest in alphabetical order in other, related but more abstract fields.\textsuperscript{72}

The system of alphabetic organisation had been refined and used by the creators of the Alexandrian glossaries: Galen’s \textit{Interpretation of the Hippocratic glosses} is considered to be the earliest example of a work that displays absolute alphabetisation. Daly has noted, however, that in his preface Galen indicates that the work will be alphabetised by the first letter or first syllable, thus raising the possibility that the absolute alphabetisation was performed by a later redactor.\textsuperscript{73}

The \textit{Oxyrhynchus Papyrus} no. 1802, dated to about AD 200, is a polyglot glossary that contains about twenty glosses all beginning with the letter \textit{mu} and arranged in absolute alphabetic order.\textsuperscript{74}

Alphabetisation was never used in Latin lexicographical works compiled in antiquity and thus the knowledge was lost to Western Europe.\textsuperscript{75}

In fact, the compilers of dictionaries and lexica were forced to re-invent

\textsuperscript{72} McArthur, \textit{Worlds of reference}, 41-43.
\textsuperscript{73} Lloyd D. Daly, \textit{Contributions to a history of alphabetisation}, 35.
\textsuperscript{74} ibid. 30.
\textsuperscript{75} Lloyd W. Daly and Bernardine A. Daly, ‘Some techniques in medieval Latin lexicography’, \textit{Speculum} 39 no. 2 (1964) 237.
1. Introduction: an overview of the history of lexicography

the technique, as will be shown below by reference to Old-English glossaries and medieval Latin dictionaries.

1.5.2 Alphabetisation in Old-English Glossaries

The Épinal glossary\textsuperscript{76} has been dated to the early eighth-century\textsuperscript{77} and comprises 3200 entries of rare and difficult words, arranged in six columns. It is based on sources such as the Hermeneumata Glossary, the Hisperica Famina and Gregory’s Dialogues.\textsuperscript{78} Two types of alphabetical ordering are employed in the work: one portion of the Glossary is arranged in A-order\textsuperscript{79} while the second section has an AB-order.\textsuperscript{80} Healey has stated ‘this glossary clearly belongs to an early stage in the development of fully-alphabetised dictionaries, for here glosses gathered from various sources are only partially assimilated into the new construct. The presence of two alphabetic systems clearly demonstrates its process of accretion’.\textsuperscript{81}

The Old-English Leiden glossary has been dated to the eighth century and is predominately written in the Mercian dialect of Old English. The Glossary consists of a collection of glosses culled from different texts.\textsuperscript{82}

\textsuperscript{76} See Healey, op. cit. 4, for a discussion of the relationship between the Épinal, Erfurt, and Corpus Glossaries.

\textsuperscript{77} Collison, Foreign-language dictionaries, 41.

\textsuperscript{78} Healey, Old English glossaries, 4.

\textsuperscript{79} A-order indicates that alphabetisation has been applied to a text only in terms of grouping together all the words that begin with the letter A, AB-order indicates the same process has been performed on the words beginning with AB, while ABC-order indicates that the process has been carried out on the first three letters i.e. ABC of each word.

\textsuperscript{80} Green, Chasing the sun, 56.

\textsuperscript{81} Healey, Old English glossaries, 4.
1. Introduction: an overview of the history of lexicography

The original context of the glosses has been preserved; most of the pages in the glossary are headed by an excerpt of the text whence the section of glosses was excerpted. The glossary reflects only A-order alphabetisation and most words in the glossary are recorded in the inflected forms in which they occur in the original text.\textsuperscript{83}

The first ABC-order has been identified in a manuscript in the British library: MS Harley 3376, which dates from approximately AD 1000. Green states that it contains '5563 entries, which end at the letter F, in the bulk of which both head-words and explanations are in Latin, although there are some 1500 definitions in Old English. Occasionally the scribe goes into ABCD-order.'\textsuperscript{84}

In terms of the Old-English glossaries, alphabetisation was adopted only in a half-hearted manner, and as late as the sixteenth century John Withal’s \textit{Shorte dictionarie of English and Latin for yonge beginners} eschewed an alphabetic arrangement for a thematic one.\textsuperscript{85}

1.5.3 Alphabetisation in Medieval-Latin Dictionaries

The \textit{Elementarium doctrinae erudimentum} of Papias the Lombard dates from approximately AD 1053 and was widely used by Latin lexicographers until the fifteenth century.\textsuperscript{86} The text presents two approaches to

\textsuperscript{82} Collison, \textit{Foreign-language dictionaries}, 42.
\textsuperscript{83} Green, \textit{Chasing the sun}, 55-56.
\textsuperscript{84} ibid. 56.
\textsuperscript{85} ibid.
\textsuperscript{86} Shaw, op. cit. §1.3.1, notes that ‘Papias was an important source for virtually all Latin lexicography until the fifteenth century. According to Bursill-Hall (1981) there are 110 existing manuscripts of the \textit{Elementarium}. It was first printed in
1. Introduction: an overview of the history of lexicography

defining words: the first a concise approach supplying short definitions or synonyms to explain words; the second involves more extensive definitions of an encyclopaedic nature.\(^87\)

In the preface to his work, Papias offers a rather perplexing account of the system of ABC-alphabetisation. He did not succeed in systematically incorporating this level of alphabetisation in the *Elementarium doctrinae erudimentum*, and indeed his account provides an insight into the undeveloped state of the system of alphabetisation at this time.

Anyone who wishes to find anything quickly must also notice that this whole book is composed according to the alphabet, not only in the first letters of the parts but also in the second, third and sometimes even in the further determinative arrangement of the letters. The first indication of division then, will be made by a.b.c. and the other letters in order. This will be subdivided in the second order of differentiation by the same letters a.b.c. etc. in larger form before any change of the letters. In the third order of subdivision, however, all that is included under one variety [combination?] of three letters will be distinguished by a third paragraph so that whatever is sought may be found within exactly this space. But once the first, second and third order of these letters has been observed, the same cannot be maintained in the following letters to the extent that one does not put them in the wrong order. Even in these first, second and third steps the relationship will sometimes vary because of the writing of different letters. For example, hyena is written by some with an i, by others with a y or an aspiration. And the plant that some call *verbena* others call *berbena*. Informed, therefore, by these and many other examples, it will not be disadvantageous to be forwarded about similar matters.\(^88\)

---

\(^87\) Ibid.

\(^88\) Daly & Daly, *Some techniques* 233.
1. Introduction: an overview of the history of lexicography

Hugo of Pisa completed the *Derivationes* some time in the late twelfth century, and despite being a successor to Papias, he did not advance Papias’s system any further. Daly and Daly observed that Hugo applied initial letter alphabetisation only to his work, and that this was not done consistently throughout. They further noticed that many lexical entries contain words that had been displaced from other entries.\(^{89}\) However, Weijers has rejected this view and argued that Hugo followed another alphabetic system, used by lexicographers such as Johannes de Mera, where:

The words starting with a consonant are first arranged according to the following vowels, then according to the consonants in second position, for instance for the letter *f*, words with *fa*, *fe*, *fi*, *fo*, *fu* are treated before those with *fl* and *fr* . . . (*fiton* is followed by *fonos* etc.; *flaneo* comes only after *fusco*).\(^{90}\)

The two medieval Latin works that achieved absolute alphabetisation were completed quite close together: the first was the *Summa* of Guillelmus Brito published in 1272, closely followed by the *Catholicon* of Giovanni di Genoa (also referred to as Balbus and Johannes Januensis de Balbis), completed on the 7th March 1286. Daly and Daly state that the *Catholicon* was probably the first Latin dictionary to be printed with movable type and was printed on vellum at Mainz in 1460, probably by Gutenberg.\(^{91}\)

\(^{89}\) See ibid. 235 for a discussion of the *repertoria* that were devised to cater for this problem in medieval dictionaries.


\(^{91}\) Daly & Daly, *Some techniques* 237.
1.6 Introduction

Paul Russell has stated ‘Glossaries probably constitute the largest area of early Irish literature to have been left untouched by the modern commentator and editor’.\(^\text{92}\) Much work remains to be done on the glossaries and new editions are badly needed. This section provides a general introduction to the major glossaries of medieval Irish by treating of Sanas Cormaic, O’Mulconry’s Glossary, the Lecan Glossary, O’Clery’s Glossary, O’Davoren’s Glossary, and Dúil Laithne in some detail. The ordering systems used in the glossaries are examined in terms of the ordering systems used in medieval Latin lexicons and Old-English glossaries, set out above. One problem inherent in the alphabetisation of Irish is that of initial mutations (lenition, nasalisation, gemination). To their credit, Irish glossators ignored initial mutations from the beginning.

1.7 Old-Irish Glossaries

Medieval Irish language glossaries treat of a wide variety of subjects and are extant in both text dependent and text independent forms. In addition to the major, text dependent, glossaries set out below, the corpus includes, inter alia, text independent, metrical, and medical glossaries.

1.7.1 **Metrical glossaries**

Russell has described metrical glossaries as being ‘lists of difficult words with very brief, usually non-etymological, explanations set in a metrical form, usually *debide*’ and cites *Forus Focal* in the *Book of Leinster* and *Egerton* in London, BL, MS Egerton 90, f. 17a, as examples of metrical glossaries.\(^93\)

1.7.2 **Text-dependent glossaries**

Examples of text dependent glossaries include the glossaries to the *Amra Choluim Chille* and the *Féilire Oenguso* in Dublin, TCL, 1337 (olim H. 3. 18) 565-610.\(^94\)

1.7.3 **Medical glossaries**

In 1900 Stokes published\(^95\) two medical glossaries contained in Dublin, TCL, 1334 (olim H.3.15), a sixteenth-century vellum, and also a further glossary found in Lord Crawford’s Irish medical manuscript.

The medical glossaries comprise vocabulary describing plant names, plant varieties, and their medicinal properties. The type of glosses contained in the medical glossaries range from one-word glosses, to more detailed explanations. For example:

---


\(^95\) Whitley Stokes, ‘Three Irish Medical glossaries’, *Archiv für celtische Lexicographie* 1 (1900) 325-347.
1. Introduction: an overview of the history of lexicography

§14. Antos i.e. in ros marina. Antos i.e. the Rosemary

§96. Gallitricum i.e. oculus Christi. Gallitricum i.e. the eye of Christ

§90. Mirbolanı. torad 7 atait .5. gnei air .i. mirbolani sitri[n]i 7 folmaigí limn ruad, mirbolani inndi[cí] folmaiges limn dubh 7 na .3. gnei [c]e folmaighí limn finn .i. cebuli 7 imbilisi [leg. emblici] 7 bellirisi [leg. bellerici]. Myrobolanus, a fruit, and there are five kinds of it, m. citrinus and it purges phlegm, m. indicus, which purges black bile, and the three other kinds which purge lymph . . . .

Neither the first glossary included in Dublin, TCL, 1334 (olim H.3.15), nor the glossary in Lord Crawford’s MS show alphabetisation. However, the second glossary of Dublin, TCL, 1334 (olim H.3.15) shows first letter alphabetisation, save a few lapses such as a lone Braccae ederae among words beginning with A and also Arilli, a hiccup in the transition from A- to B- words.

96 ibid. 325. Translation mine unless otherwise stated.
97 ibid. 328.
98 Oculus Christi was used as a synonym for Salvia Verbenaca (Vervian sage) and also for Calendula officinalis (Marigold). See Mrs M Grieve, A modern Herbal: the medicinal, culinary, cosmetic and economic properties, cultivation and folk-Lore of herbs, grasses, fungi, shrubs & trees with their modern scientific uses (New York 1931).
100 ibid. 325-330.
101 ibid. 333-337.
102 ibid. 330-332.
103 ibid. no 3, 330.
104 ibid. no. 20, 331.
1. Introduction: an overview of the history of lexicography

1.8 Scholarship on the compilation of glossaries

The process of compiling a glossary has been discussed by Meyer. He identifies three stages. The first involves the gleaning from a text sentences and phrases with marginal or interlinear glosses, and the grouping of them, in another text, in original document order. The second stage involves recasting the excerpts, with the lemma at the head, followed by a phrase such as ut est, ut dixit, amal atá, amal adeir, ocus deismirecht air, that serves to introduce the explanatory material. The third stage is the alphabetisation of the lemmata.

Russell has expanded Meyer’s second stage, and discussed the methods of explanation employed by the glossators:

1. Explanation by means of a gloss, much as in a modern dictionary.
2. Etymology by analysing into elements, the collocation of which can explain the meaning of the head-word.
3. Etymology by comparison with Latin, Greek, and Hebrew words.

Examples of such methods of explanation will be given in the discussion of glossaries below.

Regarding the ordering systems used in the glossaries, Meyer states that ‘it is but one step’ from the third stage in glossary making (when

106 ibid. 139.
107 ibid. 140.
108 ibid. 26-27.
the lemmata are arranged under a letter of the alphabet) to the fully alphabetised glossary of O’Clery. The validity of this statement will be assessed by examining the ordering systems of the major medieval Irish language glossaries. Further, neither Meyer nor Russell discuss thematic ordering systems in Irish glossaries and therefore the glossaries discussed below will also be examined for evidence of such a system.

1.9 List of manuscripts containing major medieval Irish-language glossaries

○ Sanas Cormaic\(^{110}\)


2. Dublin, TCL, 1318 (olim H. 2. 16) al. *Yellow Book of Lecan* col. 3-87

3. Dublin, RIA, 1225 (olim D. ii. 1) al. *Book of Uí Maine* f. 177-84, col. a

4. Oxford, Bodleian L, Laud Misc. 610, f. 70'-84'

5. Dublin, TCL, 1339 (olim H. 2. 18) al. *Book of Leinster* (miscellany) 179

6. Dublin, TCL, 1317 (olim H. 2. 15B), 13-37; 77-102

\(^{109}\) Meyer, op. cit. 140.

\(^{110}\) For a bibliography of edited versions of the manuscripts listed here see Russell, *Sounds of a silence*, 2-8.
1. Introduction: an overview of the history of lexicography

7. Dublin, TCL, 1337 (olim H. 3. 18), 76-83

○ O’Mulconry’s Glossary

1. Dublin, TCL, 1318 (olim H. 2. 16) al. Yellow Book of Lecan col. 88-122

○ Lecan Glossary

1. Dublin, TCL, 1319 (olim H. 2. 17), 331-35
2. Dublin, RIA, 1225 (olim D. ii. 1) al. Book of Úi Maine, 184
3. Dublin, TCL, 1337 (olim H. 3. 18), 415-16; 663a-667
4. Dublin, RIA, MS Stowe, C. i. 2, f. 8b-10b2, 29b\(^{111}\)
5. Dublin, TCL, 1363 (olim H. 4. 22), 36

○ O’Davoren’s Glossary

1. London, BL, Egerton 88, f. 79a-92c

○ H3

1. Dublin, TCL, 1337 (olim H. 3. 18), 623-28

○ Dúil Laithne

1. Dublin, TCL, 1317 (olim MS H. 2. 15B), 116

○ Dúil Dromma Cetta

\(^{111}\) This MS is incorrectly cited as C.i.z. by Whitley Stokes, ‘The Lecan Glossary’, Archiv für celtische Lexikographie 1 (1900), 50, and incorrectly cited as ‘London, BL, MS Stowe, C.i.z’ by Paul Russell, Sounds of a silence, 7.
1. Introduction: an overview of the history of lexicography

1. Dublin, TCL, 1337 (olim H. 3. 18), 63-75; 633a-638b
2. London, BL, Egerton 1782, f. 15
3. Dublin, TCL, MS H. 1.13, 361-62

○ O’Clery’s Glossary


○ The Stowe Glossaries

1. Dublin, RIA, 1234 (olim C. i. 2.) f. 8b-f. 10b; f. 29 b 1-f. 31 b 3

1.10 O’Davoren’s Glossary

O’Davoren’s Glossary was compiled by Domhnall Ó Duibh Dé Bhoirenn (O’Davoren), c. 1569. Indeed, the opening lines of the glossary allow us to pinpoint the date that O’Davoren began his work on as 4 February 1569:

Mei est incipere, Dei est infinire .i. Domnall Ó Dhubh dha boirenn in aine iar feil Bride 1569, ocus ar in Pairc atu.112 ‘Mine to begin, God’s to end i.e. Domhnall Ó Duibh Dá Bhoirenn on the Friday after the feast of St. Brigid, 1569, and I am in Park [in Galway]’

The glossary is a collection of difficult words, mostly legal terms, and one of its most valuable features, apart from citations from lost legal

---

texts, is the number of verbal forms it contains that do not occur as head-words in other glossaries.\textsuperscript{113} It draws, directly or indirectly, on Isidore as a source.\textsuperscript{114} Russell has stated that if its material overlaps with Sanas Cormaic or Dúil Droma Cetta, it does so in the second part of the gloss and not in the head-word.\textsuperscript{115}

The glossary contains a limited number of one-word glosses, for example: §190. *Axal .i. imagallaim.* axal i.e. a colloquy\textsuperscript{116} as well as etymological glosses such as §289. *Brat .i. ba ar fuit .i. ar fuacht.* brat a ‘cloak’, i.e. bá ‘good’ ar ‘for’ fuit ‘cold’.\textsuperscript{117} Most entries in the glossary are lengthy, for example:

§283. *Baislec .i. log basaighthi, ut est baislec bo-air .i. locc i mbasaighther na ba tria ar ndochur in buaile indigh.* ‘bàis-lec (lit. death-stone) a place of dying, ut est ‘the baislec of the murrain’, i.e. the place in which the cows die through the severe mortality of the ... cow stall.’\textsuperscript{118}

O’Davoren’s glossary does not have a sophisticated ordering system. In common with the medical glossaries discussed above, it shows first letter alphabetisation only, as illustrated by the first ten lemmata of A:

§1. *Aipgiter*
§2. *Airchendech*
§3. *Aithrighe*
§4. *Achtail*
§5. *Annone*

\textsuperscript{115} Russell, op. cit. 6.
\textsuperscript{116} Stokes, *O’Davoren’s glossary*, 228.
\textsuperscript{117} ibid. 243
\textsuperscript{118} ibid. 242.
1. Introduction: an overview of the history of lexicography

§6. Alt
§7. Ardracht
§8. Ard
§9. Ail
§10. Atnora

1.11 Lecan Glossary

The *Lecan Glossary* is a list of *dubfhocla* or ‘obscure words’, and most of the glossary comprises one-word glosses. The glosses from §203 to §222,\(^{119}\) as well as a small number of other glosses, break with this pattern and are more complex. For example, the curious etymological gloss:

\[
\text{§207. Gaedel .i. gæth-dul .i. fear no theiged co gaeth ar cach mbescna iar fir.}^{120}
\]

Zimmer and Stokes have identified some of the sources of the glossary: §59-94 are excerpts from a copy of the *Liber Hymnorum* or of the Irish poems in that codex, and §477-565 are excerpts from the metrical vocabulary preserved in London, BL, Egerton 90.\(^{121}\) Regarding the rest of the glossary, Stokes states ‘nothing certain can be said, save that many of them agree with the *Forus Focul* and *Derbshiur*, metrical glossaries, edited with a commentary in *Bezz. Beitr* 19, 8-31, 38-120’.\(^{122}\)


\(^{120}\) ibid. 54.

\(^{121}\) ibid. 50.

\(^{122}\) ibid. 50.
1. Introduction: an overview of the history of lexicography

The *Lecan Glossary* is an unalphabetised list of words, but it displays some evidence of thematic organisation. For example, the antonyms *clean* and *unclean* are grouped together:

§13. *Andag .i. glan*
§14. *Anandac[h] .i. nemglan*\(^{123}\)

As are the synonyms for words such as *slaughter*:

§50. *Orguin .i. guin*
§51. *Oirgther .i. guin*
§52. *Hirt .i. orguin*
§53. *Bith .i. guin*\(^{124}\)

and *harlot*:

§106. *Bleasc .i. merdrech*
§107. *Dormuin .i. merdrech*
§108. *Ben inroma .i. merdrech*
§109. *Be charna .i. merdrech*\(^{125}\)

While some evidence for thematic organisation can be found in the glossary, it is not systematically employed throughout. For example, we find synonyms of *dliged* at §19, §42, §56, §177 and §187. It is possible, therefore, that such grouping was not intentional on the part of the scribe, but merely a result of the pattern of glosses in one or more of the base texts of the glossary. Russell has drawn attention to a later stage in the development of the ordering system used in the *Lecan Glossary*. A later version, contained in Dublin, TCL, 1337 (olim H. 3. 18), displays first-letter alphabetisation.\(^{126}\)

\(^{123}\) ibid. 51.
\(^{124}\) ibid.
\(^{125}\) ibid. 52.
1. Introduction: an overview of the history of lexicography

1.12 Dúil Laithne

The colophon to *Dúil Laithne* furnishes information about the name of its scribe as well as the date and location in which he completed it:

Finit DUBaltach FIRbisigh [sic]
adomcomhnvic.
1643. 5° maij
baile mic aodha gan
molocc127 'Finit it is I D. [Mac] F., 1643, 5th May. I am at Baile Mic Aodhagáin’

The glossary was edited by Stokes in 1872,128 and published in facsimile form by Meyer in 1909.129 It contains material that can be used to expand Meyer’s second stage. *Dúil Laithne* contains two types of entry: glossary words that are contained in other glossaries, and also ‘simple words expanded and disguised by the insertion of syllables’.130 Commentators have tied the glossary with *bérla na filed*131 and argued that Virgilius Maro Grammaticus is a possible source for it.132

---

128 ibid, 75-83.
1. Introduction: an overview of the history of lexicography

The vocabulary of the glossary falls into fifteen main categories: words describing the body; words describing garments and accoutrements; terms for people; household implements; food and drink; instruments; animals; metals; synonyms and antonyms; words about the earth (such as countries and the physical landscape); colours; pronouns; numbers; directions; times of the day as well as a small number of miscellaneous terms. It consists mostly of one-word glosses and is unalphabetised. Nonetheless, it can be argued that the glossary employs an ordering system that is unique among other medieval Irish language glossaries. For the most part, categories of terms are grouped together, for example, words for food:

§81 Anrad .i. biad ‘cibus’ [food]
§82 Doib .i. deogh ‘potus’ [drink]
§83 Muincir .i. m´ir ‘frustum’, ‘bucca’ [portion]
§84 Betroisgenn .i. barigen ‘placenta’ [loaf; cake]
§85 Seeglan .i. salann ‘sal’ [salt]
§86 Anros .i. arbar ‘fruges’ [arbar can be translated as ‘grain’; Stoke’s use of fruges indicates ‘crops’]
§87 Bloa .i. vbla ‘poma’ [apple]
§88 Gortan .i. cainenn ‘porrus’ [leek]
§89 Roinn .i. coirm ‘cerevisia?’ [beer]
§90 Ailmis .i. as ‘lac’ [milk]
§91 Mincill .i. mil ‘mel’ [honey]
§92 Iodamm .i. im ‘butyrum’ [butter]
§93 Collruim .i. feoil ‘caro’ [meat]
§94 Gech .i. saillte ‘sale conditus,’ ‘lardum’? [lard?]
§95 Sailalm .i. saill ‘pinguedo’ [salted meat]
§96 Gorgruth .i. gruth ‘lac pressum’ [curd]
§97 Gortrus .i. grus . . . ?’ [curd]
§98 Mvincedhg .i. medhg ‘serum’ [whey]
§99 Mvinchidh .i. miadh ‘hydromeli’ [mead]
§100 Brasach .i. blathach ‘butyri serum’ [buttermilk]
§101 Lemocen .i. lemhnacht ‘lac dulce’ [sweet milk]
§102 Mvadhgalan .i. muillenn ‘mola’ [flour]
1. Introduction: an overview of the history of lexicography

§103 Amloicit .i. aith ‘fornax’ [drinking horn]

A further example of the thematic ordering system in *Dúil Laithne* is the grouping together of terms for animals:

§110 Eabadcoll .i. ech ‘grex equorum’ [a stud of mares]
§111 Ebandan .i. ech ‘equus’ [steed]
§112 Eabathan .i. lair ‘equa’ [mare]
§113 Bethan .i. bo ‘vacca’ [cow]
§114 Davrailm .i. damh ‘bos’ [bull]
§115 Durailbind .i. dartaid ‘bucula’ [heifer]
§116 Bviglen .i. laogh ‘vitulus’ [bull calf]
§117 Cetaimni no rosca .i. cairig {ms. cairidh} ‘oves’ [sheep]
§118 Glædmuine no Gairmnech .i. gabar ‘caper’ [goat]
§119 Mvinscvill .i. muc ‘sus’ [pig]
§120 Collar .i. cv ‘canis’ [dog]
§121 Caipist .i. cat ‘felis’ [cat]
§122 Luipist .i. luch ‘mus’ [mouse]
§123 Luathan .i. en ‘avis’ [bird]
§124 Sceman .i. sionnach ‘vulpes’ [fox]
§125 Lornan .i. patu ‘lepus’ [hare]
§126 Oraill .i. elit ‘cerva’ [doe]
§127 Roscon .i. rón ‘phoca’ [seal]
§128 Bedban .i. bradan ‘salmo’ [salmon]
§129 Snuad .i. cechlus ‘quævis herba’ [some herb]
§130 Bibe .i. bech ‘apis’ [bee]

Synonyms and antonyms are also grouped together, for example:

§279. Betchennacht .i. bennacht ‘benedictio’ [blessed]
§280. Metchennacht .i. mallacht ‘maledictio’ [cursed]
§189. Tinnichis .i. tis ‘infra’ [below]
§190. Tinnechuas .i. tuas ‘supra’ [above]


133 ibid. 77.

134 ibid. 79.

135 ibid. 78.

136 ibid. 78.
1. Introduction: an overview of the history of lexicography

Though most of the *Duíl Laithne* glossary displays a thematic ordering system, parts of the glossary deviate from this pattern. For example, terms for people are mostly found at §52-§65.\(^{137}\) However, within this group, the entry: §53. *Beth losach .i. bathach ‘moribundus* [dying],\(^{138}\) can be viewed as misplaced as it is a state of a person rather than a term for one. Further, we find terms for people at other junctures in the glossary, for example: §206. *Braininta [ms. Brainionta] .i. banchainte ‘femina quae satiras scribit* [A woman who writes satires].\(^{139}\)

The caveat regarding the pattern of glosses in a base text, discussed above in relation to the *Lecan* glossary, also needs to be borne in mind in relation to this glossary. Yet in contrast with the *Lecan* glossary, where some entries appear to be grouped thematically, most entries in *Duíl Laithne* are grouped thematically. Indeed, the ordering system that is observed here is unique in among Irish glossaries.

In regard to a comment of MacManus,\(^{140}\) Russell stated that ‘*Duíl Laithne* may well preserve the genitive singular of the earlier form of the Irish loanword from Latin *Latīna*, the common form being *Laiten* … if so the title is to be translated as as ‘collection of (from?) Latin’; however its precise relationship to anything Latin remains unclear’.\(^{141}\)

\(^{137}\) ibid. 76.

\(^{138}\) ibid.

\(^{139}\) ibid. 78.


1. Introduction: an overview of the history of lexicography

The discussion above (see § 1.5.1) established that alphabetisation was never used in Latin lexicographical works, but thematic organisation was widely employed (see § 1.4). Perhaps the ordering system that has been discussed in *Duil Laithne* can be viewed as one aspect of the glossaries ‘relationship to anything Latin’ and a starting point for further investigation.

1.13 Sanas Cormaic

Versions of *Sanas Cormaic* are found in seven manuscripts, as illustrated in § 1.9. Within this group two different versions of the text exist: a long one and a short one.

The long version is contained in *The Yellow Book of Lecan* and *Dublin, TCL, 1317 (olim H. 2. 15B)*, and the shorter version occurs in *The Leabar Breac*, *MS Laud 610* and *The Book of Leinster* and *The Book of Uí Maine*. The material contained in Dublin, TCL, 1317 (olim H. 2. 15B) is an ‘independent version, for the letters L-U at least, of the extra

---

142 ibid. 3. Russell describes the long version as having an extra block of entries at the end of each letter.

143 Whitley Stokes (ed), *Sanas Cormaic: Cormac’s glossary* (Calcutta 1868); Bergin, Best, Meyer, and O’Kefte (ed), *Anecdota from Irish Manuscripts* vol. v (Dublin 1913).


146 Whitley Stokes, ‘On the Bodleian fragment of Cormac’s glossary’, *Phil. Soc. Trans.* (1901) 149-206’.

147 R. I. Best & M. A. O’Brien (ed), *The Book of Leinster* vol. iv (Dublin, 1965) 780-81; Russell, op. cit. 2, notes part of glossary has been omitted from this edition.

material which goes to make up the longer versions', and the second part of the manuscript is unrelated to *Sanas Cormaic*.

Meyer and Stokes attributed the glossaries to Cormac mac Cuilennáin, king and bishop of Cashel until his death in battle in 908. As further evidence for the Munster origin of the glossary, Russell has argued that the term *sanas* is frequently used in connection with Cormac, King of Cashel. He suggests that the name of the glossary may be ‘a play on words; for it is being used of a collection of words and their explanations which may offer advice, possibly secret or arcane, to a king of Cashel’.

The relationship between the various manuscripts was first tackled by Thurneysen, who analysed the articles Y §883 *mughéme* and Y §1059 *prull* which occur in all of the Cormac manuscripts and also occur independently in two other manuscripts. Russell notes that scholars agree the oldest version is that preserved in the *Book of Leinster* and *MS Laud 610*. Thurneysen’s analysis has been modified by scholars such as the O’Sullivans, Ó Ríain, and most recently by Russell.

---

149 Russell, 3.
who holds that the stemma may not be a completely reliable method of establishing the relationships that exist between the different versions of *Sanas Cormaic* based on examples of manuscripts where ‘contamination between different branches’ has occurred.\(^{158}\)

The relationship between the versions of *Dúil Dromma Cetta* has also been discussed by Russell,\(^ {159}\) who argues that the glossary may have been compiled in Northern Ireland.\(^ {160}\) Furthermore, there is a concurrence of material between *Dúil Dromma Cetta* and *Sanas Cormaic* in about 11% of Cormac entries and it is likely that the two glossaries share a common archetype.\(^ {161}\)

Both *Sanas Cormaic* and *Dúil Dromma Cetta* are alphabetised, but alphabetisation does not extend beyond the first letter. The type of glosses in *Sanas Cormaic* ranges from one-word glosses to more complex glosses and the different versions of the glossary include both etymological and explanatory glosses. For example:

\[
\text{§435. } \text{Dobhrith i. dobar-ith insin i. usque 7 arbar i. cuid aossa aithrihe et pendaite.}^{162}\]

\[
\text{§673. Gaimrith quasi gaim-rith, unde dixit, Colmán mac húi Cláusaigh ni marbhnaidh Cuimíne Fota:} \\
\text{Húa Corpre, húa Cuirc,}
\]


\(^{158}\) Russell, *Dúil Dromma Cetta*, 156-7.

\(^{159}\) ibid. 147-155.

\(^{160}\) ibid. 161-62.

\(^{161}\) ibid. 160.

\(^{162}\) Bergin et al. (ed), *Anecdota*, 36.
1.14 O’Mulconry’s Glossary

A degree of pietas to his old friend O’Curry seems to have influenced Stokes’s decision to name the glossary discussed in this section after O’Mulconry. Stokes states that the glossary was pointed out to him c. 1860 by O’Curry who proposed to call it after John (son of Torna) O’Mulconry.164 Whether O’Curry held John O’Mulconry to have been the author of our glossary I cannot now remember. But if he did so, he was certainly wrong, for John flourished in 1566, and the grammatical forms in the glossary point to the thirteenth or (at latest) the fourteenth century at the time of its compilation.165

Stokes dating of the glossary to the thirteenth or fourteenth century was revised by MacNeill. He notes that while the manuscript shows signs of scribal modernization in spellings, enough Old-Irish spellings have been preserved to date parts of it not later that the middle of the eighth century and ‘probably as early as the middle of the seventh’.166

163 ibid. 54.

164 It is interesting to note that John O’Mulconry is referred to in the preface to O’Clery’s glossary ‘As aithníd dhuit sáotha maithe san cceirdí asa fós san aísmir dheighneaigh mar át Sean[sic] ua Máolchonáire prionhoide sgoile na druinge a dubhramar cheana agus [derscnaigthech]hearn eiríann a seanchas, ina aísmir fein, …’ Arthur W. K. Miller, ‘O’Clery’s glossary’, Revue Celtique 4 (1879) 1-69; 5 (1880) 349-428: 354.

1. Introduction: an overview of the history of lexicography

Most entries in the glossary are etymological, but it contains two distinct strata: the first extends as far as the list of words under L and the second from O to U. The second part possibly also contains part of L. The treatment of material is different in each stratum. In the first, most etymologies are supplied from Latin, Greek, Hebrew, and to a lesser extent from Irish. The words contained in it ‘belong to the vocabulary in general use, not a few of them being still current, for example, under A, adnacul, adrad, aess, aiged, álainn, amarc, and, ander, aslinge, anim, atenn’. The second stratum differs from the first in that it is a collection of obsolete words accompanied by later forms and explanatory glosses rather than etymologies.

Examples of glosses from the first stratum include:

§2. [Abb. i.] Abba ebraice, πατήρ, graece, genitor latine.
§129. Bás ebraice, tribulatio vel angustiae vel mors latine.
§711. Ilach ilactis graece latratus latine, ar it cosmaile cuana 7 choinhuala 7 ilach.

Examples of glosses from the second stratum include:

§813. Machæ i. badb. nó así an tres morrigan, unde mesrad Machæ i. cendae done iarna n-airlech.
§825. Memm i. póg.

---

167 Russell, Sounds of a silence, 5.
168 MacNeill, Scoticae linguae, 113.
169 ibid. 113.
170 Stokes, O’Mulconry’s Glossary, 235.
171 ibid. 240.
172 ibid. 267.
173 ibid. 271.
Stokes noted in his edition that the *Origines* of Isidore was the source for a number of entries, and MacNeill added one more article, borrowed from Isidore, to his list. MacNeill also noted that material borrowed from Isidore occurs only in the first stratum of the glossary. Russell furthermore notes that large portions of the glossary are contained in *Sanas Cormaic*, *Duíl Dromma Cetta* and in *O’Davoren’s Glossary*.

The ordering system used in *O’Mulconry’s Glossary* is more sophisticated than that used in *Sanas Cormaic* and *Duíl Dromma Cetta*. MacNeill has shown that the glossary shows alphabetisation as far as the second letter:

Under the initial A, the words numbered 1 to 10 begin with *ab*, 11 to 16 *ac*, 17 to 27 *ad*, and so on. This series ends at 69 with *at*. Then with 70 begins a second series, arranged in the same manner but apparently with some exceptions to the order. This series ends with 94. The remainder of the list under the initial *A*, 95 to 114, is in promiscuous sequence.

All of the letters in stratum one, including portions of the letter L display AB-alphabetisation to varying degrees. However, no letter exhibits consistent AB-alphabetisation throughout. Stratum two does not show such alphabetisation, and MacNeill notes that the letters M, N, O, P, R, S and T do not contain alphabetical sub-classifications.

---

174 *ibid.*
175 *ibid.* 273.
176 *MacNeill*, op. cit. 119.
1.15 O’Clery’s Glossary

O’Clery’s Glossary is a late collection. In his preface to the work, signed 28th October 1643, O’Clery describes it as a:

Foclóir no sanasán nua, in a mínighthear cáil eigin dfoclaibh cruaidhe na gaoidheilge, ar na sgriobhadh ar urch aibhitre . . . . 

177

His glossary marks a departure in Irish-language lexicography not only because it is the first fully alphabetised reference work, but furthermore because O’Clery’s is the first Irish reference work to set out a methodology of sorts that was adhered to during the compilation of the glossary. Indeed, O’Clery’s Glossary can be viewed as a Irish proto-dictionary.

