# An Analyser and Generator for Irish Inflectional Morphology Using Finite-State Transducers

Elaine Uí Dhonnchadha

Schools of Computer Applications and Fiontar
Dublin City University
Glasnevin
Dublin 9

**Supervisors**: Prof. Josef van Genabith
and Dr. Caoilfhionn Nic Pháidín

A dissertation submitted for the degree of
Master of Science

July 2002

**Declaration**

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Master of Science is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: Elaine Uí Dhonnchadha ID No.: 84106140

Date: 17/9/2002

# Abstract

Computational morphology is an important step in natural language processing. Finite-state techniques have been applied successfully in computational phonology and morphology to many of the world's major languages. Celtic languages, such as Modern Irish, present unique and challenging morphological features that to date have not been addressed using finite-state technology. This thesis presents a finite-state morphology of Irish developed using Xerox Finite-State Tools. To the best of our knowledge, such a resource does not exist.

The computational model, implemented as a finite-state transducer, encodes the inflectional morphology of nouns, adjectives, and verbs. Other parts of speech are also included in the interests of language coverage. The implementation is a strictly lexicalised design: the morphotactics of stems and affixes are encoded in the lexicon using replace rule triggers. Word mutations are then implemented as a series of replace rules written as regular expressions. Both components are compiled into finite state transducers and then combined, to produce a single two-level morphological transducer for the language.

A major advantage of finite-state implementations of morphology is their inherent bi-directionality; the same system is used for both analysis and generation of word forms in the language.

This resource can be used as a component part in parsing and generation in natural language processing (NLP) applications, such as spelling checkers/correctors, stemmers and text to speech synthesisers. It can also be used for tokenising text, lemmatising, and as an input to automatic part-of-speech tagging of a corpus.

The system is designed for broad coverage of the language and this is evaluated by comparing it with a list of the 1000 most frequently found word forms in a corpus of contemporary Irish texts.

Finally, maintainability of the system is discussed and possible extensions to the system are suggested, such as derivational morphology and the inclusion of dialectal or historical word-forms.

## Acknowledgements

Sincere thanks to my supervisors Prof. Josef van Genabith (Computer Applications) and Dr. Caoilfhionn Nic Pháidín (Fiontar) for their essential advice and direction. Thank you to Dr. Eoghan Mac Aogáin (Institiúid Teangeolaíochta Éireann) for his encouragement and support. Go raibh míle maith agat. Thank you to Ken Beesley (Xerox Research Centre Europe) for his assistance and abundant advice. Gabhaim buíochas do An tOllamh Dónall Ó Baoill agus do Donla Uí Bhraonáin as a gcuid moltaí.

Buíochas go háirithe, do Ghearóid, cara mo chléibh, agus do Bhéibhinn agus d'Éanna, gliondar mo chroí, as bhur gcuid foighne.

Last, but not least, thank you to everyone who read various drafts, for your valuable comments and suggestions.

## Abbreviations

| | |
|---|---|
| Com | Common Case |
| Gen | Genitive Case |
| Nom | Nominative Case |
| Voc | Vocative Case |

| | |
|---|---|
| Caol | *Caolú*, Slenderising |
| Coim | *Coimriú*, Syncopation |
| Lea | *Leathnú*, Broadening |
| Sé | *Séimhiú*, Lenition |
| Urú | *Urú*, Eclipsis |

| | |
|---|---|
| FST | Finite State Transducer |
| FSA | Finite State Automaton |
| TM | Turing Machine |

| | |
|---|---|
| NLP | Natural Language Processing |

## Typographical Conventions

All non-English language examples in the text are in *italic* typeface followed by the translation in single quotation marks, e.g. Irish: *cos* 'foot'. The language is identified in the example or in the accompanying text e.g. "in the following example from Irish...".

Single quotation marks are also used to highlight English words described in the text e.g. the plural of 'woman' is 'women'.

When a particular segment of a word is being discussed it is highlighted using **bold** typeface e.g. Irish: *c**athair*** 'city'.

Where the language of the example is not explicitly indicated it may be assumed to be Irish.

# Table of Contents

## APPENDICES

## List of Tables

## List of Figures

# 0. Introduction

This thesis presents an inflectional morphological analyser and generator for Irish using finite-state transducers. To the best of our knowledge such a system does not exist. The inflectional morphology of Irish verbs, nouns, adjectives, and conjugated prepositions, is modelled as a finite-state morphology. Morphological analysis is a fundamental component of many natural language processing systems, e.g. parsers, grammar-checkers, text-to-speech synthesisers etc. Morphological analysis is of particular relevance to Irish due to the phenomenon of initial mutation. To give an example, it is not obvious that *(na) huibheacha* ‘eggs’ is related to the root *ubh* ‘egg’ and that a learner or a computational system must look up the form *ubh* in a dictionary.

The thesis is arranged in two parts: part one presents the relevant background material and part two details the implementation. Chapter one gives an overview of linguistic morphology including examples of morphological phenomena found in the languages of the world. The second chapter describes the morphological phenomena of Irish, which are modelled in the current implementation. The third chapter introduces two-level and finite-state morphology, and chapter four gives an overview of finite-state technology.

Part two details the design and implementation of a finite-state lexical transducer for Irish for inflectional morphological analysis and generation. Chapter five describes the implementation of a finite-state morphology for Irish. Chapter six discusses issues such as testing during development, assessing current language coverage of the system, guidelines for adding new items to the lexicon, and suggestions for possible extensions to the system.

An online version of the system may be accessed at http://www.ite.ie/morph.htm

# PART 1 Theoretical Background

# 1. Morphology

> *"Morphology is truly a crossroads of linguistics (Bauer, 1983, p6); it is not really a field unto itself, because to understand word structure one has to understand many things that touch on areas outside morphology."*
>
> (Sproat, 1992, p123)

## *1.1 Introduction*

This chapter gives a broad overview of the range of morphological phenomena found in the languages of the world in order to provide a context for the morphological features of Irish described later in this work. The internal structure of words, the types of morphological phenomena which exist in various languages, and how morphology relates to other sub-disciplines of linguistics will be described.

## *1.2 Linguistic Morphology*

Fig. 1 shows where morphology lies in the field of linguistics. The main sub-disciplines of linguistics range from the physical production of speech sounds through to the more abstract interpretation of those sounds.

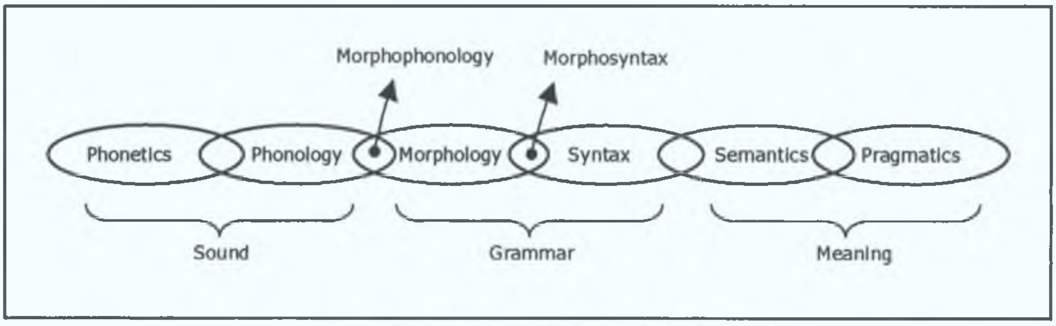


**Fig 1. Areas of linguistic study**

The study of language can be divided into three broad categories (Crystal, 1997b, p83):

a) medium of transmission
b) grammar
c) meaning.

Speech is the most common medium of language transmission. Phonetics and phonology examine speech sounds and their properties.

Grammar encompasses morphology and syntax, which deal with structure and arrangement of elements of language. Morphology is the study of the internal structure of words, and syntax examines structure at the sentence level.

Semantics deals with the meanings of words and sentences, and pragmatics focuses on language usage.

As can be seen from Fig. 1 there are not discrete areas of study with clearly defined boundaries. In fact, the overlaps themselves constitute areas of study. Morphophonology studies the interaction between phonological and morphological phenomena. Morphosyntax is the term covering features which are both morphological and syntactical in nature.

In the following sections morphological phenomena will be discussed under the following headings:

- Word structure, i.e. how a word is constructed, its constituent parts: morphemes
- Word formation, i.e. inflectional and derivational morphology
- Interactions,i.e. how morphology related to phonology, orthography, syntax and semantics.

Finally the area of computational morphology is introduced.

## 1.3 Word Structure

Morphology is the study of the internal structure of words. Some common functional words, such as 'and' and 'the', cannot be further analysed, but many words can be sub-divided into smaller meaningful parts. The smallest meaningful parts into which a word can be divided are called **morphemes**. In example (1) *cosa* 'feet' is a plural noun composed of two morphemes. The first morpheme *cos* 'foot' is a noun stem and is followed by a plurality morpheme represented by the suffix *'a'*. Morphemes are abstract units describing the composition of the word, and **morphs** are the actual realisation of those morphemes (Bauer, 1988, p11).

| | | | |
|---|---|---|---|
| (1) | *Orthographic form:* | *cosa* | |
| | *Morphs:* | *cos* | *a* |
| | *Morphemes:* | *cos*+Noun | +Pl |
| | *Gloss:* | 'feet' | |

There are two distinct types of morpheme: those found in **stems,** and those found in **affixes**. A stem morpheme can exist independently as a **free morpheme**, as in *féach* 'look' (2a), or with an affix as in (2b). An affix must be attached to a stem and is therefore known as a **bound morpheme**, e.g. the verbal suffix *-faidh* (2b) can not occur alone. An affix usually requires certain properties to be present in a stem before it can attach to it (Sproat, 1992, p79). For example, some affixes attach only to nouns and others only to verbs.

(2) a. *féach* 'look'

b. *féachfaidh* 'will look'

c. *feicfidh* 'will see'

If a particular morpheme has more than one actual realization, these are known as **allomorphs**. In (2b) and (2c) we can see that, depending on certain characteristics of the stem, different orthographic[1] forms of the future tense verbal suffix are used, i.e. either *–faidh* or *–fidh.*

There are a great many ways in which words are formed from morphemes but the processes involved can all be assigned to one of the following categories:

- Affixation of Stems
- Modification of Stems
- Compounding of Stems.

Word-forms are created using one or more of these methods. Affixation requires adding new material (affixes) to the stem, and these affixes belong to a closed class for the language. Modification entails making changes to the stem itself either through replacement, removal, or insertion of stem material. The actual form of the change is often lexically determined, i.e. it depends on the stem in question. Compounding, which involves joining stems to create a new word, will be discussed later under derivational morphology.

**Affixation of Stems**

An affix can attach to the beginning of a word as a **prefix**, or to the end as a **suffix**. An example of a prefix, *réamh-* 'pre-' is given in (3a), while (3b) and (3c) are examples of suffixes. (3b) *-faidh* is a future tense inflectional suffix, and (3c) *–án* is a diminutive derivational *suffix*.

Some affixes are inserted into the word and others surround the word. The former are known as **infixes,** e.g. *-um-* (3d) (O'Grady and de Guzman, 1997, p139), and the latter as **circumfixes**, like *ge-t* (3e) (Bauer, 1988, p23). A **transfix** is a type of affix found in Arabic and Semitic languages, shown in example (3f) (Bauer, 1988, p24) and discussed further below. An **interfix** is a morpheme added to a compound word, an example of which is given in (3g) (Bauer, 1988, p24).

(3) a. Irish: *léiriú* 'demonstration or presentation' ***réamh**léiriú, réamh+léiriú* 'rehearsal'

b. Irish: *déan* 'do' or 'make', *déan**faidh**, déan+faidh* 'will do ' or 'will make '

c. Irish: *leabhar* 'book' *leabhr**án**, leabhar+án,* 'booklet'

d. Tagalog (Philippine language): *lakad* 'walk', *l**um**akad : l+um+akad* 'walked'

e. German: *fragen* 'to ask', ***ge**frag**t**, ge+frag+t* 'asked'

f. Egyptian Arabic: *ktb* 'write, *k**a**t**a**b, k+a+t+a+b* 'he wrote'

g. German: *tag* , 'day', *licht*, 'light', Tag**es**licht, 'daylight', *tag+es+licht* .

The extent to which morphology uses affixing varies from language to language. Languages with very little use of affixes are described as **isolating** languages. In example (4) from Mandarin Chinese, the morpheme *le*, denoting past tense, is not attached to the verb *chi* (Steinbergs, 1997, p380).

(4) *Ta chi fan* ***le***
he eat meal past
He ate the meal

At the opposite end of this spectrum are **polysynthetic** languages, in which one word can represent a phrase or a whole sentence, and would be equivalent to several different words in another language. Inuktitut (5) (Steinbergs, 1997, p380) is an example of this.

(5) *Qasuiirsarvigssarsingitluinarnarpuq*
'Someone did not find a completely suitable resting place'.

**Agglutinating** languages like Turkish, add to the information conveyed by the word by attaching suffixes to the stem in a manner which is sometimes described as beads on a string (Sproat, 1992, p44), as in the following example (6) (Lewis, 2000, p39).

(6) a. el
el+Noun
'hand'
b. el**ler**
el+Noun+Pl
'hands'
c. eller**im**
el+Noun+Pl+Poss
'my hands'

In other languages, particularly Indo-European languages, including Irish, a single affix can convey several pieces of information. This type of affix is sometimes called a **portmanteau morpheme**. In example (7) one verbal suffix, *–eamar*, conveys tense, mood, person, and number of the verb *rith* 'to run'.

(7) *rith+**eamar***
rith+Verb+PastInd+1P+Pl
'we ran'

Prefixing, suffixing and circumfixing are quite straightforward operations, in that the morphemes are concatenated to the start or the end of the stem - two easily identifiable locations in text that has already been tokenised into words. This is also known as **Item and Arrangement**, or **Concatenative** morphology.

However, many languages include non-concatenative morphological processes that are more complex, particularly from a computational point of view. For example, with infixing (3d) the specific location within the stem must be determined before the morpheme can be inserted.

Another very different style of morphology is the **root and pattern** (or **templatic**) morphology of Arabic and Semitic languages. An Arabic verb stem has a root of three consonants and a vowel pattern (8) (Sproat 1992, p50-51; O' Grady and de Guzman, 1997, p139). Bauer (1988, p24-25) describes this vowel pattern as a **transfix** morpheme. Rather than forming a contiguous segment, the vowels are interspersed among the root consonants according to a specific pattern.

(8) *k (v) t (v) b*
'to write'

**Modification of Stems**

All of the morphological processes mentioned so far involve adding specific material to the base or stem. Another important category of word-building techniques found in natural languages, involves making changes to the stem itself. These processes often occur in conjunction with affixation (Bauer, 1988, p27).

**Reduplication** involves duplicating all or part of the stem. The reduplicated part can be prefixed, suffixed, or infixed to the stem. Example (9), from Indonesian, shows full reduplication while example (10), from Javanese, shows partial reduplication (Sproat, 1992, p57). Both prefix the original stem.

(9) *orang*, 'man'
***orang**orang*, 'men'

(10) *mulari*, 'initiated man'
***mula**mulari*, 'initiated men'

Some morphological processes involve **internal changes** to segments of morphemes, in particular changes to vowels. **Ablaut** and **umlaut** are types of internal vowel change common to Germanic languages (Sproat, 1992, p61). When vowel change occurs as a result of assimilation to a following vowel it is called umlaut (Bauer, 1988, p27) as in the case of some irregular plurals in English. In (11), the following vowel, which originally caused the internal change, has now disappeared from the language (O'Grady and de Guzman, 1997, p141).

(11) **foot**

foot+Noun+Nom+Sg

**feet**

feet+Noun+Nom+Pl

Ablaut is the term given to vowel alternations used to signify grammatical differences. An example of its usage in Icelandic morphology is given in (12) (Bauer, 1988, p28).

(12) ***gef*** 'I give'

gef+Verb+Pres

***gaf*** 'I gave'

gaf+Verb+Past

Internal change is a common phenomenon in Celtic languages, and there are several processes by which **vowel changes** take place in Irish. They are classified as **final mutations** in Ó Siadhail (1989, p134) since they always affect the final syllable. Some examples of final mutations are given below. In (13), the Irish stem *cos* 'foot' undergoes an internal change whereby an *'i'* is inserted to signify the **slenderising** (Ó Siadhail, 1989, p135) of the final consonant of the stem, i.e. the palatalisation of the final velarised consonant (Stenson, 1981, p35). In (14), *féar* 'grass' is also slenderised, this is shown orthographically by replacing the *'a'* with an *'i'*. Examples of similar features are given for Welsh (15) (Thomas, 1992, p303), and Breton (16) (Ternes, 1992, p415).

(13) *cos*, 'foot'

*coise,* 'of the foot'

(14) *féar*, 'grass'

*féir* 'of grass'

(15) *sant* 'saint'

*saint* 'saints'

(16) *maen* 'stone"

*mein* 'stones'

The opposite to slenderising can also occur. The Irish example (17a), shows an instance of **broadening** of the final consonant, i.e. depalatalisation (Stenson, 1981, p35), or velarisation (Ó Baoill and Ó Riagáin, 1990, p187), in this case by deleting the *'i'*. In (17b), if the *'i'* were simply removed, no vowel would remain and so the *'i'* is changed to *'ea'*.

(17)a. *athair* 'father'

*athar* 'of the father'

b. *binn* 'peak'

*beanna* 'peaks'

A third internal change (or final mutation), known as **syncope**, also involves vowel deletion. In this case a vowel in an unstressed syllable (of a polysyllabic stem) is susceptible to deletion when a nearby syllable is stressed (Murray, 1997, p322). In Irish this occurs when an unstressed syllable is sandwiched between a stressed syllable and a suffix as in (18).

(18) *cathair* 'city'

*cathracha* 'cities'

*cathr+acha*

**Initial mutations** are phenomena which are typical of Irish, and Celtic languages in general (Stenson, 1981, p18). A number of different processes are involved: **lenition** and **s-prefixing** apply to consonant-initial stems, **vowel-prefixing** affects vowel-initial stems, while **eclipsis** applies to both types of stem.

**Lenition** is a morphophonemic initial consonant change. Othographically a 'h' is inserted immediately after the initial consonant of a word to indicate this change. Originally the trigger was phonetic in nature, causing consonants occurring in intervocalic position to change, either internal to the word or in the case of initial mutations when the preceding word ends in a vowel. These triggers, such as the final vowel of the preceding word (Bammesberger, 1983, p22), have since disappeared, and lenition is now used to signify many grammatical changes (Stenson, 1981, p18). In (19), lenition is used to denote the genitive case of a masculine noun, and in (20), lenition is used when a feminine noun is preceded by the definite article.

(19) *cailín* 'girl'

*hata an chailín* 'the girl's hat'

(20) *bean* ' woman'

*an bhean* 'the woman'

Lenition can also occur word-internally. The second member of a compound word is nomally lenited as in (21).

(21) *príomhchathair*, 'capital city'

*príomh* 'main' + *cathair* 'city'

An example of **s-prefixing** is given in (22).

(22) ***sráid*** 'street'
*an* ***tsráid*** 'the street'

**Eclipsis** (like lenition) is a morphophonemic initial consonant change of initial stop consonants, the fricative 'f' (Stenson, 1981, p19), and initial vowels (Bammesberger, 1983, p23). It is often referred to as nasalisation. In Irish, it is denoted orthographically by placing an eclipsing consonant before the original consonant, as in (23a) and (23b). Unlike lenition, eclipsis never occurs word internally.