O’Clery begins his preface by setting out the criteria by which he selected interpretations of words contained in the foclóir.

An céid ní, nár chuireamar en fhocal ann so sios do mínighadh, no do ghluais ar fíoclaibh cruaidhe ar dteangtha mathartha acht na focail do chualamar féin da mínighadh, no fuaramar ag cáth eile iar na mínighadh ona maighistribh do ba fíorítle agus dibadh foghlamtha an eolas chríais na gaoidhilge in ar laithbh féin. As diobh sin go somnradhach Baothghalach Ruadh Mhac Aodháin, Torna Ua Máolchonaire, Móileachlainn Modardha Ua Máolchonaire, agus Lughaidh Ua Cleirigh. ‘First, we have not put down any word of interpretation or gloss on the difficult words of our mother-tongue, except such words as we have ourselves heard interpreted, or have found explained by the ablest and most learned masters of the knowledge of the difficulty of the Gaelic in our own days. Of these especially are Baothghalach Ruadh Mac Egan, Torna O’Mulconry, Malachy Modardha O’Mulconry, and Lugaidh O’Clery’. 

178

177 ‘A new vocabulary or glossary, in which are explained some difficult words of Gaelic, written in alphabetical order . . . ’ Arthur W. K. Miller, O’Clery’s Glossary 349.

178 ibid. 352-3.
1. Introduction: an overview of the history of lexicography

He continues by naming some of the sources that he used, though in the sanasán itself, he does not cite his sources.

An dara ní bhíodh a fhios agad gur ab iad na leabhair chruaidhe ar an chuirsead na sean ughdair gluais mhínighthe agus as ar ghlacamar na focalsi sios maille re mínighthe na druinge réumhráite dothí ag teagasg go deigheanach, Anghra Coluim Chille, Agallaimh an da Shuadh, Feilire na náomh, Felire I Ghornáin, Leabhar iomann, Sanasán, Bheithe Phatruic, Seinscrewtra maenruim agus seinteabhar paipéir ina bfrith móran dfocalbh cruaidhe gona mínighthe, Foras focal agas Deirbhshuir don Eagna an égsi agus urnfhor an leabharain o sin amach do réir na gluais do glacadh on mBáothghalach réumhraite. ‘Secondly, you must know that the following are the difficult books which the old authors have glossed, and from which we have taken these words together with the interpretation of the aforesaid persons who have taught lately: The Elegy on Colum Chille, the Dialogue of the two Sages, the Festiology of the Saints, the Festiology of O’Gormán, the Book of Hymns, Sanasán (little glossary), Life of Patrick, old manuscripts on vellum and an old paper book in which many difficult words and their interpretation were found, Foras Focal and the Deirbhshuir don Eagna an Égisi, and a great part of the book according to the gloss received from the aforesaid Baothghalach’.  

O’Clery also justifies the treatment of words in his Glossary and states, for example, that etymologies have not been provided because his target audience is young people and those with a moderate education.

Finally, aware that he had not standardised the spellings of head-words in

---

179 E. J. Gwynn ‘On a source of O’Clery’s glossary’, Hermathena 24 (1906-1907) 20, has further shown that O’Clery drew on a copy or copies of the Dindshenchas, though he states that ‘O’Clery does not seem to have followed any one of the MSS. of the Dindshenchas which now survive’; Gwynn also named the elegy on Niall Nóigíallach and the Ode to Brian na Máithra O’Ruairc as further sources that O’Clery drew on, idem, ‘O’Clery’s Glossary (ii)’, Hermathena 25 (1908-1909); Russell also states that ‘there is so much overlap with earlier glossaries that it is clear that they, rather than the texts themselves, were the sources of much of the material’ (Russell, Sounds of a silence, 9).

180 ibid. 354.

181 ibid. 355.
1. Introduction: an overview of the history of lexicography

the *sanasáin*, O’Clery gives an explanation of the orthography of Old Irish, which can be used by the reader to navigate the glossary successfully:

An ceathramhadh ní bhíodh a fhios ag an áos óg ag an áos ainbís lerach mian na seinleabhair do leághadh (ní nách bhfuil na aincheas ar eolchaibh ar tíre) gurab annamh bhíos coinhéd aca ar chaol re leathan, no ar leathan re ccaol do sgríobhadh, agas ar fírthearc cuírid uathadh ar na consainibh mar áta, bh, ch, dh, fh, etc. agas fós as annamh chuiríd síneadh fada ar fhoclaibh. Scriobhthar go minic cuid dona consainibh ar son a cheile, mar ata c, ar son g, agas t, ar son d. Ag so samhail na bfocal tre sa dtuigfheidh sin. Ar son an fhocailsi clog, ionann agas cloc, agad, acat, beag, beac, codladh, cotladh, arsd, art, etc. Cuírthear fós go minic ae, ar son ao, agas ái ar son aoi, agas fós oí ar son aoi. Sompla airsin mar sgríobhthar go minic aedh ionann agas aodh, agas cael as ionann agas caol. Agas bóí agas fós bái as ionann agas báí. Sgríobhthar go minic E ar son A sna seinleabhraíbh, mar ata die, as ionann agas dia, cie as ionann agas cia, etc. Sgríobhthar go minic I ar son A, mar ata so dochuaidh, as ionann agas dochuaidh. Sgríobhthar go coitcheann a, o, u ar son a cheile a ndeireadh focail mar ata sompla, somplo, somplu, ceardcha, ceardcho, ceardchu, etc. ‘Fourthly, the young and ignorant who wish to read the old books, (a thing which is not difficult for the educated of our country), must know that they rarely guard against writing slender with broad, or broad with slender, and that they very rarely put the aspirate upon the consonants, as bh, ch, dh, fh, etc, and also that they seldom put the long accent on vowels. Some of the consonants are often written one for another, as c for g, and t for d. Here are some examples of words by which this will be understood: clog the same as cloc; agad, ... etc. Also ae is often put for ao, and ái for doi and also oí for aoi. For example aed, is often written for aodh, and cael is the same as caol. And bóí and also bái is the same as báí. E is often written instead of A in the old books, as die, which is the same dia, cie which is the same as cia etc. I is often written instead of A, as dochuaidh, a, o, u are commonly written one for another at the end of a words as sompla, somplo, somplu ... ’.182

Indeed, as will be discussed below, the editors of the *Dictionary of the Irish language* would have done well to heed O’Clery’s approach to the complex issue of Irish language orthography.

182 ibid. 355-6.
1. Introduction: an overview of the history of lexicography

*O’Clery’s Glossary* contains a range of explanatory glosses, from one-word explanations to more lengthy glosses, for example:

\[
\text{ail } .i. \text{ cloch ‘stone}^{183}
\]

\[
\text{allabhar } .i. \text{ oll arbhar } .i. \text{ sluagh mór, óir as ionann arbhar } 7 \text{ sluagh } 7 \text{ as ionann oll } 7 \text{ mór. ‘A great army because arbhar and sluagh (army) are the same, and oll and mór (great) are the same.}^{184}
\]

1.16 **Alphabetic and thematic classification in Irish glossaries**

This examination of Irish-language glossaries has established that the most sophisticated glossary—in terms of its ordering system and date—is *O’Mulconry’s Glossary*. Considering the ordering system of the glossary in a comparative context with Old-English glossaries, based on MacNeill’s dating of it, the glossary displays a sophistication equal to the eighth-century *Épinal glossary* and greater sophistication than the early eighth-century *Leiden glossary* (See § 1.5.2)

Meyer’s statement that it is ‘but one step’ from glossaries with initial letter alphabetisation to fully alphabetised glossaries is called into question by a closer examination of the ordering systems used in the glossaries discussed above. *O’Mulconry’s Glossary* is the only pre-printing-press glossary that displays a more sophisticated ordering system that that of initial letter alphabetisation, but this is not employed systematically throughout. Indeed MacArthur’s view (see § 1.5.1), that the printing press

---

\[^{183}\text{ibid. 360}\]

\[^{184}\text{ibid. 364.}\]
was the dominant factor in the development of absolute alphabetisation is borne out in the Irish context. Significantly, the first evidence of the employment of absolute alphabetisation is in O’Clery’s *Sanasán*, first published at Louvain in 1643. Thus, we see that absolute alphabetisation was not achieved in the indigenous Irish context but was a product of Renaissance print culture and the Louvain experience. Further, in the context of medieval Irish glossaries, the progression from initial letter alphabetisation to absolute alphabetisation was achieved only when the medium the glossary was recorded in changed from that of hand-written manuscript to printed edition.

Thematic organisation in Medieval Irish-language glossaries has not been hitherto discussed by scholars. The evidence presented here indicates that *Dúil Laithe* is unique in its ordering system. Both this glossary and, to a lesser extent, the *Lecan Glossary*, merit further investigation in order to determine whether such a system was intentional on the part of the scribe or simply a result of the pattern of glosses in one or more of its base texts.

### 1.17 Medieval glossaries in *DIL*

The glossaries illustrated above continued to have relevance to Irish language lexicographers and many entries in *DIL* draw heavily on them. In some instances, *DIL* depends solely on glossary entries for citations of certain words. For example:
1. Introduction: an overview of the history of lexicography

Example 1

blinn, blind (AS or ON blind. For ON see Bidrag 93, 101, 155. Cormac’s blindauga, see ex. below, prob. a Norse personal name, 155). Expld. as dead person’s spittle, filament (nerve ?) of a dead person’s eye: blind .i. saile mairb, unde dicitur: bás mblindach, Corm. Y 136. blinn .i. snáithe ruisc mairb, unde dicitur blind (.i. auga, MS.) .i. dall nó cōech. Blindauga .i. dallshúilech in ling[u]a Galleorum, 138. So also O’Mulc. 183. blind .i. saile, O’Dav. 337. blinn .i. seile dhuine mhairbh, O’Cl. blinn .i. snáithe ruisg mairbh, ib. ‘blim’ (a corruption of blinn, blind, ed.) sili, Met. Gl. 34 §26.185

Example 2

cauil losevill .i. c. ‘testiculus’, Goidelica 75 §21 (Dúil Laithne).186

Example 3

1 rach a variety of baldness (or shaving the head?), Corm. Y 1091, defined: as e rōt na maile ðthā ind ētan na maile corricci in mullach (i.e. extending from forehead to crown).187

Example 4

feid .i. dliged, Lec. Gl. 43= feith .i. doilig M 83. Stowe Gl. 394. Cf. fēid.188


186 ibid. fasc. C 93.11.

187 ibid. fasc. R 1.77.

188 ibid. fasc. F 61.85.
1. Introduction: an overview of the history of lexicography

Example 5

gabáilche iä, f. (gábalch) prehensiveness [sic], rapacity: cosmail fri séigh é ... ar a gabáilchi, Corm. Y 1165.¹⁸⁹

Example 6

1 nathan illustrious, pre-eminent? nathan .i. ordeirc . . . nathan quasi nath in-aon [.i.] ind ollaman, Corm. Y 982. n. .i. oirdheirc, O’Cl. O’Brien.¹⁹⁰

1.18 List of some printed Irish language dictionaries before 1913


¹⁸⁹ ibid. fasc. G 1.37.
¹⁹⁰ ibid. fasc. N 17.33.
¹⁹¹ Lhuyd had Plunkett’s Latin-Irish manuscript dictionary (which remains unpublished to this day) transcribed and it became the main source for the dictionary in the Archaeologia, see Séan Ua Súilleabhair ‘Dinneen’s dictionaries of 1904 and 1927: background, use of historical dictionaries and of contemporary informants’, Pádraigín Riggs (ed), Dinneen and The Dictionary 1904-2004 (Dublin 2005) 62-77.
¹⁹² Tomás de Bhaldraithe, ‘Diolaim Focal’, Celtica 6 (1963) 262-7 contains citations from this work and other seventeenth- and eighteenth-century published and unpublished collections; idem, ‘Foclóir Úi Bheaglaoich’, A. Ahlqvist and Vera
1. Introduction: an overview of the history of lexicography

3. John Keogh, *Botanologia universalis Hibernica* or, a general Irish Herbal calculated for this kingdom, giving an account of the herbs, shrubs, and trees, naturally produced therein in English, Irish and Latin (Cork 1735).

4. John Keogh, *Zoologia medicinalis Hibernica*, or a Treatise of birds, beasts, fishes, reptiles, or insects, giving an account of their medicinal virtues and their names in English, Irish and Latin (Dublin 1739).

5. John O’Brien, *Focalóir Gaoidhilge-Sax-Bhéarla* or an Irish-English Dictionary. Whereof the Irish part hath been compiled not only from various Irish vocabularies, particularly that of Mr. Edward Lhuyd; but also from a great variety of the best Irish manuscripts now extant; especially those that have been composed from the 9th & 10th centuries, down to the 16th: besides those of the Lives of St. Patrick & St. Brigit, written in the 6th & 7th centuries (Paris 1768, 2nd ed. revised and corrected by Robert Daly, Dublin 1832).\(^{193}\)

6. William Shaw, *A Galic and English Dictionary*, containing all the words in the Scotch and Irish dialects of the Celtic that could be collected from the Voice and old books and MSS (Vol 1. London, 1780; vol 2 entitled *An English and Galic Dictionary*, containing the most usual and necessary Words in the English Language, explained by the correspondent Words in the Galic (London 1780)).

---


1. Introduction: an overview of the history of lexicography

7. Thaddeus Connellan, *An English-Irish dictionary intended for the use of schools, containing upwards of eight thousand words, with their corresponding explanation in Irish* (Dublin 1814).

8. Edward O’Reilly, *An Irish-English dictionary containing upwards of twenty thousand words that never appeared in any former Irish lexicon*. In their proper places in the Dictionary, are inserted, the Irish names of our indigenous plants, with the names by which they are commonly known in English and Latin. The Irish words are first given in the original letter, and again in italic, for the accommodation of those who do not read the language in its ancient character. To which is annexed, a compendious *Irish Grammar* (Dublin 1817; 2nd ed. Dublin 1921).<sup>194</sup>


10. Thomas De Vere Conneys, *Foclóir gaoidhilge-sacs-bearla, or an Irish-English dictionary, intended for the use of students and teachers of Irish* (Dublin 1849).<sup>195</sup>

11. Daniel Foley, *An English-Irish dictionary*: intended for the use of students of the Irish language, and for those who wish to translate their English thoughts, or the works of others, into language intelligible to the present Irish-speaking inhabitants of Ireland (Dublin 1855).

<sup>194</sup> Patrick O’Leary ‘A list of words not found in O’Reily’s Dictionary’, *Gaelic Journal* 28 (1902) 107-8.

12. John O’Donovan and Edward O’Reilly, *An Irish-English Dictionary* with copious quotations from the most esteemed ancient and modern writers, to elucidate the meaning of obscure words, and numerous comparisons of Irish words with those of similar orthography, sense, or sound in the Welsh and Hebrew languages. By Edward O’Reilly. A new edition, carefully revised and corrected, with a supplement, containing many thousand Irish words, with their interpretations in English, collected throughout Ireland, and among ancient unpublished manuscripts (Dublin 1864).

13. John Cameron, *The Gaelic names of plants, (Scottish, Irish and Manx)*, collected and arranged in scientific order, with notes on their etymology, uses, plant superstitions, etc., among the Celts, with copious Gaelic, English and scientific indexes (London 1883, 2nd ed. Glasgow 1900).


1. Introduction: an overview of the history of lexicography

20. L. McKenna, *English-Irish phrase dictionary, compiled from the works of the best writers of the living speech* (Dublin 1911).


---

CHAPTER 2

The history and limitations of the

*Dictionary of the Irish Language*

2.1 Abstract

The history and limitations of the *Dictionary of the Irish Language* are discussed in Chapter 2, in order to establish the value of the application of XML to the lexicography of Old, Middle and Early Modern Irish. In the first part of this Chapter, a detailed discussion of the history of the dictionary is presented, in order to understand how many of its limitations came about. In the second part, the limitations of the dictionary are discussed and categorised.

2.2 Description of the *Dictionary of the Irish Language*

The *Dictionary of the Irish Language* (hereafter DIL) is the most authoritative reference book available to scholars of Old and Middle Irish, and to a lesser extent Early Modern Irish. *DIL* can be broadly defined, in line
2. The history and limitations of the *Dictionary of the Irish Language*

with the criteria that have been established by Merkin, as a historical dictionary:

... the most comprehensive type of academic dictionary, usually covering a national language with a long recorded history. Its aim is to trace the development of each word from its first known appearance by means of a series of quotations and references selected from literary and non-literary sources and arranged in chronological order as to give evidence of:

1. Each word’s existence in different periods, places and genres
2. The changes in the word’s meaning, use, form and spelling
3. Its idiomatic phrases and habitual collocations
4. Its external, or comparative, etymology as well as its internal derivation
5. Its morphological and syntactic features.
6. Its stylistic and statistical characteristics.¹

While *DIL* broadly fulfils these criteria, criterion 1 and criterion 4 require qualification. With regard to criterion 1 (‘Each word’s existence in different periods, places and genres’), one can observe that a diachronic view of lemmata is presented, although with varying degrees of success in different fascicles: the fascicle D-Dégóir, for example, contains 46 references to Modern-Irish words and usage,² while the second fascicle of D contains only 13 such references and the letter E has only 5 such references. Words from different disciplines and areas are noted such as medical, legal, and botanical terms. The geographical distribution of a word is not a concern in *DIL* to the same degree that it is in other languages because Old Irish is peculiar in that it provides little recorded

² See especially the entry *debuith*. 
2. The history and limitations of the *Dictionary of the Irish Language*

evidence of dialectical variation.\(^3\) The only dialectical information I have observed in *DIL* is set out below.

### 2.2.1 Dialectical information supplied in the *Dictionary of the Irish Language*

**Fascicle B** has two references to dialectical variants in the words *bech* and *bocán*

**Fascicle D** as three references to dialectical variants in the words *dardóin* (sic), *dígu* and *dúalgas*

**Fascicle P** has one reference to dialectical variants in *partán*

**Fascicle R** supplies information about dialectical variants in the general entry on the letter *R*

**Fascicle S** has two references to dialectical variants in the words *seic* and *simlér*

**Fascicle U** has one reference to dialectical information in the word *úaibrech*

With regard to criterion 4 (‘Its external, or comparative, etymology as well as its internal derivation’), one notes that in most entries in *DIL*, words are etymologised only within Irish itself. Fascicle *P* has one of the highest instances of external etymology because most of the words in

---

2. The history and limitations of the Dictionary of the Irish Language

the letter P are loan words and their origin is obvious in most cases. The words beginning with P that are not loan words, for example, partán ‘a crab’ are likely to be substratum words, but this is nowhere indicated in DIL.

2.3 Early history of the Dictionary of the Irish Language

DIL has a protracted editorial history; its period of compilation spanned very many years and it includes contributions from several generations of scholars. It will be argued below that many of DIL’s limitations result from its complicated gestation. In order to understand how such limitations came about, the history of DIL is now examined in some detail.

The proposal to compile a dictionary was first mooted at a meeting of the Irish Archaeological Society held in the rooms of Dr Todd, Secretary of the Royal Irish Academy, in November 1852. A Dictionary

---

Committee was formed, and it was decided that Eugene O’Curry and John O’Donovan ‘who were alone capable of compiling the dictionary’\(^5\) should be employed as lexicographers on the dictionary project. However, this initial proposal was unsuccessful. Pledges of support for the work were not forthcoming; and this difficulty was compounded by the employment of O’Curry and O’Donovan by the Brehon Law Commission.\(^6\)

When the work of the Commission ceased seven years later, another proposal for a dictionary of the Irish language was put forward. O’Donovan, Whitley Stokes, and Rudolph Siegfried (Lecturer in Sanskrit in Trinity College Dublin), furnished memoranda of support. A memorandum from Mr Pigott, Secretary of the Committee, dated 12 December 1859, formalised the views expressed by the three scholars and set out a detailed plan of compilation for the dictionary. Pigott’s memorandum proposed that Peter O’Connell’s dictionary should be the base text for the Royal Irish Academy’s dictionary, and that this text should be further supplemented with words selected by O’Donovan from the dictionaries of O’Reilly and O’Brien and from Plunkett’s Collections. Material should also be selected from the collections of O’Curry, O’Donovan, Zeuss and Stokes. Then the Brehon Law Glossaries should be consulted, followed by the ‘more valuable of the great vellum MSS’.\(^7\) Then the words contained in all the existing printed books should be excerpted. After all this material had been collected, it was proposed that the *Book of Leinster*, the

\(^5\) R. I. Best, ‘On recent studies in the Academy’, *Proc Roy Ir Acad* (C) 51 (1945-1948) 19.

\(^6\) ibid. 19-20.

\(^7\) ibid. 33.
2. The history and limitations of the Dictionary of the Irish Language

*Leabhar Breac* and the *Leabhar na h-Uidhre* should be particularly studied. Indeed, it was estimated at the time that these MSS contained more than 10,000 words that had not been included in any printed reference work. In addition to setting out a detailed work schedule, Pigott’s memorandum also notes the possibility that Stokes and Siegfried would append a dictionary of very ancient words to the dictionary proper.\(^8\)

This second proposal to begin work on a dictionary of the Irish Language was unsuccessful because of the deaths of the scholars who had been selected to work on the project: O’Donovan and O’Curry died in 1861 and 1862 respectively, and Siegfried died young in 1863.

Efforts to compile a dictionary were again made in 1880 when Robert Atkinson was appointed editor. However, his twenty-seven year tenure did not result in any lexicographical publications as such, though he did edit (in a thoroughly unsatisfactory manner)\(^9\) the Glossary to the *Ancient Laws of Ireland*.\(^10\) Further, he appended a more successful glossary to his *Passions and Homilies*.\(^11\) When Kuno Meyer succeeded him in 1907 as director, with Osborn Bergin as editor, it became clear that the dictionary was in a most disorganised state. The material collected by Atkinson

---

\(^8\) ibid. 32-34. The memorandum is reproduced in full by Best.


was mostly defective and the ‘task of sifting, arranging, and classifying the meanings of the words excerpted was beyond the power of the few workers then available’. After a meeting with the Royal Irish Academy Committee the sense of exasperation that Meyer felt is palpable in a letter he wrote to R. I. Best:

As to Dictionary matters I know not what to think or write, or do. … The truth is that none of these gentlemen understands the difficulties of the task, and I do not understand their interpretation of the bequest.

Such were the demands on Meyer’s time that he was able to act in a supervisory role only, and Bergin was given the actual work of dictionary making. The task of excerpting so large a body of material was an impossible one for a single scholar, even if he were energetic, and Bergin’s progress was slow. The decision was now taken to leave the publication of the letters A-C until a later date and to begin publication of the dictionary with the letter D, since A to Dn had been covered (after a fashion) in Kuno Meyer’s Contributions to Irish Lexicography. Meyer’s Contributions, the result of ‘twelve years desultory reading of Middle- and early Modern-Irish books and manuscripts’ will be discussed in greater detail in § 2.3.6

12 Best, op. cit. 23.
14 Best, Academy, 23.
15 Kuno Meyer, Contributions to Irish lexicography (i & ii) (London 1906).
16 ibid. 1.
2. The history and limitations of the *Dictionary of the Irish Language*

2.3.1 The dictionary in 1910

By 1910 it was clear to Meyer that the dictionary project was facing serious difficulties. When he reviewed preliminary model sheets, prepared by Bergin, he was most unhappy:

> The meaning must be much better given. What troubles me most is that hardly anything has been properly excerpted. I relied on Bergin to see to that, sent him everything as soon as it was printed (*Anecdota* 3 for example) and find that the work has either not been done at all, or imperfectly. I am now asking Bergin to let me have a further instalment which I will go through. But dear me, it is slow work.\(^{17}\)

Charles Plummer also expressed dissatisfaction with the preliminary sheets, and he complained that the definitions were ‘given indiscriminately’.\(^{18}\)

In addition to the problems of substandard work carried out by Atkinson and Bergin, and a lack of staff and finance, other factors hindered the lexicographical work. One of the most intractable difficulties the project faced was a dearth of printed editions of texts. Furthermore, many published texts were not of an acceptable scholarly standard—making the excerpting of information from them difficult and unreliable.

The significance of this had been impressed on Meyer during a visit he made to the Egyptian Museum in 1907. There he observed that like Irish scholars, Egyptian scholars possessed a number of unprinted texts ‘thought not nearly as much as we have’.\(^{19}\) He also observed, in common

---

\(^{17}\) Ó Luing, *Kuno Meyer*, 81 (Meyer to Best, 3 Feb. 1910. NLI, Ms No 11002).

\(^{18}\) ibid.

\(^{19}\) ibid. 46.
with the Irish situation, that the Egyptian scholars had been limited by
the number of badly edited texts that had been published. In an effort to
solve this problem, the Egyptian dictionary staff transcribed many texts
afresh, and then edited and published them. This was made possible by
the level of assistance they received from the government, as well as from
scholars and institutions all over the world. Meyer realised that it would
not be possible for the scholars working on the Irish dictionary to follow
such a lead because no comparable level of assistance was available in
Ireland at that time (and in regard to digital text projects one sadly notes
that today the situation is little better). After the visit he wrote to Best:

It made me realise once more that the R.I.A. do not understand
the task upon which they have embarked, and that I have to work
with wholly insufficient means.\(^{20}\)

### 2.3.2 The appointment of Marstrander

In 1910 Bergin was appointed to the chair of Old Irish in the National
University, and Carl J. S. Marstrander, took over the position of editor
of *DIL*. Marstrander, who was given the sobriquet *An Lochlannach* by
his Irish colleagues, was considered a brilliant young scholar and would
go on to be appointed professor of Celtic Philology in the University of
Kristiania in 1913, at the age of 30.

It is clear from correspondence between Meyer and Best that they
felt a corner had been turned with his appointment. Equally clear,
from their correspondence, is the unrealistic nature of the expectations

\(^{20}\) ibid.
of both Marstrander and the dictionary project. In light of both men’s knowledge of the fundamental problems that existed, such expectations are astonishing.

On October 26 1909 Best wrote to Meyer:

... P[raeger] was curious to know who Marstrander was. I praised him to the utmost, of course, and said were he to work on the Dictionary, things would begin to materialise wonderfully fast. He was greatly impressed, and hoped you would be able to carry the Council, who must now begin to grow anxious. ... I have no doubt whatever if a sure-footed rapid worker like M. were on the work, he would under your directions complete the whole work in 10 years, and then be able to return home, with a great reputation, to fill a chair in Norway. Nothing he could do in the meantime would be so likely to spread his name as the Dictionary and he is not the person to allow it to exclude him from throwing off many and varied studies, such as he has just done for ZCP.\(^{21}\)

When Meyer took Marstrander through dictionary proofs and slips that Bergin had prepared, Meyer was encouraged to observe that Marstrander identified many errors. His tone is ebullient in a letter to Best:

I am delighted with him. He is wonderfully keen and eager and looks forward to his work with zest. He will reform the dictionary. Nothing escapes him. He went through the proof and Bergin’s slips and he at once spotted every weak point and had the right thing ready. I feel that I can absolutely trust him and a great burden is off my mind. \(^{22}\)

Similar sentiments are expressed in another letter to Best:

The young man has the ball at his feet.\(^{23}\)

\(^{21}\) ibid. 78. (Best to Meyer, 26 Oct. 1909 NLI, Ms No 11002 (58)). Emphasis mine in the above citation.

\(^{22}\) ibid. 139 (Meyer to Best, 25 Feb. 1910 NLI, Ms No. 11002).

\(^{23}\) ibid. 83 (Meyer to Best, 4 April 1910. NLI, Ms No. 11002).
2. The history and limitations of the *Dictionary of the Irish Language*

2.3.3 Marstrander’s *D*

When Marstrander was appointed editor in 1910, fifty-eight years had elapsed since the dictionary had been first proposed; and not a single line had appeared. This was due partly to the practical problems I have discussed above—among them underfunding and a lack of willing and competent lexicographers—that beset the project during the years 1852-1910. In addition, the project was badly managed and badly thought out. When Marstrander became editor in 1910, not even a proper plan for the dictionary existed, and the work that had been carried out to lay the foundations of *DIL* was, for the most part, substandard. Marstrander himself described the collection of dictionary slips that had been gathered up to the time of his appointment as ‘defective and heterogeneous, gathered from east to west without any general plan, and to some extent by people who lacked the most elementary knowledge of older Irish’.24

To complicate this situation further, Marstrander found himself working to a close deadline that had no connection with the lexicographical work as such. The Reverend Maxwell Close, a member of the Society for the Preservation of the Irish Language, had granted the Academy £1000 (a very large sum in those days) on the condition that a portion of the dictionary be published within ten years of his death (which occurred in 1903).25 The Academy, ever impoverished, laid great store on not losing

---


2. The history and limitations of the *Dictionary of the Irish Language*

this endowment. Thus, from the time of his appointment, Marstrander had less than three years to realise what the other scholars who had worked on the project had not achieved in fifty-three.

On August 14 1913, the first fascicle of the *Dictionary of the Irish Language*, D-dégoir—to become known as Marstrander’s *D* due to the controversy that surrounded it—was published by the Royal Irish Academy. The ensuing controversy was not only intensely personal but has been described as a ‘victory for the men of caution over the man of vision, for the conservatives over the pioneer of adventurous yet soundly-based scholarship.’

In order to explore this controversy further, references to Marstrander in the correspondence between Best and Meyer will now be discussed.

### 2.3.4 The correspondence between Meyer and Best

Despite the praise Meyer and Best had lavished on Marstrander upon his appointment as editor, the correspondence that passed between them indicates that the working relationship between Meyer and Marstrander had broken down at an early stage, for reasons that are now unclear.

By 1911, just a year into his editorship, Marstrander was so overworked that he got pleurisy, and was forced to have a lung operation. Ó Lúing recounts that Meyer was extremely worried about him, and that he showed Marstrander great care during this period:

> Nobody could have shown more concern for Marstrander during his illness and convalescence than Kuno Meyer, who gave him

2. The history and limitations of the Dictionary of the Irish Language

good advice and stayed with him for company in Monte Carlo where he went for some weeks in April to recuperate, watching over him with something like paternal solicitude.27

Such ‘paternal solicitude’ did not last. After their stay in Monte Carlo, Marstrander returned to work in haste, and Meyer wrote to Best urging him to take care of the young scholar, stating: ‘if he fell ill again now we would all blame ourselves very much’.28 Meyer expressed similar concern on hearing of Marstrander’s return to Norway, and wrote to Marstrander ‘no doubt advising him to be prudent’.29

Nevertheless, by December 1911, the working relationship between Meyer and Marstrander was beginning to deteriorate. In a letter from Meyer to Alice Stopford Green, Meyer states, further to her comments about Marstrander that, like her, he had not heard from him in months.30 Later that month, Meyer wrote to Marstrander, urging him to reply and to write also to Mrs Green.31

When Marstrander did reply, Meyer’s attitude to him has changed remarkably. Marstrander’s letter, ‘written in a dissatisfied spirit’,32 stated that he could not continue to juggle the dictionary project, teaching and the editorship of Ériu. Meyer wrote to Best ‘I am afraid he

27 Ibid. 96.
28 Ibid. 97 (Meyer to Best, 20 May 1911 from Endsleigh Palace Hotel, London. NLI, Ms No. 11002).
29 Ibid. 100.
30 Ibid. 106.
32 Ibid. 109.
[Marstrander] is eaten up with conceit’.\textsuperscript{33} In a letter to Best, dated 25 March 1912, further evidence of the deteriorating communication between Meyer and Marstrander is evident, when Meyer seeks Best’s help with the forthcoming volume of Ériu because Marstrander had accepted, unknown to Meyer, material that was not suitable for publication.\textsuperscript{34}

Progress was made at a meeting in April 1912 when it was decided that Marstrander should temporarily relinquish his teaching duties and Ériu, and concentrate on the dictionary.\textsuperscript{35} Meyer’s correspondence with Best after this meeting indicated that he felt Marstrander had made the right decision. But again, in a letter dated 9 November 1912, Meyer stated that he had heard nothing from Marstrander since his return to Norway.\textsuperscript{36} The next mention of Marstrander is in early in 1913, when Meyer received a card from him stating that he had requested three more assistants from the Academy.\textsuperscript{37}

In February 1913, Meyer wrote to Best: ‘I have to write to the Norse Storthing to recommend a chair for C. M.—but this entre nous’.\textsuperscript{38} A reference to Marstrander in a May 1913 correspondence is disparaging: ‘Marstrander remains the same queer fish he always was’.\textsuperscript{39}

\textsuperscript{33} ibid. 109 (Meyer to Best, 2 March 1912 from Charlottenburg, Niebuhrstrasse 11A. NLI, Ms No. 11002).

\textsuperscript{34} ibid. 113 (Meyer to Best, 25 March 1912. NLI, Ms No. 11002).

\textsuperscript{35} ibid. 112.

\textsuperscript{36} ibid. 123.

\textsuperscript{37} ibid. 126.

\textsuperscript{38} ibid. 131.

\textsuperscript{39} ibid. 134 (Meyer to Best, postcard 26 May 1913 from Charlottenburg. NLI, Ms No. 11002).
2. The history and limitations of the *Dictionary of the Irish Language*

The reasons for the lack of correspondence on the part of Marstrander are difficult to interpret. It is likely that he became so immersed in work that he let communication slide. Perhaps, too, something happened to upset him in Monte Carlo. At this remove, we can only conjecture, but it is clear that he was bearing an overwhelming workload.

Whatever the reasons for the poor communication between the two, on the 23 June 1913, not a full three years after Marstrander’s appointment, Meyer wrote to Best that he had ‘at last’ received some proofs from Marstrander. 40  It is interesting to note that Meyer expressed some reservations about proofs of the fascicle, but nothing comparable to the criticisms that he would later express:

> Of course he has amassed a lot of information and nothing comes amiss; but it comes in a very questionable form. However, all that is now too late to mend. The great thing is to bring it out. I suppose it will remain a torso. 41

### 2.3.5 Controversy

The first controversy regarding Marstrander’s *D* broke even before the fascicle was published. During 1913, Best had been growing irritated that the only people who were allowed to see the dictionary proofs that had been arriving at the Academy were Maud Joynt and Edward Gwynn. His irritation gave way to anger when Maud Joynt showed him the preface Marstrander intended to publish. He was so annoyed that he showed it to Robert Macalister and Lloyd Praeger, members of the Dictionary

---

40 ibid. 139.

41 ibid. 140 (Meyer to Best, from Charlottenburg, 23 June 1913. NLI, Ms No. 11002).
2. The history and limitations of the *Dictionary of the Irish Language*

Committee. The Academy refused to publish the preface. Best wrote to Meyer describing aspects of the preface that so irked him:

Then he proceeds to define the scope of the Dictionary which, beginning with the dawn of the Irish Language, traces the shape, meaning and grammatical inflexion etc. of all the words in use among the *Celts of Ireland*. He stated that his predecessors had your *Contribb.* as a model. He however thought that the time had come for a bigger thing. He winds up by thanking his 3 assistants in alphabetical order, and Miss Joynt for indefatigable help in proof reading, and then couples you and Gywnn together for the proofs of several slips and supplying many illustrations, which called for the most grateful recognition. Fancy not a word about Osborn Bergin, not a mention of Purton. Your published D-Dno not mentioned. I was so disgusted that I expressed my mind to Miss Joynt. She, poor girl was greatly worried over this preface, and felt something was wrong, but her [task?] was merely to correct English. When Praeger saw it he agreed it could not be published. So we have called the Dictionary Committee for Friday. We shall print instead a bald business-like statement, explaining why this Fasciculus 1 (not Part 1) begins with D rather than A, and what its price will be, and how it will be sold. The Introduction should come at the end of Volume 1 and Praeger and I think there should be a sort of historical or official preface stating how the Dictionary arose, and who worked upon it. But fancy ignoring Bergin and his work upon it, and bracketing you with Gwynn. He had send me two paragraphs for insertion, explaining his use of etymology, and referring to the dawn, the broad daylight of Old and Middle Irish and the gloomy eve of Modern Irish !!! all of which are represented in the Dictionary. He writes like a school boy. All of this *entre nous.*

In Meyer’s response to Best, his appraisal of the proofs is harsher than in his earlier letter: he refers to Marstrander as a ‘great sinner’

---

42 ibid. 140. Ó Lúing’s comments on this incident are very insightful. He comments that Marstrander stated the aim of the ‘R.I.A Dictionary of the Irish Language was to introduce order into the chaos of Irish Lexicography, and to mark his astonishment Best put down a treble exclamation mark at this statement, which [Ó Luing comments] seems no more than a plain recital of fact’.