(23) a. ***bád*** 'boat'
*ar an* ***mbád***, 'on the boat'
b. ***tír***, 'country'
*i* ***dtír***, in a country'

Vowel eclipsing is shown in (24), which also has the plural suffix *'-eanna'* added. An example for Welsh is given in (25) (Davies, 1993, p110). In Welsh, the initial letter is replaced rather than prefixed as in Irish.

(24) ***áit***, 'place'
*na* ***n-á****iteanna*, 'of the places'

(25) ***cath*** 'cat'
*y* ***gath*** 'the cat'

Another morphological process involving change to a stem morpheme is **suppletion**. Here the morpheme is replaced by a phonologically unrelated morpheme, as in example (26), where the past tense of 'go' is 'went', and in Irish (27) the plural of *bean* 'woman' is *mná* 'women'.

(26) **go**
go+Verb+Present
**went**
go+Verb+Past

(27) ***bean,*** 'woman'
bean+Noun+Com+Sg
***mná***, 'women'
bean+Noun+Com+Pl

## *1.4 Word Formation*

Morphology is traditionally divided into two branches; inflectional morphology and derivational morphology. Inflectional morphology tells us about grammatical relationships between words, whereas derivational morphology tells us about lexical relationships.

### Inflectional Morphology

A common noun, such as ‘boat’, denotes a class of objects. This word can appear in various forms: boat, boats, boat's or boats'. Each form conveys different information about the noun and its relationship to other parts of the sentence. Likewise, a verb-form, as well as describing an action which took place, may also tell us when it took place, how many were involved, and whether the action is complete or ongoing. Words which change their appearance depending on the particular context in which they are used are said to be **inflected**. Inflectional morphology is the study of the forms that words in a word-class, (e.g. nouns), can assume. The particular word-form used depends on the context and the underlying grammatical and morphological rules. A **paradigm** is a template for the set of possible inflected forms for a word-class.

*Verbs*

The main inflections of verbs are **tense**, **mood**, **aspect**, **voice**, **person** and **number** . Tense describes the relative time at which the eventuality occurred. A verb can be marked as Past, Present or Future, as in English or Irish, but the degree of contrast can vary from language to language. Dyirbal, (Australia) distinguishes between Future and non-Future only, whereas ChiBemba (Zambia) distinguishes degrees of pastness and futurity with tenses such as Remote Past, Near Past, Immediate Past etc. (O' Grady and de Guzman, 1997, p170).

Mood indicates whether the verb expresses a statement (indicative mood), a wish or desire (subjunctive mood), or an order or instruction (imperative mood).

Aspect describes whether the eventuality is completed (perfect), incomplete or ongoing (imperfect), or happens regularly (habitual).

Voice indicates whether the subject took an active or passive role in the eventuality.

Verbal morphology often includes **agreement** between the verb and its arguments, in terms of person, number or gender. For example, in Indo-European languages, the verb agrees with the subject of the sentence, in number and person, but does so to varying degrees, as in (28). The Italian example is from O' Grady and de Guzman (1997, p168).

(28) English: I speak, you speak, she/he speaks
we speak, you speak, they speak

Italian: *parlo* 'I speak', *parli* 'you speak', *parla* 'she/he speaks',
*parliamo* 'we speak', *parlate* 'you speak' , *parlano* 'they speak'

Irish: *labhraím* 'I speak', *labhraíonn tú* 'you speak', *labhraíonn sí/sé* 'she/he speaks',
*labhraímid* 'we speak', *labhraíonn sibh* 'you speak' , *labhraíonn siad* 'they speak'

In Italian, since each form of the verb is marked differently for person and number, a subject (pronoun) is in effect discernible from the verb-form. Therefore *'Parlo Italiano'* means 'I speak Italian', and an explicit pronoun, as in *'Io parlo Italiano'*, 'I speak Italian', need only be used for emphasis. Languages with this ability to omit the pronoun are known as **pro-drop** languages (Sproat, 1992, p28). In Irish, a seperate pronoun, e.g. *mé* 'I', can never be used with the **synthetic form** *labhraím* 'I speak' (Sproat, 1992, p29), as *labhraím* is the synthesis of *labhraíonn + mé* and therefore already includes a pronoun. The separated form *labhraíonn mé* is also used.

*Nouns*

Nouns may be inflected for **number** or **case**. The number may be singular, plural or dual, and this is a semantic property of the noun. Case is a syntactic issue, where the form of the noun changes depending on its role in the sentence. For instance, the **nominative case** is used if the noun is the subject of the sentence. In Latin, the **accusative case** is used if the noun is the direct object of the verb, and the **dative case** is used if it is the indirect object. There are a wide variety of cases used in different languages. Finnish has fourteen cases: nominative, genitive, accusative, partitive, innessive, abessive, adessive, ablative, elative, illative, allative, prolative, translative and instrumental (Sproat, 1992, p31).

The gender of a noun tends to be an inherent lexical property of the word. Nouns can be masculine, feminine or neuter. Although in Irish, lexical gender usually follows semantic gender, there are exceptions. An example is given in (29) where the lexical gender is masculine, while the semantic gender is feminine.

(29) *cailín* 'girl'
noun, masculine

*Inflectional Morphology of Adjectives*

Adjectival inflection usually either marks contrast, or ensures agreement (gender, case or number) with the noun which it is qualifying. Example (30), gives the comparative degrees of 'hot'. In Irish, the comparative and superlative of adjectives have the same form, and are distinguished by use of the comparative particle *níos*, and superlative particle *is* (31).

(30) hot, base form
hotter, comparative
hottest, superlative

(31) ***te***, 'hot', base form
***níos teo***, 'hotter', comparative
***is teo***, 'hottest', superlative'

In the following example from Irish, the adjective agrees with the gender of the noun. After a feminine singular noun, e.g. *bean*, 'woman', the adjective *beag* 'small' is lenited (32a), but after a masculine singular noun, e.g. *fear* 'man', the adjective remains unchanged (32b).

(32) a. *bean* ***bheag*** 'small woman'
b. *fear* ***beag*** 'small man'

In general, inflectional morphology is productive[2] (Sproat,1992, p24), in that all new members of a word-class will undergo the standard inflections for that class. Some new words, and new uses of existing words, are given in (33).

(33) email (noun, singular), email**s** (plural), email**ing** (present participle)
fax (noun, singular), fax**es** (plural), fax**ing** (present participle)
junket (noun, singular), junket**s** (plural), junket**ing** (present participle)
text (noun, singular), text**s** (plural), text**ing** (present participle)

If nouns are marked for plural in a language then all new nouns must be capable of having a plural form. In some cases, however, the plural may have the same form as the singular – known as **syncretism** - as in (34).

(34) sheep (singular and plural)
information (singular and plural)

**Derivational Morphology**

Morphology can also tell us about word-formation rules for new words in a language. New words can be created through the addition of affixes to existing stems (possibly accompanied by modification to the stem) or through the joining of stems to create compounds. The derived word may belong to the same word-class, as in (35) and (36), or to a different word-classes as in (37). (In inflectional morphology the word-form always belongs to the same word-class).

(35) 'duck**ling**', noun,

'duck' noun + suffix '-ling'

(36) '**re**move' verb

'move' verb + prefix 're-'

(37) 'colour**ful**', adjective

'colour' noun + suffix '-ful'

Compounding combines stems to form new words as shown in example (38). In a compound, the new word assumes the word-class of the head of the word. In English, this is the last member of the compound (39). This is also the case in Irish (40).

(38) 'handbag'

'hand' noun, + 'bag' noun

(39) 'blackbird', noun

'black' adjective + 'bird' noun

(40) *príomhchathair*, 'capital city', noun

*príomh* 'main' adjective + *cathair* 'city' noun

Derivation is not productive to the same extent that inflection is. Derivational rules are optional within a word class; they may only apply to some members of a class (41), or they may be very productive, as in the English suffix '–less' (42) (Sproat, 1992, p35), and the Irish example '*–ach'* (43).

(41) host (noun)

host**ess** (feminine noun)

(42) penny (noun)

penni**less** (adjective)

leg (noun)

leg**less** (adjective)

(43) *dóchas*, 'hope' (noun)

*dóchas**ach***, 'hopeful' (adjective)

In Irish, verbs are commonly derived from nouns by adding verbal suffixes (Bráithre Críostaí, 1999, p250). The two most commonly used verbal suffixes are *–(e)áil* and *-(a)igh*.

(44) *idirdhealú* 'differentiation' (noun)

*idirdheal**aigh*** 'to differentiate' (verb)

(45) *plean* 'plan' (noun)

*plean**áil*** 'to plan' (verb)

The suffix *–(e)áil* is particularly productive and is frequently used with loan words to derive a verb or verbal noun (Stenson, 1981, p 18), as in (41).

(46) *clic 'click' (noun)*

*cliceáil* 'to click' (verb)

*ag clic**eáil*** 'clicking' (verbal noun)

A.J. Hughes (2001, p119) cites the following lines from a poem by Cathal Ó Searcaigh (1997, p134), which satirises the tendency to use loan words in Irish and the frequent use of the *-(e)áil* suffix.

> *Rinne sé an t-arasán a hoover**áil**,*
> *na boscaí bruscair a jeyes-fluid**eáil**,*
> *an* loo *a harpic**eáil**, an* bath *a vim**eáil**.*
> *Ansin rinne sé an t-urlár a flash**áil**,*
> *na fuinneoga a windowlen**eáil***
> *agus na leapacha a eau-de-cologn**eáil***[3].

Many prefixes and suffixes are used to derive nouns and adjectives in Irish. Some examples of the use of the prefix *frith-* 'anti-' are given in (47).