43 ibid. 140 (Best to Meyer, date unavailable, NLI, Ms No. 11002).
and he contemplates letting mistakes he considers Marstrander to have made stand in order to teach him a lesson. Meyer also sent Best a proof of the Dictionary that he had revised, and Ó Lúing concludes that Meyer must indeed have made numerous changes to the proof, based on Best’s reaction. Nevertheless, the personal nature of Best’s comments on Marstrander are startling:

The proof is indeed awful! I hope this headstrong youth will have the sense to adopt your corrections. We can do nothing here. It is ludicrous that the members of the Dictionary Committee are not permitted to see the proofs of the Dictionary. The great mistake was for the Academy to regard M. as master of the situation and to pander to his petty vanity . . . . It [Marstrander’s D] has all the appearance of profound learning without the reality. It looks more like a hotch-potch. . . . M[arstrander] has no sense of what is fitting. He must be swollen with vanity, and a belief in the importance of his methods and analyses . . . if you do not succeed in bringing him to reason. It will be known as Marstrander’s Folly.

After the publication of the fascicle, the criticism continued unabated. The vituperative tone of another private correspondence between Best and Meyer indicates how bad relations must have been between them and Marstrander:

Bergin also continues to discover fresh Marstrandiana. From what you tell me of Pokorny’s labours and your own in the same field this great North Sea bubble will be pricked and rent in a very short time.

---

44 ibid. 140.
45 ibid. 141.
46 ibid. 140 (Best to Meyer, NLI, Ms No. 11002).
47 ibid. 151 (Best to Meyer, postcard 12 Nov. 1913, from 57 Upper Leeson St., Dublin. NLI, Ms No. 11002).
The controversy did not stay in the private arena. Pokorny published a list of corrections of the fascicle in 1915 and 1921. Kuno Meyer also published two articles expressing dissatisfaction with Marstrander’s treatment of the material. While Edward Gwynn of Trinity College and John MacNeill expressed their approval of the fascicle, Meyer’s ‘influence and authority attracted the agreement … of other Celtic scholars [with his views]’.49

In 1913, Meyer drew attention to a number of personal names that Marstrander had included, but that he considered to be ghost names. Further, in 1915, he received a list of errors from Pokorny and Meyer published another article in which he stated:

\[
\text{Since it is not to be taken that this dictionary will be continued—or indeed that there will be a continuation—in the way that it is laid out I do not deal here with a critique of its plan or manner in which it is carried out but I confine myself rather to giving a list of errors, mistakes, and mis-apprehensions …} \quad 51
\]

He instead criticised the fascicle on a number of grounds from printer’s errors and what he considered an over-ambitious approach to the work, to more specific criticisms such as mistakes he considered Marstrander to have made in the treatment of some words.

---


49 ibid. 146.


In his reply to Meyer’s article (‘Das Wörterbuch’), Marstrander discussed the difficult circumstances that he worked under, and rebutted some of Meyer’s criticisms. Marstrander furthermore argued that some of the limitations Meyer notes are in fact based upon mistakes that Meyer himself made in the treatment of words, for example, to ‘dagsad, fälschlich aus munadagsad erschlossen’.\(^{52}\)

Marstrander states:

The mistake is Meyer’s, who did not know what to do with the form when editing the poem in question CZ VIII 310 ff. and printed munadagsad. This was faithfully reproduced in the Dictionary by the naìve Editor, who, however, printed *dagsad so as to warn the reader. But I frankly admit that Meyer’s mistakes are no excuse for mine.\(^{53}\)

In rebutting Meyer’s claims regarding his overambitious approach to the work, Marstrander claims ‘… I bring a great deal that is new. But fresh ground is hard to till, and no one could be more ready to admit it than myself that I have often been mistaken in respect to details’.\(^{54}\)

Indeed Binchy remarked on Meyer’s critique of the work:

Although it included some valuable material, it was nothing more than a long list of corrigenda many of them indeed genuine, but some doubtful, some clearly unjustified, some trivial and some (though to be fair, most of these were not contributed by Meyer himself) just silly. And there was not a word of praise for the ingenuity and the wide learning displayed by the pioneer of this new advance in Irish studies.\(^{55}\)

---

\(^{52}\) ibid. 364 §16, 36.


\(^{54}\) ibid. 5.

2. The history and limitations of the *Dictionary of the Irish Language*

But his remarks were made in 1966, when everybody knew that Marstrander’s approach was the right one.

**2.3.6 Comparison of Marstrander’s *D* with Meyer’s *Contributions***

While Marstrander certainly made mistakes, when his work is contrasted with Meyer’s *Contributions* it is clear that Marstrander ploughed a new furrow in Irish lexicography and that the scheme he set out had a lasting influence on the form that the dictionary was to take.

The primary differences between Marstrander’s *D* and Meyer’s *Contributions* are as follows:

1. Marstrander gives, where possible, a full treatment of the inflexion of nouns and adjectives. In the *Contributions*, information about the inflexion of lemmata is not given, but gender is supplied for most words, for example, ‘*ab* f, a river’.\(^{56}\)

2. Meyer presents an achronological view of gender (which changes chronologically in Irish), but Marstrander presents, where possible, a diachronic view of gender and inflexion. For example:

   (a) Marstrander: ‘*daig*, i later t, t’,\(^ {57}\) cf. Meyer: ‘*daig* f.’\(^ {58}\)

   (b) Marstrander: ‘*dairthech* s[- stem], n. later m.’,\(^ {59}\) cf. Meyer: ‘*dair-thech* n.’\(^ {60}\)

---


\(^{57}\) *DIL*, Fasc. D-Degóir, 17.42.


\(^{59}\) *DIL*, Fasc. D-Degóir, 38.39-38.41.

\(^{60}\) Meyer, *Contributions*, 580.
2. The history and limitations of the Dictionary of the Irish Language

(c) Marstrander: ‘damthain n[-stem], later i, f. vn of daimid,‘61 cf.
Meyer: ‘damthain inf. of damim q.v. [no inflexion or gender supplied in damim].‘62

3. Meyer etymologises the lemma from a derivation within Irish itself and to a much lesser extent from Latin and Welsh. For example: ‘2 abacc a beaver (W. afanc)‘63 ‘baislech (baslach) a stag‘64 canthus (Lat. cantus) a chant‘65

4. Marstrander follows a similar practice in regard to etymology: the majority of etymologies he supplies are from derivation within Irish, for example: ‘daidbir … from do- and adbar‘66 ‘dardöin … from *etar-dâ-öin‘67

He also supplies some Latin etymologies, for example: ‘2 dechtaid … An early loan from Latin dictare‘68

5. Meyer records a large number of the verbal entries under the first person singular present indicative form of the verb, for example, ‘ad-annaim‘69 ‘ad-cotadaim‘70 ‘ainmim‘71

---

61 DIL, Fasc. D-Degóir, 70.12.
62 Meyer, Contributions, 588.
63 ibid. 4.
64 ibid. 169.
65 ibid. 315.
67 ibid. 102.25.
68 DIL, 202.62.
69 Meyer, Contributions, 14.
70 ibid. 18.
2. The history and limitations of the *Dictionary of the Irish Language*

However, he also has entries to verbs in other tenses, numbers and moods, as illustrated by: ‘ad-cúaid’;\(^{72}\) ‘ad-cuitecht’.\(^{73}\)

6. Marstrander, on the other hand, consistently records all forms of a verb under the third singular present indicative form, as is conventional in very many dictionaries of inflected languages.\(^{74}\)

7. Meyer gives every compound form of a noun, adjective, verb and verbal noun a separate lemma. For example, ‘bán-chath’;\(^{75}\) ‘ban-dál’;\(^{76}\) ‘bél-derg’;\(^{77}\) ‘dáiscor-duine’.\(^{78}\)

Marstrander, on the other hand, supplies compound forms under the lemma, for example, all compound forms such as ‘deg-ainm’\(^{79}\) and ‘dag-mór’\(^{80}\) are recorded under ‘1 dag’ in DIL.

8. Meyer’s *Contributions* and Marstrander’s *D* were planned on completely different scales. As cited above, Meyer notes that the *Contributions* was the outcome of twelve years desultory reading of Middle and Early Modern Irish material.\(^{81}\) Marstrander’s plan was

\(^{71}\) ibid. 42.
\(^{72}\) ibid. 19.
\(^{73}\) ibid. 19.
\(^{74}\) Notable exceptions being Patrick S. Dinneen’s *Folcór Gaedhilge agus Béarla* and *Oxford Latin Dictionary*.
\(^{76}\) ibid. 175.
\(^{77}\) ibid. 195.
\(^{78}\) ibid. 580.
\(^{79}\) *DIL*, fasc. D-Dégóir. 10.42
\(^{80}\) ibid. 9.34.
2. The history and limitations of the Dictionary of the Irish Language

to ‘… trace, not only every word, but if possible every characteristic expression on its way forwards, whether it be first met with in the hard lines of the Ogam stone or in Keating’s graceful style …’.  

If one compares Marstrander’s D with Meyer’s Contributions it is evident that Marstrander’s treatment and categorisation of the material is unparalleled in the field of Irish lexicography—medieval and modern. Despite the criticisms made of his work when published, he had a lasting influence on the shape that the dictionary was to take. His was the plan followed and where it was not followed, later editors were mistaken.

2.3.7 History of DIL from 1913 to 1975

Marstrander resigned the editorship of the Dictionary in 1913 and the remaining part of the letter D, in two fascicles, was not published until 1959-60. After his departure, Bergin took the position of general editor, and nineteen years later, in 1932, the letter E was published. It was edited by Maud Joynt and Eleanor Knott. This is the least satisfactory fascicle of DIL.

Quin states that ‘if the dictionary had continued at this rate the date of completion would have been so remote as to be of very little interest’. No doubt recognising this, the Department of Education intervened in

---

81 Meyer, Contributions, 1.
82 Marstrander, Dictionary, 5.
84 DIL, v.
1936, at the behest of Eamon De Valera, then President of the Executive Council of the Irish Free State. It was proposed that the remaining ten letters should be published within five years (i.e. 1936-41), under the new title Contributions to a dictionary of the Irish language. What was proposed was ‘a dictionary of somewhat lower standard of meticulous scholarship than the Royal Irish Academy “Dictionary”, and with the ‘minimum of editorial revision’. A sum of £5,200 was furnished, of which £2,000 was allocated for printing. Even this much reduced target was not met by the dilatory Academy. The publications of the remaining fascicles spanned thirty-nine years, finishing with the publication of B in 1975.

Numerous scholars were involved in the lexicographical process during the period of compilation of DIL, as illustrated by § 2.3.8 and many did not live to complete their work or see its final publication. Maud Joynt, for example, arranged the letters M; N-O-P and R, but died after the fascicles M and N-O-P had been issued; Myles Dillon and Eleanor Knott prepared the letter R for publication.

Quin has discussed the reasons why it took so long to finish the Dictionary, even after the intervention of De Valera. The letters M; N-O-P; U; and R had already been prepared for publication in the style and scope that had been initiated by Marstrander, and so the publication

---

85 ibid. v.
86 ibid. This was of course not including F and G which were near completion, or the letters A-D1.
87 DIL, v. Quin, quoting a report in the Academy archives.
88 Best, Academy, 23.
2. The history and limitations of the Dictionary of the Irish Language

of the fascicles under the title of Contributions to a dictionary of the Irish Language, represented only a name change, and not a change in the treatment of the material. It is clear from Quin’s discussion that many of the problems that beset the Dictionary in its earlier years remained unsolved: the shortage of qualified staff\(^\text{89}\) was still a problem, and in any case, the RIA was not prepared to pay staff for lack of funds, but also for lack of commitment. The problems of bad management, poor control, lack of planning and general disorganisation, still persisted. Furthermore, not all of the materials that the lexicographers required were at their disposal. Due to the glacial progress of Irish studies, and perennial underfunding, the lack of good printed editions, and the presence of poorly edited editions were problems still as relevant in the 1930s as they had been in the early years of the project.

### 2.3.8 Editors and assistants of the Dictionary of the Irish Language

**Fascicle A(1) (A—allbrúach)** 1964.

Arranged by Anne O’Sullivan and E. G. Quin.

**Fascicle A(2) (allcaille—attcuinti)** 1967.

Arranged by Anne O’Sullivan and E. G. Quin.

**Fascicle B** 1975.

Arranged by Maura Carney and Mairín O Daly. General Editor: E. G.

---

\(^{89}\) This situation can be contrasted with the editors of the Dutch Dictionary who obtained the status of civil servants in 1922. Furthermore, the Dictionary was fully subsidised by the Belgian and Dutch governments and incorporated as the Instituut voor Nederlandse Lexicologie that was established in 1967. Merkin, The historical/academic dictionary (1983) 126.
2. The history and limitations of the *Dictionary of the Irish Language*

Quin.

**Fascicle C(1)** (*C*—1 *clúain*) 1968.

Arranged by Próinséas Ní Chatháin, Maírín O Daly, P. Ó Fiannachta and Anne O'Sullivan. General editor: E. G. Quin.

**Fascicle C(2)** (*clúain*—*con-ling*) 1970

Arranged by Seán Connolly, Próinséas Ní Chatháin, Maírín O Daly, and P. Ó Fiannachta. General Editor: E. G. Quin.

**Fascicle C(3)** (*con-luí*—*cutum*) 1974.

Arranged by Maura Carney, Próinséas Ní Chatháin, Maírín O Daly, and P. Ó Fiannachta. General editor: E. G. Quin.

**Fascicle D(1)** (*D*—*degóir*) 1913.

Under the editorship of Carl J. S. Marstrander. (Dictionary)

**Fascicle D(2)** (*degra*—*dodelbtha*) 1959.

Arranged by Mary E. Byrne and Maud Joynt.

**Fascicle D(3)** (*dodénta*—*dúus*) 1960.

Arranged by Mary E. Byrne and Maud Joynt.

**Fascicle E** 1932.

Arranged by Maud Joynt and Eleanor Knott. General editor: Osborn Bergin. (Dictionary)

**Fascicle F(1)** (*F*—*fochrait* 1950.

Edited by Maud Joynt and Eleanor Knott. (Dictionary)

**Fascicle F(2)** (*fochratae*—*futhu*) 1957.

Edited by Maud Joynt and Eleanor Knott. (Dictionary)

**Fascicle G** 1955.

Arranged by Mary E. Byrne.
2. The history and limitations of the *Dictionary of the Irish Language*

**Fascicle H** 1976.
Arranged by Maura Carney. General editor: E. G. Quin.

**Fascicle I** 1952 (issued in parts in 1952; reprinted 1966).
Arranged by Mairín O Daly and Anne O’Sullivan.

**Fascicle L** 1966.
Arranged by Mairín O Daly, and P. Ó Fiannachta. General editor: E. G. Quin.

**Fascicle M** 1939.
Arranged by Maud Joynt.

**Fascicle N-O-P** 1940.
Arranged by Maud Joynt.

**Fascicle R** 1944.
Arranged by Maud Joynt.

**Fascicle S** 1953.
Arranged by Maud Joynt, Teresa Condon, Mairín O Daly and Anne O’Sullivan and E. G. Quin. General editor E. G. Quin.

**Fascicle T(1) (T–tnúthaigid)** 1943.
Arranged by David Greene and E. G. Quin.

**Fascicle T(2) (to–tu)** 1948.
Arranged by David Greene and E. G. Quin.

**Fascicle U** 1942. Arranged by Teresa Condon.
2.4 Limitations of the *Dictionary of the Irish Language*

As the section above set out, *DIL* has a protracted gestation and complex editorial history. Here, it will be argued that despite being widely accepted as the most authoritative reference work of its kind for the Irish language, it exhibits serious limitations. It will further be argued that many of these result from the lack of a clear editorial policy, and on occasion, any editorial policy at all.

2.5 Navigating *DIL*: inadequate cross-referencing

2.5.1 *air-*

Irish experienced complex sound changes at various stages of its development,\(^90\) and the language therefore displays a complicated orthography.\(^91\) Words beginning with *air-* for example, can appear as *aur-* , *er-* , *ir-* , and *ur-* . Thus, the word *airthach*, is also attested as *aurthach*, *urthach* and *erthach*. A student who encounters the forms *urthach* and *erthach* in a text is unlikely to be aware of such spelling variants—especially in light of the fact that a table explaining such sound changes is not provided in *DIL*—and so is wholly dependent upon cross-references to find the word in question. In the case of the word *airthach*, however, no cross-reference is provided from the forms *erthach*, *aurthach* or *urthach*. The only form

---


that is cross-referenced is *urrach*. Furthermore, the form of the lemma *airthach* is not cited in the lexical entry itself as an attested form, and so the form must represent a form reconstructed by an editor, though this is not indicated in any way (for example by an asterisk as is customary). Indeed, this limitation in terms of cross-referencing and failure to indicate the provenance of the lemma is not limited to this entry, as § 2.5.1.1 illustrates.

### 2.5.1.1 Analysis of airaibind to airbaid

**Figure 2.1**

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airaibind attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>uráibinn no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>uráoibhinn no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>uráibhind no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>éraibind no cross-reference</td>
</tr>
</tbody>
</table>

**Figure 2.2**

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airáibnius attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>uráibnius no cross-reference</td>
</tr>
</tbody>
</table>
2. The history and limitations of the *Dictionary of the Irish Language*

**Figure 2.3**

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airaiccecht</th>
<th>attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>airaiccept</td>
<td>cross-referenced</td>
</tr>
<tr>
<td>VARIANT</td>
<td>auraiccept</td>
<td>cross-referenced</td>
</tr>
<tr>
<td>VARIANT</td>
<td>uraisecht</td>
<td>no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>uracecht</td>
<td>no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>eraicept</td>
<td>no cross-reference</td>
</tr>
</tbody>
</table>

**Figure 2.4**

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airáilde</th>
<th>attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>eráildi</td>
<td>no cross-reference</td>
</tr>
</tbody>
</table>

**Figure 2.5**

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airain [sic]</th>
<th>attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>urain</td>
<td>cross-referenced</td>
</tr>
<tr>
<td>VARIANT</td>
<td>errain</td>
<td>no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>erain</td>
<td>no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>aurain</td>
<td>no cross-reference</td>
</tr>
</tbody>
</table>

**Figure 2.6**

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airairde</th>
<th>attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>urairdi</td>
<td>no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>urarda</td>
<td>no cross-reference</td>
</tr>
</tbody>
</table>
2. The history and limitations of the *Dictionary of the Irish Language*

*Figure 2.7*

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airairgna [sic]</th>
<th>attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>erérna</td>
<td>no cross-reference</td>
</tr>
</tbody>
</table>

(Furthermore, a lexical entry for the word *érérna* exists, which notes ‘cf. 2 ér, and érgna, airgna’. The intensive form of *airgna, airairgna* ‘very conspicuous’, with variant *erérna*, is not drawn to the reader’s attention).

*Figure 2.8*

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airamnas</th>
<th>attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>eramnas</td>
<td>no cross-reference</td>
</tr>
</tbody>
</table>

*Figure 2.9*

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airamra</th>
<th>attested form not cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>eramra</td>
<td>no cross-reference</td>
</tr>
</tbody>
</table>

*Figure 2.10*

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>airard</th>
<th>attested form cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>urard</td>
<td>no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>irard</td>
<td>no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>erard</td>
<td>no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>aurardai</td>
<td>no cross-reference</td>
</tr>
</tbody>
</table>

2.5.2 *ar-fo-

The orthography of lemmata where *ar-* and *fo-* fall together is also complicated. The number of cross-references from variant forms is also unsatisfactory in this case.
2. The history and limitations of the Dictionary of the Irish Language

2.5.2.1 Analysis of ar-fócair and ar-fogni

Figure 2.11

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>ar-fócair attested form cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS SIMPLE VERB</td>
<td>airfhócran cross-referenced</td>
</tr>
<tr>
<td>VARIANT</td>
<td>aurfogair no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>erfhuagradh no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>roerfhúacair no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>ro hirfuagrad no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>ro urocair no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>urfocra no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>urfhograidi no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>erfuagair no cross-reference</td>
</tr>
</tbody>
</table>

Figure 2.12

<table>
<thead>
<tr>
<th>LEMMA</th>
<th>ar-fogni attested form cited in lexical entry</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIANT</td>
<td>urgnaiter no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>arufognot no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>aurfogonotes no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>aurgnames no cross-reference</td>
</tr>
<tr>
<td>VARIANT</td>
<td>urgnaid cross-referenced</td>
</tr>
<tr>
<td>VARIANT</td>
<td>érgnai no cross-reference</td>
</tr>
</tbody>
</table>

2.5.3 Internal orthography

Other aspects of Old-Irish orthography require a degree of specialist knowledge of the part of the student of the language who attempts to locate a form in DIL. To consider some examples: orthographically, in
Old-Irish *t* and *d* are often interchangeable. Thus, the variant éccraide\(^92\) is recorded under the lemma éccraite; the variant éccraides\(^93\) is recorded under the lemma éccraites; and the variant nemhéilnidhe\(^94\) is recorded under the lemma éilnithe. Furthermore, -mb- is often interchangeable orthographically with -mm- and thus the variant fembur\(^95\) is recorded under the lemma fem(m)ar. In the letter *P*, which has a high instance of loan words the influence of Latin orthography can also be seen in the variant forms. For example, the variant forms prouens\(^96\) and provinsi\(^97\) are recorded under the lemma proibinse and the variant prouirbecha\(^98\) is recorded under the lemma proibeirb. The letters *p* and *b* are often interchangeable orthographically in the letter *P* and thus the variant bobba\(^99\) is recorded under the lemma popa and the variant puball\(^100\) recorded under the lemma pupall.

### 2.5.4 Lemmata of complex verbs

In the case of the more difficult verbs, issues of orthography and morphology mean that locating the lemma for a verbal form that is

\(^{92}\) *DIL*, 22.53.

\(^{93}\) *ibid.* 23.6.

\(^{94}\) *ibid.* 82.50.

\(^{95}\) *ibid.* 75.33.

\(^{96}\) *ibid.* 208.82.

\(^{97}\) *ibid.* 205.83.

\(^{98}\) *ibid.* 205.76.

\(^{99}\) *ibid.* 195.60.

\(^{100}\) *ibid.* 210.29.
2. The history and limitations of the *Dictionary of the Irish Language*

encountered in works of Old-Irish and Middle-Irish literature can be an extremely difficult and time-consuming undertaking. In practice, students often fail to find the objects of their searches.