(47) *ábhar* 'matter' (noun)

***frith**ábhar* 'antibody' (noun)

*caitheamh* 'throw, cast' (noun)

***frith**chaitheamh* 'reflection' (noun)

*cosúil* 'like' (adjective)

***frith**chosúil* 'paradoxical' (adjective)

The diminutive suffixes *–án* and *–ín* are used in (48), and (49) shows some adjectives which are derived from nouns. A detailed list of derivational affixes can be found in Bráithre Críostaí (1999, p242-249).

(48) *leabhar* 'book' (noun)

*leabhr**án*** 'booklet' (noun)

*bád* 'boat' (noun)

*báid**ín*** 'little boat' (noun)

(49) *ór* ‘gold’ (noun)
*ór**ga*** ‘golden’ (adjective)
*áit* ‘place’ (noun)
*áiti**úil*** ‘local’ (adjective)

Verbal adjectives and verbal nouns are two frequently used derived word-forms in Irish. A number of different derivational suffixes are used, samples of which are given in (50) and (51). Bráithre Críostaí (1999, p250) may be consulted for a detailed description.

(50) *buail* ‘to beat, hit’ (verb)
*buail**te*** ‘beaten’ (verbal adjective)
*bual**adh*** ‘beating’ (verbal noun)

(51) *gluais* ‘to move’ (verb)
*gluais**te*** ‘moved’ (verbal adjective)
*gluais**eacht*** ‘movement, moving’ (verbal noun)

Derivational rules may be subject to etymological restrictions. In English, certain Latinate affixes can only attach to Latinate stems, as in (52), where a noun is derived by attaching the Latinate suffix ‘-ity’ to Latinate adjectives only (Sproat, 1992, p35).

(52) rare (adjective)
rar**ity** (noun)

Some derivational affixes are completely productive and apply to all stems. ‘Pro-‘ and ‘anti-‘ can be placed before most nouns given the right context (53) (Sproat, 1992,p25).

(53) **pro**-agreement
**anti**-war

A word may be both derived and inflected (54c). Derived words are subject to inflectional morphology. Derivational affixes or processes therefore must occur (54b) before inflectional rules can be applied.

(54) a. *bád* ‘boat’ (noun)
b. *báid**ín*** ‘little boat’ (derived diminutive noun)
c. *báidín**í***  ‘little boats’ (derived diminutive noun with plural inflection)

Some examples of Irish compounds involving two stems rather than a stem and affix(es) are given in (55) and (56) (Bráithre Críostaí,1999, p242).

(55) *fíon* 'wine' (noun)
*gort* 'field' (noun)
*fíonghort* 'vineyard' (compound noun)

(56) *úr* 'fresh, new' (adjective)
*scéal* 'story' (noun)
*úrscéal* 'novel' (compound noun)

As is the case with derivation, inflection (57c), where possible, takes place after compounding (57b).

(57) a. *luach* 'value', *liosta* 'list' (nouns)
b. *luachliosta* 'price-list' (compound noun)
c. *luachliostaí* 'price-lists' (compound noun with plural inflection)

## *1.5 Interactions*

In this section, the interactions between morphology and related linguistic levels of description are touched upon. 'Sandhi'[4] is a general term, covering a variety of linguistic phenomena, whose common factor is that they straddle two or more linguistic sub-disciplines. According to Andersen, the term sandhi, "refers to the interfaces between phonetics and phonemics, and between phonology and morphology..."(1987, p1). He cites Bloomfield's (1935) discussion of sandhi phenomena, ranging from "phonetics through morphophonemics and lexicalised 'included forms' to the expressions of grammatical content in the Celtic initial mutations" (Andersen, 1987, p1).

The interfaces and related sandhi phenomena which are discussed in the following sections are:

- morphology-syntax interface: **morphosyntax**
- morphology-phonology interface: **morphophonology**
- morphology-orthography interface: **morphographemics.**

Morphosyntax and morphophonology are indicated in Fig.1; morphographemics is the orthographic representation of morphophonology.

### Morphology-Phonology Interface

Morphology is closely linked to phonology, as ease of articulation has a direct bearing on the form of a word. Morphophonology (or morphophonemics) is the analysis of the phonological features which affect the appearance of morphemes (Crystal, 1997a, p250). Morphological rules cannot always be enforced due to phonological considerations. In Irish, the initial consonant of a noun is lenited (orthographically a 'h' is inserted) when preceded by the possessive determiner *mo* (also known as a possessive pronoun or possessive adjective) meaning 'my' (58a), but certain initial consonants, e.g. 'r', cannot be lenited resulting in exceptions, as in (58b).

(58)a. *cóta* 'coat'
*mo* ***ch****óta* 'my coat'
b. *rothar* 'bicycle'
*mo rothar* 'my bicycle'

The genitive case of some Irish nouns is formed by appending an '*e*' suffix as in (59) and also by slenderising (palatalising) the final consonant if necessary, as in (60). The slenderising of the final '*s*' of *cos* 'foot' is shown orthographically by inserting an '*i*' before the '*s*'. The final '*n*' of *seachtain* is already slender so no change is required. In such cases, the morphology and the phonology of the word are inter-related.

(59) *seachtain* 'week'
*na seachtain***e** 'of the week'

(60) *cos* 'foot'
*na cois***e** 'of the foot'

In some cases, there are alternative forms of a morpheme, **allomorphs**, depending on the phonetic environment (Dobrovolsky, 1997, p245). Example (61) gives two plural allomorphs of English. In the first case 's' is suffixed to the noun. However, in (61b) the sequence 'ch+s' would be too difficult to articulate, so the sequence 'ch+es' is articulated (and written) instead.

(61)a. street, street**s**
b. church, church**es**

**Morphology-Orthography Interface**

Many computational morphology systems process textual input, which means orthographic forms and not phonological forms are being analysed. The term morphographemics (Coates, 1994, p2603) is also used for the study of **orthographic morphophonology.**

The correspondence between orthography and phonology varies greatly from language to language. In Finnish, the orthography is closely related to phonology, so that in analysing the orthographic forms the system is also modelling the phonological rules. English orthography deviates from the phonetic realisation in many cases, as demonstrated in (62) (Sproat, 1992, p93). The final silent 'e', removed when '-ing' is suffixed, is not accompanied by a corresponding change in pronunciation.

(62) bake
bak**ing**

The vowel system of Irish "is highly redundant" according to Campbell (2000, p765). The long vowel /o:/ is represented not only by *ó* but also by *ói, eo, eoi, omh* and *omha(i).* These vowel combinations therefore must be considered in morphographemic rules relating to long vowels. For example, as well as accented vowels, these vowel combinations are stressed and will, in many cases, also resist syncopation.

**Morphology-Syntax Interface**

The form a word takes in a particular context is determined by agreement with other elements of the phrase or sentence. This interaction of inflectional morphology with syntax is called **morphosyntax**. According to Stenson (1981, p17), "the [Irish] language has an elaborate morphophonemic system; rules are operative primarily across word-boundaries" .

Phonological words and syntactic words do not always have a one-to-one correspondence. A phonological word can represent more than one syntactic word, as in the examples from English (63), German (64) (Sproat, 1992, p73), and Irish (65).

(63) **wasn't** : was + not

(64) ***am*** 'on the' : *an + dem*

(65) ***chugam*** 'to me' : *chuig + mé* 'to' + 'me'
***labhraímid*** 'we speak' : *labhraíonn + muid* 'speak' + 'we'

A **clitic** is a syntactically separate word which functions *phonologically* as an affix (Sproat, 1992, p73). Where it attaches to the end of the word it is known as an **enclitic**, as in example (63) above. In example (66), *d'* is a **proclitic** in Irish, i.e. a clitic which attaches to the start of the word.

(66) ***d'fhéach*** 'looked' : *do + fhéach*

Clitics can also attach to phrases, as in (67) (Bauer, 1988, p99), where the possessive marker "**'s**" relates to the whole phrase – not just the word 'white'.

(67) the woman in white**'s** face.

Some compounds are not hyphenated, or joined together to denote that they belong to one syntactic unit. A phonological word may represent just part of one syntactic unit, as example (68) (Sproat, 1992, p38) in English, and examples (69) and (70) in Irish show.

(68) **spark** as in 'spark plug'

(69) ***bunachar*** 'base' as in *bunachar sonraí* 'database'

(70) ***síos*** 'down' as in *cur síos* 'description'

The above morphological phenomena are just some of the issues which must be addressed by a natural language processing (NLP) system when parsing or generating text.

## *1.6 Summary*

In this chapter, I have outlined the internal structure of words, the types of morphological phenomena which exist in various languages, and how morphology relates to other sub-disciplines of linguistics.

# 2. Modern Irish

*A people without a language of its own is only half a nation.*
(Davis, 1914)

## 2.1 Introduction

The purpose of this chapter is to give some background information on Irish, and to outline the fundamental characteristics of the language. The chapter is divided into the following sections:

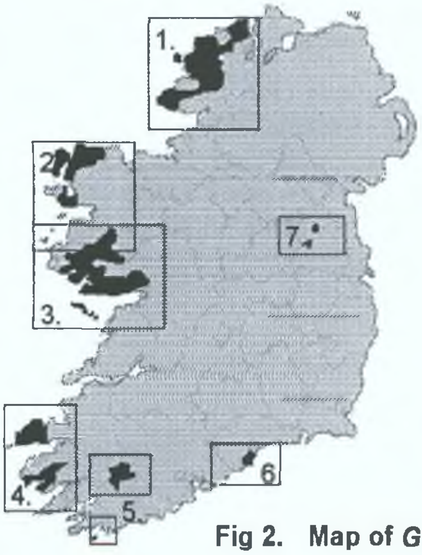
- background – the current status of Irish as well as the origins of the language
- phonology – vowels, consonants, diphthongs
- morphophonology – syllables, stress, word mutations
- inflectional morphology – how and when the various parts of speech are inflected.