Example 1 *An excerpt from the saga Fingal Rónáin*

```
<TEXT TYPE="poem">
<BODY>
<LG N="1" TYPE="quatrains">
<lb n="129"><L>Is úar fri cloí ngaíthe</L>
<lb n="130"><L>do neoch in-gair <PN>Bú Aife</PN>.</L>
<lb n="131"><L>Iss ed ingaire mada, or si-se</L>
<lb n="132"><L>cen bú, cen nech no chara.</L>
</LG>
</BODY>
</TEXT>

Example 1 is an excerpt from the electronic edition of the saga *Fingal Rónáin*,\(^\text{101}\) which based on internal evidence took its shape towards the end of the Old-Irish period.\(^\text{102}\) In the phrase *sech ní ranac-sa, ní thucais-siu na bú lat*, the verb *ní thucais-siu* is the second singular conjunct form of the perfective verb *do-ucc* with an attached emphasising suffix. In

---


\(^{102}\) ibid. 2.
2. The history and limitations of the *Dictionary of the Irish Language*

*DIL* the verb is, of course, recorded under the third singular present indicative form *do-beir*. There is, however, no cross-reference to *do-beir* from *do-ucci*.\(^{103}\) Thus, a student who encounters the form *ni thucais-siu* in a text and does not know, independently of *DIL*, the present indicative form is unlikely to ever find it in *DIL*.

Further difficulties can also be illustrated when one attempts to use *DIL* in conjunction with a piece of literature. For example, the editors, for the most part—though not exclusively—observe the rule of *caol le caol agus leathan le leathan* in their choice of lemma, but this can be at variance with the orthographic practices of early Irish literature. In the saga *Cath Cairnd Chonaill*, *méit* is listed in *DIL* under the lemma *méid* with *méit* as a secondary lemma.

While the lack of cross-referencing illustrated above is due to the lack of a clear editorial policy in *DIL*, the problem of having to know a third person present indicative form of a verb in order to locate it in *DIL* is a limitation of the hard copy medium itself. As will be illustrated below, the electronic Lexicon is a much more satisfactory medium to present such complex works as a dictionary or lexicon. Indeed it will be illustrated in Chapter 5, that the electronic format remedies most of the limitations that are inherent in hard-copy format of *DIL* as well as many of those caused by the lack of proper editorial procedures in *DIL*.

\(^{103}\) There is a cross-reference to *do-beir* under the first person singular present indicative form *do-ucci*, again illustrating the lack of consistency in *DIL*. 

87
2.6 Inconsistent presentation and treatment of information in *DIL*

2.6.1 Abbreviations

The use of abbreviations throughout *DIL* is very inconsistent. For example, the imperfect tense can be marked by any one of four abbreviations: *Impf, impf., imperf., ipf.;* and the Indicative mood can be referred to by any of the following variations: *Indic., indic., Ind., ind.*. A word from the Welsh language can be introduced with either one of ‘*Welsh, We., WE., and W.*’ and a disyllabic value noted with either *disyll* or *dissyll.* The abbreviations of works cited in *DIL* are also inconsistent, and the implications of this are treated of in detail in § 5.3.3.

These errors and inconsistencies in the use of abbreviations are not so grave as to affect adversely the work in a major way, but they are nonetheless very off-putting for the careful reader and the student of early Irish.

2.6.2 Structure

A further limitation of *DIL* is the lack of an ordered and repeatable structure from one entry to another. This is especially marked in the treatment of the verbs. In the verbs *ar-fócair, ar-fogni, ar-foichlea,* the verbal noun is recorded at the end of the entry and on a separate line. In the verb *ar-fóim,* however, the verbal noun is recorded at the end of the first paragraph, but it is not given a new line of text and so is not
easily distinguishable from the rest of the verbal entry. The same can be observed in the verb *do-roich* where the verbal noun is recorded at the end of the ninth paragraph and in the case of the verb *fo-cíallathar* where the verbal noun is recorded at the end of the first paragraph.

The lack of an ordered structure is particularly unsatisfactory in the case of complex verbs. Lexical entries for verbs such as *téit* span seven pages and if users of the dictionary are seeking only one piece of information, such as a verbal noun form, they are nonetheless forced to read through the entire lexical entry in order to locate the required information. Thus, it is not possible to use the hard-copy version of *DIL* as a quick reference guide.

### 2.6.3 Words treated of in incorrect lexical entries

Cases can also be illustrated in *DIL* where words are not given separate lexical entries, but are described under another lemma. This is especially evident in the case of the personal pronoun. There is no separate lexical entry for the first person plural personal pronoun *sní*: it is described in the entry for the first person singular form *mé*. The same can be observed in the case of the second person plural form *sí*, which is described under the second person singular *tú*.104 By way of contrast, the same system is not adopted for the possessive pronouns: the first person singular form *mo* and the first person plural form *ar* are treated of in separate lexical entries.

---

104 A cross-reference is, however, provided for the words *sní* and *sí*.
2.6.4 Subjective information labels in DIL: the terms earlier and later

DIL is a diachronic dictionary and it employs eight terms to describe the evolution of a word: archaic, earlier, later, classical Old Irish, Middle Irish, Early Modern Irish and Modern Irish. The terms Old Irish, Middle Irish, and Modern Irish have a frame of reference that is widely accepted by scholars in the discipline.\textsuperscript{105} ‘The language of the earliest sources is called Old Irish, that from about A.D. 900 Middle Irish, and that from the beginning of the seventeenth century Modern Irish’.\textsuperscript{106}

Within the context of DIL, however, the terms earlier and later cannot be defined in any objective sense. DIL is a dictionary of Old and Middle-Irish and, to a lesser extent, of Early Modern Irish, and so the lemmata are mostly from the Old and Middle-Irish periods. Because the terms are used in reference to a particular lemma, and the lemmata can refer to the Old and Middle-Irish (and to a lesser extent Early Modern Irish) periods, the definition of these terms may change in any lexical entry. Thus, the terms earlier and later do not refer to a chronological scheme in the Dictionary as a whole, but can only be defined in reference to the lemma in which they are used. This limitation reveals one of the conceptual differences that exist between a hard-copy edition and a digital edition and will be explored to greater detail later in Chapter 5.


\textsuperscript{106} Thurneysen, GOI, 1.
2.7 Categorisation of limitations

The limitations that have been discussed in this section can be divided into two categories: limitations that result from a poor editorial policy, and limitations that are contingent upon the hard copy medium of DIL.

To consider, firstly, limitations that are the result of a poor editorial policy: sections § 2.5.1 and § 2.5.2 illustrate that cross-referencing in DIL is at best haphazard, and at worst non-existent. Section § 2.5.3 illustrates that in order to navigate DIL successfully, users require specialised knowledge of the orthography of early Irish. The failure to provide a table setting out a proper description of orthography is an oversight of successive editors of DIL. Further evidence of poor editorial policy is presented in sections § 2.6.1, § 2.6.3, § 2.6.2.

Section § 2.5.4 (lemmata of complex verbs), can be categorised as a limitation that is contingent upon the hard-copy medium of DIL. Because it would be unreasonable to expect lexicographers to provide a cross-reference to every form of every word in a language, an appropriate lemma must be chosen in a hard copy format. Thus, regardless of the amount of information that exists for a particular word, the reader can access it through one avenue only. In the case of a complex verb, the reader may not know the lemmatic form and in this way, much of the valuable information contained in DIL remains hidden from non-specialist users. In Chapter 5, the suitability of the eXtensible mark-up language to supersede the hard-copy format and its reliance upon the lemma to access information will be assessed.
2. The history and limitations of the Dictionary of the Irish Language

The final limitation, discussed in section § 2.6.4, illustrates that DIL conceives of itself not as a single hierarchical unit, but rather as a series lexical entries. This is made clear by the fact that information labels such as earlier and later do not apply to DIL as a whole, but apply only to the lexical entries in which they are used. In addition to the conceptual differences between a hard-copy format and an electronic edition that this brings out, the importance of such issues when creating an electronic edition based on a hard-copy format, rather than an electronic edition in the first instance, will be discussed in Chapter 5.
CHAPTER 3

Mark-up and XML: preliminary theoretical and historical considerations

3.1 Abstract

This chapter begins with a summary of the views of electronic text that have been put forward by various scholars. A more detailed examination of mark-up is then presented, and the different types of mark-up that can be applied to a text are treated of. The Ordered hierarchy of content objects (hereafter OHCO) theory of textuality is then discussed, and the insights into the limitations of this theory that the electronic Lexicon prompted are pointed to. The history of XML is then outlined and the question ‘What is XML?’ answered. Finally, the Text Encoding Initiative is discussed.
3. Mark-up and XML: preliminary theoretical and historical considerations

3.2 Views of electronic text

Electronic text has been both censured and lauded from many perspectives. Robinson has argued that despite the ‘… articles hymning the glories of electronic editions as we then imagined them’¹ in terms of scholarly editions of texts, the hard-copy print edition still dominates. Nonetheless, an astonishing event at a project in the US patent office, whose remit it was to produce digital editions of eighteenth-century and nineteenth-century patent documents, reveals the fervour with which the process of digitisation has been embraced by some. When a member of the public recovered four original x applications of some of Thomas Edison’s patents from a skip outside the patent office, it emerged that the original documents had being digitised, and then disposed of. ²

Varying interpretations of electronic text are reflected in scholarship, and Sutherland in her introduction to *Electronic Text: Investigations in Method and Theory* has stated ‘The challenge posed by electronic technology to the technology of the printed book, which has dominated Western textual practices since the end of the fifteenth century, is voiced by both enthusiasts and critics alike in extravagant, even apocalyptic terms’.³

Theories about the perceived impact of electronic text are expressed

---


3. Mark-up and XML: preliminary theoretical and historical considerations

in general and specific terms. The more general arguments treat of the impact of electronic text on society and on the printed book. MacArthur has argued that society is fragmenting into those who understand technology and use it in their daily lives, and those who are frightened and suspicious of it. He compares the impact of digital technology to the impact of the printing press, telephone, handgun and radio and states that it is ‘one of history’s great equalisers’. Critics such as Landow and Lanham have argued that electronic technology is a positive development that will make a greater number of texts more accessible to a greater number of people. Steiner and Birkerts view it as a harbinger of chaos and predict that our current society—where the book is in a central position—will be destroyed. Birkert imagines a world in which knowledge is no longer learned, built upon and exchanged, but is merely refashioned and manipulated in the many ways that computers make possible. Sutherland has characterised the opposing views of critics such as Landow and Lanham, and Steiner and Birkerts, and noted that, while they may be opposed in the details of their arguments, they nonetheless occupy ideologically similar territories. Both sets of arguments are moralistic and political, based on interpretations of good

7 George Steiner, Real presences: is there anything in what we say? (London 1989).
versus bad, and are not amenable to empirical analysis.\(^9\)

More specific criticisms have also been levelled against electronic text, and such arguments are frequently characterised by the way they try to shape the evolution of electronic text to serve a particular purpose. Sperberg-McQueen has analysed the standards that should be achieved in electronic scholarly editions and, in an unfavourable comparison of the text collection Project Gutenberg with the types of digital texts that are prepared in accordance with the Text Encoding Initiative, he has stated that ‘we must reject out of hand proposals to create electronic scholarly editions in the style of Project Gutenberg, which objects in principle to the provision of apparatus, and almost never indicates the sources, let alone the principles which have governed the transcription of its texts’.\(^10\)

In a similar vein, Schreibman reviewed sites publishing the works of Oscar Wilde (encoded in HTML or SGML) and identified collections such as Bibliomania, which publishes 30 of Wilde’s works but provides absolutely no bibliographical or editing details.\(^11\)

Another perspective on this issue has been expounded by Sutherland. She has noted that digital text can contain information such as hyper-links to other documents, appendixes, source materials and bibliographical material. Because it is often not possible to include such documents in traditional hard-copy editions, the electronic edition forces us to re-


evaluate our concept of the boundaries of electronic text and ask ‘do texts have a beginning?’ Such a view has been echoed by Jerome McGann who states that a hyper-text need never be ‘complete’.

While Schillingsburg has argued that features of electronic texts such as hyper-links have ‘… extended the textual world; … it has added dimensions and ease of mobility to our concepts of textuality’ Tanselle has argued that this point is overblown and while he concedes that the ease of mobility in an electronic text is greater than in a hard-copy format, he finds the suggestion that the readers ‘concepts of textuality’ are extended in an electronic text preposterous.

Other scholars have argued that electronic texts are still constrained by the limitations of the printed book. The observation that the encoding applied to an electronic text is not held in a file separate from that of the base text, but is usually applied directly to the document file, and that the resultant electronic text, published on the web or on CD-ROM, is usually a finished version, has prompted Eggert to state that the process is ‘thus representing in an essential way the paradigm of the printed book’.

With this in mind, as well as the lack of technical support for establishing

---

12 Sutherland, Introduction, 9.
authenticity of e-documents, *Just in Time Mark-up* was developed. The designers of the system claim that it heralds a ‘second-stage expression’ of humanities computing.\(^{17}\) The system consists of a base transcription, and a number of external tag sets that can be applied to the base transcription file. The system then authenticates the content of the base file before and after the external tag set has been applied, thus establishing authenticity of the base file and allowing end-users to engage with the text as a continually evolving work rather than an end product. Eggert visualises that:

> The reader will be able to accept the guidance of the editor whose reading text and collations can be treated, if the reader so chooses, as the advisable pathway through the work-site. Provided the various states of the text are available in transcription and facsimile, the reader will be at liberty to seek untrodden paths—to reject the editor’s advice—and, if the reader so chooses, to leave a record of a different journey.\(^{18}\)

### 3.3 What is mark-up?

Mark-up has been defined in a pithy statement by the Text Encoding Initiative (hereafter TEI) ‘as any means of making explicit an interpretation of a text’.\(^{19}\) Allen Renear has described the process of marking up a digital text as ‘storing, in the computer’s memory, codes that represent the linguistic content (typically alphabetic characters and punctuation)

---

\(^{17}\) ibid. 433.

\(^{18}\) ibid.

3. Mark-up and XML: preliminary theoretical and historical considerations

and additional information related to this content, such as intended or observed formatting or layout effects and explicit identification of sections of text as being footnotes, titles... . A text-encoding system is the system of codes which effect such a representation'.20

The concept of mark-up is not a new or modern one, and it has been employed in one form or another since texts were first committed to hard-copy representations. Indeed, McGann has argued that ‘There is no such thing as an unmarked text, and the mark-up system laid upon documents to facilitate computerised analyses are marking orders laid upon already marked up material. (Thus all texts implicitly record a cultural history of the artifactuality).’21

In the earliest western manuscripts mark-up was applied to texts in the form of different coloured inks: brick-red ink was frequently used to distinguish or emphasise notable portions of text and in the early middle ages the main text of a manuscript was sometimes written in red ink with the accompanying commentary clearly delineated in black.22 Daly and Daly have described presentational features in a thirteenth-century text of the *Elementarium doctrine erudimentum* of Papias. They note that first letter alphabetisation is indicated by a large decorated initial that


stands at the front of each section and smaller capitals indicate that alphabetisation has been extended as far as the second or third letter in a word.\textsuperscript{23}

Such presentational features clearly function as mark-up. Scribes also used various kinds of mark-up to paginate the quires of a codex: such as the letters in a name, a number of points or in the case of the Italian scribe of Vat. Reg. Lat. 1308 a series of Arabic numerals and their names.\textsuperscript{24} In the medieval period, a form of mark-up was also applied to metrical forms—if they were still recognised as such by the scribe. Bischoff, for example, quotes the example of Otfrid and the anonymous author of the \textit{Ludwigslied} who indicated metrical forms by writing the verses on separate lines.\textsuperscript{25}

Scribes also used a form of presentational mark-up in Irish manuscripts.

---

\textsuperscript{23} Lloyd W. Daly and Bernardine A. Daly, ‘Some techniques in mediaeval Latin lexicography’, \textit{Speculum} 39 no. 2 234.

\textsuperscript{24} Bischoff, \textit{Latin paleography}, 23.

\textsuperscript{25} ibid. 29.
For example, in the lower right hand corner of Dublin RIA 1229 (olim 23 E 25 al. *Lebor na hUidre*), 20, presentational mark-up is clearly visible, in the form of a decorated initial that clearly stands out from the rest of the text and introduces a new section.

Another example of this practice is seen on page 6 of the same manuscript and is illustrated in example 2.
3. Mark-up and XML: preliminary theoretical and historical considerations

Example 2 *Decorated initial in Dublin RIA 1229, 6*
Presentational mark-up is also clear in Dublin TCD 1321, 48, where the scribe has indented a section of text to allow the reader to readily locate it and to differentiate it from the rest of the text.

Example 3 Presentational mark-up in Dublin TCD MS 1321, 48

In the next example from Dublin RIA, 1219 (olim Ciii 1) f. 73r. the *Annals of Connacht*, the scribe has indented different sections of text in order to draw the reader’s attention to new sections.
With regard to more modern literary settings, the following quotation from one of the fathers of SGML emphasises that the practice of applying mark-up to a text was as habitual in the twentieth century as it was in earlier times:

In 1966 I was an attorney practising in Boston, MA, two years out of Harvard Law school. I knew nothing about computers, but I knew there had to be a better way to produce documents than dictating them, reviewing a draft, marking up the draft with corrections, reviewing the retyped draft, and then, in frustration, seeing that the typist had introduced more errors while making the corrections.\textsuperscript{26}

3.4 Types of mark-up that can be applied to electronic documents

3.4.1 Mark-up applied to electronic texts

With regard to electronic texts, three main divisions of mark-up have been identified and expounded upon by theorists: presentational, procedural and descriptive.\(^{27}\)

3.4.2 Presentational and Procedural mark-up

Presentational mark-up is applied to a text to describe its appearance\(^ {28}\) and can be used to indicate, for example, white space and font changes.\(^ {29}\)

In many text processing systems, for example, LaTeX\(^ {30}\) and word processing systems such as OpenOffice, presentational mark-up is frequently indicated through the use of procedural mark-up.\(^ {31}\) Procedural mark-up usually comprises a set of instructions given to a piece of software to spe-


\(^{29}\) Renear, Three (Meta)Theories, 113.

\(^{30}\) Cournane has noted that LaTeX uses both procedural mark-up and declarative mark-up. Declarative mark-up specifies structural aspects of a text without giving any details about how that mark-up should be processed. Mavis Cournane, The application of SGML/TEI to the processing of complex multilingual historical texts, unpublished PhD diss. University College Cork (1997) 26.

\(^{31}\) Coombs et al. Mark-up systems, 4.
cify how a portion of a document should be processed. From an academic encoding perspective, procedural mark-up is frequently unsatisfactory: it does not describe the semantic content of a document, and consequently, such a document cannot easily be searched for specific information, for example, all personal names.\textsuperscript{32} Users who encode documents with procedural mark-up in one editor will frequently encounter considerable difficulties if they try to edit their document in another. Furthermore, if a user decides to change the style of a document it is likely that the entire document will have to be marked up again.\textsuperscript{33}

\textbf{3.4.3 Descriptive mark-up}

Descriptive mark-up is applied to a document to indicate what each part of a document is, for example, a lemma, a quotation or a sobriquet. Charles Goldfarb has stated in regard to descriptive mark-up that ‘the mark-up process stops at the first step: the user locates each significant element of the document and marks it with the mnemonic name … that he feels best characterised it’.\textsuperscript{34} The process can be described as stopping at the first step because all further processing that is applied to a document such as formatting, is specified in another document by naming each part of a file that has been marked up and indicating how it should be processed.

\textsuperscript{32} Steven J. DeRose and David G. Durand and Eli Mylonas and Allen H. Renear, ‘What is text, really?’, \textit{Journal of Computing in Higher Education}, 1 no. 2 (1990) 7.

\textsuperscript{33} Charles G. Goldfarb, \textit{Annex A of ISO 8879, the SGML international standard Introduction to generalised markup}, (Geneva 1986) 2.

\textsuperscript{34} ibid. 3.
3. Mark-up and XML: preliminary theoretical and historical considerations

Descriptive mark-up has many advantages over presentational and procedural mark-up: if applied correctly, in most cases it ensures that documents are searchable, portable, usually vendor independent, easily transformed into other formats, and support multiple views of data without ever changing the master file.\textsuperscript{35}

3.5 Theories of textuality

In an article that was to have important implications for the theory of text and the applications of declarative mark-up,\textsuperscript{36} the ontological units of text were identified and described as ‘content objects’.\textsuperscript{37} Indeed, Caton has stated that this observation ‘underpins all the major mark-up schemes for written text like the TEI’s Guidelines for Text Encoding and Interchange, the Davenport Group’s DocBook and the ISO 12083 DTD (document type definition)’.\textsuperscript{38}

This categorization was based on the observation that when a document is formatted in two different ways, despite the superficial differences that exist between each version, the essence of the document remains unchanged. On the other hand, if a document is prepared with the first

\textsuperscript{35} See especially: Renear Three (Meta)Theories.

\textsuperscript{36} Sperberg-McQueen et al. Guidelines, 21. One such application is the TEI, who seem to obliquely acknowledge the influence this theory has exerted on the TEI guidelines when it is acknowledged that the ‘grossly simplified’ view presented in the What is text, really? article ‘turns out to be very effective for a large number of purposes’

\textsuperscript{37} DeRose et al. ‘What is text?’, 4.

\textsuperscript{38} Paul Caton, ‘Mark-up’s current imbalance’, Markup Languages: Theory & Practice 3 no. 1 Winter (2001) 5.
line of every paragraph omitted, a different document is created in the process. The key factor that resulted in the interpretation of one version of a document being substantially different from another—after it had been reproduced in another form or format—was expressed as ‘largely one of form versus content’.\textsuperscript{39}

Thus, the parts of a document that were essential to a text and imparted the same meaning despite being, for example, formatted in numerous different ways, were described as ‘content objects’. Examples of content objects include sentences, paragraphs, and interrogative phrases. It was furthermore observed that content objects were discrete units of text that did not overlap with one another and text was therefore described as an OHCO or an ‘ordered hierarchy of content objects’.\textsuperscript{40}

This theory was modified later, when it became clear that overlapping hierarchies did, in fact, occur in texts.\textsuperscript{41}

The revision of the OHCO theory was called OHCO-2 and the authors acknowledged that overlapping hierarchies did occur within texts, but stated that ‘text encoders dealing with overlapping objects found that although objects from different analytical perspectives would overlap with

\textsuperscript{39} DeRose et al. ‘What is text?’, 3.
\textsuperscript{40} ibid. 4.
\textsuperscript{41} Renear has classified and discussed three theories of textuality that emerged among the members of the text encoding community in response to content-based text processing: the first of these is the Platonic view which states that text is an ordered hierarchy of content objects. The second theory, which he describes as Pluralism states that texts can contain a multiplicity of structures which are not necessarily hierarchical. Finally, Antirealism views texts not as objective entities, but as the products of the subjective theories and analytical processes we apply to them when inter alia, we transcribe, normalise, edit, annotate and encode them. Renear, Three (Meta)Theories, 117-124.
3. Mark-up and XML: preliminary theoretical and historical considerations

each other, pairs of objects from within a single analytical perspective seemed never to overlap’.42

Counter examples to this argument, such as the case of enjambment in verse, were raised, and the authors were again forced to revise their theory, which was now described as OHCO-3. This revision stated that content objects may overlap, but when they do they can be viewed as sub-hierarchies of the same hierarchy.43

However, further counter examples such as objects that can overlap with themselves (for example, variant readings) and non-hierarchical objects, such as lists broken across paragraphs, brought the acknowledgement from the authors that not all texts could be described in terms of ordered hierarchies.

The theory of OHCO has come under further scrutiny in the intervening years since the publication of What is text, really? Huitfeldt, due to his perception of the limitations of the theory has developed an alternative encoding system for the Wittgenstein archives: the Multi-Element Code System (MECS) in which any element may overlap another.44 McGann has argued that the question ‘What is text?’ will not be answered by focusing exclusively on establishing the ontological units of text and he argues that the ‘performative’ intent of imaginative works of literature and poetry should also be considered.45


43 ibid. 10.

Caton has criticised **OHCO** theory on two fronts: he calls for a reappraisal of the form versus content distinction by pointing out that we experience the written text as a visual object before we read it and determine its message.\(^4^6\) He furthermore illustrates that the **OHCO** model of text allows us to remain at one remove from a text, because it focuses on encoding the atomic units of text such as paragraphs or chapters, but completely neglects the communicative intent of a document.

It is the content *object* that is important for **OHCO**-1 encoding, not the content itself. Encoders interest themselves in content only so far as to identify objects pertaining to their particular scheme, and there is otherwise no need for encoders to care what the content actually says, only to transcribe it accurately.\(^4^7\)

An issue that complements Caton’s field of inquiry was highlighted in the course encoding the Lexicon and will be discussed in § 5.2.1.

### 3.6 The History of XML

XML was created because a mark-up language was required that could supersede HTML in the encoding of documents, and also in terms of the standards that could be enforced when a document was processed and served on the World Wide Web. Flynn\(^4^8\) has identified factors that can be interpreted as contributing to or fostering the need for XML. The texts that were served on the World Wide Web in its early years were

---

\(^{4^5}\) Jerome McGann, *Radiant textuality*, 206; 193-207.

\(^{4^6}\) Caton, *Mark-up’s imbalance*, 7.

\(^{4^7}\) ibid. 8.

\(^{4^8}\) Peter Flynn, ‘Is there life beyond the web?’, *Literary and Linguistic computing* 17 no. 1 April (2002) 49-60.
3. Mark-up and XML: preliminary theoretical and historical considerations

marked up in a simple and non-rigorous way. HTML did not achieve full ISO status until the year 2000 and the large number of browsers that were released in the early years of the web mirrored this somewhat haphazard, standard-less, situation.49

This situation persisted in many second-generation browsers because—despite what would appear to have been the best philanthropic intentions—CERN made the source code of their browsers freely available. Thus, ‘the major component of a browser unfamiliar to programmers (the parsing of arbitrary text for tags in angle brackets) had already been written for them’.50 The result was that many second generation browsers displayed an erroneous handling of certain types of mark-up.

Two further factors, attributed to economic considerations by the developers and programmers who were designing the browsers, exacerbated this situation. Firstly, many of the HTML documents that were in circulation were invalid in terms of the requirements of the HTML standard, because many of the authors did not completely grasp what was required by the standard. This situation was legitimised when browsers were developed that accepted documents with invalid mark-up or even

49 In March 1991 the first line mode browser was released to a limited usership, by 1993 an update of this first browser had been released and two new browsers had been developed including Marc Andree’s Mosaic for X. For a brief, chronological treatment of the developments that led to the World Wide Web from the device called Memex, proposed by Vannever Bush in 1945, which he visualised as being able to make and follow links between documents on microfilm, to a treatment of the development of the Web up to the founding of the Web society in 1995 see Vannevar Bush, ‘As we may think’, The Atlantic Monthly, July (1945). http://www.press.umich.edu/jep/works/vbush/vbush-all.html; Robert Cailliau A Little History of the World Wide Web (c. 1995). http://www.w3.org/History.html.

50 Flynn, op. cit. 1.
3. Mark-up and XML: preliminary theoretical and historical considerations

mark-up that was not HTML. The second factor was that the graphical browser Netscape began inventing its own mark-up independently of HTML. The result was that by 1996 HTML had become an ‘often abused barrier to communication’ and a remedy to this situation was sought.

While in theory SGML was the perfect solution, in practice this was not the case. People found it difficult to learn, and many of the processors of the time were not sophisticated enough to process it at a reasonable speed.

At a conference in August 1996, held in Seattle, industry experts, led by Jon Bosak of Sun Microsystems held discussions about:

1. Classes of software applications for which HTML was an inadequate information format
2. Aspects of the SGML standard itself that impeded SGML’s

---

51 ibid. 53.
52 ibid.
53 ibid. 49.
54 There is an interesting parallel in XML and the XML Binary Characterisation Working Group has been set up to gather ‘information about uses cases where the overhead of generating, parsing, transmitting, storing, or accessing XML-based data may be deemed too great for a particular application, characterizing the properties that XML provides as well as those that are required by the use cases, and establishing objective, shared measurements to help judge whether XML 1.x [sic] and alternate (binary) encodings provide the required properties.’ http://www.w3.org/XML/Binary/\#Introduction. See the following reports from the XML Binary Characterization Working Group: XML Binary Characterization Use Cases published on-line at http://www.w3.org/TR/2005/NOTE-xbc-use-cases-20050331/; XML Binary Characterization Properties published on-line at http://www.w3.org/TR/2005/NOTE-xbc-properties-20050331/; XML Binary Characterization Measurement Methodologies published on-line at: http://www.w3.org/TR/2005/NOTE-xbc-characterization-20050331/ and also Kendall Grant Clark Binary XML, again, published online at http://www.xml.com/pub/a/2003/08/13/deviant.html, and Leigh Dodds, Intuition and Binary XML, http://www.xml.com/pub/a/2001/04/18/binaryXML.html.
3. Mark-up and XML: preliminary theoretical and historical considerations

acceptance as a widespread information technology.\textsuperscript{55}

Around this time, a group of SGML experts approached the World Wide Web Consortium (hereafter W3C) to propose the creation of a working group dedicated to the effort of putting SGML on the web. The working group was established, but decided that pure SGML was not a feasible technology for the world wide web, and another solution should be sought. On the 10 February, 1998 these efforts came to fruition and the first edition of the \textit{XML 1.0 Specification}, a meta-mark-up language that is essentially a subset of SGML, was published.\textsuperscript{56}

### 3.7 What is XML?

The acronym XML stands for \textit{eXtensible Mark-up Language}. It is a set of guidelines, as outlined in the XML 1.0 Specification\textsuperscript{57} that can be used to identify information both within and about texts. It is essentially a subset of SGML. In terms of mark-up languages, HTML on one hand, and SGML and XML on the other, occupy different conceptual layers. Thus, HTML is described as a mark-up language, while XML and SGML are described as meta-mark-up languages.\textsuperscript{58}


\textsuperscript{56} ibid. 3-5.


\textsuperscript{58} The history of the meta-mark-up language called the Generalized Mark-up Language (GML) from whence SGML was derived has been well documented, see especially: Cournane, op. cit.; Charles F. Goldfarb, \textit{The roots of SGML} (1996); idem.
The mark-up language HTML consists of a fixed corpus of tags that can be applied to a document in order to reflect three different types of information that have been categorised by Flynn as structural, content descriptive and visual.\textsuperscript{59}

Thus, an encoder can use elements to describe the structure of a document, for example, its headings, paragraphs and lists. Content descriptive tags are also available to the encoder, who can use the element \texttt{<STRONG>} to mark information as noteworthy. Formatting tags can also be used to indicate that a portion of text should be rendered in bold type or Times New Roman font, or whatever.

These three divisions have not always been acknowledged in discussions of HTML and it has been described incorrectly by many as a technology that specifies only details pertaining to the appearance of a document. For example, Todd Freter has stated ‘HTML is a presentational technology only’.\textsuperscript{60} Flynn has pointed out that such statements have become ‘popular myths’ and that HTML 2.0 contains only seven elements related to presentation and HTML 3.0 contains only eleven.\textsuperscript{61}

Languages such as SGML and XML are more properly described as meta-mark-up languages.\textsuperscript{62} Thus, XML is not a language in itself, but

\textsuperscript{59}Peter Flynn, \textit{Making more use of markup}, SGML ‘95 Boston, MA. http://imbolc.ucc.ie/~pflynn/articles/moreuse.html.

\textsuperscript{60}Freter, \textit{Mastering information}, 2.

\textsuperscript{61}Flynn, \textit{Making more use}, 4.

\textsuperscript{62}Bosak notes that this is a generalisation and that the SGML layer ‘is not as abstract as a true meta-language like Bachus/Naur Form (BNF) which is used to define programming languages’. Jon Bosak, ‘Media-independent publishing: four myths
rather a set of instructions that can be combined in order to create specialised mark-up languages. A project using XML rather than HTML as a format to encode its primary documents is not limited to the number of tags that are set out in the HTML specification, but rather can create a specialised mark-up language to reflect the unique requirements of that project. All the advantages of declarative mark-up outlined above are also associated with XML. It is vendor and platform independent, using XSLT it can be output in formats such as pdf, HTML, TEI-XML, and XHTML. Formatting and transformations of structure are applied to a document in a separate file and thus the information can be processed in numerous different ways without ever changing the data contained in the master file. Jon Bosak, who chaired the W3C’s XML working group, has stated:

> XML derives from a philosophy that data belongs to its creators and that content providers are best served by a data format that does not bind them to particular script languages, authoring tools, and delivery engines but provides a standardised, vendor-independent, level playing field upon which different authoring and delivery tools may freely compete.\(^{64}\)

### 3.8 The Text Encoding Initiative

One of the most successful scholarly applications of SGML/XML has been developed by the Text Encoding Initiative, established in 1987.

---


3. Mark-up and XML: preliminary theoretical and historical considerations

The TEI has been funded by the Association for Computers in the Humanities (ACH); the Association for Computational Linguistics (ACL), and the Association for Literary and Linguistic Computing (ALLC); as well as the U.S. National Endowment for the Humanities (NEH), the European Community, the Mellon Foundation, and the Social Science and Humanities Research Council of Canada.\textsuperscript{65} Up to 1987 texts created by museums, libraries, publishing houses as well as Universities and individual scholars were being marked up in a variety of computer languages. The goal of the TEI was the creation of an encoding scheme that was stable, international, and interdisciplinary.\textsuperscript{66}

Members of the TEI had clear ideas and theories about the criteria that an electronic scholarly text should meet\textsuperscript{67} and guidelines were written to define the mark-up of, \textit{inter alia}, a text’s physical appearance, content, structure, and bibliographical information.\textsuperscript{68} As the name suggests, the guidelines are not a series of prescriptive statements about how mark-up can be applied to a text. While the notion of TEI conformance has been clearly defined, scholars are able to select the mark-up they require from a number of sets of encoding schemes and it is possible for projects to extend their TEI mark-up if necessary.\textsuperscript{69}

\textsuperscript{65} http://www.tei-c.org/ accessed on 1/10/05.
\textsuperscript{66} Cournane, \textit{The application of SGML}, 44.
\textsuperscript{67} See especially Sperberg-McQueen, \textit{Text in the electronic age: textual study and text encoding, with examples from medieval texts}, Literary and Linguistic Computing 6, (1991) 34-46.; idem. \textit{Textual criticism and the Text Encoding Initiative.}
\textsuperscript{68} ibid. 46.
\textsuperscript{69} See for example Susan Rennie, ‘The Electronic Scottish National Dictionary (eSND): Work in Progress’, Literary and Linguistic Computing 16 no. 2 (2001) 153-160; Gregory
The TEI has been criticised by some scholars, Olsen argues that it simply allows too much variation and flexibility: ‘The editors of the TEI are writing a data interchange format while at the same time working out a mechanism to support theoretically informed encoding specifications for just about any textual object that scholars in a wide variety of disciplines might encounter. Unfortunately, the resulting drafts of the TEI specification(s) reflect this underlying confusion of the task at hand’. McGann, on the other hand has argued that the TEI’s focus and interpretation of text is too narrow. Hockey has used the somewhat perplexing phrase ‘powerful and flexible dinosaur’ to refer to the TEI, and one can only conclude that dinosaur is a reference to the size of the guidelines rather than their obsolescence.

Despite such criticisms, the TEI is a widely accepted international standard and a large number of academic projects adhere to it because of the significant advantages that are associated with using it. In addition to the advantages of conforming to a standard that is unlikely to become obsolete at any time in the near future, projects that encode texts at the most superficial level and projects that encode texts to a very detailed

---

71 Jerome McGann, Radiant textuality, 187-191; 193-207.
73 The projects page at http://www.tei-c.org/Applications/ (accessed on 1/12/05) lists 125 individual projects registered as using the guidelines.
level can both use it, thus enabling the exchange of information and
tools between all kinds of projects. Furthermore, the TEI boasts a large
support community and members are able to draw on existing tools and
documentation to assist in the implementation of the TEI guidelines as
well as subscribe to mailing lists and attend conferences and specialist
seminars.

In § 4.7 section an example of a TEI-conformant Lexicon entry is given.
CHAPTER 4

XML: the foundation of the electronic Lexicon

4.1 Abstract

Chapter four presents examples of ten different types of lexical entry, which have been excerpted from the electronic Lexicon and are accompanied by a brief description. The mark-up used in the electronic Lexicon is then considered in more detail, and the elements and attributes used to mark up the Lexicon are described. The structure and properties of a document type definition are then discussed with reference to ‘lex.dtd’, the dtd of the Lexicon. Finally, diagrams illustrate the inter-relationships of the various units of the DTD with one another.
4. XML: the foundation of the electronic Lexicon

4.2 Examples and description of ten different types of encoded lexical entry

4.2.1 Brief description of ecal

1. The lexical entry for the adjective ecal is illustrated in Example 1

2. This adjective displays a-stem inflexion, and it is an exception to the rule governing the formation of adjectival compounds with a-stem inflexion

3. The reconstructed word *ek(s)gal is cited as the etymology of ecal

4. The adjective is polysemous and is frequently used in the phrase is ecal lim(m) with the conjugated preposition la and the suffixed first person singular personal pronoun

5. ecal is attested in simple forms and it also occurs as a compound adjective

Example 1 A lexical entry for an adjective

```xml
<entry id="4567">
  <lemma>ecal</lemma>
  <gramgrp pos="adj">
    <itype>a-stm, o-stm</itype>
    <note>an exception to the rule that in adjectival compounds a-stems become i-stems</note>
  </gramgrp>
</entry>
```
XML: the foundation of the electronic Lexicon

<gramgrp>
<etym> *ek(s)gal</etym>
<senses>
  <sense n="1.0">fearful, timorous, cautious</sense>
  <sense n="1.1" type="substantive">a coward</sense>
  <sense n="2" type="substantive">something to be feared, a peril, a danger</sense>
  <sense n="3"><note ref="usage">common in phrase</note> I fear, I apprehend</sense>
  <sense n="4"><note ref="usage">with the connotation</note> likelihood, probability</sense>
</senses>
<form gr type="simple">
  <form type="simple">ecal</form>
  <form type="simple">egal</form>
  <form type="simple">ecol</form>
  <form type="simple">ecgal</form>
  <form type="simple">ecil</form>
  <form type="simple">ecla</form>
  <form type="simple">eagal</form>
  <form type="simple">eccal</form>
  <form type="simple">ecail</form>
</form>
4. XML: the foundation of the electronic Lexicon

4.2.2 Brief description of *didiu*

1. The lexical entry for the conjunction *didiu* is given in Example 2

2. the etymology of *didiu* is from *di* + *suide*

3. a number of definitions of *didiu* are presented

4. the simple forms of *didiu* are also listed

Example 2 A lexical entry for a conjunction

```xml
<entry id="30675">
  <lemma>didiu</lemma>
  <gramgrp pos="conj">
    <etym><xref ptr="D1435">di</xref> + shuidiu
    (<xref ptr="S25967">suide</xref>)</etym>
    <senses>
      <sense n="1">now, therefore, then, hence,
```
4.2.3 Brief description of clár

1. The lexical entry for the noun clár is given in Example 10 (p 156)

2. the gender of clár was in its earliest stages neuter, at a later stage
   the neuter gender was replaced by a masculine gender.¹

3. clár is polysemous

¹Thurneysen, GOI, 154. ‘...the neuter was destined to be largely superseded by the
masculine and feminine in the ninth century and to disappear almost completely
by the tenth’.
4. XML: the foundation of the electronic Lexicon

4. it occurs as a simple word and also forms compound nouns and compound adjectives

4.2.4 Brief description of mo

1. The lexical entry for the possessive pronoun is given in Example 12 (p 164)

2. mo occurs in simple forms and is also suffixed to prepositions such as la and co

3. mo² is affixed to personal names to make hypocoristic forms

4.2.5 Brief description of mé

1. The lexical entry for the personal pronoun mé is given in Example 13 (p 167)

2. the simple forms of the personal pronoun are listed and the mutations that occur in words that are used immediately after or near mé are indicated

3. the infixed forms of the personal pronoun are also presented

---


³ For a discussion of the origins and nature of Irish personal names and their encoding in SGML/TEI see Cournane, Application of SGML/TEI, 167-198.
4. XML: the foundation of the electronic Lexicon

4.2.6 Brief description of \textit{prím}

1. The lexical entry for the prefix \textit{prím}- is given in Example 14 (p 173)

2. it is indeclinable

3. the etymology of \textit{prím}- is the Latin word \textit{primus}

4. the senses of the prefix are presented

5. numerous compounds such as compound nouns and compound adjectives that \textit{prím} forms are listed

4.2.7 Brief description of \textit{Gall}

1. The lexical entry for the o-stem masculine noun \textit{Gall} is given in Example 15 (p 182). This entry is given in addition to the lexical entry for the noun \textit{clár} (Example 10), because the information it contains is somewhat different to that of a more standard noun entry, such as \textit{clár}

2. The earliest meaning is “a Gaul”

3. In addition to the simple forms of this noun, it occurs in a number of compounds, and some of these function as personal names and place names.\textsuperscript{4}

\footnote{\textsuperscript{4} ibid. see 195-98 for a discussion encoding of place names with SGML/TEI.}
4. XML: the foundation of the electronic Lexicon

4.2.8 Brief description of cen

1. The lexical entry for the preposition *cen* is given in Example 11 (p 160)

2. The word is polysemous

3. The simple and compound forms of *cen* are presented, as are the forms with infixed pronouns

4.2.9 Brief description of the definite article

1. The lexical entry for the definite article is given in Example 16 (p 188)

2. The etymology of the definite article is shown to be Indo-European

3. The forms of the definite article across the different numbers, cases and genders are presented and the mutations that occur after each form are indicated

4.2.10 Brief description of téit

1. The lexical entry for the verb *téit* is given in Example 17 (p 195)

2. It is polysemous

3. *téit* is an extremely complex verb and all the attested moods, tenses, persons, and numbers forms of the verb are given in the paradigm.
4.2.11 Description of elements and attributes

ELEMENT multilex is the root element and it contains all the entries in the Lexicon. It can be used to encode a unified text that contains numerous texts; see section § 5.2.1.3.

ELEMENT entry indicates the beginning of a lexical entry and contains all the information pertaining to each head-word.

ATTRIBUTE id is attached to the entry element and is the unique identifier for that element (it is automatically generated using XSLT).

ELEMENT lemma is the head-word(s) of each entry.

ATTRIBUTE htype on the lemma indicates that the lemma is a homograph and specifies the number of the homograph, for example, <lemma htype="4">bus</lemma>.

ELEMENT syntax.gr is the container element for all syntactical information pertaining to the lemma and the meta-data that further classifies syntactical information.

ELEMENT syntax contains grammatical information about the way the lemma is used in different types of sentences. The syntax element can make reference to many categories of information, for example, sense information is frequently presented with syntactical information. Consequently the syntax element may contain a number of other elements such as <note>, <form> and <sense>.
4. XML: the foundation of the electronic Lexicon

ATTRIBUTE n on the syntax element indicates the number of each syntax element.

ELEMENT gramgrp is the container element for all the grammatical information pertaining to the lemma, and the meta-data that classifies that information.

ATTRIBUTE pos of the element gramgrp specifies the part of speech of the lemma, for example, an adjective.

ATTRIBUTE gen of the element gramgrp indicates the gender of the part of speech, for example, feminine.

ATTRIBUTE genshift indicates that the gender of a word has changed from, for example, neuter to feminine. It should be noted that the gen and genshift attributes are not mutually exclusive stages in the development of a word. For example the masculine noun pupall is also attested in a feminine form pupaill and the gender of the word is described as ‘capricious, often varying in same text.’

ELEMENT itype is the inflexion type of the lemma, for example, o-stem, in the case of nouns such as bann (o-stem, also a-stem) it is possible for more that one inflexion to be recorded.

ATTRIBUTE resp on the element itype indicates that the inflexional information has been supplied by the editor of the particular fascicle in DIL, or the editor of the electronic edition.

ATTRIBUTE cert combined with the attribute resp on the itype element indicates that the editor has indicated that the information he has

---

5 DIL, 210 Fas. P.
supplied is uncertain.

**ELEMENT** etym labels the etymology of the lemma, if it is known.

**ELEMENT** senses is a container element for all the senses of the lemma in addition to the meta-data that describes the various types of senses and any additional information that may exist.

**ELEMENT** sense is the definition of the lemma.

**ATTRIBUTE** n on the element sense indicates the number of the definition.

**ATTRIBUTE** type on the sense element classifies semantic information, for example as, “figurative” or “botanical”.

**ELEMENT** gloss introduces explanatory information.

**ATTRIBUTE** lang on the gloss element specifies the language that the gloss is recorded in, for example, Latin or Old Norse.

**ELEMENT** forms is the container element for all of the various forms of the lemma, the classification of the various types of forms and any additional information pertaining to them.

**ELEMENT** form.gr is used to group individual categories of forms.

**ELEMENT** type on the form.gr provides further information about that element, for example that the type of form group is a compound noun.

**ELEMENT** form contains the various attested forms of the lemma.
ATTRIBUTE type on the form element specifies the types of forms that are attested, for example, a compound verbal noun.

ATTRIBUTE number on the form element indicates whether it is singular, plural or dual.

ATTRIBUTE case on the form element can be one of nominative, accusative, genitive or dative.

ATTRIBUTE gen on the form element can be masculine, feminine, neuter or common.

ATTRIBUTE person on the form element indicates whether the form is the first, second or third person (and the number element refines this information by specifying either singular or plural). The attribute value 0 is also available on the person element and can be used to encode passive forms, which do not have a number explicitly associated with them.

ATTRIBUTE tense on the form element indicates whether the verb is, for example, in the present or future tense.

ELEMENT cert is used to indicate various degrees of certainty about a piece of information.

ATTRIBUTE level on the element cert is used to indicate the degree of certainty with which the information in question can be interpreted.

ATTRIBUTE resp on the element cert indicates that this information was indicated to be uncertain by the editor of DIL or
4. XML: the foundation of the electronic Lexicon

the electronic edition.

**ELEMENT** later specifies information, usually a form, that has been described as *later* in *DIL*.\(^6\)

**ELEMENT** earlier specifies information, usually a form, that has been described as *earlier* in *DIL*.

**ELEMENT** archaic specifies information, usually a form, that has been described as *archaic* in *DIL*.

**ELEMENT** oldir specifies information, usually a form that has been described as *Old Irish* in *DIL*.

**ELEMENT** midir specifies information, usually a form that has been described as *Middle Irish* in *DIL*.

**ELEMENT** modir specifies information, usually a form that has been described as *Modern Irish* in *DIL*.

**ELEMENT** resp indicates that the information encoded with *resp* has been supplied by the editor of the fascicle (in *DIL*) or other work that the information has been excerpted from. It can also be used to indicate that the information has been supplied by the editor of the electronic edition.

**ATTRIBUTE** inits on the element *resp* allows the initials of the person who provided the information marked with *resp*, to be

\(^6\) The elements *later* and *earlier* are used, but with the caveat discussed in chapter 1 clearly indicated.
specified.

**ELEMENT** *ps* indicates a personal name.

**ATTRIBUTE** *type* on the element *ps* allows the type of personal name in question to be specified such as "comic" or "fictitious".

**ELEMENT** *pn* indicates a place name.

**ATTRIBUTE** *type* on the element *pn* allows the type of place name in question to be specified, for example, "fictional".

**ELEMENT** *frn* indicates word(s) in a foreign language i.e. any language other than Irish, for example, German, French or Greek.

**ATTRIBUTE** *lang* on the element *frn* specifies the type of language that has been specified as foreign.

**ELEMENT** *note* encodes additional or noteworthy information.

**ATTRIBUTE** *ref* allows information to be classified as general in nature, as a phrase, as syntactical, as semantic or to indicate that the information pertains to the way the lemma is used.

**ELEMENT** *cit* is the container element for the elements that describe bibliographical information.

**ELEMENT** *author* specifies the author of a scholarly contribution or primary source.

**ELEMENT** *bibl* specifies the title of a scholarly contribution or primary source.
4. XML: the foundation of the electronic Lexicon

**ELEMENT biblscope** is the pagination range of the element `<bibl>`.

**ELEMENT super** can be used in the `biblscope` element to encode a superscript number.

**ELEMENT sic** indicates the information was *thus* in the hard-copy edition and is not an error that occurred during digitisation.

**ELEMENT mut** is used to indicate that a word causes lenition, nasalisation, or gemination.

**ATTRIBUTE person** is used to specify whether the first, second or third person form causes mutation.

**ATTRIBUTE number** is used to specify whether the person is singular or plural.

**ELEMENT paradigm** is the container element for all the moods of a verb, the classification of the types of moods and any additional information specified.

**ELEMENT mood** indicates that a form is in one of three moods.

**ATTRIBUTE type** is used to specify one of three moods: the Imperative, Indicative or Subjunctive.\(^7\)

**ELEMENT tense** is used to indicate a tense.

**ATTRIBUTE type** of the element `tense` is used to specify, for example, “future”.

\(^7\) Thurneysen, *GOI*, 329.
4. XML: the foundation of the electronic Lexicon

ELEMENT number indicates that the information is a number.

ATTRIBUTE type of the element number specifies whether the form is singular or plural.

ELEMENT person introduces a person.

ATTRIBUTE n specifies either first, second or third.

ATTRIBUTE voice on the person element indicates that the verb is either active or passive.

ELEMENT xref is cross-reference to another word in the Lexicon.

ATTRIBUTE ptr is a pointer to the unique reference number for that form.

ATTRIBUTE doc on the xref element allows an external document to be specified in a link.

4.3 The Document Type Definition

The acronym DTD stands for Document Type Definition, and is described in the XML 1.0 specification as a ‘grammar’.\(^8\) Thus, as the term grammar suggests, the DTD lists the elements, attributes, entities and notations that form an XML document, and specifies the various types of relationships that exist between them.

---

\(^8\) Tim Bray et al., *XML Spec.* §2.8.
4.3.1 The Document Type Declaration

An XML file must consist of a processing instruction, and a root element. While the processing instruction tells the parser which version of XML is being used and specifies the character encoding of a particular document, the root element completely contains all the other elements in a file.

A project may create a number of DTDs in order to encode and describe different document instances. In order to associate a specific DTD with a specific XML document, the Document type declaration is invoked, ‘the XML document type declaration contains or points to mark-up declarations that provide a grammar for a class of documents. This grammar is known as a document type definition, or DTD.’

Example 3 An example of an embedded DTD