## 2.2 Background

### Irish Today

Irish is the national language of Ireland; it is the first official language and English is recognised as a second official language (Bunreacht na hÉireann, 1937, Art. 8). In practice, Irish is the first language of a small percentage of the population, and surviving monolingual speakers of Irish, if remaining, are advanced in years.

According to the 1996 census of population (Central Statistics Office, 1998, p80) over forty percent of the population (3.5 million approx.) report an ability to speak Irish, and it is spoken on a daily basis by up to ten[5] percent of the general population (Central Statistics Office, 1998, p82). Irish is a required subject on the school curriculum at first and second level, and is studied by most students up to approximately the age of eighteen.



**Na Gaeltachtaí**

1. Donegal
2. Mayo
3. Galway
4. Kerry
5. Cork
6. Waterford
7. Meath

**Fig 2. Map of *Gaeltacht* areas**

(Source: Coimisiún na Gaeltachta, 2002)

The areas in which Irish is the primary language of communication are known officially as *Na Gaeltachtaí* and are located mainly on the western fringes of Ireland in Donegal, Mayo, Galway, Kerry and Cork. There are also the smaller Irish speaking areas of Ráth Cairn and Baile Ghib in County Meath, and An Rinn in County Waterford. Broadly speaking, there are  three main dialects; Donegal (Fig 2, area 1), Connemara (Fig 2, areas 2, 3 and 7) and Munster (Fig 2, areas 4, 5 and 6).

The number of Irish medium schools in the country has been steadily increasing over the last thirty years to the point where there are now as many such schools outside of the Gaeltacht areas as within[6]. In fact, because many of the non-Gaeltacht schools are located in more densely populated urban areas, in 1999/2000, they accounted for over seventy percent of pupils attending such schools (An Roinn Oideachais agus Eolaíochta, 2001, p27). It should however be remembered, that Irish medium primary schools account for only seven percent of the overall number of primary schools in the country. (An Roinn Oideachais, 2001, p27).

**Celtic and Indo-European Origins**

The known languages of the world have been categorised into language families (Ruhlen, 1987, p3). Irish belongs to the Indo-European family. The Indo-European family of languages covers territories stretching from Ireland to Assam, and from Norway and central Russia, to the Mediterranean, the Persian Gulf and Central India (Campbell, 2000, p738), as indicated on the map in Fig. 3.

Although Indo-European languages account for only approx 3% of five thousand or more recorded languages, they are spoken by half of the world's population (Ruhlen, 1987, p35), with the Chinese languages being the other major grouping.

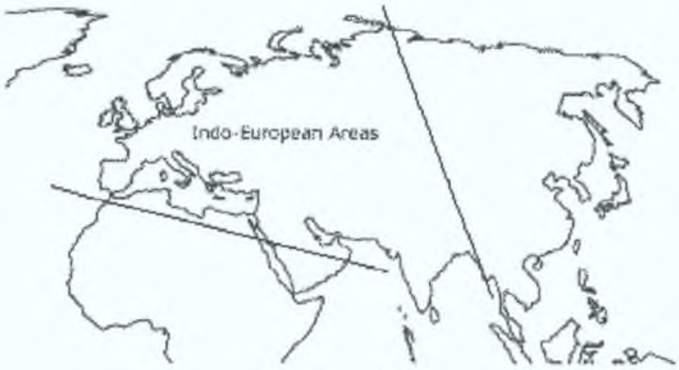


**Fig 3.   Map of indo-european territories**

(Source: http://www.worldatlas.com/clipart.htm)

Irish belongs to the Celtic branch of the Indo-European family of languages, as shown in Table 1. The name Celtic comes from the word "Keltoi" found in Greek texts dating from 500 BC, which was used to describe the people then occupying Central Europe (Fife, 1993, p3).

The Celtic languages are divided geographically into Continental and Insular Celtic. The Continental Celtic languages are now extinct, and all of the modern Celtic languages belong to the Insular Celtic group. This group is further sub-divided into Goidelic which includes Irish, Scots Gaelic and Manx, and Brittonic which includes Welsh, Cornish and Breton.

It is thought that the Gauls brought the ancestors of modern Irish speakers to Ireland circa 300 BC (Ó Siadhail, 1988, p1). Scottish Gaelic and Manx (Gaelic) resulted from changes, which occurred to the language of Irish Gaelic speakers, who first migrated to western Scotland and the Isle of Man around the fifth century (Russell, 1995, p9; Palmer, 1972, p388).

Breton originated with British speakers from southern England migrating to northern France in the fifth and sixth centuries (hence the area of Brittany) (Ternes, 1992, p371; Palmer, 1972, p388).

**Table 1. Indo-European Language Families**

1. Albanian<br>
2. Anatolian (extinct)<br>
3. Armenian<br>
4. Baltic<br>
5. Celtic<br>
5.1. Continental (extinct)<br>
5.2. Insular<br>
5.2.1.Brittonic<br>
5.2.1.1. Breton<br>
5.2.1.2. Cornish<br>
5.2.1.3. Welsh<br>
5.2.2.Goidelic<br>
**5.2.2.1. Irish (Gaeilge)**<br>
5.2.2.2. Manx (Gaelic)<br>
5.2.2.3. Scottish Gaelic<br>
6. Germanic<br>
7. Hellenic<br>
8. Indic<br>
9. Iranian<br>
10. Italic<br>
11. Slavonic<br>
12. Trochaic (extinct)

### Characteristics of Celtic Languages

Celtic languages share many Indo-European traits, such as having inflected parts of speech, a gender system and similar word roots. But they also have characteristics that set them apart from other Indo-European sub-groups. Some of the most distinguishing characteristics of Irish, and Celtic languages in general, are as follows: (Fife, 1993, p22)

- initial mutation of words
- verb – subject – object sentence word order
- prepositions inflected for person and number.

The sentence word order verb – subject – object (VSO) is relatively uncommon among the worlds languages, and no other Indo-European languages outside of the Celtic grouping share this word order (Fife, 1993, p23). A simple example of VSO syntax in Irish is given in (1).

(1) *Chaith Seán an liathróid.*
Threw Seán the ball.
'Seán threw the ball'.

## 2.3 Phonology

### Orthography

Traditionally eighteen letters represent fifty-one phonemes (An Roinn Oideachais, 1986a, p xii-xiii):

*a b c d e f g h i l m n o p r s t u.*

The remaining eight letters of the Roman alphabet,

*j k q v w x y z*

have been introduced in loan words, such as those in (2) to (6), and in Gaelicised versions of some foreign placenames. They do not undergo any of the initial mutations discussed below.

(2) *jíp* 'jeep'
(3) *quinín* 'quinine'
(4) *veilbhit* 'velvet'
(5) *xileafón* 'xylophone'
(6) *zú* 'zoo'

The information in Table 2, relating to the phonemes of Irish, is based on data in on *Foclóir Póca* (An Roinn Oideachais, 1986a, p. xv).

**Table 2. Phonemes of Irish**

| | | |
|---|---|---|
| 5 short vowels | 3 broad (back) | a o u |
| | 2 slender (front) | e i |
| 5 long vowels | 3 broad (back) | a: o: u: |
| | 2 slender (front) | e: i: |
| 1 neutral vowel | | ə |
| 36 consonants | 18 broad | b k d f g h l m n p r s t v (w) z ŋ γ x |
| | 18 slender | b′ k′ d′ f′ g′ l′ m′ n′ p′ r′ s′ t′ v′ z′ ŋ′ γ′ x′ d′z′ |
| 4 diphthongs | | uə, iə, ai, au, |

### Vowels

*Short and Long*

As shown in Table 2, the short vowels (unstressed vowels) are represented as: *a e i o u*.
The long vowels (stressed vowels) are most commonly represented as: *á é í ó ú* , but in certain contexts they are represented in a number of other ways as shown in Table 3 (Christian Brothers, 1988, p1).

**Table 3. Long Vowels**

| *Long Vowel* | *Orthography* | *Example* |
|---|---|---|
| /a:/ | *á*<br>*a* before *rd*, *rl*, *rn*<br>*a* before final *rr* | b*á*d 'boat'<br>ard 'high', *tharla* 'happen, *cearnóg* 'square'<br>carr 'car' |
| /i:/ | *í*<br>*i* before *á* or *ó* | *mín* 'smooth'<br>*fiáin* 'wild', *sióg* 'fairy' |
| /e:/ | *é*<br>*ae*, *ao* | *cé* 'who, quay'<br>*Gael* 'person of Irish race', *faocha* 'periwinkle' |
| /o:/ | *ó*<br>*omh*<br>sometimes *eo* | *mór* 'big'<br>*chomh* 'as'<br>*ceo* 'fog' |
| /u:/ | *ú*<br>*umh*<br>*u* before *á* or *ó* | *rún* 'secret'<br>*cumhacht* 'power'<br>*fuáil* 'sewing', *ruóg* 'cord' |

**Consonants**

*Broad and Slender*

Each consonant has a broad and slender version denoted orthographically by its adjacent vowel. For example, in (7), *'h'* and *'t'* of *hata* are broad, and in (8), the *'t'* and *'n'* of *tine* are slender. In example (9), *'b'* is broad and *'l'* is slender. In general, the vowels preceding and following a consonant must match, as is the case with *'d'* in (10) and (11).

(7) *hata* 'hat'
(8) *tine* 'fire'
(9) *buail* 'beat'
(10) *báidín* 'little boat'
(11) *éadach* 'cloth'

Phonetically, the slender consonants are shown with an accent following the letter, as is shown in Table 2.

*Broad and Slender*

The consonants are classified as either "broad" (back), or "slender" (front), which relates to the airspace created when articulating them (see Fig . 4 which is based on Annunciata le Muire and Ó Huallacháin (1966, p18,22). The broad vowels (back and centre vowels) are: *a á o ó u ú,* and the slender vowels (front vowels) are: *e é i í.*

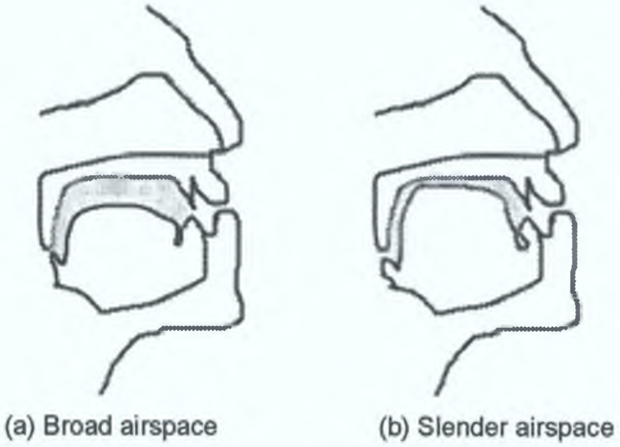

(a) Broad airspace (b) Slender airspace

**Fig 4. Articulation of broad and slender vowels**

Orthographic vowel harmony plays a central role in Irish morphographemics. When a suffix is added to a stem, either the stem is adjusted to match the broad or slender character of the suffix, or an allomorphic suffix (broad or slender) is used.

In (12), an example of verbal inflection, the future tense suffix *–faidh* is broad to agree with the broad stem *las* 'light'. In (13), the slender allomorph *–fidh* of the morpheme *–faidh* is used to agree with the slender stem *suí* 'sit'.

(12) las 'light'
*las+**faidh*** 'will light'

(13) suigh 'sit'
*suí+**fidh*** 'will sit'

In (14), an example of noun inflection, the genitive case of *súil* is *súile*; the slender '*e*' suffix agrees with the stem *súil.* In (15), adding the slender '*e*' requires that the broad stem *cos* be made slender to agree with the '*e*' suffix.

(14) súil 'eye'
*na **súile*** 'of the eye'

(15) cos 'foot'
*na **coise*** 'of the foot'

Plural noun suffixes, likewise, must have agreement with the preceding syllable. In (16) and (17), two allomorphs of the plural morpheme *–anna* are used, rather than adjusting the stem to accommodate a single suffix.

(16) carr 'car'
*carr**anna*** 'cars'

(17) seit 'set'
*seit**eanna*** 'sets'

**Diphthongs**

The four diphthongs in Irish (An Roinn Oideachais,1986a, p xiii; An Roinn Oideachais,1986b, p7; Russell, 1995, p108) are given in Table 4.

**Table 4. Diphthongs**

| *Diphthongs* | *Orthography* | *Example* |
|---|---|---|
| /ai/ | *adh(a)* | ***radh**arc* 'view', *T**adh**g* - a personal name |
| | *a(i)gh(a)* | *l**agh**ad* 'fewness', *s**aigh**diúir* 'soldier' |
| | *eidh* | *v**eidh**lín* 'violin' |
| /au/ | *abh(a)* | ***abh**ainn* 'river', *g**abh**a* 'smith', |
| | *(e)abh(a)* | *l**eabh**ar* 'book |
| | *amh(a)* | ***amh**arc* 'look', *s**amh**radh* 'summer' |
| | *ogh(a)* | *t**ogh**a* 'best' *r**ogh**a* 'choice' |
| /iə/ | *ia* | *sc**ia**n* 'knife', *b**ia*** 'food', *s**ia**d* 'them' |
| /uə/ | *ua* | *br**ua**s* 'lip', *díom**ua**ch* 'defeated', *s**ua**s* 'up' |

## *2.4 Morphophonology*

**Syllable Structure**

A syllable, in Irish, consists of at least one vowel (maximum three), which is preceded by up to three consonants, and followed by up to three consonants. The structure of Irish is given below (Ó Dochartaigh, 1992, p89) and parentheses indicate optionality.

$(C_1)$ $(C_2)$ $(C_3)$ $V_1$ $(V_2)$ $(V_3)$ $(C_1)$ $(C_2)$ $(C_3)$.

The words in examples (18) to (23) contain syllables ranging from minimum to maximum vowel and consonant clusters.

(18) *í* 'she, her, it (fem.)'
(19) *ae* 'liver'
(20) *treoir* 'direction'
(21) *sí* 'she, it (fem.)'
(22) *leanbh* 'child'
(23) *strogán* 'a stocky person'.

### Word Stress

It is usual for the first syllable of a word to be stressed (An Roinn Oideachas,1986a, p xvi), although in the Munster dialect, the stress is frequently on the final syllable (Ó Siadhail, 1989, p28-9). Also, in compound words (e.g. *anseo* 'here' *ansin* 'there') or loan words, stress is often on the second syllable.

### Inflectional Morphology

In common with most Indo-European languages, Irish is an inflectional language, i.e. it displays grammatical relationships morphologically. The suffix is the predominant type of affix used, although there are a number of proclitics used in verbal inflection. Prefixes are mainly used derivationally (Stenson, 1981, p17). Inflections also frequently include modification to the stem. These modifications are divided into two distinct categories - initial mutations, which affect the initial letter of the word, and final mutations, which affect the vowels of the final syllable, and in a few cases the consonants also. Initial and final mutation processes are listed below:

- Initial mutation
    - Lenition
    - Eclipsis
    - Prefixing of vowel-initial and s-initial words
- Final mutation
    - Slenderising
    - Broadening
    - Syncopation
    - Syllable replacement (change).

*Initial Mutations*

Many languages have phonological accommodation at the juncture of two words, but it is usually the end of the first word which is affected. In English, the definite article 'a' becomes 'an' before a word starting with a vowel sound. In French, the pronunciation of the word *les* also depends on whether the following word begins with a vowel or a consonant, as in (25) (Campbell, 2000, p324).

(24) a ball
an apple

(25) *les femmes* 'the women' /le:/
*les enfants* 'the children' /le:z/

Mutations in Irish also originate in phonological accommodations, but it is the initial syllable of the second word which is affected rather than the last syllable of the first word. As the language changed over time the conditions causing the mutations disappeared, but the mutations remained and became grammaticalised (Campbell, 2000, p324; Ó Cuív, 1987, p395-400; Russell, 1995, p237).

Irish morphology and syntax are inextricably linked. In many cases the inflected form of a word is as dependent on its syntactic relationship with a preceding word as on the phonological or lexical characteristics of the word itself.

Irish has a large number of nominal and verbal particles, most of which trigger initial mutations in the following word-form. (26) shows how articles, possessive determiners, numerals and simple prepositions all trigger initial mutation of a noun (Stenson, 1981, p 32).

(26)***b**róg* 'shoe'
*an **bh**róg* 'the shoe'
*mo **bh**róg* 'my shoe'
*seacht **mb**róg* 'seven shoes'
*i **mb**róg* 'in a shoe'
*ar an **mb**róg* 'on the shoe'

Nouns themselves can trigger mutations in other nouns (in compounds or genitive relations) and in adjectives. Example (27) shows how the gender of a noun influences initial mutation of a following adjective. *Bróg* 'shoe' is a feminine noun and therefore causes lenition of the adjective *beag* 'small', whereas *cat* 'cat', which is a masculine noun, does not.

(27)*bróg **bh**eag* 'small shoe'
*cat **b**eag* 'small cat'

There are a number of verbal particles[7] which trigger initial mutation in verbs, as shown in example (28). *Ní* is a negative particle requiring lenition of the following verb-form, and *an* is an interrogative particle requiring eclipsis of the following verb-form. (For a listing of the particles included in this implementation see Appendix I.)

(28)déanfaidh 'will do'
*ní **dh**éanfaidh* 'will not do'
an **nd**éanfaidh 'will do?'

*Final Mutations*

The final syllable of inflected stems (e.g. nouns, adjectives, verbs) are also subject to change. These changes (or mutations) involve altering the broad or slender character of the syllable, the removal of an unstressed syllable (syncopation), or changes to the final consonants.

Final mutations tend to be grammatically triggered. For example, a first declension noun (see Table 15) in the genitive case must end in a slender consonant. The exact method of slenderising depends on the particular word, as shown in (29) below. The most common method, that of inserting an '*i*' after the broad vowel, is given in (29a), and an alternative method is given in (29b).

(29) a. *bád* 'boat'
*báid* 'of boat' or 'boats'
b. *fear* 'man'
*fir* 'of man' or 'men'

In the following section, initial mutations (lenition, eclipsis and prefixing) and final mutations (slenderising, broadening and syncopation) are described. In each case there is a brief description of the phenomena together with some examples and an outline of the main grammatical contexts in which it is used.

**Lenition**

Lenition is the most common type of initial consonant mutation. Lenition is a phonological term used to describe softening (*séimhiú*) or weakening of a sound (Ó Siadhail, 1989, p340). Nine of the thirteen consonants of Irish are subject to lenition, which is orthographically denoted by placing a '*h*' immediately after the initial consonant. Lenition of the broad consonants only is shown Table 5.

**Table 5. Lenition**

| *Unlenited* | | *Lenited* | |
|---|---|---|---|
| *Orthographic* | *Phonetic* | *Orthographic* | *Phonetic* |
| b | /b/ | bh | /v/ or /w/ |
| c | /k/ | ch | /x/ |
| d | /d/ | dh | /ɣ/ |
| f | /f/ | fh | silent |
| g | /g/ | gh | /ɣ/ |
| m | /m/ | mh | /w/ or /v/ |
| p | /p/ | ph | /f/ |
| s | /s/ | sh | /h/ |
| t | /t/ | th | /h/ |

Lenition is not indicated orthographically for the following four consonants, although initial *l*, *n* and *r* can be lenited in speech (Connaught/Ulster dialects):

'*h*', '*l*', '*n*' and '*r*'.

Also, nouns beginning with the following letters are not lenited following the definite article *an* 'the'.

'*d*', '*t*', '*sl*', '*sn*', '*sr*' or *s*+vowel .

*Examples*

The following are some examples of lenition. In (30), a feminine noun is lenited after the definite article. In (31), a personal name is lenited in the vocative case.