```xml
<?xml version="1.0" encoding="iso-8859-1"?>
<!DOCTYPE verse [ 
<!ELEMENT verse (lb+)>
<!ELEMENT lb (#PCDATA)> 
<!ATTLIST lb n CDATA #REQUIRED>
]> 
<verse>
<lb n="1">In esser dam to á?</lb>
<lb n="2">Tó, mani má mo á</lb>
<lb n="3">Ara tairí mo á mó?</lb>
```

9 ibid.
Depending on the requirements of a particular project, an encoder can associate a DTD with a document instance in one of three ways. Example 3 illustrates a document\textsuperscript{10} with an embedded DTD, known as an \textit{internal subset}.

Figure 4.1 (p 135) is an excerpt of the Processing Instruction and document type declaration from the DTD that is associated with the electronic Lexicon of Old Irish. In this case, the DTD: \textit{lex.dtd}, is stored as a separate file from that of the Lexicon and is described as an external subset.

In the case of the Lexicon, the method of storing the DTD as an external subset was adopted for a number of reasons, the primary one being the need for consistency of mark-up across all the files. During the period of compilation of the Lexicon, the DTD was continually refined. As each letter of the alphabet was prepared in a separate file, had the DTD been stored as an internal subset of each of the documents as they were created, it would have created an extra editorial layer in order to ensure that every file was being validated against the same DTD. In addition to this practical decision, other issues such as maximising the contribution of the Lexicon to both the field of humanities computing and Old-Irish scholarship also came into play. By storing the DTD as an external subset of the Lexicon it will be possible to associate it with numerous XML files.

4.4 Commentary on and explanation of lex.dtd

4.4.0.1 Element declaration

A DTD contains a number of declarations and each declaration consists of two parts: a generic identifier and a content model. The generic identifier\(^\text{12}\) gives the name of the element being declared: in this case `<sense>`, as illustrated in Figure 4.3. The content model\(^\text{13}\) defines what is permitted to be contained in that element. In this case the element

\[^{11}\text{Elliott Rusty Harold, XML Bible (USA 1999) 243-245.}\]
\[^{12}\text{Tim Bray et al. XML spec, see §3.}\]
\[^{13}\text{ibid. §3.2.1.}\]
4. XML: the foundation of the electronic Lexicon

Figure 4.2: List of symbols and terms used in the DTD

```
<!-- begin statement -->
--> end statement
+ indicates one or more occurrence of an element
* indicates zero or more occurrences of an element
? indicates zero or one occurrence of the element
| union operator facilitates 'or'
(... ) groups
#FIXED the specified value cannot be changed
#REQUIRED the attribute cannot be omitted
#IMPLIED if a value is not supplied it will be
inserted by the parser
#PCDATA Parsed Character Data (text only)
CDATA
```

Figure 4.3: Element declaration

```
<!ELEMENT senses (sense+)>
<!--generic identifier--> <!--content model-->
```

<senses> is not permitted to contain any other elements such as <lemma>
and it may only contain one or more <sense> elements.

4.4.0.2 Content model

Figure 4.4 illustrates a slightly more complex content model, and is an
example of a model group. The declaration specifies the elements that
are permitted to occur in the <gramgrp> element. It furthermore describes
the relationships that are permitted between the elements, by harnessing
connectors such as the union operator | and the occurrence indicator +.

In this example, the model group permits the <gramgrp> element to contain a resp element or a group that consists of an <itype> element, followed by, at most, one occurrence of a <note> element and specifies that this group may occur one or more times.\(^{14}\)

### 4.4.0.3 Declaring an attribute

An element can have numerous attributes, as illustrated by Figure 4.5. In this case, the element gramgrp has six different attributes, all of which must be defined in the same declaration. The syntax for declaring an attribute is very similar to that of declaring an element, one of the obvious differences between the two types of declarations is the use of the term !ATTLIST for attributes.

After specifying the name of the attribute the next necessary piece of information pertains to the type of attribute. In the case of genshift it is specified that the value can be either masculine or feminine. In the case of the genresp attribute, where the likely gender has been supplied by an editor, the initials of his/her name are not specified: but the keyword

\(^{14}\) If \#PCDATA\ were included in the content model content model XML states that it must be listed first.
4. XML: the foundation of the electronic Lexicon

Figure 4.5: Attribute declaration

```xml
<!ELEMENT gramgrp (resp| (itype, note?) + )>
<!ATTLIST gramgrp pos
(adj|adv|art|causal-conj|conj|indep.neg.part|
interj|nn|part|poss.adj|prefix|prep|pronom|pron|
suffix|vb|vn|unclear) #REQUIRED
  gen (m|f|n|c) #IMPLIED
genresp CDATA #IMPLIED
cert CDATA #IMPLIED
genshift (m|f) #IMPLIED
genshiftresp CDATA #IMPLIED>
<!--name--> <!--type--> <!--default value-->  
```

CDATA states that this attribute must contain character data i.e. [A-Za-z], white space and punctuation.

Finally the default values for the attributes are specified, the #REQUIRED in the case of the attribute pos indicates that one of the list of types such as adj, adv etc must be included in the attribute.
4.4.0.4 Internal parameter entity

An internal parameter entity is illustrated in Figure 4.6. XML Entities can be thought of as a ‘storage unit that contains particular parts of an XML document…An entity may consist of a file, a database record or any other item that contains data’.\textsuperscript{15}

While different types of entities can be used in different ways in XML files, an internal parameter entity can only be used in a DTD, and it must be declared in a DTD before it is used. Figure 4.6 illustrates that such an entity begins with a percentage sign, followed by a name, and in this case the name of the entity is ‘type’.

The fact that the compilation of lex.dtd was an ongoing process has already been indicated. The use of parameter entities was one of the factors that made that DTD easier to maintain. The attribute type is used on many elements in the DTD, Figure 4.7 illustrates that it is used with form, but it can also be used with the elements <ps>, <pn>, <syntax>, <form.gr>, <form> and <sense>. By storing the possible values of the type attribute as an entity, it was possible to emend or add to the list of values in one place, and ensure that same values applied in every part of the DTD, thus ensuring the required standard of consistency.

\textsuperscript{15} Harold, *XML Bible*, 247.
4. XML: the foundation of the electronic Lexicon

Figure 4.6: Internal parameter entity

```xml
<!ENTITY % type "absolute | absolutegen | adverbial | attributive-genitive | comic | comparative | cpd | cpdadj | cpdnn | cpdvb | cpdvbnn | cpdpart | emphaspart | emphpart | figurative | equative | grammatical | hypocoristic | infixed-A | infixed-B | infixed-C | intens-prefix | legal | metaphorical | medical | nominal | participle-necessity | passive | passive-perfective | passive-relative | past-participle-passive | perfective | poss-pron | possessive | prep-art | prepositional | regular | relative | relative | relative-perfective | sobriquet | simple | substantive | suffix-pron | suffixed | superlative | verbal-noun | with-do | with-ro | with-suffix-pron | vocative " />
```

Figure 4.7: Invoking an internal parameter entity

```xml
<!ELEMENT form (#PCDATA|mod|cert|resp|ps|pn|frn|note|later|earlier|archaic|oldir|midir|modir)*>
<!ATTLIST form type (%type;) #IMPLIED
   number (%numtype;) #IMPLIED
   case (%case;) #IMPLIED
   gen (m|f|n|c) #IMPLIED
   person CDATA #IMPLIED >
```
4. XML: the foundation of the electronic Lexicon

Figure 4.8: Character references

```
<!ENTITY amacr "&#x00E1;">  
<!ENTITY emacr "&#x00E9;">  
<!ENTITY imacr "&#x00ED;">  
<!ENTITY omacr "&#x00F3;">  
<!ENTITY umacr "&#x00FA;">  
<!ENTITY auml "&#x00E4;">  
<!ENTITY Auml "&#x00C4;">  
<!ENTITY euml "&#x00EB;">  
<!ENTITY iuml "&#x00EF;">  
<!ENTITY ouml "&#x00F6;">  
<!ENTITY uuml "&#x00FC;">  
<!ENTITY amp "&#x0026;">  
```

4.5 Character references

Figure 4.8 illustrates some of the character references that are used in the DTD. By default, all XML documents are encoded in Unicode, which currently defines more than 40,000 different characters in a range of languages.\(^{16}\)

However, when an XML document is used in a variety of ways in different organisations and across a number of platforms, one of the issues that can arise is the possibility that not all applications support Unicode.\(^{17}\)

The TEI consortium give the example of the character é, that can


\(^{17}\) See especially, Markus Kuhn, *UTF-8 and Unicode FAQ for Unix/Linux*, (2005).  
be represented in an XML document with default encoding as the Unicode character with value OOE9. Representing character references with numeric values ensures that documents can be passed between different applications, that may or may not support Unicode, without resulting in a loss of data. For example, the character é may nor be available in a non-Unicode character set, but loss of data can be prevented by representing the character with its hexadecimal value or character entity reference: &#x00E9; or value OOE9.\textsuperscript{18}

A document with a number of hexadecimal references embedded in it is not very legible\textsuperscript{19} and an encoder may find themselves having to refer continually to the Unicode book in order remind themselves which values represent which character. XML, therefore, defines a set of entities, that can be used to represent characters from the Unicode set. In Figure 4.8 above, for example, the mnemonic Auml is mapped to its hexadecimal reference, and the entity can then be used in the document in the form of &Auml; to represent, Ä, a capital A with an umlaut, hexadecimal value &#x00C4;.


\textsuperscript{19} Legibility is a stated goal of XML, 'XML documents should be human-legible and reasonably clear', Tim Bray et al. \textit{XML 1.0}. §1.1.
4. XML: the foundation of the electronic Lexicon

4.6 Interrelationship of elements and attributes in lex.dtd

For the purposes of these diagrams, the <lexicon> root element is the same as the <multilex> root element, which will be discussed below.

Example 4 Graphical view of the lemma in lex.dtd
Example 5 *Graphical view of the gramgrp in lex.dtd*
Example 6: Graphical view of syntax in lex.dtd

4. XML: the foundation of the electronic Lexicon
Example 7 Graphical view of senses in lex.dtd
Example 8 Graphical view of forms in lex.dtd

4. XML: the foundation of the electronic Lexicon
4. XML: the foundation of the electronic Lexicon

Example 9  Graphical view of paradigm-note in lex.dtd

4.7 XSLT examples

As discussed in Chapter 1, when work on the Lexicon began, TEI P4 had not yet been released. In the light of this, and in order to increase the speed of data entry, the Lexicon was written in XML and it was decided to also achieve TEI-XML conformance at a later stage. § 4.9 illustrates a simple set of XSLT transformations and Figure 4.10 illustrates the TEI-XML file that is created after the XSLT transformations illustrated below have been processed.

The input text for this XSLT transformation is illustrated in Example 1 (p 118).
Figure 4.9

XSLT transformation templates

```xml
<?xml version="1.0" encoding="utf-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform" version="1.0">

  <xsl:output method="xml"/>

  <xsl:template match="/">
    <xsl:apply-templates/>
  </xsl:template>

  <xsl:template match="lemma">
    <form>
      <orth>
        <xsl:value-of select="."/>
      </orth>
    </form>
  </xsl:template>

  <xsl:template match="gramgrp">
    <gramGrp>
      <pos>
        <xsl:value-of select="@pos"/>
      </pos>
    </gramGrp>
  </xsl:template>
</xsl:stylesheet>
```
4. XML: the foundation of the electronic Lexicon

</pos>
<xsl:apply-templates/>
</gramGrp>
</xsl:template>

<xsl:template match="text()">
[name(preceding-sibling::note/@ref[.='phrase'][1])]"
<def>
<xsl:value-of select="."/>
</def>
</xsl:template>

<xsl:template match="note/@ref[.='general']">
<xsl:attribute name="type">
  <xsl:text>general</xsl:text>
</xsl:attribute>
</xsl:template>

<xsl:template match="note/@ref[.='phrase']">
<xsl:attribute name="type">
  <xsl:text>phrase</xsl:text>
</xsl:attribute>
</xsl:template>
4. XML: the foundation of the electronic Lexicon

```
<xsl:template match="note/@ref[.='usage']">
  <xsl:attribute name="type">
    <xsl:text>usage</xsl:text>
  </xsl:attribute>
</xsl:template>

<xsl:template match="forms">
  <form>
    <xsl:apply-templates/>
  </form>
</xsl:template>

<xsl:template match="senses">
  <xsl:apply-templates/>
</xsl:template>

<xsl:template match="form.gr">
  <xsl:apply-templates/>
</xsl:template>

<xsl:template match="@*|node()">
  <xsl:copy>
    <xsl:apply-templates select="@*|node()"/>
  </xsl:copy>
</xsl:template>
```
4. XML: the foundation of the electronic Lexicon

```xml
</xsl:template>

</xsl:stylesheet>
```
Figure 4.10: Output text: TEI conformant mark-up

<entry id="4567">
  <form><orth>ecal</orth></form>
  <gramGrp><pos>adj</pos>
    <itype>a-stm, o-stm</itype>
    <note ref="general">an exception to the rule that in adjectival compounds a-stems become i-stems</note>
  </gramGrp>
  <etym>< *ek(s)gal</etym>
    <sense n="1">fearful, timorous</sense>
    <sense n="2" type="substantive">something to be feared, a danger, a coward</sense>
    <sense n="3"><note ref="usage">common in phrase</note> <note ref="phrase">is ecal limm</note><def>
      I fear, I apprehend</def></sense>
    <sense n="4"><note ref="usage">with the connotation</note> likelihood, probability</sense>
  </form>
  <form type="simple">ecal</form>
  <form type="simple">egal</form>
  <form type="simple">ecol</form>
  <form type="simple">ecgal</form>
  <form type="simple">ecil</form>
  <form type="simple">ecla</form>
  <form type="simple">eagal</form>
  <form type="simple">eccal</form>
  <form type="simple">ecail</form>
  <form type="simple">eagail</form>
  <form type="cpdadj">ecal-mór</form>
</entry>
4.8 Examples of lexical entries

Example 10 A lexical entry for a noun

<entry id="1789">
  <lemma>cíl</lemma>
  <gramgrp pos="nn" gen="n" genshift="m">
    <itype>o-stm</itype>
  </gramgrp>
  <senses>
    <sense n="1">board, plank</sense>
    <sense n="2"><note ref="general">of a variety of specific flat articles normally made of wood</note> breastwork of a chariot, a stave <note ref="general">of a vat or cask</note>, wooden cover, gaming board, table, writing tablet, inscribed tablet</sense>
    <sense n="3"><note>by extension applied to other materials</note> strip, plate <note ref="general">of metal</note> pane <note ref="general">of glass, crystal</note></sense>
    <sense n="4"><note ref="general">used of</note> flat parts of the body</sense>
    <sense n="5">level expanse, plain, surface</sense>
    <sense n="6">table of contents, index</sense>
  </senses>
</entry>
4. XML: the foundation of the electronic Lexicon

</senses>
</forms>

<form.gr type="simple">
    <form type="simple">cláraib</form>
    <form type="simple">clair</form>
    <form type="simple">clara</form>
    <form type="simple">cláraibh</form>
    <form type="simple">cláir</form>
    <form type="simple">claraib</form>
    <form type="simple">claruibh</form>
    <form type="simple">clar</form>
</form.gr>

<form.gr type="cpdnn">
    <form type="cpdnn">cláradba</form>
    <form type="cpdnn">clarbharr</form>
    <form type="cpdnn">clarbla</form>
    <form type="cpdnn">clárbhórd</form>
    <form type="cpdnn">clarcolbadaibh</form>
    <form type="cpdnn">clárchiste</form>
    <form type="cpdnn">cl&auml;rc[h]oil&iacute;r</form>
    <form type="cpdnn">clarcraithe</form>
    <form type="cpdnn">clarchris</form>
    <form type="cpdnn">clárdroicheat</form>
    <form type="cpdnn">clárdhroicheat</form>
</form.gr>
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">clárfiacail</form>
<form type="cpdnn">fíanchlár</form>
<form type="cpdnn">cláirfiacail</form>
<form type="cpdnn">clarfhiacla</form>
<form type="cpdnn">clarland</form>
<form type="cpdnn">clairleibind</form>
<form type="cpdnn">clairleibind</form>
<form type="cpdnn">clarlestraib</form>
<form type="cpdnn">clármachaire</form>
<form type="cpdnn">clármaige</form>
<form type="cpdnn">clarmhuige</form>
<form type="cpdnn">clarmíasa</form>
<form type="cpdnn">clarmruig</form>
<form type="cpdnn">clárbruigi</form>
<form type="cpdnn">cláiirrneamh</form>
<form type="cpdnn">clárnimh</form>
<form type="cpdnn">clárrothaibh</form>
<form type="cpdnn">clarsceimelta</form>
<form type="cpdnn">clarsciathaibh</form>
<form type="cpdnn">clárshnoighe</form>
<form type="cpdnn">clarshoileach</form>
<form type="cpdnn">cláirthigh</form>
</form.gr>
<form.gr type="cpdadj">
</form.gr>
<form type="cpdadj">clárchorcra</form>
<form type="cpdadj">clarfhind</form>
<form type="cpdadj">clárghorma</form>
<form type="cpdadj">cláirleabhar</form>
<form type="cpdadj">cláirmhíne</form>
<form type="cpdadj">cláirniamhdha</form>
<form type="cpdadj">clártana</form>
<form type="cpdadj">cláirthe</form>

</form.gr>
<form.gr type="cpdadj">
  <note>with adjective derived from nouns</note>
  <form type="cpdadj"><ps>Connil
    Cláiringnig</ps></form>
  <form type="cpdadj"><ps>Congail
    Chláiringnig</ps></form>
  <form type="cpdadj">clarthairrngeacha</form>
</form.gr>
<form.gr type="cpdadj">
  <note>with past participle adjective</note>
  <form type="cpdadj">clardedaighthi</form>
  <form type="cpdadj">claruaighte</form>
</form.gr>
</entry>
Example 11  A lexical entry for a preposition

<entry id="B2162">
  <lemma>cen</lemma>
  <gramgrp pos="prep">
    <itype></itype>
  </gramgrp>
  <etym></etym>
  <senses>
    <sense n="1"><note ref="general">originally</note> on this side of <note ref="general">hence</note> apart from, besides <lbl>etc</lbl> <note ref="general">the usual meaning at all stages of the language is</note> without, deprived of, not having, -less <note ref="syntax">freq. with following verbal noun with which, in its verbal function, it serves as negative particle</note></sense>
    <sense n="2">apart from, besides, except</sense>
    <sense n="3"><note ref="general">with noun as in</note> without, not having, -less <note ref="syntax">and following preposition indicating position, accompaniment, possession</note></sense>
    <sense n="4"><note ref="general">frequent in verse in chevilles</note> <lbl>and</lbl> <note ref="syntax">quasi-adjectival phrases</note></sense>
  </senses>
</entry>
4. XML: the foundation of the electronic Lexicon

<sense n="5"><note ref="usage">with following acht</note> except that, only <lbl>etc</lbl></sense>

<sense n="6"><note ref="syntax">with verbs of forgetting, preventing, etc. the preposition merely reinforces the meaning of the main verb</note></sense>

<sense n="7"><note ref="general">in more or less literal sense</note> in addition to it, without him or it, besides, apart from this <note ref="general">As shown by occasional examples of cenae referring to feminine or plural nouns the word had even in Old Irish begun to have the force of an independent adverb</note> besides, in addition, moreover, anyway</sense>

<sense n="8"><note ref="general">hence of actions and states of being</note> already, previously</sense>

<sense n="9"><note ref="syntax">freq. with following</note> <xref ptr="C495034">co</xref> as conjunction</sense>

<sense n="10">without that, apart from the case that, because of the fact that not <note ref="usage">with indicative and subjunctive</note>
4. XML: the foundation of the electronic Lexicon

</senses>

<forms>

<form.gr type="simple">
    <form type="simple">cin</form>
    <form type="simple">can</form>
    <form type="simple">gen</form>
    <form type="simple">gin</form>
    <form type="simple">gan</form>
    <form type="simple">cene</form>
    <form type="simple">cenco</form>
    <form type="simple">gion</form>
    <form type="simple">génco</form>
</form.gr>

<form.gr type="suffix-pron">
    <form type="suffix-pron" number="pl" person="3">cenaib</form>
    <form type="suffix-pron" number="sg" person="2">cenut</form>
    <form type="suffix-pron" number="sg" person="3">cenaie</form>
    <form type="suffix-pron" number="pl" person="3">cenuib</form>
    <form type="suffix-pron" number="pl" person="3">cenaib</form>
</form.gr>
4. XML: the foundation of the electronic Lexicon

<entry>

<form type="suffix-pron" number="sg" person="3">cena</form>
<form type="suffix-pron" number="sg" person="3">cene</form>
<form type="suffix-pron" number="sg" person="3">cenae</form>
<form type="suffix-pron" number="sg" person="3">ceni</form>
<form type="suffix-pron" number="sg" person="3">cén</form>
<form type="suffix-pron" number="sg" person="3">céni</form>
<form type="suffix-pron" number="sg" person="3">céin</form>
</entry>
Example 12 A lexical entry for a possessive pronoun

<entry id="2406">
  <lemma>mo</lemma>
  <lemma></lemma>
  <gramgrp pos="pron">
    <note ref="gramgrp">possessive pronoun, genitive of first person singular personal pronoun</note>
  </gramgrp>
  <etym/>
  <senses>
    <sense n="1">my</sense>
  </senses>
  <forms>
    <form.gr type="simple">
      <form type="simple">mu</form>
      <form type="simple">mo</form>
      <form type="simple">m</form>
      <form type="simple" person="1" number="sg">M<ps>Aedóc</ps></form>
      <form type="simple" person="1" number="sg">M<ps>Aedocc</ps></form>
      <form type="simple" person="1" number="sg">M<ps>Aodhocc</ps></form>
      <form type="simple" person="1" number="sg">Mo<ps>edocc</ps></form>
    </form.gr>
  </forms>
</entry>
4. XML: the foundation of the electronic Lexicon

</ps></form>

<form type="simple" person="1" number="sg">Mo<i>e</i>́og</form>

<form type="simple" person="1" number="sg">Mo<i>b</i>í</form>

<form type="simple" person="1" number="sg">Mo<i>chaem</i>´oc</form>

<form type="simple" person="1" number="sg">M´<i>Ern</i>óc</form>

<form type="simple" person="1" number="sg">Mo<i>fheca</i></form>

<form type="simple" person="1" number="sg">Mo<i>nanna</i></form>

<form type="simple" person="1" number="sg">M<i>unnu</i> <note ref="general">abbreviated form of</note></form>

<form type="simple" person="1" number="sg">Mo<i>Fhinnu</i></form>

</ps><note ref="general">abbreviated form of</note></form>

</ps></form.gr>

<form.gr type="infixed-pron">

<form type="infixed-pron" person="1" number="sg">-m-</form>

<form type="infixed-pron" person="1" number="sg">-mm-</form>

</form.gr>
4. XML: the foundation of the electronic Lexicon

```xml
<forms>
</entry>
```

166
Example 13 A lexical entry for a personal pronoun

<entry id="B1066">
  <lemma>mé</lemma>
  <gramgrp pos="pron">
    <itype></itype>
    <note ref="gramgrp">personal pronoun first singular</note>
  </gramgrp>
  <etym></etym>
  <senses>
    <sense n="1">I, me, myself</sense>
  </senses>
  <forms>
    <form.gr type="absolute">
      <form type="simple">mé</form>
      <form type="simple">mi</form>
      <form type="simple">muí</form>
      <form type="simple">mui</form>
      <form type="simple">maoi</form>
      <form type="simple">moe</form>
      <form type="simple">mé</form>
    </form.gr>
    <form.gr type="emphas-suffix">
      <form type="emphas-suffix" number="sg" person="1">mé</form>
    </form.gr>
  </forms>
</entry>
4. XML: the foundation of the electronic Lexicon

meisse</form>

<form type="emphas-suffix" number="sg" person="1"> messe</form>

<form type="emphas-suffix" number="sg" person="1"> mise</form>

<form type="emphas-suffix" number="sg" person="1"> misse</form>

<form type="emphas-suffix" number="sg" person="1"> messi</form>

<form type="emphas-suffix" number="sg" person="1"> mése</form>

<form type="emphas-suffix" number="sg" person="1"> misiu</form>

<form type="emphas-suffix" number="sg" person="1"> mese</form>

<form type="emphas-suffix" number="sg" person="1"> mése</form>

<form type="emphas-suffix" number="sg" person="1"> <midir>mise</midir></form>

<form type="emphas-suffix" number="sg" person="1"> <midir>misse</midir></form>

<form type="emphas-suffix" number="sg" person="1"> <midir>misse</midir></form>

<form type="emphas-suffix" number="sg" person="1"> muísse</form>

<form type="emphas-suffix" number="sg" person="1">
4. XML: the foundation of the electronic Lexicon

muí-se</form>
   <form type="emphas-suffix" number="sg" person="1">
   misiu</form>
   <form type="emphas-suffix" number="sg" person="1">
   misi</form>
   <form type="emphas-suffix" number="sg" person="1">
   misse</form>
   <form type="emphas-suffix" number="sg" person="1">
   meise</form>
   <form type="emphas-suffix" number="sg" person="1">
   missi</form>
   <form type="emphas-suffix" number="sg" person="1">
   muí-sse</form>
   <form type="emphas-suffix" number="sg" person="1">
   muí-se</form>
   <form type="emphas-suffix" number="sg" person="1">
   muisse</form>
   <form type="emphas-suffix" number="sg" person="1">
   mmui-sea</form>
   <form type="emphas-suffix" number="sg" person="1">
   mésse</form>
 </form.gr>
 <form.gr type="infixed-A">
   <form type="infixed-A" number="sg" person="1">
   169
 </form.gr>
4. XML: the foundation of the electronic Lexicon

<mut type="l">-m</mut></form>
<form type="infixed-A" number="sg" person="1">
<mut type="l">-mm</mut></form>
</form.gr>
<form.gr type="infixed-B">
<form type="infixed-B">
<mut type="l" number="sg" person="1">-dom</mut></form>
</form>
<form type="infixed-B">
<mut type="l" number="sg" person="1">-dum</mut></form>
</form>
<form type="infixed-B">
<mut type="l" number="sg" person="1">-tom</mut></form>
</form>
<form type="infixed-B">
<mut type="l" number="sg" person="1">-tum</mut></form>
</form>
<form type="infixed-B">
<mut type="l" number="sg" person="1">-dam</mut></form>
</form>
<form type="infixed-B">
<mut type="l" number="sg" person="1">-damm</mut></form>
</form>
4. XML: the foundation of the electronic Lexicon

<mut type="l" number="sg" person="1">-tam</mut>
</form>

<form type="infixed-B">
  <mut type="l" number="sg" person="1">-tamm</mut>
</form>
</form.gr>

<form.gr type="infixed-C">
  <form type="infixed-C" number="sg" person="1">
    <mut type="l">-dom</mut>,
    <mut type="l" number="sg" person="1">-dum</mut>,
    <mut type="l" number="sg" person="1">-dam</mut>,
    <mut type="l" number="sg" person="1">-damm</mut>
  </form>
</form.gr>

<form.gr type="suffixed">
  <form type="suffixed" number="sg" person="1">-m</form>
</form>

<form type="suffixed" number="sg" person="1">-mm</form>
</form>

<form type="suffixed" number="sg" person="1">-am</form>
</form>

<form type="suffixed" number="sg" person="1">-amm</form>
</form.gr>
4. XML: the foundation of the electronic Lexicon

```xml
<form.gr type="possessive">
    <form type="possessive" number="sg" person="1">
        <mut type="l">mo</mut>
    </form>
    <form type="possessive" number="sg" person="1">
        <mut type="l">mu</mut>
    </form>
    <form type="possessive" number="sg" person="1">muı</form>
</form.gr>
</forms>
</entry>
```
Example 14 A lexical entry for a prefix

```
<entry id="864">
    <lemma htype="4">prím-</lemma>
    <gramgrp pos="prefix">
        <note ref="gramgrp">used as indeclinable prefix
            before nouns and adjectives, generally forming
            unstable compounds</note>
    </gramgrp>
    <etym><frn lang="la">primus</frn></etym>
    <senses>
        <sense n="1"><note ref="syntax">with noun</note>,
            first <note ref="usage">in time</note>,
            primary</sense>
        <sense n="2">foremost, chief, principal
            <note ref="general">without implied comparison</note>
            prime <note ref="usage">in quality</note>,
            eminent, <note ref="general">a commoner use than
            preceding.</note></sense>
        <sense n="3"><note ref="usage">of persons</note>,
            <note ref="general">especially with titles, names of
            callings, etc.</note> prime-</sense>
        <sense n="4"><note ref="syntax">forming compound
            adjectives with abstracts generally with
            emphasizing or intensifying force</note>
    </senses>
</entry>
```
4. XML: the foundation of the electronic Lexicon

very prime</sense>
<sense n="5"><note ref="syntax">with compound adjectives</note> possession
<note ref="general">of quality in superlative degree</note> high-quality, prime
</sense>
</senses>
<form>
<form.gr type="cpdnn">
<form type="cpdnn">prímairec</form>
<form type="cpdnn">prímarigid</form>
<form type="cpdnn">primairaigid</form>
<form type="cpdnn">prímathair</form>
<form type="cpdnn">priomhathair</form>
<form type="cpdnn">prímbérla</form>
<form type="cpdnn">primbelraib</form>
<form type="cpdnn">primchúis</form>
<form type="cpdnn">primchuitech</form>
<form type="cpdnn">primchutig</form>
<form type="cpdnn">prímfid</form>
<form type="cpdnn">primgein</form>
<form type="cpdnn">primgentecht</form>
<form type="cpdnn">prímleis</form>
<form type="cpdnn">primleis</form>
</form.gr>
</form>
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">primless</form>
<form type="cpdnn">primheas</form>
<form type="cpdnn">primhide</form>
<form type="cpdnn">primtecht</form>
<form type="cpdnn">primabstal</form>
<form type="cpdnn">primamait</form>
<form type="cpdnn">primara</form>
<form type="cpdnn">primaraid</form>
<form type="cpdnn">primchalladoir</form>
<form type="cpdnn">primchomairlid</form>
<form type="cpdnn">primchonsul</form>
<form type="cpdnn">primchríochaire</form>
<form type="cpdnn">primdalta</form>
<form type="cpdnn">primechlach</form>
<form type="cpdnn">primechlaig</form>
<form type="cpdnn">primechlach</form>
<form type="cpdnn">primepscop</form>
<form type="cpdnn">primháith</form>
<form type="cpdnn">primfaith</form>
<form type="cpdnn">primfatha</form>
<form type="cpdnn">primfiadu</form>
<form type="cpdnn">primfiada</form>
<form type="cpdnn">primfhlaith</form>
<form type="cpdnn">primfulachtóir</form>
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">prim fulachtore</form>
<form type="cpdnn">prímgaiscedach</form>
<form type="cpdnn">prim gapsedaig</form>
<form type="cpdnn">prímlaech</form>
<form type="cpdnn">prímlaích</form>
<form type="cpdnn">prímlaeches</form>
<form type="cpdnn">prímlaichas</form>
<form type="cpdnn">prímliaig</form>
<form type="cpdnn">prímmartir</form>
<form type="cpdnn">prímmind</form>
<form type="cpdnn">príminmid</form>
<form type="cpdnn">prímmollam</form>
<form type="cpdnn">prímpechtach</form>
<form type="cpdnn">prim pechtach</form>
<form type="cpdnn">prim pecht thaig</form>
<form type="cpdnn">prímpreceptoir</form>
<form type="cpdnn">prímpreceptoire</form>
<form type="cpdnn">prímrannaire</form>
<form type="cpdnn">prim rannaire</form>
<form type="cpdnn">prímréchaire</form>
<form type="cpdnn">prim reto sup resp="x" i sup ridh</form>

<form type="cpdnn">prímsacar t</form>
<form type="cpdnn">prímsacairt</form>
<form type="cpdnn">prímsaer</form>
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">prímsoir</form>
<form type="cpdnn">prímsáeraib</form>
<form type="cpdnn">prímshair</form>
<form type="cpdnn">prímshenchaide</form>
<form type="cpdnn">príomhsheanchaidh</form>
<form type="cpdnn">prímstaraide</form>
<form type="cpdnn">prímsuí</form>
<form type="cpdnn">prímthoísech</form>
<form type="cpdnn">prímuughdair</form>
<form type="cpdnn">phríomhughdair</form>
<form type="cpdnn">prímaba</form>
<form type="cpdnn">prímairecht</form>
<form type="cpdnn">prímaiste</form>
<form type="cpdnn">prímaisti</form>
<form type="cpdnn">prímaltóir</form>
<form type="cpdnn">prímbile</form>
<form type="cpdnn">prímcaindel</form>
<form type="cpdnn">prímcainnel</form>
<form type="cpdnn">prímcathair</form>
<form type="cpdnn">primh cathair</form>
<form type="cpdnn">primcathrach</form>
<form type="cpdnn">primchruachait</form>
<form type="cpdnn">primcubat</form>
<form type="cpdnn">primchubai</form>
4. XML: the foundation of the electronic Lexicon

</form>
  <form type="cpdnn">prímhuit</form>
  <form type="cpdnn">prímdorus</form>
  <form type="cpdnn">primdúni</form>
  <form type="cpdnn">primdún</form>
  <form type="cpdnn">prímeclais</form>
  <form type="cpdnn">primfhéil</form>
  <form type="cpdnn">primfhéile</form>
  <form type="cpdnn">prímgæth</form>
  <form type="cpdnn">primgaetha</form>
  <form type="cpdnn">primínad</form>
  <form type="cpdnn">primhined</form>
  <form type="cpdnn">primlepaid</form>
  <form type="cpdnn">primleathaib</form>
  <form type="cpdnn">primloch</form>
  <form type="cpdnn">primlocha</form>
  <form type="cpdnn">primlong</form>
  <form type="cpdnn">prímlongphort</form>
  <form type="cpdnn">prímmag</form>
  <form type="cpdnn">prímaigib</form>
  <form type="cpdnn">prímoenach</form>
  <form type="cpdnn">primrand</form>
  <form type="cpdnn">prímréilce</form>
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">primreilec</form>
<form type="cpdnn">primhreileag</form>
<form type="cpdnn">primrót</form>
<form type="cpdnn">primroid</form>
<form type="cpdnn">primscél</form>
<form type="cpdnn">primscéoil</form>
<form type="cpdnn">primhscol</form>
<form type="cpdnn">primhscola</form>
<form type="cpdnn">primshliab</form>
<form type="cpdnn">primslige</form>
<form type="cpdnn">primsligeda</form>
<form type="cpdnn">primshlóg</form>
<form type="cpdnn">primsráid</form>
<form type="cpdnn">primshuide</form>
<form type="cpdnn">primtonn</form>
<form type="cpdnn">primthonna</form>

<form type="cpdnn">primacallam</form>
<form type="cpdnn">primacallaime</form>
<form type="cpdnn">primchéssad</form>
<form type="cpdnn">prímcinnecht</form>
<form type="cpdnn">primchinnecht</form>
<form type="cpdnn">primgéim</form>
<form type="cpdnn">primgoibnecht</form>
<form type="cpdnn">primlocht</form>
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">prímlot</form>
<form type="cpdnn">prímshorthan</form>
<form type="cpdnn">primordan</form>
<form type="cpdnn">primgeinde</form>
<form type="cpdnn">primgin</form>
<form type="cpdnn">prímgeindi</form>
<form type="cpdnn">primgeinné</form>
<form type="cpdnn">primgindi</form>
<form type="cpdnn">primerani</form>
<form type="cpdnn">primgene</form>
<form type="cpdnn">primerani</form>

</form.gr>
<form.gr type="cpdadj">
<form type="cpdadj">prímalaind</form>
<form type="cpdadj">primalaind</form>
<form type="cpdadj">prímarrachta</form>
<form type="cpdadj">primarrachta</form>
<form type="cpdadj">prímcoitchenn</form>
<form type="cpdadj">primhcoitcheann</form>
<form type="cpdadj">priméchtach</form>
<form type="cpdadj">priméchtach</form>
<form type="cpdadj">primelathnach</form>
<form type="cpdadj">primelathnach</form>
<form type="cpdadj">primeolach</form>
<form type="cpdadj">primeolach</form>
<form type="cpdadj">primfáthacdae</form>
<form type="cpdadj">primfhathacdae</form>
<form type="cpdadj">primmaith</form>
4. XML: the foundation of the electronic Lexicon

<entry>
    <form type="cpdadj">primmaithiu</form>
    </form.gr>
    </forms>
    </entry>
Example 15 A lexical entry for a noun

<entry id="576">
  <lemma htype="1">Gall</lemma>
  <gramgrp pos="nn" gen="m">
    <itype>o-stm</itype>
  </gramgrp>
  <senses>
    <sense n="1"><note ref="general">oldest meaning</note>
       a Gaul</sense>
    <sense n="2">a Scandinavian invader
       (<xref ptr="F45668">finn-gaill</xref>, Northman,
       <xref ptr="D95064">dub-gaill</xref>, Dane)</sense>
    <sense n="3">an Anglo-Norman, an Irishman of
       Norman descent, an Englishman</sense>
    <sense n="4">Geraldines
       <cit><bibl>Ériu ix</bibl>
          <biblscope>163 &sect; 12</biblscope></cit></sense>
    <sense n="5">a foreigner</sense>
  </senses>
  <forms>
    <form gr type="simple">
      <form type="simple">Gollu</form>
    </form>
  </forms>
</entry>
4. XML: the foundation of the electronic Lexicon

<form type="simple">Golla</form>
<form type="simple">gaillu</form>
<form type="simple">gullu</form>
<form type="simple">gulla</form>
<form type="simple">Gaill</form>
<form type="simple">Gallaib</form>
<form type="simple">Gallaibh</form>
<form type="simple">Gallu</form>
<form type="simple">Gallaibh</form>
<form type="simple">Gall</form>
<form type="simple">Goll</form>
</form.gr>

<form.gr type="cpdnn">
  <form type="cpdnn">Galliathaib</form>
  <form type="cpdnn">gallassa</form>
  <form type="cpdnn">gallasu</form>
  <form type="cpdnn">gallat</form>
  <form type="cpdnn">gallbádach</form>
  <form type="cpdnn">gallbadhacha</form>
  <form type="cpdnn">gallbailti</form>
  <form type="cpdnn">gallbéire</form>
  <form type="cpdnn">guillbelrai</form>
  <form type="cpdnn">gallberla</form>
  <form type="cpdnn">gallbaile</form>
</form.gr>

183
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">gallbiail</form>
<form type="cpdnn">gaillbiail</form>
<form type="cpdnn">gallbela</form>
<form type="cpdnn">gallbrat</form>
<form type="cpdnn">gallbraide</form>
<form type="cpdnn">gallbróć</form>
<form type="cpdnn">gallbrocc</form>
<form type="cpdnn">gallcerd</form>
<form type="cpdnn">gallcét</form>
<form type="cpdnn">gallced</form>
<form type="cpdnn">gallcliar</form>
<form type="cpdnn">gailchliar</form>
<form type="cpdnn">gallcloch</form>
<form type="cpdnn">gallcochal</form>
<ps>Gallch&umacr;u</ps>
</form>

<form type="cpdnn">gallcúirt</form>
<form type="cpdnn">galldírna</form>
<form type="cpdnn">gallécosc</form>
<form type="cpdnn">Gaillecuscaib</form>
<form type="cpdnn">gallécoscda</form>
<form type="cpdnn">gaillecasgda</form>
<form type="cpdnn">gaillespuc</form>
<form type="cpdnn">gallfhocal</form>

184
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">gallfothannán</form>
<form type="cpdnn">gallfhuil</form>
<form type="cpdnn">gallga</form>
<form type="cpdnn">gallgabul</form>
<form type="cpdnn">gallghabhla</form>
<form type="cpdnn">gallglas</form>
<form type="cpdnn">gallgóídel</form>
<form type="cpdnn">gallghoidelaibh</form>
<form type="cpdnn">GallGaoidheal</form>
<form type="cpdnn">Gallgaidhel</form>
<form type="cpdnn">Gallghaedheal</form>
<form type="cpdnn">Gallgaidelu</form>
<form type="cpdnn">Gallgaidelaib</form>
<form type="cpdnn">GallGaoidhel</form>
<form type="cpdnn">Gallgaeidel</form>
<form type="cpdnn">gallgruitne</form>
<form type="cpdnn">gallgruitni</form>
<form type="cpdnn">galliall</form>
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">gaillialla</form>
<form type="cpdnn">galliasc</form>
<form type="cpdnn">gallíath</form>
<form type="cpdnn">gallmanach</form>
<form type="cpdnn">gallmerge</form>
<form type="cpdnn">gaillmeirgib</form>
<form type="cpdnn">gallmías</form>
<form type="cpdnn">gaillmias</form>
<form type="cpdnn">gallmíl</form>
<form type="cpdnn">gaillmíd</form>
<form type="cpdnn">gaillmias</form>
<form type="cpdnn">gallóclach</form>
<form type="cpdnn">galloglacha</form>
<form type="cpdnn">gallógláchaibh</form>
<form type="cpdnn">gallóclachus</form>
<form type="cpdnn">gallógláchaibh</form>
<form type="cpdnn">gallóg</form>
<form type="cpdnn">gallogam</form>
<form type="cpdnn">gallráith</form>
<form type="cpdnn">gallsail</form>
<form type="cpdnn">gallshaileach</form>
<form type="cpdnn">gallsalann</form>
<form type="cpdnn">gallsalaind</form>
<form type="cpdnn">Gallsax</form>
4. XML: the foundation of the electronic Lexicon

<form type="cpdnn">gallseoit</form>
<form type="cpdnn">gallsgiam</form>
<form type="cpdnn">gaillsgéimh</form>
<form type="cpdnn">gallsciath</form>
<form type="cpdnn">gallsochraite</form>
<form type="cpdnn">galltech</form>
<form type="cpdnn">Galltór</form>
<form type="cpdnn">galltrumpa</form>
<form type="cpdnn">gallurad</form>
<form type="cpdnn">gallgruitni</form>

<cert level="low">gallblach</cert>
</form>
<form type="cpd">
<cert level="low">gallsemind</cert>
</form>
</form.gr>
</forms>
</entry>
Example 16 A lexical entry for the definite article

```xml
<entry id="1034">
  <lemma>in</lemma>
  <gramgrp pos="art">
  </gramgrp>
  <etym>
    <frn lang="ie">sindos</frn>,
    <frn lang="ie">sinda</frn>,
    <frn lang="ie">sindom</frn></etym>
  <senses>
    <sense n="1">the</sense>
  </senses>
  <forms>
    <form gr type="simple">
      <form number="sg" case="nom" gen="m">in</form>
      <form number="sg" case="nom" gen="m">int
      <note ref="general">before initial vowel only</note></form>
      <form number="sg" case="nom" gen="n">a</form>
      <form number="sg" case="nom" gen="f">ind</form>
      <form number="sg" case="nom" gen="f">int <note ref="general">before initial s only</note></form>
    </form gr type="simple">
```
4. XML: the foundation of the electronic Lexicon

<form number="sg" case="acc" gen="m">
  <mut type="n">in</mut></form>

<form number="sg" case="acc" gen="m">
  <mut type="n">-sin</mut></form>

<form number="sg" case="acc" gen="n">
  <mut type="n">a</mut></form>

<form number="sg" case="acc" gen="n">
  <mut type="n">-sa</mut></form>

<form number="sg" case="acc" gen="f">
  <mut type="n">in</mut></form>

<form number="sg" case="acc" gen="f">
  <mut type="n">-sin</mut></form>

<form number="sg" case="gen" gen="m n">
  <mut type="l">in</mut></form>

<form number="sg" case="gen" gen="m n">
  <mut type="l">ind</mut></form>

<form number="sg" case="gen" gen="m n">
  <mut type="l">int <note ref="general">before initial s only</note></mut></form>

<form number="sg" case="gen" gen="f">
  <mut type="g">inna</mut></form>

<form number="sg" case="gen" gen="f">
  <mut type="g">na</mut></form>

<form number="sg" case="dat" gen="m">
4. XML: the foundation of the electronic Lexicon

<mut type="l">-in</mut></form>
<form number="sg" case="dat" gen="m">
  <mut type="l">-ind</mut></form>
<form number="sg" case="dat" gen="m">
  <mut type="l">-sin</mut></form>
<form number="sg" case="dat" gen="m">
  <mut type="l">-sind</mut></form>
<form number="sg" case="dat" gen="m">
  <mut type="l">-sint <note ref="general">before initial s only</note></mut></form>
<form number="sg" case="dat" gen="m">
  <mut type="l">-sint <note ref="general">before initial s only</note></mut></form>
<form number="sg" case="dat" gen="n">
  <mut type="l">-in</mut></form>
<form number="sg" case="dat" gen="n">
  <mut type="l">-ind</mut></form>
<form number="sg" case="dat" gen="n">
  <mut type="l">-sin</mut></form>
<form number="sg" case="dat" gen="n">
  <mut type="l">-sind</mut></form>
<form number="sg" case="dat" gen="n">
  <mut type="l">-sint <note ref="general">before initial s only</note></mut></form>
<form number="sg" case="dat" gen="n">
  <mut type="l">-sint <note ref="general">before initial s only</note></mut></form>
<form number="sg" case="dat" gen="n">
  <mut type="l">-sint <note ref="general">before initial s only</note></mut></form>
XML: the foundation of the electronic Lexicon

<form number="sg" case="dat" gen="n">
  <mut type="l">-int <note ref="general">before initial s only</note></mut></form>

<form number="sg" case="dat" gen="f">
  <mut type="l">-in</mut></form>

<form number="sg" case="dat" gen="f">
  <mut type="l">-sin</mut></form>

<form number="sg" case="dat" gen="f">
  <mut type="l">-sind</mut></form>

<form number="sg" case="dat" gen="f">
  <mut type="l">-int <note ref="general">before initial s only</note></mut></form>

<form number="sg" case="dat" gen="f">
  <mut type="l">-sint <note ref="general">before initial s only</note></mut></form>

<form number="pl" case="nom" gen="m">
  <mut type="l">ind</mut></form>

<form number="pl" case="nom" gen="m">
  <mut type="l">in</mut></form>

<form number="pl" case="nom" gen="m">
  <mut type="l">int <note ref="general">before initial s only</note></mut></form>

<form number="pl" case="nom" gen="n">
  <mut type="g">inna</mut></form>
4. XML: the foundation of the electronic Lexicon

<form number="pl" case="nom" gen="n">
  <mut type="g">na</mut>
</form>

<form number="pl" case="nom" gen="n">
  <mut type="g">sna</mut>
</form>

<form number="pl" case="nom" gen="f">
  <mut type="g">inna</mut>
</form>

<form number="pl" case="nom" gen="f">
  <mut type="g">na</mut>
</form>

<form number="pl" case="nom" gen="f">
  <mut type="g">sna</mut>
</form>

<form number="pl" case="acc" gen="c">
  <mut type="g">inna</mut>
</form>

<form number="pl" case="acc" gen="c">
  <mut type="g">na</mut>
</form>

<form number="pl" case="acc" gen="c">
  <mut type="g">sna</mut>
</form>

<form number="pl" case="gen" gen="c">
  <mut type="n">inna</mut>
</form>

<form number="pl" case="gen" gen="c">
  <mut type="n">na</mut>
</form>

<form number="pl" case="dat" gen="c">
  -naib
</form>

<form number="pl" case="dat" gen="c">
  -snaib
</form>
<form number="pl" case="dat" gen="c">-sna</form>

<form number="dual" case="nom" gen="m">
  <mut type="l">in dá</mut>
</form>

<form number="dual" case="nom" gen="n">
  <mut type="n">in dá</mut>
</form>

<form number="dual" case="nom" gen="f">
  <mut type="l">in dá</mut>
</form>

<form number="dual" case="acc" gen="m">
  <mut type="l">in dá</mut>
</form>

<form number="dual" case="acc" gen="n">
  <mut type="n">in dá</mut>
</form>

<form number="dual" case="acc" gen="f">
  <mut type="l">in dá</mut>
</form>

<form number="dual" case="gen" gen="m">
  <mut type="l">in dá</mut>
</form>

<form number="dual" case="gen" gen="n">
  <mut type="n">in dá</mut>
</form>

<form number="dual" case="gen" gen="f">
  <mut type="l">in dá</mut>
</form>

<form number="dual" case="dat" gen="c">
  <mut type="n">-n d́b</mut>
</form>

</form.gr>

</forms>
4. XML: the foundation of the electronic Lexicon

</entry>
Example 17 A lexical entry for a verb

```xml
<entry id="34789">
  <lemma htype="1">t´eit</lemma>
  <gramgrp pos="vb">
    <itype></itype>
  </gramgrp>
  <etym></etym>
  <senses>
    <sense n="1.0">proceeds, goes, goes forward, turns out</sense>
    <sense n="1.1">goes away, departs</sense>
    <sense n="1.2">passes away, vanishes</sense>
    <sense n="1.3">dies</sense>
    <sense n="2.0">follows <note ref="general">a route</note></sense>
    <sense n="2.1">goes <note ref="general">on an expedition</note></sense>
    <sense n="2.2">dies, perishes</sense>
    <sense n="2.3">extends to, passes through</sense>
    <sense n="3.0"><note ref="general">with prep <xref ptr="A14053">ar</xref> and <xref ptr="F487523">for</xref></note> goes upon, set out on</sense>
    <sense n="3.1">has sexual intercourse with, mates</sense>
  </senses>
</entry>
```
with</sense>

<sense n="3.2">resorts to, indulges in, becomes subject to</sense>

<sense n="3.3">is inflicted upon, comes upon</sense>

<sense n="3.4">attacks, overcomes</sense>

<sense n="3.5">prevails over</sense>

<sense n="3.6">changes into, is exchanged for, becomes</sense>

<sense n="3.7">inspires, causes, gives rise to</sense>

<sense n="3.8">goes towards, seeks protection of, comes to nought</sense>

<sense n="4.0"><note ref="general">with prep <xref ptr="A457346">ass</xref></note> goes out, leaves</sense>

<sense n="4.1">departs, escapes, vanishes, dies</sense>

<sense n="4.2" type="gramm">mutate, eclipse</sense>

<sense n="5.0"><note ref="general">with prep <xref ptr="C76545">cen</xref></note> is left undone</sense>

<sense n="6.0"><note ref="general">with prep <xref ptr="C495034">co</xref>, <xref ptr="D876987">dochum</xref></note> goes to, goes towards</sense>
<sense n="6.1" has sexual intercourse with"/></sense>
<sense n="6.2" dies"/></sense>
<sense n="7.0" sense n="7.0" note ref="general">with prep
<xref ptr="D596748">de</xref> departs from, leaves, abandons</sense>
<sense n="7.1" outstrips"/></sense>
<sense n="7.2" escapes from"/></sense>
<sense n="7.3" fails to accomplish"/></sense>
<sense n="7.4" becomes of"/></sense>
<sense n="7.5" passes away"/></sense>
<sense n="8.0" note ref="general">with prep
<xref ptr="D946757">do</xref> goes to, goes towards"/></sense>
<sense n="8.1" goes to, goes for the purpose of"/></sense>
<sense n="8.2" dies"/></sense>
<sense n="8.3" applies to"/></sense>
<sense n="8.4" is given to"/></sense>
<sense n="8.5" fares with, turns out for"/></sense>
<sense n="8.6" suits, becomes"/></sense>
<sense n="9.0" note ref="general">with prep
<xref ptr="E098002">eter</xref> </note> goes between, comes to pass between"/></sense>
<sense n="10.0" note ref="general">with prep
4. XML: the foundation of the electronic Lexicon

<note ref="general">under</note> goes under</sense>

<sense n="10.1">undergoes, submits to</sense>

<sense n="10.2">dies <note ref="general">by a weapon</note></sense>

<sense n="10.3">goes towards, goes throughout</sense>

<sense n="10.4">attacks, assails</sense>

<sense n="11.0"><note ref="general">with prep</note> goes towards, goes against, opposes</sense>

<sense n="11.1">devotes oneself to, adopts, takes a certain course</sense>

<sense n="11.2">guarantees, goes surety for</sense>

<sense n="11.3">takes after, resembles</sense>

<sense n="12.0"><note ref="general">with prep</note> goes into, enters</sense>

<sense n="12.1">penetrates, invades, attacks</sense>

<sense n="12.2">goes into <note ref="general">of numbers etc.</note>, fits into</sense>

<sense n="12.3">goes to, sets about, indulges in</sense>
becomes, results in</sense>
<note ref="gramm">ends in</note>
<note ref="general">of words</note>
<note ref="gramm">amalgamates with, is assimilated to</note> <note ref="general">of sounds</note></sense>
<note ref="general">with prep</note>
<xref ptr="I54656">iar</xref></note>
go along, come to pass</sense>
<note ref="general">with prep</note>
<xref ptr="I56780">imm</xref></note> go about</sense>
<note ref="general">with prep</note>
<xref ptr="L45867">la</xref></note> goes with, accompanies</sense>
<note ref="general">is let go</note></sense>
<note ref="general">succeeds</note></sense>
<note ref="general">pleases</note></sense>
<note ref="general">with prep</note>
<xref ptr="O57646"> ´o</xref></note> goes from, departs from</sense>
<note ref="general">is sent by</note></sense>
<note ref="general">outstrips</note></sense>
<note ref="general">escapes from</note>
<sense n="16.4">gives up, forsakes</sense>
<sense n="16.5">is deprived of, loses</sense>
<sense n="17"><note ref="general">with prep</note> <xref ptr="009894">´os</xref> surpass</sense>
<sense n="18.0"><note ref="general">with prep</note> <xref ptr="R73498">re</xref> goes before, precedes</sense>
<sense n="18.1">goes on, fares forward</sense>
<sense n="19.0"><note ref="general">with prep</note> <xref ptr="S987608">sech</xref> goes past</sense>
<sense n="19.1">misses</sense>
<sense n="20.0"><note ref="general">with prep</note> <xref ptr="T349854">tar</xref> goes over, goes across, goes past</sense>
<sense n="20.1">overcomes, overwhelms</sense>
<sense n="20.2">exceeds, surpasses</sense>
<sense n="20.3">transgresses, violates</sense>
<sense n="20.4">becomes necessary, becomes due</sense>
<sense n="20.5">passes by, is omitted, lost, forgotten by</sense>
<sense n="21.0"><note ref="general">with prep</note> <xref ptr="T456745">trí</xref>
4. XML: the foundation of the electronic Lexicon

goes through</sense>
</senses>
<paradigm>
  <mood type="indicative">
    <tense type="present">
      <number type="sg">
        <person n="1">
          <form type="regular">tíagu</form>
          <form type="regular">-tiág <note ref="general">in future sense</note></form>
        </person>
        <person n="2">
          <form type="regular">tégí</form>
          <form type="regular">-téig</form>
        </person>
      </number>
      <number type="pl">
        <person n="1">
          <form type="regular">tígim</form>
          <form type="regular">-téigim</form>
        </person>
        <person n="2">
          <form type="regular">déig</form>
          <form type="regular">-déigim</form>
        </person>
      </number>
    </tense>
  </mood>
</paradigm>
4. XML: the foundation of the electronic Lexicon

```xml
<form type="regular">téigi</form>
<form type="regular">téighi</form>
<form type="regular">-tír</form>
</person>
<person n="3">
<form type="regular">téit</form>
<form type="regular">tét</form>
<form type="regular">-tét</form>
<form type="regular">-tet</form>
<form type="regular">-téd</form>
<form type="regular">-téd</form>
<form type="regular">-téd</form>
<form type="regular">-téd</form>
<form type="regular">-tít</form>
<form type="regular">tiat</form>
<form type="regular">-tiat</form>
<form type="regular">-tíge</form>
<form type="regular">teit</form>
<form type="regular">téd</form>
</form type="regular">
<cert level="50%">-teid</cert></form>
<form type="relative">tête</form>
<form type="relative">téite</form>
<form type="relative">tét</form>
<form type="with-suffix-pron">tête</form>
<form type="with-suffix-pron">téite</form>
```

202
4. XML: the foundation of the electronic Lexicon

<form type="perfective">da-chubat</form>
<form type="perfective">da-cumat</form>
<form type="perfective">dos-cuat</form>
<form type="perfective">dos-cuet</form>
<form type="perfective">dos-cuaid</form>
<form type="perfective">do-cuat</form>
<form type="perfective">di-cuat</form>
<form type="perfective">di-cead</form>
<form type="perfective">di-chuat</form>
<form type="perfective">-dichet</form>
<form type="perfective">degas</form>

</person>

<person n="0">
<form type="passive">tíagair</form>
<form type="passive">-tíagar</form>
<form type="passive-relative">tíagar</form>
</person>

</number>

<number type="pl">
<person n="1">
<form type="regular">tíagmai</form>
<form type="regular">tíagmími
<note ref="general"><equals/>tiags-mi</note>
</form>
<form type="regular">-tíagam</form>
<form type="regular">tíagma</form>
<form type="regular">tíagmait</form>
<form type="regular">lodma</form>

<note ref="general">Middle Irish re-formation</note>

</form>

<form type="regular">tiamaid</form>
<form type="regular">-téimid</form>
<form type="relative">tíagmae</form>
<form type="relative">tiagme-ni</form>

</person>

<person n="2">
<form type="regular">-téit</form>
<form type="regular">teigthi</form>
<form type="regular">tegthi</form>
<form type="regular">teidchi</form>
<form type="perfective">-digtith</form>
</person>

<person n="3">
<form type="regular">tíagait</form>
<form type="regular">tíagad</form>
<form type="regular">tíagat</form>
<form type="regular">tiagat</form>
4. XML: the foundation of the electronic Lexicon

<form type="regular">tiagait</form>
<form type="regular">tiaghaid</form>
<form type="regular">téighit</form>
<form type="regular">téidhid</form>
<form type="regular">tíat</form>
<form type="regular">tiad</form>
<form type="regular">-tiad</form>
<form type="regular">-tiaid</form>
<form type="regular">-tíagat</form>
<form type="regular">-tiagat</form>
<form type="regular">-tiaguid</form>
<form type="regular">-tiaguit</form>
<archaic><form type="archaic">tégot</form></archaic>
<person n="0">
<form type="passive">tíagair</form>
<form type="passive">tiagar</form>
</person>

205
4. XML: the foundation of the electronic Lexicon

<form type="passive">tiaghair</form>
<form type="passive">tiaghar</form>
<form type="passive">-tiaghar</form>
<form type="passive">tiachar</form>
<form type="passive">-tiachair</form>
<form type="passive-relative">tíaig</form>
<form type="passive-relative">tíagthar</form>
<form type="passive-relative">tiagbar</form>
<form type="passive-relative">tiaghdur</form>
<form type="passive-relative">tiaghtur</form>

</person>
</number>
</tense>

<tense type="imperfect">

<number type="sg">

<person n="1">

<form type="regular">-teiginn</form>
<form type="regular">-luidinn</form>
</person>

<person n="2">
</person>

<person n="3">

<form type="regular">-téged</form>
<form type="regular">-teghedh</form>
</person>
</number>
</tense>
4. XML: the foundation of the electronic Lexicon

\[
\begin{align*}
&\text{<form type="regular">-teidheadh</form>}
&\text{<form type="regular">}</form>
&\text{<cert level="50%">-thegaid</cert></form>}
&\text{<form type="perfective">-dichtheth</form>}
&\text{<form type="relative">}
&\text{<cert level="50%">-th \text{\textsuperscript{e}iged}</cert></form>}
&\text{<form type="relative">}
&\text{<cert level="50%">-tiagad</cert></form>}
&\text{</person>}
&\text{<person n="0">}
&\text{<form type="passive">-t\text{\textsuperscript{e}igthe</form>}
&\text{</person>}
&\text{</number>}
&\text{<number type="pl">}
&\text{<person n="1">}
&\text{<form type="regular">-t\text{\textsuperscript{e}igmis</form>}
&\text{</person>}
&\text{<person n="2">}
&\text{</person>}
&\text{<person n="3">}
&\text{<form type="regular">-t\text{\textsuperscript{e}igtis</form>}
&\text{<form type="regular">-t\text{\textsuperscript{e}gtis</form>}
&\text{<form type="regular">t\text{\textsuperscript{e}gtis</form>}
&\text{<form type="regular">t\text{\textsuperscript{e}gtis</form>}
&\text{<form type="regular">teigdis</form>}
\end{align*}
\]
4. XML: the foundation of the electronic Lexicon

```xml
<form type="regular">-teigtis</form>
<form type="regular">théighdis</form>
<form type="regular">teighdis</form>
<form type="regular">tecdis</form>
</person>

<form type="passive">téigthea</form>
<form type="passive">teigthi</form>
<form type="passive">-teghthi</form>
<form type="passive">-teiththi</form>
</person>
</number>
</tense>

<tense type="preterite">
<number type="sg">
<person n="1">
<form type="regular">lod</form>
<form type="regular">lud</form>
<form type="perfective">do-coad</form>
<form type="perfective">do-choad</form>
<form type="perfective">do-coat</form>
<form type="perfective">do-cood</form>
<form type="perfective">docuad</form>
</person>
</number>
</tense>
```
4. XML: the foundation of the electronic Lexicon

<form type="perfective">dochûadusa</form>
<form type="perfective">do-cûadas</form>
<form type="perfective">tochuadhus</form>
<form type="perfective">-dechud</form>
<form type="perfective">-dechad</form>
<form type="perfective">-dechuds</form>
<form type="perfective">-deochads</form>
<form type="perfective">-deachad</form>
<form type="perfective">-deachodhus</form>

</person>

<person n="2"/>

<form type="regular">lod</form>
<form type="perfective">do-cood</form>
<form type="perfective">do-chûad</form>
<form type="perfective">dochodai</form>
<form type="perfective">dochuadais</form>
<form type="perfective">dochuadhs</form>
<form type="perfective">dodechadais</form>
<form type="perfective">-dechud</form>
<form type="perfective">-dechais</form>
<form type="perfective">-dechais</form>
<form type="perfective">-dechadais</form>

</person>

<person n="3"/>
4. XML: the foundation of the electronic Lexicon

<form type="regular">luid</form>
<form type="regular">luith</form>
<form type="regular">lauid</form>
<form type="regular">lud</form>
<form type="regular">do-luid</form>
<form type="regular">
<form type="relative">luide</form>
<form type="relative">lude</form>
<form type="relative">luidi</form>
<form type="relative">ludi</form>
<form type="relative">ludai</form>
<form type="relative">laide</form>
<form type="relative">luid</form>
<form type="with-ro">ra-luid</form>
<form type="with-ro">-rulaid</form>
<form type="with-ro">roloduir</form>
<form type="perfective">do-co"{i}d</form>
<form type="perfective">do-coith</form>
<form type="perfective">do-choid</form>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">du-coid</form>
<form type="perfective">du-choid</form>
<form type="perfective">di-choid</form>
<form type="perfective">du-chooid</form>
<form type="perfective">du-choaid</form>
<form type="perfective">du-cuaid</form>
<form type="perfective">do-cuaid</form>
<form type="perfective">do-chuaid</form>
<form type="perfective">doch´oid</form>
<form type="perfective">dochaidh</form>
<form type="perfective">dochuaid</form>
<form type="perfective">dochuaidh</form>
<form type="perfective">doch´oidh</form>
<form type="perfective">doch´aidh</form>
<form type="perfective">dodechaidh</form>
<form type="perfective">dodheachaidh</form>
<form type="perfective">roch´oid</form>
<form type="perfective">roch´uaid</form>
<form type="perfective">rochuaid</form>
<form type="perfective">-dechuid</form>
<form type="perfective">-dechuith</form>
<form type="perfective">-dechaid</form>
<form type="perfective">-deachaidh</form>
<form type="perfective">-deochaid</form>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">-dechuidh</form>
<form type="with-do">dodechaid</form>
<form type="with-do">dodeochaid</form>
</person>

<person n="0">
<form type="passive">ethae</form>
<form type="passive-perfective">do-coas</form>
<form type="passive-perfective">do-cuas</form>
<form type="passive-perfective">-dechas</form>
</person>
</number>

<number type="pl">
<person n="1">
<form type="regular">lodmar</form>
<form type="regular">lodomar</form>
<form type="regular">lodamur</form>
<form type="regular">lodsam</form>
<form type="regular">lodsamar</form>
<form type="perfective">do-commar</form>
<form type="perfective">do-chuamar</form>
<form type="perfective">docuamairni</form>
<form type="perfective">docuamar</form>
<form type="perfective">docuamair</form>
<form type="perfective">dochamar</form>
</person>
</number>
<form type="perfective">dochuadhmuir</form>
<form type="perfective">rochuammar</form>
<form type="perfective">-dechummar</form>
<form type="perfective">-deachamair</form>
</person>
</number>
<number type="pl">
<person n="2">
<form type="regular">-lodsaid</form>
<form type="regular">-ladsaid</form>
<form type="perfective">do-chuabair</form>
<form type="perfective">do-chuabar</form>
<form type="perfective">deochabair</form>
<form type="perfective">dochuabhair</form>
</person>
</number>
<number type="pl">
<person n="3">
<form type="regular">lotir</form>
<form type="regular">lotar</form>
<form type="regular">lodur</form>
<form type="regular">lautar</form>
<form type="regular">lottar</form>
<form type="regular">lodor</form>
</number>
4. XML: the foundation of the electronic Lexicon

<form type="regular">lotair</form>  
<form type="regular">lodatar</form>  
<form type="regular">lodadar</form>  
<form type="regular">luiditar</form>  
<form type="regular">lotarset</form>  
<form type="regular">lotursat</form>  
<form type="regular">luidset</form>  
<form type="regular">luidsetd</form>  
<form type="regular">luighsi</form>  
<form type="regular">luidhisiot</form>  
<form type="regular">luidsiut</form>  
<form type="regular">lutsat</form>  
<form type="regular">luidhsidad</form>  
<form type="with-do">dollotar</form>  
<form type="with-do">do-lotar</form>  
<form type="with-do">tollotar</form>  
<form type="with-do">tulatar</form>  
<form type="with-ro">ra-lotar</form>  
<form type="with-ro">ar-rolotar</form>  
<form type="relative">
  <cert level="50%">lotar</cert>
</form>  
<form type="perfective">do-cotar</form>  
<form type="perfective">do-chotar</form>  
<form type="perfective">dochódar</form>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">dochádair</form>
<form type="perfective">da-chotar</form>
<form type="perfective">do-cuatar</form>
<form type="perfective">du-cuatar</form>
<form type="perfective">do-chuatar</form>
<form type="perfective">do-chótar</form>
<form type="perfective">dochótár</form>
<form type="perfective">dochotar</form>
<form type="perfective">dochuatar</form>
<form type="perfective">dochuadar</form>
<form type="perfective">rochuadar</form>
<form type="perfective">rochuatar</form>
<form type="perfective">rachuatar</form>
<form type="perfective">chuadar</form>
<form type="perfective">dochoidset</form>
<form type="perfective">dodeachatar</form>
<form type="perfective">ro-digset</form>
<form type="perfective">-dechutar</form>
<form type="perfective">-deachatar</form>
<form type="perfective">-deachadar</form>
<form type="perfective">-deochatar</form>
<form type="perfective">-deachsad</form>
<form type="perfective">-deachadair</form>
<form type="perfective">-dechsat</form>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">-dechsatar</form>
<form type="perfective">-dechaddar</form>
<form type="perfective">-dechatar</form>
<form type="perfective">-deochotar</form>
</person>

<form type="passive-perfective">do-choas</form>
<form type="passive-perfective">do-chós</form>
<form type="passive-perfective">do-chúas</form>
<form type="passive-perfective">dochuas</form>
<form type="passive-perfective">docoas</form>
<form type="passive-perfective">docos</form>
<form type="passive-perfective">docós</form>
<form type="passive-perfective">cocoass</form>
<form type="passive-perfective">dochuas</form>
<form type="passive-perfective">dociuas</form>
<form type="passive-perfective">docuas</form>
<form type="passive-perfective">rocúadhus</form>
<form type="passive-perfective">-dechus</form>
<form type="passive-perfective">-deachas</form>
<form type="passive-perfective">-dechas</form>
</person>
</tense>
4. XML: the foundation of the electronic Lexicon

<tense type="future">
  <number type="sg">
    <person n="1">
      <form type="regular">rega</form>
      <form type="regular">riga</form>
      <form type="regular">-rig</form>
      <form type="regular">-reg</form>
      <form type="regular">raga</form>
      <form type="regular">rag</form>
      <form type="regular">rach</form>
      <form type="regular">regad</form>
      <form type="regular">reghat</form>
      <form type="regular">regut</form>
      <form type="regular">ragat</form>
      <form type="regular">ragut</form>
      <form type="regular">ragaid</form>
      <form type="regular">regaid</form>
      <form type="regular">rachat</form>
      <form type="regular">rachad</form>
      <form type="regular">reachat</form>
      <form type="regular">-rig</form>
      <form type="regular">-reg</form>
      <form type="regular">-rach</form>
    </person n="1">
  </number type="sg">
</tense type="future">
4. XML: the foundation of the electronic Lexicon

</person>
<person n="2">
  <form type="regular">regae</form>
  <form type="regular">regai</form>
  <form type="regular">ragai</form>
  <form type="regular">rogai</form>
  <form type="regular">rage</form>
  <form type="regular">raga</form>
  <form type="regular">racha</form>
  <form type="regular">racair</form>
  <form type="regular">raghair</form>
  <form type="regular">-regae</form>
  <form type="regular">-rega</form>
  <form type="regular">-raga</form>
  <form type="regular">-rada</form>
  <form type="regular">-rado</form>
</person>
<person n="3">
  <form type="regular">regaid</form>
  <form type="regular">-riga</form>
  <form type="regular">-rega</form>
  <form type="regular">ragaid</form>
  <form type="regular">rachaidh</form>
  <form type="regular">raghaidh</form>
</person>
4. XML: the foundation of the electronic Lexicon

<form type="regular">-raga</form>
<form type="regular">-racha</form>
<form type="regular">-ragha</form>
<form type="regular">-rachaidh</form>
<form type="regular">rachuidh</form>
<form type="relative">rigas</form>
<form type="relative">regas</form>
<form type="relative">regus</form>
<form type="relative">ragas</form>
<form type="relative">ragus</form>
<form type="relative">rachas</form>
<form type="relative">raghas</form>
<form type="relative">reachbas</form>
</person>

<number type="pl">
<person n="1">
<form type="regular">rigmi</form>
</person>
</number>
4. XML: the foundation of the electronic Lexicon

<form type="regular">regmi</form>
<form type="regular">regmai</form>
<form type="regular">rechmi</form>
<form type="regular">regma</form>
<form type="regular">ragma</form>
<form type="regular">ragmai</form>
<form type="regular">raigma</form>
<form type="regular">régmaït</form>
<form type="regular">ragmai</form>
<form type="regular">ragmaït</form>
<form type="regular">ragmaït</form>
<form type="regular">radhmaït</form>
<form type="regular">rachmaid</form>
<form type="regular">-regam</form>
<form type="regular">-ragam</form>
<form type="regular">-ragam</form>
<form type="regular">-ragum</form>
<form type="regular">-ragham</form>
<form type="regular">ludfamar</form>
<form type="relative">
<cert level="50%">regmae</cert></form>
</person>
<person n="2">
<form type="regular">
<cert level="50%">regthae</cert></form>
<form type="regular">ragthai</form>
</person>
4. XML: the foundation of the electronic Lexicon

<form type="regular">
<cert level="50%">ragaid</cert>
<form type="regular">-regaid</form>
<form type="regular">-rachtai</form>
</person>
<person n="3">
<form type="regular">regait</form>
<form type="regular">-regat</form>
<form type="regular">ragait</form>
<form type="regular">rechaid</form>
<form type="regular">regat</form>
<form type="regular">regut</form>
<form type="regular">rechtait</form>
<form type="regular">raghdait</form>
<form type="regular">rágdaít</form>
<form type="regular">ragdíait</form>
<form type="regular">ragtháit</form>
<form type="regular">rachdaeid</form>
<form type="regular">rachaid</form>
<form type="relative">regthaè</form>
<form type="relative">regda</form>
<form type="relative">regthar</form>
</person>
<person n="0">

221
4. XML: the foundation of the electronic Lexicon

<form type="passive">rigthir</form>
<form type="passive">regthair</form>
<form type="passive">rechtair</form>
<form type="passive">ragthair</form>
<form type="passive">rachthair</form>
<form type="passive">ragthar</form>
<form type="passive">ragar</form>
<form type="passive">rachar</form>
<form type="passive">rachtar</form>
<form type="passive">rachtair</form>
<form type="passive">raghar</form>
<form type="passive">-raghthar</form>
<form type="passive">-raigther</form>
</person>
</number>
</tense>
<tense type="secondary-future">
<number type="sg">
<person n="1">
<form type="regular">-regainn</form>
<form type="regular">-raththa</form>
<form type="regular">-rachaind</form>
</person>
<person n="2">
</number>
</tense>
4. XML: the foundation of the electronic Lexicon

<form type="regular">-regtha</form>
<form type="regular">-rigtha</form>
<form type="regular">-rachthá</form>
<form type="regular">rachtha</form>
<form type="regular">raghtha</form>

</person>
<person n="3">
<form type="regular">-regad</form>
<form type="regular">-rigad</form>
<form type="regular">-rígad</form>
<form type="regular">-reched</form>
<form type="regular">-ragad</form>
<form type="regular">-rachadh</form>
<form type="regular">-raghadh</form>
<form type="regular">-rachfadh</form>
<form type="perfective">do-coísed</form>
</person>
</number>
<number type="pl">
<person n="1">
<form type="regular">-regmais</form>
</person>
<person n="2">
</person>
</number>
4. XML: the foundation of the electronic Lexicon