(30) *cos* 'foot'
*an* ***chos*** 'the foot'

(31) ***Dónall***
*a* ***Dhónaill*** 'Dónall!' (vocative case)

*Grammatical Contexts*

Lenition occurs in a variety of grammatical contexts (Bráithre Críostaí, 1999, p 22-32), including the following:

a) a feminine noun preceded by the definite article *'an'*, (except for nouns starting with *'d', 't', 'sl', 'sn', 'sr'* or *'s'*+vowel), e.g. *an* ***bhróg*** 'the shoe'
b) a masculine noun in the genitive case preceded by the definite article *'an'*, (except for nouns starting with *'d', 't', 'sl', 'sn', 'sr'* or *'s'*+vowel), e.g. *eireaball an* ***chait*** 'the cat's tail'
c) a proper noun following another noun in genitive case relative to the first, e.g. *cóta* ***Ghráinne*** 'Gráinne's coat', *foireann* ***Shasana*** 'the English team', *Tom* ***Sheáinín*** personal name
d) nouns following the possessive pronouns such as *mo* (my), *do* (you) and *a* (his) etc., e.g. *mo* ***chóta*** 'my coat'
e) nouns following the numerals one to six i.e. *aon, dó, trí, ceathair, cúig, sé,* e.g. *sé* ***chapall*** 'six horses'
f) nouns in the vocative case (they are always preceded by the vocative particle *a*), e.g. *Tar anseo a* ***Sheáin*** (Come here Seán)
g) verbs after some verbal particles, e.g. *ní* ***bhíonn*** 'is not usually', *ar* ***bhris*** *(tú)*, 'did (you) break'
h) adjectives following a feminine noun, e.g. *an bhean* ***mhór*** 'the big woman'
i) adjectives following plural nouns ending in a slenderised consonant, e.g. *na cait* ***bheaga*** 'the small cats'
j) after some prepositions, e.g. *ar* 'on', *ar* ***bharr*** 'on top'.
k) the second part of compound words, e.g. *sráid**bhaile*** 'village', *an-**mhór*** 'very big', *droch-**chaoi*** 'bad condition'.

**Eclipsis**

Eclipsis is the second type of initial mutation to be described. It is denoted by prefixing a particular consonant to the initial consonant, or *'n-'* before a vowel. The prefixing consonant is pronounced, and the original consonant becomes silent (except for '*ng*'), though it remains in the orthographical form.

The vowels, five short and five long, and seven consonants, which undergo eclipsis are shown in Table 6. (Only the broad consonants are shown in the table.)

**Table 6. Eclipsis**

| *Uneclipsed* | | *Eclipsed* | |
|---|---|---|---|
| *Orthographic* | *Phonetic* | *Orthographic* | *Phonetic* |
| b | /b/ | mb | /m/ |
| c | /k/ | gc | /g/ |
| d | /d/ | nd | /n/ |
| f | /f/ | bhf | /w/ or /v/ |
| g | /g/ | ng | /ŋ/ |
| p | /p/ | bp | /b/ |
| t | /t/ | dt | /d/ |
| | | | |
| a | /a/ | n-a | /na/ |
| á | /a:/ | n-á | /na:/ |
| e´ | /e/ | n-e´ | /ne/ |
| é´ | /e:/ | n-é´ | /ne:/ |
| i´ | /i/ | n-i´ | /ni/ |
| í´ | /i:/ | n-í´ | /ni:/ |
| o | /o/ | n-o | /no/ |
| ó | /o:/ | n-ó | /no:/ |
| u | /u/ | n-u | /nu/ |
| ú | /u:/ | n-ú | /nu:/ |

The following consonants cannot be eclipsed:

'h', 'l', 'm', 'n', 'r' and 's'.

*Examples*

Some examples of the use of eclipsis are given below. The genitive plural form of the noun *cailín* is eclipsed, as in (32), nouns after the numerals seven to ten are eclipsed, as in (33), and nouns after some possessive adjectives are eclipsed, as in (34).

(32) cailín 'girl'
*na **g**cailíní* 'of the girls'

(33) *éan* 'bird'
*seacht **n-**éan* 'seven birds'

(34) *gluaisteán* 'car'
*a **ng**luaisteáin* 'their cars'

*Grammatical Contexts*

The following are some of the principal contexts in which eclipsis is used (Christian Brothers, 1988, p20-22):

a) nouns after plural possessive determiners i.e. *ár* 'our', *bhur* 'your', *a* 'their', e.g. *ár **d**teach* 'our house', *bhur ngáirdín* 'your garden', *a **g**cótaí* 'their coats'
b) nouns after the numerals seven to ten i.e. *seacht* 'seven, *ocht* 'eight', *naoi* 'nine', *deich* 'ten', e.g. *seacht **g**capall* 'seven horses'
c) nouns after the simple preposition *i* 'in', e.g. *i **d**teach* 'in a house, *i **m**bliana* 'in (this) year'. Note that *i* becomes *in* before a vowel-initial noun rather than eclipsing the noun, e.g. *in oráiste* not *i n-oráiste.* Likewise *in* is used before proper nouns e.g. *in Fiontar* not *i bhFiontar.*
d) genitive plural of nouns, e.g. *scoil na **g**cailíní* 'the girls' school', *ceol na **n-**éan* 'music of the birds'.
e) some prepositional phrases contain eclipsis, e.g. *ar **n**dóigh* 'indeed', *ar **g**cúl* 'behind', *ar **d**tús* 'at first', *go **bh**fios dom* 'to my knowledge'.
f) simple prepositions in combination with the article *an* 'the' cause eclipsis in a following noun, e.g. *as an **b**páirc* 'out of the field' *leis an **g**cat* 'with the cat'.
g) many pre-verbal particles cause eclipsis, e.g. *an **bh**faca* 'did you see', *cá **bh**fuil* 'where is', *nach **d**tuigeann tú* 'don't you understand', *mura **d**téann tú* 'if you don't go' *sula **n**geallann tú* 'before you promise', *dá **m**beadh sé agam* 'if I had it'.

**The Prefix t**

The letter *t* is prefixed to vowel-initial masculine nouns and *s*-initial nouns after the singular article *an.* Examples and grammatical contexts are given below. *t-* is used when prefixing a lowercase vowel.

*Examples*

(35) *Aire* 'Minister'
*An **t**Aire* 'The Minister'

(36) *ainmhí* 'animal'
*an **t-**ainmhí* 'the animal'

(37) *seachtain* 'week'
*an **t**seachtain* 'the week'

*Grammatical Contexts*

Vowel-initial and *s*-initial nouns are prefixed by '*t*' in the following situations (Christian Brothers, 1988, p22-23):

a) a masculine vowel-initial noun, in the common form, is prefixed with '*t-*' when it follows the definite article *an,* e.g. *ainmhí* 'animal', *an **t-**ainmhí* 'the animal'
b) a feminine s-initial noun followed by *l*, *n*, *r* or a vowel, in the common form, is prefixed by '*t*' when it follows the definite article *an*, e.g. *sláinte* 'health' *an **t**sláinte* 'the health', *seachtain* 'week' *an **t**seachtain* 'the week', but *scoil* 'school', *an scoil* 'the school'
c) a masculine s-initial noun followed by *l*, *n*, *r* or a vowel, in the genitive form, is prefixed by '*t*' when it follows the definite article *an*, e.g. *barr sléibhe* 'mountain-top' *barr an **t**sléibhe* 'top of the mountain'.

**The Prefix h**

There are a number of words (nouns, adjectives, pronouns and verbs) ending in a vowel (or a vowel sound) which cause a following vowel-initial word to be prefixed by *'h'*, if it is not already lenited or eclipsed (Christian Brothers, 1988, p23-24). (34) to (40) show some instances of h-prefixing.

*Examples*

(38) *an ubh* 'the egg'
*na **h**uibheacha* 'the eggs'

(39) *í* 'she'
*cé **h**í* 'who is she'

(40) *áit* 'place'
*ainm na **h**áite* 'name of the place'

*Grammatical Contexts*

The following are some of the principal contexts in which h-prefixing takes place (Christian Brothers, 1988, p23-24).

a) nouns after *a* meaning 'her', e.g. *aois* 'age', ***a haois*** 'her age'

b) nouns after *dhá* 'two', e.g. *iníon* 'daughter', *a dhá* ***hiníon*** 'her two daughters'

c) feminine nouns in genitive singular after the definite article, e.g. *eala* 'swan' *muineál* ***na heala*** 'the swan's neck'

d) common plural of nouns after the definite article, e.g. *éan* 'bird' ***na hé****in* 'the birds'

e) adjectives after *go* 'to' and *chomh* /cho:/ 'as' and *le* 'with', e.g. *iontach* 'wonderful' ***go hiontach*** 'wonderfully', *ard* 'tall/high', ***chomh hard*** 'as tall/high', eagla 'fear' ***le he****agla* 'with fear'

f) pronouns after *cé* 'who', *ní* 'not', *le* 'with', e.g. *é* 'he' ***cé hé*** 'who is he', *iad* 'them' *ní* ***hiad*** 'not them', *í* 'her/her' ***le hí*** *a cháineadh*' to criticise her'

g) verbs after *ná* 'not', e.g. *imigh* 'go' ***ná himigh*** 'do not go'.

**Slenderising**

This involves making the final syllable of a broad stem slender, and is usually achieved by inserting an '*i*' after the last broad vowel. Slenderising vowel-changes may be found in Table 7.

In (41) a broad (masculine) noun *cat* 'cat' is made slender in the genitive case, i.e. *cait* 'of the cat'. (42) is also a case of slenderising used to mark the genitive case.