```xml
<person n="3">
  <form type="regular">-regthae</form>
  <form type="regular">regadáis</form>
  <form type="regular">-ragdaí</form>
  <form type="regular">-rachdaí</form>
  <form type="regular">do-ragtaí</form>
  <form type="regular">dorachdaois</form>
  <form type="regular">-raghduí</form>
  <form type="regular">rachdaí</form>
  <form type="regular">-rachdaois</form>
  <form type="perfective">do-cóestí</form>
</person>
<person n="0">
  <form type="passive">-ragtha</form>
  <form type="passive">-rachtá</form>
  <cert level="50%">-rachtaí</cert>
  <form type="passive">ragtháí</form>
</person>
</tense>
</mood>
```
4. XML: the foundation of the electronic Lexicon

<mood type="imperative">
  <number type="sg">
    <person n="1">
      <form type="regular">t íag</form>
      <form type="regular">eirgear</form>
      <form type="regular">eargar</form>
    </person>
    <person n="2">
      <form type="regular">eirg</form>
      <form type="regular">ercc</form>
      <form type="regular">airc</form>
      <form type="regular">aircc</form>
      <form type="regular">airg</form>
      <form type="regular">airgg</form>
      <form type="regular">ercc</form>
      <form type="regular">t éig</form>
      <form type="regular">t éighuid</form>
    </person>
    <person n="3">
      <form type="regular">t ét</form>
    </person>
  </number>
</mood>
<form type="regular">teit</form>
<form type="suffix-pron">téiti</form>
<form type="regular">tíat</form>
<form type="regular">téiged</form>
<form type="regular">éirgidh</form>
<form type="regular">tiagá</form>
<form type="regular">téidheadh</form>
<form type="passive">tíagar</form>
<form type="regular">eirgthear</form>
<form type="regular">eargthar</form>
</person>
</number>
<number type="pl">
<person n="1">
<form type="regular">tíagam</form>
<form type="regular">tiagum</form>
<form type="regular">tiagmai</form>
<form type="regular">eirgeam</form>
<form type="regular">eargam</form>
</person>
<person n="2">
<form type="regular">ergid</form>
<form type="regular">ergid</form>
</form type="regular">
4. XML: the foundation of the electronic Lexicon

<cert level="50%">aircid</cert></form>
<form type="regular">airgid</form>
<form type="regular">argib</form>
<form type="regular">eirgid</form>
<form type="regular">eirgeid</form>
<form type="regular">
<cert level="50%">´eirgid</cert></form>
<form type="regular">´eirghidh</form>
<form type="regular">t´eighidh</form>
</person>
<person n="3">
<form type="regular">t´ıagat</form>
<form type="regular">t´ıagat</form>
<form type="regular">tiagaid</form>
<form type="regular">
<cert level="50%">tighed</cert></form>
<form type="regular">eircett</form>
</person>
<person n="0">
<form type="passive">tiagar</form>
<form type="passive">t´ıagair</form>
<form type="passive">t´íachar</form>
<form type="passive">eirgtheir</form>
</form>

227
4. XML: the foundation of the electronic Lexicon

<form type="passive">eargthar</form>
</person>
</number>
</mood>
<mood type="subjunctive">
<tense type="present">
<number type="sg">
<person n="1">
<form type="regular">tíasu</form>
<form type="regular">tíasa</form>
<form type="regular">-tías</form>
<form type="regular">-tias</form>
<form type="regular">-theis</form>
<form type="regular">-tiasur</form>
<form type="regular">tíighedh</form>
<form type="perfective">do-cous</form>
<form type="perfective">-dichius</form>
<form type="perfective">-digus</form>
<form type="perfective">-digius</form>
<form type="perfective">-dechos</form>
<form type="perfective">-dechus</form>
<form type="perfective">-deachar</form>
<form type="perfective">-deach</form>
<form type="perfective">-dighear</form>
<form type="perfective">-deach</form>
</number>
</person>
</mood>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">conicius</form></person>

<person n="2">
<form type="regular">tési</form>
<form type="regular">téisi</form>
<form type="regular">-téis</form>
<form type="regular">teisiu</form>
<form type="regular">-téis</form>
<form type="regular">-tiasair</form>
<form type="regular">-téidhe</form>
<form type="regular">théighe</form>
</person>

<person n="2">
<form type="perfective">do-cois</form>
<form type="perfective">dige</form>
<form type="perfective">díchi</form>
<form type="perfective">do-cui</form>
<form type="perfective">-dichis</form>
<form type="perfective">-digis</form>
<form type="perfective">-dighes</form>
<form type="perfective">-dechais</form>
<form type="perfective">-dechair</form>
<form type="perfective">-dechuir</form>
<form type="perfective">-deachair</form>
<form type="perfective">-docuissiu</form>
</person>
XML: the foundation of the electronic Lexicon

</person>

<cert level="50%">docoid</cert></form>

<cert level="50%">docoid</cert></form>

</form type="passive">tíasair</form>
4. XML: the foundation of the electronic Lexicon

<form type="regular">tíastaisi</form>
<form type="regular">-tésid</form>
<form type="regular">-téssid</form>
<form type="perfective">do-coísid</form>
<form type="perfective">-digsid</form>
<form type="perfective">-dechsaid</form>
</person>

<form type="regular">tíasait</form>
<form type="regular">tíasuit</form>
<form type="regular">-tíasat</form>
<form type="regular">-tíassat</form>
<form type="regular">-tigsad</form>
<form type="regular">-téset</form>
<form type="relative">tíastae</form>
<form type="relative">tiasda</form>
<form type="perfective">do-coíset</form>
<form type="perfective">do-coísat</form>
<form type="perfective">do-cóesat</form>
<form type="perfective">do-chósat</form>
<form type="perfective">(cert level="50%")do-choiset</form>
<form type="perfective">-dichset</form>
<form type="perfective">-digset</form>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">-digsit</form>
<form type="perfective">-dechsat</form>
<form type="perfective">dechuid</form>
<form type="perfective">deachud</form>
<form type="perfective">-dechsat</form>
</person>
<person n="0">
<form type="passive">dighthir</form>
</person>
</number>
</tense>
</tense type="imperfect">
<number type="sg">
<person n="1">
<form type="regular">-t ´eisinn</form>
<form type="perfective">do-co ´ısinn</form>
<form type="perfective">-dechsainn</form>
<form type="perfective">-deochsaind</form>
<form type="perfective">-dechaind</form>
<form type="perfective">-dighear</form>
<form type="perfective">-dechsaind</form>
<form type="perfective">-deachainn</form>
<form type="perfective">dechar</form>
<form type="perfective">-digusa</form>
</form type="perfective">-dighear</form>
<form type="perfective">-dechsaind</form>
<form type="perfective">-deachainn</form>
<form type="perfective">dechar</form>
<form type="perfective">-digusa</form>
<form type="perfective">-dechus</form>
<form type="perfective">-deochus</form>
<form type="perfective">-dechar</form>
<form type="perfective">-dechor</form>
<form type="perfective">-digser</form>
<form type="perfective">-diccsiur</form>
<form type="perfective">-dechaind</form>
</person>
<person n="2">
<form type="regular">-t īast</form>
<form type="perfective">do-co īsta</form>
<form type="perfective">-digesta</form>
<form type="perfective">-digis</form>
<form type="perfective">-digese</form>
<form type="perfective">-d īchiseo</form>
<form type="perfective">-dechais</form>
<form type="perfective">-dechis</form>
<form type="perfective">-deachadas</form>
<form type="perfective">-dechasu</form>
<form type="perfective">-deochais</form>
<form type="perfective">-dechair</form>
<form type="perfective">-digthea</form>
<form type="perfective">-dechta</form>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">-deachtha</form>
</person>
<person n="3">
<form type="regular">-téised</form>
<form type="regular">-tésed</form>
<form type="regular">-tésad</form>
<form type="regular">-tésed</form>
<form type="regular">-tésad</form>
<form type="regular">-téssed</form>
<form type="regular">-téised</form>
<form type="regular">tesed</form>
<form type="regular">tesad</form>
<form type="regular">tessed</form>
<form type="regular">-theissed</form>
<form type="regular">-tesid</form>
<form type="regular">-thaesad</form>
<form type="regular">-tíasad</form>
<form type="regular">don-iasadh</form>
<form type="regular">-toised</form>
<form type="perfective">do-coisèd</form>
<form type="perfective">do-coisead</form>
<form type="perfective">do-coiset</form>
<form type="perfective">-dichsèd</form>
<form type="perfective">-digsèd</form>
<form type="perfective">-diccèad</form>
<form type="perfective">-digseth</form>
<form type="perfective">-dechsad</form>
<form type="perfective">-dig</form>
<form type="perfective">-dech</form>
<form type="perfective">-deoch</form>
<form type="perfective">-disghsedh</form>
<form type="perfective">dhigsid</form>
<form type="perfective">dichsead</form>
<form type="perfective">dechsad</form>
<form type="perfective">dheachsad</form>
<form type="perfective">-digsith</form>
<form type="perfective">-dechad</form>
<form type="perfective">-deachadh</form>
<form type="perfective">-digsith</form>
<form type="perfective">-dechad</form>
<form type="perfective">-deachadh</form>
<form type="perfective">-deachaid</form>
<form type="perfective">-deachaidh</form>
<form type="perfective">-téstae</form>
<form type="passive">-téstae</form>
</person>
</number>
<number type="pl">
<person n="1">
<form type="perfective">do-coísms</form>
</person>
</number>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">tiasmais</form>
<cert level="50%">-dichesmis</cert>
<form type="perfective">-dechsum</form>
<form type="perfective">-deachmaois</form>
</person>
<person n="2">
<form type="regular">-tiastae</form>
<form type="regular">-tésstae</form>
<form type="perfective">do-coíste</form>
<form type="perfective">-dicheste</form>
</cert level="50%">-dicheste</cert><form type="perfective">-dechaiste</form>
<form type="perfective">-dechastaí</form>
<form type="perfective">-dechaisti</form>
</person>
<person n="3">
<form type="regular">-tiastais</form>
<form type="regular">-tiastais</form>
<form type="regular">-teostaiss</form>
<form type="perfective">do-coístis</form>
<form type="perfective">-digsitis</form>
</cert level="50%">-teostaiss</cert>
4. XML: the foundation of the electronic Lexicon

<form type="perfective">-dechsaitis</form>
<form type="perfective">-dechsitis</form>
<form type="perfective">-dighset</form>
<form type="perfective">-diccsed</form>
<form type="perfective">-dechsat</form>
<form type="perfective">-digsitís</form>
<form type="perfective">-dechsaitís</form>
<form type="perfective">-dechdais</form>
<form type="perfective">-dechtais</form>
<form type="perfective">-dechtásí</form>
<form type="perfective">-deachdáis</form>
<form type="perfective">-deachdáois</form>
</person>

<person n="0">
<form type="passive">-tíastais</form>
<form type="passive-perfective">-dechasta</form>
</person>
</number>
</tense>
</mood>

<mood type="non-finite">
<form type="participle-necessity"></form>
<form type="verbal-noun">techt</form>
<form type="verbal-noun">techtæ</form>
<form type="verbal-noun">dul</form>
</mood>
4. XML: the foundation of the electronic Lexicon

```xml
<form type="verbal-noun">dula</form>
<form type="past-participle-passive">dulta</form>
<form type="past-participle-passive">dolta</form>
</mood>
</paradigm>
</entry>
```
Chapter 5

Advantages and disadvantages of XML

5.1 Abstract

This chapter treats of the advantages and disadvantages of using XML to encode an electronic Lexicon of Old, Middle and Early Modern Irish and to create tools to link the Lexicon with other XML encoded dictionaries and corpora. The main disadvantage, described as the problem of date and context is specific to the Lexicon, but also is relevant in the wider discipline of humanities computing. The advantages of using XML will be shown to be manifold, and to reflect the tenets that lie at the heart of XML. Further, the advantages discussed in this chapter represent innovations in the field of humanities computing in Irish as they are the first tools of their kind to be made available to students and scholars of the language.
5. Advantages and disadvantages of XML

5.2 Advantages and disadvantages

5.2.1 The problem of date and context

The use of XML to encode the Lexicon has drawn attention to the deficiencies of some aspects of lexicography in DIL, while contributing further to our understanding of electronic text. This problem is revealed by the power of XML and its related technologies such as XSLT and XPath because they can enable transformations of text that are not, in linguistic terms, meaningful. This issue is particularly relevant in the case of a document that has been migrated from hard-copy to digital format, and lies at the heart of the conceptual differences that exist between these two formats of text.

5.2.1.1 Inadequate lexicographical criteria

As discussed in Chapter 1, the diachronic terms earlier and later used in lexical entries in DIL are much less specific—at least in intent, perhaps not in reality—than the labels Old Irish, Middle Irish and Modern Irish. This is a problem of linguistic delimitation in a field in which most manuscripts are not contemporary with the texts they contain. As discussed in Chapter 1, the use of these terms in DIL is unsatisfactory because they are not used objectively, but have a subjective meaning based on the chronology of the lemma they occur in, and perhaps on nothing more than the general subjective impressions of the individual lexicographer. To revise this system and replace all the earlier and later references with more specific and objective terms such as Old Irish is beyond the scope
of anything but a fundamental long-term lexicographical undertaking. Therefore, the elements `<earlier>` and `<later>` are used here, but the limitations of these terms clearly illustrated.

The problem of date and context became apparent when XSLT templates, such as those illustrated in Example 1, were written to select all the earlier and later elements in a file, extract the contents of the elements, and output them to a new file, transformed, somewhat arbitrarily, to the elements `<OldIr>` and `<ModIr>`.

Example 1 Examples of simple XSLT templates

```
<xsl:template match="earlier">
  <OldIr>
    <xsl:value-of select="."/>
    <xsl:apply-templates/>
  </OldIr>
</xsl:template>

<xsl:template match="later">
  <ModIr>
    <xsl:value-of select="."/>
    <xsl:apply-templates/>
  </ModIr>
</xsl:template>
```

The purpose of such extraction of all earlier and later references from
the Lexicon was to produce a test stand-alone glossary to see whether it could illustrate the development of every word form from Old to Modern Irish. However, it was realised that while it was possible to extract such information, it was not meaningful to do, because in DIL every earlier reference is relative to the individual lemma. Therefore, the list of earlier forms, for example, includes words from radically different periods in the history of the Irish language and, clearly, the use of earlier and later in DIL is an inadequate criterion for establishing the historical development of Irish.

5.2.1.2 Limitation of OHCO theory of textuality

The problem of date and context also makes clear a limitation of the OHCO theory of text, and based on this, an extension to the mark-up that can be applied to some TEI conformant dictionaries will be proposed.

In order to examine this limitation, a page of DIL was digitised and marked up (the Lexicon was not used because it is a subset of DIL) and compared with the same page in the hard-copy edition of DIL. Once the xml page was served, the two pages were, in terms of their content identical, thus fulfilling the ‘form versus content’ criterion that had been used to determine whether one text differed from another by DeRose et al (see § 3.5). Nonetheless, it was observed that the hard-copy text of DIL and the electronic version of it function in completely different ways. Indeed, certain units of the hard-copy text such as earlier and later

---

1 DeRose et al. What is text? 3.
develop a new significance when represented in a digital environment.

Unlike a book or article, from the perspective of the end user, DIL is non-linear in structure. The user will not read all five volumes of DIL from cover to cover, but will look up lexical entries on a ‘need to know’ basis. Indeed, the use of terms earlier and later indicate that DIL was not conceived of as a single document with completely objective reference criteria, but as a series of discrete lexical entries containing some subjective and objective reference criteria. Due to the hard-copy format and non-linear structure of DIL, the limitations of the earlier and later criteria do not become apparent to the user, or if they do, they do not impinge on the end-user because DIL only has one search path to retrieve information: the lemma.

The electronic version of DIL is also non-linear in structure, due to the hyperlinks to information—both within and outside the document—that it can contain and the multiple search paths that can be used to retrieve information from it. However, when the text is encoded in line with TEI, it is enclosed within a <document> start and end tag and therefore it must be viewed as a single, coherent document, rather than a series of discrete lexical entries bound between two covers. Indeed, it is not possible to conceptualise a single electronic <document> in such a disjointed way because non-linear texts can have no tangible beginning or end. In her discussion of electronic text, Sutherland notes that an electronic text does not ‘exist’ in an ongoing continuous and tangible sense, but has to be regenerated from a combination of zero and one every time a file is opened. Thus, a page on a computer screen is not a page, but the
illusion of a page; besides, it is not one in a series of pages, but the only ‘page’.\textsuperscript{2} Therefore, electronic text with its commonplace technologies such as hyperlinking and possibilities for searching the macro-structure and micro-structure\textsuperscript{3} of a document means that ‘...every text or even every portion of a text (i.e. every logical unit in the hypertext) has an absolute value within the structure as a whole unless its absolute character is specifically modified’.\textsuperscript{4}

In the electronic version of DIL, an encoder who engages with the text as an ordered hierarchy of content objects (or even an OHCO-2 or OHCO-3) would encode earlier and later references as information labels, and most likely link all such labels together. Implicit in this encoding is the assumption that such reference criteria have an objective meaning in the electronic document. Once the text is marked up, the information contained in the electronic document can be manipulated and transformed in as many different ways as the document engineer can think of. Thus, information that has been migrated from the hard-copy environment can be manipulated in ways that, most likely, were never envisaged by its original compilers. As the experiment described above illustrated, the results of such data manipulation can be, in linguistic terms, meaningless.

The Platonic view of the OHCO theory states that text is an ordered hierarchy of content objects. The early criticism of the theory rested on

\begin{itemize}
\item \textsuperscript{2} Sutherland, \textit{Introduction} (1997), 11,12.
\end{itemize}
5. Advantages and disadvantages of XML

the structure of content objects and the theory was expanded to take account of overlapping hierarchies. In recent years, calls for the theory to be further modified have been convincing argued by McGann and Caton who made clear the role of the performative intent of a text as well as the communicative act of a text and the need for encoders to engage with their texts at a deeper level than simply encoding a series of content objects (see § 3.5). The problem of date and context furthermore makes clear, in line with Caton and McGann, that explaining a text as an ‘ordered hierarchy of content objects’ is an impoverished theory of text. In the case of a hard-copy text that is being transformed to a digital text, this observation becomes all the more pertinent, as Caton has stated:

Metaphorically encoders must be down at what would be the lowest level of an OHCO tree, completely immersed in the #PCDATA, because content generates interactional encoding far more that any content object

5.2.1.3 Multilex root element

The implications of the medium in which the earlier and later reference criteria are communicated is clear: in the hard copy environment, they underline the non-linear structure of DIL but do not impinge on the reader who is not in a position to transform the text; in the electronic environment, however, they can be employed in transformations of DIL that are not linguistically meaningful. But how can such a reservation be encoded in a dictionary? Further, where all such features of a text have not been discovered, how can such reservations be encoded?

The TEI guidelines state that the <text> element ‘may contain a single
5. Advantages and disadvantages of XML

text of any kind, whether unitary or composite, for example a poem or
drama, a collection of essays, a novel, a dictionary, or a corpus sample.\(^5\)
The *problem of date and context* draws attention to DIL in particular (and
possibly hard copy dictionaries in general) as being unique among the
documents listed because they display a non-linear structure in hard
copy form. To bring this to the attention of both encoders and document
engineers, the Lexicon is contained in a `<multilex>` element, which is
defined as a unified text that contains numerous texts. Therefore, the
textual features and reference criteria used in a lexicon are not assumed
to be objective, but rather their objective nature is something that needs
to be established. I also hold that the `<multilex>` element should be
included as an extension to the TEI guidelines because it signals a
different methodological approach to encoding and transforming texts.
Indeed it inverts McGann’s assumption, quoted above, that every logical
unit within a hypertext has an absolute value unless specifically modified.

The conceptual differences that lie at the heart of the hard-copy DIL
and the electronic Lexicon are brought into sharp focus by the use of
XML to mark-up the data and present it in a different medium. While
McGann has stated that ‘...it is clear, *prima facie*, that digital tools can
execute many of the tasks of scholarly editing much better, much more
thoroughly, and much more precisely than books can’,\(^6\) the *problem of
context* makes clear the need for the encoders and editors of scholarly

\(^5\) Sperberg-McQueen et al, Guidelines, 280.

\(^6\) Jerome McGann ‘Visible and invisible books: hermetic images in n-dimensional
texts to be very much aware of the difficulties that may be encountered when attempting to transport text from one container to another. In the case of a hard-copy work that is migrated to digital format, it is likely that encoders will create both hyperlinks and tools based on the existing textual features of the hard-copy document. Thus, an understanding of the way textual features are re-defined in different media is vital.

5.3 Advantages

5.3.1 The Lexicon as a stand-alone research tool

HTML is an appropriate mark-up language for encoding simple documents, multimedia and creating simple links between web pages.\textsuperscript{7}

XML on the other hand, is much better suited to encoding complex scholarly documents because it describes a document’s structure and semantics and its text is, therefore, much more amenable to computer analysis than a document that has been prepared in HTML. For example, while a computer may not be able to extract any more information from a HTML document than that it contains a number of text strings that have been rendered in italics, it is possible for a computer to locate and extract all the text strings that have been marked with the elements $<$pn$>$ or $<$etym$>$ from an XML document.

Thus, because the Lexicon has been encoded in XML, it can be used as a fully searchable text base, the first such fully searchable linguistic text base in medieval Irish studies.

\textsuperscript{7} Bosak, XML and Java 1.
5. Advantages and disadvantages of XML

Characteristic features of Old Irish such as its complicated morphology and irregular orthography that have compounded some of the inherent limitations of DIL will no longer prove a problem for students. It is possible to search the Lexicon for any part of speech by entering any one of its attested forms, whether nouns, adjectives, adverbs, conjunctions, causal conjunctions, independent negative particles, letters, interjections, particles, personal and possessive adjectives, prefixes, suffixes, infixed forms, prepositions, substantives, pronomial forms, verbal forms in all attested moods, tenses, voices and numbers, the definite article, verbal nouns, and forms encoded as unclear.

A broad range of languages such as Old Norse, Latin, Anglo-Saxon and Hebrew are represented as etyma or comparanda in DIL and while the paper copy can be used as an Irish-English dictionary only, it is possible to search the Lexicon for materials in any of these languages. It is also possible to use the Lexicon as an English to Old-Irish dictionary. This in itself represents a contribution to scholarship in that no such reference work is otherwise available. It is also possible to construct specialised vocabularies, for example, in relation to metalwork or the technology of writing, by working from English equivalents. This can be used to trace gaps in the vocabulary, for example, missing terms for technologies, established by other disciplines (for example, archaeology, epigraphy, history).8

The Lexicon will also be of use to those involved in lexicostatistics

5. Advantages and disadvantages of XML

and it will be possible to retrieve, for example, statistics relating to the number of Irish words with a Hebrew etymology, the number of words that originally had a neuter gender or the number of words that have been influenced by Latin orthography.

Scholars involved in textual reconstruction\(^9\) will, for the first time, be able to harness the results of wild card searches across the entire corpus of Irish. If a manuscript has been damaged and, for example, the first and last letters in a word are illegible, its editor must attempt to reconstruct what is missing. Until the creation of the Lexicon, scholars have only been able to draw on their memory and knowledge of Irish as retained in memory, or memory aided by non-systematic collections of forms, to reconstruct what the word was likely to have been. The Lexicon is the first tool of its kind that can be harnessed by scholars of Irish to perform searches—refined with occurrence indicators—to retrieve a list of all the possible words (insofar as the list is complete) that the damaged word could have been. Depending on the context of the word, the scholar can also use other tools such as the lexicostatistics function to establish usage profiles for the word and then draw on their expertise in metrics, rhyme, genre, or dialect etc to choose the most likely word.

It will also be possible to use the Lexicon as the base text for a thesaurus. Following the method used by Roget, each sense could be grouped into a wider group of semantically related terms, for example, botanical terms or nautical terms. When all the senses had been encoded with this

additional meta-data, it is but one step to write an XSLT template to extract all required fields from the Lexicon and use this to generate a thesaurus. Further encoding could also be incorporated at this stage to create a dictionary of synonyms and antonyms.

Another advantage of encoding the Lexicon in XML is that it supports advances in pedagogy in both education and traditional courses. It will be possible to use this tool as a teaching aid in the classroom, or as part of an online course or for private study as XSLT stylesheets allow instructors to generate many different views of texts depending on their requirements. For example, an instructor who wishes to demonstrate the effect that inflexion has on various forms of nouns will be able to use the CELT corpus and the Lexicon to produce a view of a text where, for example, every inflexion of the noun *ech* or every mood, tense and voice and number of the verb *as-beir* is highlighted.

Furthermore, it is possible to generate different views of the information, based on the requirements of a single end-user. A student of history, for example may not need to know the inflexional information of a noun and a series of XSLT scripts could be used to generate different views of the information. For example, a student who does not require inflexional or syntactic information could choose a transformation of the information that contains only a head-word and its definition. A visually impaired user may wish to generate a view where syntactical information is in a font two sizes bigger than the rest of the text, or blind students may use the Lexicon with a programme such as Jaws, or a server-side XSLT file to transform the mark-up to VoiceML, and use an application that reads
the text out to them.

5.3.2 Lexicon and Text interlinked

The advantages set out above illustrates that the Lexicon represents a sophisticated research tool in its own right. A further advantage of encoding the Lexicon in XML is that it can be readily used in conjunction with other XML conformant text corpora.

CELT, the Corpus of Electronic Texts at the History Department, University College Cork, is a scholarly corpus of multi-lingual texts of Irish literature and history. The texts in the CELT corpus are written in Irish of all periods, Latin, English, Anglo-Norman French and both English and English translation of original texts are represented. The corpus, which is the largest of its kind available freely to members of the public, is conformant with the TEI guidelines, encoded in SGML and will soon switch to XML as the metalanguage of choice for its master files.

To quote C. M. Sperberg-McQueen ‘one reason our libraries no longer chain books to the shelves is that scholarship is materially easier when we can have both books out on our desk at the same time for direct comparison’. In the case of the Lexicon and CELT it was recognised that the concept of having ‘both books out on our desk at the same time…’ could not only be achieved, but furthermore improved upon.

---

As has already been discussed, locating a word in the Dictionary of the Irish Language can be a difficult task that frequently requires some degree of specialist knowledge on the part of the student. By enabling end-users to harness the resources of CELT and the electronic Lexicon simultaneously, students who encounter a word they do not understand in the CELT corpus no longer have to locate a copy of the *DIL* and spend what can frequently amount to a considerable time for many people locating that word, but are able to retrieve it within seconds. In order to enable ease of use for the end-user it was decided that it should be possible to highlight any word or phrase in the CELT corpus and retrieve it in the Lexicon. The nature of Internet development over the past decade encourages this approach, although there are misgivings in related fields such as Library Science about the levels of software and skills needed for anything more that occasional use of real-time referencing.  

---

A robust mechanism for arbitrary reference that could be adapted to enable textual connections between CELT and the Lexicon, or indeed vice versa was kindly brought to my attention by a colleague.\textsuperscript{12} The piece of Javascript that is illustrated in Example 3 is the work of authors or contributors of blogging systems like Wordpress\textsuperscript{13} and others, where it was designed for the capture of URIs.

Suitably adapted, the code must be entered all on one line (it is broken over several lines in Example 3 for convenience) and installed in browsers that support bookmarklets. The code initiates a DOM function getSelection() that captures the highlighted area to a variable Q, and a window is then opened to a suitable URI, with the escaped value of Q appended as a value to the field text. Once installed, the user can highlight a word in the CELT corpus in the browser, and click on the new bookmarklet. The text gets sent to the URI specified, which in this implementation is the development lookup function for the Lexicon. Finally the XSLT stylesheet that has been developed generates the lookup display.

\textsuperscript{12} I am especially grateful to Peter Flynn, manager of the Electronic Publishing Unit, UCC, for bringing this mechanism to my attention.

\textsuperscript{13} See: http://wordpress.org.
5. Advantages and disadvantages of XML

Example 3 Javascript code for a bookmark toolbar button to implement arbitrary links

```javascript
javascript:
Q=document.selection?document.selection.createRange().text:
document.getSelection();void(window.open(' http://www.ucc.ie:8080/cocoon/lexicon/find?text=' +escape(Q), 'textselectionbookmarklet', ' scrollbars=no, width=480, height=300, left=100, top=150, status=yes ' ));
```

An example of a user having highlighted a word in the saga *Fingal Ronáin* and clicked on the lookup button is illustrated in Example 4

Example 4 Fingal Ronáin

![Image of Fingal Ronáin example](image-url)
In addition to enabling students, scholars and lay readers to use existing resources such as CELT and the Lexicon more effectively, this method of arbitrary textual linking enables existing scholarly resources to be examined from different perspectives. Words highlighted in the CELT corpus that are not retrieved in the Lexicon will most likely represent words, that for one reason or another, were not included in the Dictionary, whether as head-words or variants thus providing information for future supplements to both the Lexicon and DIL.

By using the Lexicon in conjunction with the corpus, it will also be possible to address other questions such as the degree to which lexical entries are indicative of the level of use of certain words. For example, if one head-word has thirty citations from literary works and another head-word has only five citations from literary works does that mean that the word with the larger number of citations was a more commonly used word? Or could it be related to factors such as the materials that were at the disposal of the lexicographers or the limitation of their sampling?
5. Advantages and disadvantages of XML

5.3.3 Lexicon, Dictionary, and Text interlinked

Example 5 Linking CELT and the Electronic Lexicon

The Centre for Irish and Celtic Studies at the University of Ulster with funding from the Arts and Humanities Research Board, is currently producing a digital edition of the Dictionary of the Irish Language.\textsuperscript{14}

The electronic *Dictionary of the Irish Language* is a reproduction of the hard-copy version of the work, but in digital format. The encoders are, therefore, obliged to reproduce the work exactly.

While the Lexicon is based upon the *Dictionary of the Irish Language*, there can be no doubt that it differs from the digital and hardcopy editions of the Dictionary in important ways.

The Lexicon is a re-edited subset of the information contained in *DIL* and should additional funding be obtained, it is intended to supplement the Lexicon with the results of further and more recent scholarship. In the Lexicon, most of the limitations of *DIL* that have been highlighted in

\textsuperscript{14} See Toner & Fomin, *The eDIL project.*
Chapter 1 have been remedied. The structure of the Lexicon is ordered and in contrast with *DIL* and *eDIL*, the etymological information, for example, is always recorded immediately after the `<gramgrp>` element. Furthermore, abbreviations are used consistently throughout the Lexicon, for example, *vb* for verbal noun and entries that are treated of in incorrect lexical entries have been identified and new entries prepared, for example, *sní*.

The most marked difference between the electronic version of the Dictionary and the Lexicon is that of content. The Lexicon contains only a portion of the information that is contained in *DIL* and this information has been carefully selected and marked-up to allow refined searches of the material. With regard to the forms of the lemma for example, the Lexicon does not reproduce complete citations from texts as in *DIL* and *eDIL*, but rather culls the particular form from the citation, categorises it and then marks it up accordingly.

With regard to the bibliographical information that accompanies citations, the Lexicon reproduces these in only a limited number of cases because it was recognised at an early stage in the research that these would represent a needless duplication of information contained in the electronic *Dictionary of the Irish Language*. Also recognised at an early stage was that the effectiveness of CELT, the Lexicon and the *eDIL* as research tools could be maximised by cooperation. It was envisaged that this cooperation would take the form of conventional predetermined (fixed) links, either encoded in a static HTML file or using HTML generated from an SGML or XML source. This would enable users to click on a
5. Advantages and disadvantages of XML

bibliographical citation (in the eDIL or to a lesser extent the Lexicon) that would resolve to the specified text in the CELT website.

The technology that facilitates this is commonplace and well established on the Internet and involves the citational information being appended to the URI in the conventional manner and is illustrated in Example 6.

Example 6 Fixed Linking

```
<cit><bibl>ZCP</bibl> <biblscope>xiii 169.36</biblscope></cit>,
```

the link would be in the format

```
<xptr to="http://celt.ucc.ie/cgi-bin/lookup?cite=ZCP\%40xiii\%40169.36"/>
```

After the citational information has been appended to the URI, a script at the server uses a table of the known bibliographical data (based on the document abbreviation and the format of the bibliographical numeration) to determine whether the required document is on-line or not. If it is, the server will return the document fragment surrounding the reference; if not, it will return an explanation. In this way, as documents are added to the CELT textbase over the coming years, further references from the eDIL and the Lexicon will be satisfied, because the citation data required for recognition can be stored in the TEI header of the CELT documents themselves, and extracted for use by the server on an automated basis.
5. Advantages and disadvantages of XML

5.4 Partial automation of encoding of citations

In order to facilitate such linking between documents, the bibliographical citations in both the Lexicon and eDIL must be consistently encoded in both documents. When research into building the tool described here began, both eDIL and the Lexicon already existed as XML files, that had been marked up to varying degrees. To employ researchers to encode the bibliographical citations manually would have involved a large commitment in terms of time and money. Furthermore, it would have resulted in staff working on different projects duplicating work, and so research focused on developing a script to automate the encoding of the bibliographical citations.

XML is an important component in the powerful tool described here—the first tool of its kind to be made available to scholars and students of the Irish language. Nonetheless, the use of XML in this tool can be viewed as both an advantage and a disadvantage. From the perspective of this project, the use of XML to encode eDIL and the Lexicon ensures that the documents are stable; interoperable with server side and client side applications; and extensible. Nonetheless, the use of XML also represents a disadvantage because the type of text processing required to automate such encoding is not fully supported by XML.

Indeed, Sam Wilmott has stated:

Mark-up language description languages, like XML, are an important component of a wide range of applications. However they are not everything. General text processing functionality is also an important component of a wide range of applications, both in
5. Advantages and disadvantages of XML

conjunction with and separate from the use of mark-up languages.\textsuperscript{15}

Text processing that would support the automation of encoding of bibliographical citations was considered a desideratum for this project because the documents in it are large, complex, and display little consistency in the structure and presentation of data. Automation was also considered a desideratum from the perspective of project funds and time, because in order to develop the tool discussed above, it was necessary to encode both the eDIL and Lexicon documents in ways that had not initially been envisaged and therefore not included in the budget.\textsuperscript{16}

The support for such processing improved somewhat recently, with the publication by the W3C\textsuperscript{17} on 3 November 2005 of the Candidate Recommendation for XSLT 2.0, which supports regular expressions.\textsuperscript{18} But because it is not yet a full recommendation\textsuperscript{19} and is supported only by three processors\textsuperscript{20} that I am aware of, it was decided to develop the

\textsuperscript{15} ibid. §2.

\textsuperscript{16} ibid. §2.1. In his discussion of types of text processing that are not supported by XML Wilmott has drawn attention to scenarios ‘there all the potential information in a document is not marked up, and later uses of the data require revision or enhancement of the mark-up. This is the normal state of any rich textual data for reasons of incomplete knowledge—who knows what future users of data will want of it?— and economics— who can afford to investigate and allow for all future uses?’.

\textsuperscript{17} Candidate Recommendation status means that the XSLT working group is putting the XSLT 2.0 specification into the public arena in order to get feedback. After any issues have been ironed out, the document will become a proposed recommendation and members of the public will again be invited to give feedback. Following this XSLT 2.0 will become a full recommendation.

\textsuperscript{18} Available at http://www.w3.org/TR/xslt20/.

\textsuperscript{19} Further, it had not reached Candidate recommendation at the time this research was being undertaken in 2004.
5. Advantages and disadvantages of XML

regular expressions and process them in *sed* script, leaving the possibility open of incorporating the regular expressions into an XSLT 2.0 script at a later date, should it be deemed necessary and should the cross-platform issues that are pertinent in an cross-border project between University College Cork and the University of Ulster, Coleraine, such as this, be resolved.

5.4.1 Patterns of citation in *DIL*

Initial research established fifteen common patterns of citation in *DIL*, as illustrated in Example 7.

Example 7 *Early Regular Expressions*

<table>
<thead>
<tr>
<th>Reference</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cin Lae Ó Meall</td>
<td>48.4.</td>
</tr>
<tr>
<td>Misc. Celt. Soc.</td>
<td>332.22.</td>
</tr>
<tr>
<td>Ir. Nen.</td>
<td>118. 3, 4.</td>
</tr>
<tr>
<td>Buile Suibni</td>
<td>28. 6.6</td>
</tr>
<tr>
<td>IGT Decl. \S38 (87.17)</td>
<td></td>
</tr>
<tr>
<td>O’Curry MS. Mat.</td>
<td></td>
</tr>
<tr>
<td>Fel. O. Ep.</td>
<td></td>
</tr>
<tr>
<td>Aisl. MC 116.20.</td>
<td></td>
</tr>
<tr>
<td>Corm. Y 408.</td>
<td></td>
</tr>
<tr>
<td>CCath. 5367.</td>
<td></td>
</tr>
<tr>
<td>PRIA xix 532</td>
<td></td>
</tr>
<tr>
<td>TBC p. 267, n. 4</td>
<td></td>
</tr>
<tr>
<td>FM v 1862.</td>
<td></td>
</tr>
</tbody>
</table>

However, further research made clear, due to the protracted gestation and editorial history of DIL, as discussed in Chapter 2, that the editors of DIL showed little consistency in the way they abbreviated titles of works in citations. Thus, each of the fifteen common patterns of citation illustrated above can show a number of permutations, because it is not unusual for a work to be cited by many different abbreviations. In addition to this, the pattern matches could not be developed based solely on the hard-copy edition of DIL. As both eDIL and the Lexicon already existed as XML documents that had been marked up to varying degrees, mark-up contained in these documents would also have to be taken into account when writing pattern matches.

5.4.2 Limitations of early regular expressions

An early attempt to automate the encoding of citations was not particularly promising. A sed script was written that consisted of extremely general regular expressions, as illustrated in Example 8.21

\begin{example}
\textbf{Early Regular Expressions}

\begin{verbatim}
\end{verbatim}
\end{example}

\footnote{Regular expressions have been split over several lines to aid legibility.}
These scripts are extremely general and permit many combinations of characters, numerals, punctuation markers and white space. As is frequently the case with regular expressions that are too general, these scripts caused unforeseen problems. The most common error was the capture of extraneous data, for example, an upper-case letter and comma immediately preceding the title portion of the citation was frequently captured and output as part of a title.
5. Advantages and disadvantages of XML

5.5 Predictable and unpredictable data patterns

In order to automate the encoding of citations more accurately, it was realised that a more nuanced understanding of the patterns of citation in DIL, eDIL, and Lexicon files was required.

5.5.1 Predictable exceptions type A

In the Lexicon, eDIL and DIL, citations deviate from the patterns illustrated in Example 7 in predictable and unpredictable ways. For the most part, the predictable deviations involve structural mark-up in the XML documents occurring in a citation.

An example of a predictable deviation is illustrated by the citation to the *Miscellany of the Celtic Society* (O’Donovan, 1846) in Example 9.

Example 9 *Miscellany of the Celtic Society*

```
Misc. <br column="5" line="24"/>Celt. Soc. 54. 14.
```

In this example, the reference to the *Miscellany* spans a line break, and the citation therefore includes an empty element that splits the title portion of the bibliographical citation.

5.5.2 Predictable exceptions type AA

Another example of structural mark-up intervening in a citation that spans a line break is illustrated in Example 10.
5. Advantages and disadvantages of XML

Example 10 Miscellany of the Celtic Society

Misc. Celt. Soc. <br column="30" line="30"> 166. 3.

In this example the line break occurs immediately after the title, and the empty element splits the title and pagination ranges of the bibliographical citation.

5.5.3 Predictable exceptions type B

The second type of exception involves the element <sup> (encoding a superscript numeral) occurring in the pagination section of the citation.

Example 11 Windisch and Stokes’s Irische Texte

IT ii<sup>2</sup> 245. 10

In DIL, superscript characters usually occur immediately after roman numerals as in a citation of Windisch and Stokes’ Irische Texte illustrated in Example 11.

Example 12 Glosses on Priscian, St Gall

Sg. 204<sup>b</sup>6.

A citation of the Glosses on Priscian, St Gall (Example 12), illustrates that the <sup> element can also be placed after Arabic numerals.
5. Advantages and disadvantages of XML

5.5.4 Predictable exceptions type AB

It is also possible, as Example 13 illustrates, to have a combination of the predictable deviations illustrated in Examples 10-12.

Example 13 Citation of the Codex Paulinus Würzburgensis MS

\[\text{Wb. <br column="1" line="34"/>10<sup>a</sup>10.}</verbatim

This citation of the Codex Paulinus Würzburgensis MS illustrated above, spans both a line break and a column break and has a superscript numeral in its pagination range.

5.5.5 Mark-up that cannot feasibly be automated

In DIL, citations can also deviate from the patterns illustrated above in a way that is in theory predictable but in practice difficult to cater for. Examples of such citations are given in Example 14

Example 14 Citations with empty elements

\[\text{C 2.65, <br column="1" line="29"/>75, 3.4.2}
\text{I 184.79-81, 185.19, <br column="1" line="27"/>20, 21, 24, 37.}
\text{B iii 1, 78b<br column="3" line="22"/>6}
\text{RV iv <br column="3" line="22"/>252.23}
\text{GJ 1905, <br column="14" line="56"/>715.116}

In such citations, the empty element can intervene at any point in the pagination range. Because the pagination range is thoroughly unpredictable it was decided that attempts to write regular expressions
to cover all the possible pagination ranges and all the places that an empty element could occur in such ranges would be an unreasonable dedication of time and resources.

5.5.6 Variations within types of citation

In addition to deviations illustrated in § 5.5.1; § 5.5.2; § 5.5.3; § 5.5.4, which are to some extent predictable, the editors of DIL did not adopt a consistent editorial policy. Thus, there is little consistency in the way a work is cited in the different fascicles, and in some instances even in the same fascicle. Example 15 illustrates the inconsistent use of abbreviations of titles of works. It is not unusual for the same work to be cited in many ways in the same fascicle.

Example 15 Inconsistencies in bibliographical citations

| L. Fl. 33 (88)          |
| LFl. 56 (77)            |
| Lib. Flav. ii 6 b 2. 8; |
| Lib. Fl. Fergusiorum 674|
| L.Flav.Ferg ii f. 52 (105)<sup>r</sup>a |
| L.Flav. Ferg. iii 567  |

As illustrated in Example 15, the *Liber Flavus Fergusiorum* can be cited in six ways that are known to me, and possibly more.
5.6 Methodology

As a more nuanced understanding of the patterns of citations was developed, it became clear that a methodology that could reflect the patterns illustrated above had to be developed. As the methodology was developed it furthermore became apparent that recognising the manner in which patterns overlapped with one another was an integral part in successfully developing a routine to automate such mark-up.

5.6.1 Routine 1.0

Based on the abbreviation page in DIL a number of predictable citation types were identified and scripts were written to capture these.

Example 16 *Buile Suibni* (I)

```xml
s'\(Buile Suibni)\ \{([$]$*\ *[0-9][0-9]*)\ *\$*\ * \\
[0-9]*\.*\$*[a-z]*\*\ *[0-9][0-9]*)\'
<bibl><title>\1</title> <biblScope>\2</biblScope></bibl>`

s'\(Buile Suibni)\ \{([$]$*\ *[0-9][0-9]*)\ *\$*\ * \\
[0-9]*,\$*[a-z]*\*[0-9][0-9]*)\'
<bibl><title>\1</title> <biblScope>\2</biblScope></bibl>`

s'\(Buile Suibni)\ \{([$]$*\ *[0-9][0-9]*)\ *\$*\ * \\
[0-9]*\.*\$*[a-z]*\*[0-9][0-9]*)\'
<bibl><title>\1</title> <biblScope>\2</biblScope></bibl>`

s'\(Buile Suibni)\ \{([a-z][a-z]*\.*\ *[0-9][0-9]*)\ *\$*\ * \\
[0-9]*\*\*[0-9][0-9]*)\'
```
In routine 1 the title portion of the citation was known. The pagination section was more variable and could contain references to sections of the text (these are marked with §) or to sections of the text that are specified with Roman numerals or Arabic numerals. Based on the lessons learned at an early stage of the project when the regular expressions were too general and captured information that was not part of the bibliographical
5. Advantages and disadvantages of XML citations, a more restricted pattern match was developed. As illustrated in Example 16 this involved specifying three different types of pagination ranges and permitting three slight variations within each range such as different combinations of punctuation.

5.6.1.1 Routine 1.1

Routine 1.1 is based on Example 15 illustrated above.

Example 17 Buile Suibni (II)
5. Advantages and disadvantages of XML

In this step, the title range of the citation is also unpredictable due to the varying ways that citations can be supplied throughout DIL. In order to reduce the number of pieces of text that could be encoded incorrectly by the routines, it was decided to make this routine as specific as possible and in the case of, for example, Buile Suibni the upper case B and upper-case S were made obligatory, as illustrated in Example 17.
5. Advantages and disadvantages of XML

5.6.1.2 Routine 1.2

As illustrated in Example 9 empty elements can also occur in a title of a work.

Example 18 Buile Suibni (III)

```
s'\([B\][a-zA-Z]*\ <[^>]*>\[S\][a-zA-Z]*\.*\)\ <\([^>]*\)[S\][a-zA-Z]*\.*\)
\ *\$*\ *[0-9]*\ *\$*[a-zA-Z]*\#*\ *[0-9]*\ *[0-9]*\)
<bibl><title>\1</title> <biblScope>\2</biblScope></bibl>
```
5. Advantages and disadvantages of XML

In this example, the title portion of the regular expressions will match an upper case B, followed by zero or more lower case letters, followed by a space, followed by an empty element and then an upper-case S followed by zero or more lower case letters and zero or more fullstops.

5.6.1.3 Routine 1.3

As illustrated in section Example 10 empty elements could also occur after the title of a work and immediately before the pagination range.

Example 19 Buile Suibni (IV)
5. Advantages and disadvantages of XML

s'\([B][a-z]*\ [S][a-z]*\.*\) \(</[^>]*>\)\(([S]\$*\ *[0-9][0-9]*) \*\$*\ *[0-9]*,\*$*[a-z]*#[ ]*[0-9]*\)*\n\<$[a-z]*\*$*[a-z]*#[0-9]*\ *[0-9]*\)*\n<bibl><title>\1</title> \2<biblScope>\3</biblScope></bibl> g

s'\([B][a-z]*\ [S][a-z]*\.*\) \(</[^>]*>\)\(([S]\$*\ *[0-9][0-9]*) \*\$*\ *[0-9]*\ *\$*[a-z]*#[0-9]*\ *[0-9]*\)*\n<bibl><title>\1</title> \2<biblScope>\3</biblScope></bibl> g

s'\([B][a-z]*\ [S][a-z]*\.*\) \(</[^>]*>\)\(([a-z][a-z]*\.*\ *[0-9]*\ *[0-9]*\ *\$*[a-z]*#[0-9]*\ *[0-9]*\)*\n<bibl><title>\1</title> \2<biblScope>\3</biblScope></bibl> g

s'\([B][a-z]*\ [S][a-z]*\.*\) \(</[^>]*>\)\(([a-z][a-z]*\.*\ *[0-9]*\ *[0-9]*\ *\$*[a-z]*#[0-9]*\ *[0-9]*\)*\n<bibl><title>\1</title> \2<biblScope>\3</biblScope></bibl> g

s'\([B][a-z]*\ [S][a-z]*\.*\) \(</[^>]*>\)\(([a-z][a-z]*\.*\ *[0-9]*\ *[0-9]*\ *\$*[a-z]*#[0-9]*\ *[0-9]*\)*\n<bibl><title>\1</title> \2<biblScope>\3</biblScope></bibl> g

s'\([B][a-z]*\ [S][a-z]*\.*\) \(</[^>]*>\)\(([a-z][a-z]*\.*\ *[0-9]*\ *[0-9]*\ *\$*[a-z]*#[0-9]*\ *[0-9]*\)*\n<bibl><title>\1</title> \2<biblScope>\3</biblScope></bibl> g

s'\([B][a-z]*\ [S][a-z]*\.*\) \(</[^>]*>\)\(([a-z][a-z]*\.*\ *[0-9]*\ *[0-9]*\ *\$*[a-z]*#[0-9]*\ *[0-9]*\)*\n<bibl><title>\1</title> \2<biblScope>\3</biblScope></bibl> g

s'\([B][a-z]*\ [S][a-z]*\.*\) \(</[^>]*>\)\(([a-z][a-z]*\.*\ *[0-9]*\ *[0-9]*\ *\$*[a-z]*#[0-9]*\ *[0-9]*\)*\n<bibl><title>\1</title> \2<biblScope>\3</biblScope></bibl> g
5. Advantages and disadvantages of XML

In this example, the title portion of the regular expressions will match an upper case B, followed by zero or more lower case letters, followed by a space, followed by an upper-case S, followed by zero or more lower case letters, followed by zero or more spaces, zero or more full stops and an empty element.

5.6.1.4 Routine 1.4

In order to prevent the scripts becoming too long to work with from the perspective of the developer, elements such as `<sup>` occurring in the pagination ranges of citations (see § 5.5.3) were replaced with characters that were not used anywhere else in the text. In this case the element `<sup>` was replaced with a dollar sign and `</sup>` was replaced with a hash sign. These characters were also incorporated into the regular expressions to ensure that the encoded superscript character was included in the portion of the pagination range, as is illustrated in the examples above.

5.6.2 Routine 2

Routine 2 focused on developing regular expressions to match the citation patterns illustrated in Example 7. These regular expressions were further developed to match the predictable deviations illustrated in Example 9 and Example 10 for example:
Example 20 Step 2 Regular Expressions (I)

```
s'\([A-Z][A-Z][a-z][a-z]*\.*\)\[(\$)*\*[0-9]*\] \((\$)*\[0-9]*\)'
<bibl><title>\1</title> <biblScope>\2</biblScope></bibl>'
```

The regular expression illustrated in Example 20 matches references to
inter alia, CCath. (In cath catharda, Stokes, IT iv), TSh. (Keating’s Three
shafts of death, Atkinson, Ir. MSS. Ser. ii, 1890. Bergin 1931) TTr. (Togail
troy, (Stokes 1881)), BColm. (Betha Colmáin maic Lúacháin, Meyer, Todd
Lect. xvii, 1911), BCr., (Glosses on the Carlsruhe Beda, Thes. ii) PCR.
(Glosses on Priscian, Carlsruhe, Thes. ii) TFrAg. (Annals of Ireland, Three
Fragments, O’Donovan 1860)

Example 21 Step 2 Regular Expressions (II)

```
s'\([A-Z][a-z][a-z]*\.*\)*\([A-Z][a-z]\)\[(\$)*\*[0-9]*\] \((\$)*\[0-9]*\)'
<bibl><title>\1</title> <biblScope>\2</biblScope></bibl>'
```

Example 21 matches references to inter alia, Misc.Arch.Soc. (Miscellany
of the Archaeological Society, O’Donovan, 1846), Misc.Hag.Hib (Miscellanea
hagiographica Hibernia, Plummer, 1925) Misc. Celt. Soc the (Miscellany of
the Celtic Society, O’Donovan, 1849) Ir. Astr. Tr. (An Irish Astronomical
Tract, Power, ITS 14, 1914) Thr. Ir. Gl. (Three Irish Glossaries, Stokes
1862) Bard. Synt. Tr. (Bardic Syntactical Tracts, McKenna 1944) Phil. Soc.
Trans. (Transactions of the London Philological Society, 1854-)

277
Regular expressions were also developed to match ‘orphans’ or pagination ranges that are separated from their titles due to intervening commentary in DIL. Unique id numbers were generated for titles, and id pointers were generated for orphans (with XSLT) to ensure that pagination ranges can be explicitly linked with titles.

5.6.3 Routine 3 and sed’s pattern space

The way sed executes a script has important implications for the order in which the sed scripts are processed and thus for the desired accuracy levels that will be achieved in the output file. This is due to sed’s pattern space, ‘a workspace or temporary buffer where a single line of input is held while editing commands are applied’.22

In practice this means that if a sed script contains twenty regular expressions, each regular expression will be applied sequentially to the single line of input that is being held in the pattern space until all of the routines specified have been run. The line of input that was in the pattern space is then output and the next line from the input file is read into the pattern space. This process repeats itself until the end of the input file is reached.

Thus, tests showed that the order in which the sed scripts were applied to the input file was very important because it recognised that some of these scripts may overwrite each other.

For example, the Irish-English Dictionary prepared by O’Brien and

---

first published in 1768 is usually cited in the *DIL* as O’Br. However, depending on the order in which the scripts are processed, a script that matches patterns such as *Gault.*, an Irish version of *Gaulterus de dosibus* would catch the Brien portion of the O’Brien citation. In order to prevent errors such as this, a set of regular expressions was written that specify an initial capital O followed by an apostrophe, and furthermore in light of the pattern matching mechanism of *sed*, these regular expressions are applied before the other completely symbolic matches.

### 5.6.4 Routine 4

Patterns were also identified that matched a bibliographic citation in terms of the way it was constructed, but had a different function in the text. For example, the common cross-reference in *DIL See 3b* was identified as a pattern that would be matched by the routine, but a machine cannot recognise that this is not a bibliographical citation. Therefore *sed* scripts were placed at the head and foot, for example, Example 22 of the routine which replaced words such as ‘See’ with characters that were not used in the text and were not included in Regular Expressions. Thus minimising further the amount of text that can be incorrectly encoded by the routines.

**Example 22 Character replacement**

```
s
  See
  /
/
  ¬
  &
  /
  g

s
  see
  /
/
  ¬
  &
  /
  g
```
5. Advantages and disadvantages of XML

5.6.5 Accuracy rates

In C.xml, for example, the encoding of 41421 titles and 5018 orphans was automated. If it is estimated that it would take an encoder approximately one minute to read through a paragraph of text, find a bibliographical citation and encode it, and approximately thirty seconds to locate an orphan and encode it then the amount of time that was saved through this research, based on thirty-three hours of work per week, equals approximately twenty-two weeks, for the letter C alone.

At this stage of the project the accuracy level is 91%. While this can only be described as partial automation, it has nonetheless considerably reduced the number of citations that have to be encoded manually.

---

23 This percentage was calculated by using a linux utility to count the number of bibliographical elements in C.xml that were generated by the routine. This figure was then subtracted from the final copy of C.xml that included both generated mark-up and manually applied mark-up.
CHAPTER 6

Conclusion

The work has sought to assess the technical and academic advantages and disadvantages of using the meta-mark-up language XML to encode the lexicography of Old, Middle and Early Modern Irish. In the course of this research, an electronic Lexicon of Old, Middle and Early Modern Irish was created. The advantages and disadvantages of XML, from both the academic and technical perspectives, are discussed in the context of the Lexicon. In addition, the technical advantages and disadvantages associated with XML are set out by demonstrating the extensibility and interoperability of XML, and describing tools that were designed to link the Lexicon with other XML conformant corpora and dictionaries.

By way of introduction, chapter 1 outlined the history of lexicography in Greece, Rome and Byzantium and also looked at some Old-English glossaries and medieval Latin dictionaries. Special attention was paid to the ordering systems used in such works. The major medieval Irish language glossaries: O’Mulconry’s Glossary, Sanas Cormaic, the Lecan Glossary, O’Davoren’s Glossary, Dúil Laithne and O’Clery’s Glossary were
then described, and particular emphasis was placed on establishing the ordering systems used in such works. It was argued that O’Mulconry’s *Glossary* has the most developed alphabetic ordering system, and that *Duil Laithne* is unique among Irish glossaries because it employs a thematic ordering system. The continuing importance and relevance of medieval Irish language glossaries was then illustrated by citing examples from *DIL*, where the editors depended on the glossaries for a citation or explanation of a word. The final section of this chapter set out a list of the printed dictionaries that were published between the manuscript glossaries and the publication of the first fascicle of *DIL*.

When work on this thesis began, *DIL* was identified as being particularly suitable for xml encoding. In effect it is the only historical dictionary of the Irish language available to scholars. Furthermore, the information contained in it is complex, and the structure and presentation of its information is complicated, unpredictable, ill-organised and unwieldy. Chapter 2 argued that some of the serious limitations of *DIL* result from its complicated gestation and editorial history. Therefore, its history was treated of in detail to gain a better understanding of the environment and circumstances that surrounded its compilation. The main limitations of the dictionary were identified and divided into two categories: limitations integral to the hard copy medium of the publication, and limitations that resulted from poor editorial policies.

Chapter 3 begins with a summary of the views of electronic text that have been put forward by various scholars. An examination of the different kinds of mark-up that can be applied to electronic texts follows.
6. Conclusion

The development of the OHCO theory of textuality—the theoretical foundation of mark-up schemes such as the TEI and the Davenport group’s Docbook—was then discussed. The history of XML is outlined and the question ‘What is XML?’ is answered. Finally, a well-known application of SGML/XML, the TEI is discussed.

Chapter 4 presents a detailed description of the xml mark-up used in the Lexicon. Examples of ten different types of lexical entry are presented, namely: a noun, possessive pronoun, personal pronoun, adjective, conjunction, verb, prefix, definite article, encyclopaedic entry, preposition. The elements and attributes that were created to describe the meta-data of the Lexicon are then discussed in detail. The grammar that defines the elements and attributes used in the Lexicon, the DTD, is then set out. Finally, diagrams give a graphic illustration the inter-relationships of the various units of the DTD with one another.

The advantages and disadvantages of using XML to encode an electronic Lexicon of Old, Middle and Early Modern Irish and to create tools to link the Lexicon with other XML encoded dictionaries and corpora are set out in chapter 5.

This thesis shows that the advantages of using XML to encode Irish language lexicographical material far outweigh the disadvantages, and it can be concluded that XML is especially suited to encoding data of this nature.

From the perspective of the academic, the main advantages of the use of XML is that it is both flexible and powerful. As a result, XML fully supports the encoding of complex meta-data, such as that illustrated in
Chapter four. The XML encoded Lexicon discussed above is the first of its kind to be made available to scholars and students of early Irish and it contains the essential data encountered by the scholar and researcher in the first instance. The ability of XML to encode its meta-data has resulted in the creation of a tool that offers unprecedented opportunities to scholars to search and analyse both the Lexicon itself, and also the texts they work with.

The main disadvantage of the application of XML to Irish lexicographical material has been described as the problem of date and context. This disadvantage means that, due to the power of XML and its related technologies, an encoded document can be transformed in ways that are not linguistically meaningful. Yet it is possible to overcome this disadvantage, and a remedy to this problem was proposed by creating an element that brings this possibility to the attention of the document engineer.

Nonetheless, depending on one’s perspective, the disadvantages associated with XML can also be viewed as advantages. The problem of date and context prompts an important insight into the influential OHCO theory of textuality, and a fundamental limitation of this theory has been identified. Though this limitation is established in the context of the Lexicon, it also has applications in the wider discipline of humanities computing and for anybody involved in migrating a hard copy dictionary to digital format. The problem of date and context makes clear the importance of establishing the context of content objects and the problem that their properties can be altered when text is transported from one container to
another. Furthermore, not only did the XML encoding of the Lexicon highlight some deficiencies in Irish language lexicography, but it also helped to establish some of the conceptual differences that exist between a hard copy medium and an electronic document.

From the technical perspective, one of the main advantages of using XML is its extensibility and therefore its ability to support transformations of data. In the case of the Lexicon, it was demonstrated that transforming the specially created XML mark-up illustrated in chapter 3 to TEI conformant mark-up is easily achieved with the programming language XSLT.

Another advantage of XML, one that also demonstrates its extensibility, is that it can be readily used in conjunction with other XML conformant corpora. The tools that have been described here represent innovations in Humanities Computing in Old Irish because they are the first tools of their kind to be made available to the discipline. Indeed, a diversity of tools has been designed for a diversity of uses.

The lookup mechanism described in § 5.3.2 was created to resolve some of the key limitations of DIL, such as the lack of cross references and some of the integral limitations of the hard-copy medium itself, such as the necessity of knowing a third person singular present indicative form in order to locate any verb. The result is a tool that resolves such problems and furthermore enables students and scholars to harness the resources of two digital text corpora simultaneously. While a piece of Javascript enables such an application, the role of XML is also vital. The interlink with the corpus is successful because swift answers to queries
are provided from an array of data that is highly structured. Precise and concise processing demands mean that answer time for the end-user is shortened.

A further tool has also been created, one that allows the resources of the eDIL project, CELT and the Lexicon to be used simultaneously. As was established in § 5.3.3, the way in which bibliographical data is cited in DIL is highly inconsistent and most unsatisfactory. Furthermore, many of the books and manuscripts cited in DIL are only available in the best reference libraries and this frequently forces researchers to invest much time and money accessing such resources. The tool discussed above enables researchers to follow a link from the eDIL, and to a lesser extent from the Lexicon, that resolves to the text in the CELT corpus. Furthermore, the longevity of this resource is ensured by the use of fixed linking and a lookup table on the server side which means that texts added to the CELT corpus over the coming years will also satisfy links in the eDIL and Lexicon.

The main technical disadvantage of using XML is that it does not yet offer full support for some aspects of text processing, as illustrated in section § 5.4. Nonetheless, it was possible to develop a program, using the Linux utility sed to achieve the required transformation. One can also argue that this lack of support is, in fact, an advantage, because it makes clear to humanities computing scholars the dangers of depending too heavily on one encoding or programming language. After all, humanities computing scholars should have a selection of programming languages and utilities in their tool box and be willing to be creative in the ways
they implement them.

To conclude, this work has sought to demonstrate that the application of XML to the lexicography of Old, Middle and Early Modern Irish is a fruitful one. Not only has an electronic Lexicon of Old, Middle and Early Modern Irish been created, but some important steps towards creating a first class electronic infrastructure for those involved in all aspects of Irish studies have been taken.
CHAPTER 7

Bibliography

7.1 Bibliographies consulted


McCarty, Willard, 1996 Selective bibliography for humanities computing
7. Bibliography

(Second edition)

http://www.cch.kcl.ac.uk/legacy/tmp/bib/index.html.


7.2 Guidelines, Technical Reports and Working Group notes


Goldman Oliver & Lenkov Dmitry, 2005. *XML Binary Characterization Measurement Methodologies*  

Kay, Michael, 2005. *XSL Transformations (XSLT) Version 2.0*  
http://www.w3.org/TR/xslt20/.


Williams, Stephen D. & Haggar, Peter, 2005. *XML Binary Characterization Measurement Methodologies*  
http://www.w3.org/TR/2005/NOTE-xbc-measurement-20050331/.

### 7.3 Manuscripts

Dublin, RIA, 1225 (olim D ii 1) al. *Book of Uí Maine.*


Dublin, RIA, 1234 (olim C. i. 2.).

Dublin, TCL, 1317 (olim H. 2. 15B).
7. Bibliography

Dublin, TCL, 1318 (olim H. 2. 16) al. *Yellow Book of Lecan*.

Dublin, TCL, 1319 (olim H. 2. 17).

Dublin, TCL, 1337 (olim H. 3. 18).


Dublin, TCL, 1363 (olim H. 4. 22).

Dublin, TCL, MS H. 1.13, 361-62.

Dublin, RIA, MS Stowe, C. i. 2.


London, BL, Egerton 1782.


7.4 Dictionaries & printed lexicographical sources


7. Bibliography


Dwelly E., 1901-11. *Illustrated Gaelic to English dictionary* Glasgow.


7. Bibliography


Pearson, Agnes Isabella, 1940. ‘A medieval glossary’, *Ériu* 8 61-87.


Stokes, Whitley, 1900. ‘Three Irish Medical glossaries’, *Archiv für celtische Lexicographie* 1, 325-347.


7. Bibliography

Whitehead, David, (ed). *Suda on-line: Byzantine lexicography*

http://www.stoa.org/sol/.

7.5 Secondary works


Adams, G. B., 1980. ‘Place-names from pre-Celtic languages in Ireland and Britain’, *Nomina* 4 43-63.


Atkinson, Robert, 1901. *Ancient laws of Ireland VI: glossary to volumes I-IV* Dublin.


7. Bibliography

to Kuno Meyer Halle a. S.

Best, R. I., 1945-1948. ‘On recent studies in the Academy’, Proc Roy Ir Acad (C) 51 14-34.


7. Bibliography


Daly Lloyd D., 1967. ‘Contributions to a history of alphabetisation in Antiquity and the middle ages’ (*Collection Latomus* 90) Brussells.

de Bhaldraithe, Tomáis, 1986. ‘Ker. mar nod i bhfoclóirí an Duinnínigh’, *Ériu* 47 13-16.


di Paolo Healey, Antonette, 1997. ‘Old English glossaries: creating a
vernacular’, *Computing in the Humanities Working Papers* B.36
http://www.chass.utoronto.ca/epc/chwp/healey/.


Dodds, Leigh, 2001. *Intuition and Binary XML.*


Flynn, Peter, 2002. ‘Is there life beyond the web?’, *Literary and Linguistic computing* 17 no. 1 April 49-60.

Flynn, Peter, 1995. *Making more use of markup*

Freter, Todd, 1998. *XML: mastering information on the web*
7. Bibliography


Grieve, Mrs M., 1931. *A modern herbal: the medicinal, culinary, cosmetic and economic properties, cultivation and folk-Lore of herbs, grasses, fungi, shrubs & trees with their modern scientific uses* New York.


Harold, Elliottte Rusty, 1999. *XML Bible* USA.

7. Bibliography


Hockey, Susan, 1999. Is there a computer in this class?.
http://www.iath.virginia.edu/hcs/hockey.html


Kuhn, Markus, 2005. *UTF-8 and Unicode FAQ for Unix/Linux*


Lemerle, Paul., 1986. *Byzantine humanism: the first phase; notes and remarks on education and culture in Byzantine from its origins to the 10th century*, tr. Helen Lindsay and Ann Moffatt Canberra.


Merrilees, Brian, 1996. ‘The shape of the medieval dictionary entry’
7. Bibliography

*Computing in the Humanities Working Papers.* B [CHWP.B], 15


Meyer, Kuno., 1907. ‘The sources of some Middle-Irish glossaries’, *Archiv für celtische Lexikographie* 3, 138-44.


Piez, Wendell, 2001. ‘Beyond the “descriptive vs. procedural”’
7. Bibliography

distinction’, *Markup Languages: theory and practice* 3 no. 2, 141-172.


Pokorny, Julius, 1930. ‘Keltisch-Baskisch-Hamitisches’, *Z Celt Philol* 18, 111.


http://dx.doi.org/10.1162/109966200750428722.


Stokes, Whitley, 1903. *A criticism on Dr. Atkinson’s glossary to volumes I-V*
of the Ancient laws of Ireland. London.


Thurneysen, Rudolf, 1892. ‘La terme iarmérla dans la grammaire Irlandaise’Revue Celtique 8, (1892) 267-74.