*Examples*

(41) *an* ***cat*** 'the cat'
*eireaball an* ***chait*** 'tail of the cat' or 'the cat's tail'

(42) *an* ***ghrian*** 'the sun'
*solas na* ***gréine*** 'light of the sun' or 'sunlight'

**Table 7. Vowel-Changes: Signifying Slenderising of Consonants**

| *Base Vowel(s)* | *Orthographic Changes* |
|---|---|
| a | ai |
| á | ái |
| ao | aoi |
| ea | ei / i |
| éa | éi |
| eo | eoi |
| ío | í |
| ia | éi / iai / ea |
| io | i |
| o | oi |
| ó | ói |
| u | ui |
| ú | úi |
| ua | uai |

*Grammatical Contexts*

The following are some of the contexts in which slenderising takes place (Bráithre Críostaí, 1999, p55-58, 94):

a) genitive singular of masculine nouns of the first declension, e.g. *crann* ‘tree’, *crainn* ‘of the tree’

b) plural of nouns, e.g. *an cat* ‘the cat’, *na cait* ‘the cats’

c) some adjectives after genitive singular of masculine nouns, e.g. *an cat bán* ‘the white cat’, *eireaball an chait bháin* ‘the white cat’s tail’.

**Broadening**

Broadening involves making the final syllable of a slender stem broad. Broadening vowel changes may be seen in Table 8. The most common way of broadening a stem is by removing a final ‘*i*’ as shown in (43). Another method of broadening is given in (44) where a broad vowel is inserted after a sender vowel.

*Examples*

(43) *súil* ‘eye’

*radharc na súl* ‘sight of the eye’ or eyesight’

(44) *feadall* ‘whistling’

*port feadaíola* 'whistling tune'

**Table 8. Vowel-Changes: Signifying Broadening of Consonants**

| *Base Vowels* | *Orthographic Changes* |
|---|---|
| a e i | a e |
| a i | a |
| á i | á |
| a o i | a o |
| e i | e a |
| é i | é a |
| i | e a |
| í | í o |
| o i | o |
| ó i | ó |
| u a i | u a |
| u i | o |
| ú i | ú |

*Grammatical Contexts*

The following are some of the contexts in which broadening takes place (Bráithre Críostaí, 1999, p55-58, 95):

a) genitive singular of feminine nouns of the second declension, e.g. *súil* 'eye', *radharc na súl* 'eyesight'

b) genitive singular of feminine nouns of the fifth declension, e.g. *abhainn* 'river', *bruach na habhann* 'riverbank'.

**Syncopation**

Syncope is a term used to describe the dropping of unstressed vowel(s) in the final syllable of a polysyllabic word when a suffix is added. It is phonologically conditioned, rather than grammatically conditioned. (46) and (47) are instances of noun syncopation, and (47) shows an instance of verb syncopation.

*Examples*

(45) *saghas* 'type'
*saghsanna* 'types'

(46) *cathair* 'city'
*cathracha* cities

(47) *imir* 'to play' (a sport)
*imreoidh* 'will play'

## *2.5 Inflectional Morphology*

The system described in later chapters focuses on the analysis and generation of the inflectional morphology of Irish. The inflected parts of speech may be divided into two categories: open and closed.

The **open inflected classes** include:

- Verbs
- Nouns
- Adjectives.

New lexical items are routinely added to these categories and these new members are subject to the general inflection rules appropriate to their class.

New items are rarely added to the following inflected classes and are therefore considered to be **closed inflected classes**:

- Pronouns (personal, contrastive)
- Articles
- Prepositional pronouns (conjugated prepositions).

The number of items in these classes is small, and their inflected forms are usually listed in full in reference grammars. In this implementation they are treated in a similar manner by listing all inflected forms and their descriptions in the lexicon.

Derivational morphology relating to:

- Derived verbs, adjectives and nouns
- Verbal Nouns and Verbal Adjectives
- Compounds
- Others

is not addressed in the current implementation.

The following is a list of the lexical, syntactic, and semantic properties reflected in Irish inflectional morphology:

- Gender: feminine, masculine
- Case: common (nominative, accusative, dative), genitive, vocative

- Number: singular, plural
- Person: first, second, third
- Definiteness: presence of article or possessive pronoun
- Emphasis: presence of emphatic suffix
- Tense: past, present, future
- Mood: indicative, conditional, subjunctive, imperative
- Aspect: perfect, habitual (imperfect)
- Voice: active, autonomous.

In the following sections, the formation of inflected forms will be described. A summary of the parts-of speech to be described and the features reflected in inflection, is given in Table 9. The table is based on information in Graiméar Gaeilge na mBráithre Críostaí (1999, p40-1).

**Table 9. Inflections in Irish**

| *Word-Class* | *Features reflected in inflection* |
|---|---|
| Verbs | Broad/Slender Stem<br>Tense and Mood<br>Aspect<br>Voice<br>Number<br>Person |
| Nouns | Broad/Slender stem<br>Gender<br>Case and Number<br>Definiteness<br>Emphasis |
| Adjectives | Broad/Slender/Vowel-final/Syncopated stem<br>Gender<br>Case and Number |
| Personal Pronouns | Gender, Number and Person |
| Articles | Gender, Case and Number |
| Prepositional pronouns | Gender , Number and Person |

**Verbal Inflection**

Verbal inflection in Modern Irish is quite regular, apart from some irregular verbs and defective verbs. The irregular verbs contain many suppletive forms, and the defective verbs are those for which many of the verb forms are missing. Verbs are inflected by means of initial mutations, and the addition of a suffix. In some instances there is also final mutation There are allomorphic suffixes for broad and slender stems.

*Tense/Mood*

There are five tenses and four moods giving the nine tense/mood combinations shown in Table 10. Eight apply to all verbs and one (Habitual Present) is found only in the irregular verb *bí* 'to be' (Christian Brothers, 1988, p92-3).

**Table 10. Verb Tenses and Moods**

| *Tense* | *Mood* |
|---|---|
| Habitual Past (Imperfect) | Indicative |
| Past | Indicative<br>Subjunctive |
| Present | Indicative<br>Imperative<br>Subjunctive |
| Habitual Present | Indicative |
| Future | Indicative |
| Past/Present/Future | Conditional |

*Person/Number*

Person and number are indicated in one of two ways. Either a pronoun accompanies the verb form, or the pronoun is incorporated into the inflected verb form. (48a) and (49a) incorporate a subject pronoun and are known as **synthetic forms**. When person and number are expressed separately using the appropriate pronoun, e.g. *mé* 'I', *muid* 'we', as in (48b) and (49b), it is termed the **analytic form** (Christian Brothers, 1988, p94). Where there is a synthetic form, the analytic form may also be used, though one or other form is usually more common in practice, e.g. depending on the dialect.

(48) a. *táim* 'I am' (synthetic form)
b. *tá mé 'I am'* (analytic form)

(49) a. *táimid* 'we are' (synthetic form)
b. *tá muid* 'we are' (analytic form)

Verbs are commonly used without person or number. The **autonomous form**, as shown in example (50a), expresses a verbal action without any mention of the agent (subject), person or number (Christian Brothers, 1988, p94).

(50) *a. moltar é* ‘it is praised’ (autonomous form)

*b. molaim é* ‘I praise it’ (synthetic form)

*c. molann mé é* ‘I praise it’ (analytic form)

*Regular Verbs*

Regular verbs are divided into two categories based on the inflectional suffix for the third person future indicative which is either *–faidh/-fidh* or *–óidh/-eoidh* (Bráithre Críostaí, 1999) (see Tables 11 and 13). The first **conjugation** contains monosyllabic roots (and some polysyllabic roots), and the second conjugation contains polysyllabic roots only. There are different sets of verb endings (suffixes) for each conjugation for the various tense/mood combinations, as shown in Tables 11 and 13 below. Verb stems are either broad or slender, therefore each individual suffix has a broad and a slender alternative (allomorph).

In Table11, the suffixes for synthetic forms are shown where they exist. Where no special suffix for person and number is given, the suffix in the default column is used with the appropriate pronoun. (Pronouns are given in Table 16). Where “no suffix” is indicated, the stem is used without attaching any suffix.

*First conjugation of regular verbs*

The information in Table 11 is based on pages 95-96 of *New Irish Grammar* (Christian Brothers, 1988). In (51), in order to form the first person singular (1P Sg.) of the Present Indicative of *cuir* ‘to put’, the slender suffix *–im* is appended to the stem *cuir*, to form *cuirim*.

(51) *cuir**im***, ‘I put’, *cuir*+Verb+PresInd+1P+Sg

In (52), in order to form the second person singular (2P Sg.) of the Present Indicative of *cuir* ‘to put’, the default suffix *–eann* is used together with the second person singular pronoun *tú* ‘you’.

(52) *cuir**eann tú***, ‘you put’, *cuir*+Verb+PresInd; *tú*+Pron+1P+Sg

In the first conjugation, a suffix is attached directly to the monosyllabic stem, if possible, as in (53) and (54).

(53) *seas* ‘stand’

*seas**faidh***, ‘will stand’, *seas*+Verb+FutInd

*seas**faimid*** ‘we will stand’, *seas*+Verb+FutInd+1P+Pl

*seas**far***, ‘it stands’, *seas*+Verb+FutInd+Auto

(54) *cuir* 'put'

*cuir**fidh**,* 'will put', *cuir*+Verb+FutInd

*cuir**fimid*** 'we will put', *cuir*+Verb+FutInd+1P+Pl

*cuir**fear***, 'it is put' *cuir*+Verb+FutInd+Auto

Roots ending in *-áigh, -eoigh, -óigh, -uaigh or -úigh*, such as *dóigh* 'burn' or *luaigh* 'mention' phonetically consist of two syllables. Therefore the ending is removed to give a stem to which a suffix may be attached, as in (55) and (56).

(55)*dóigh*, 'burn'

*dó**faidh***, 'will burn', *dóigh*+Verb+FutInd

*dó**itear***, ' it burns', *dóigh*+Verb+PresInd+Auto

(56) *luaigh* 'mention'

*lua**faidh***, 'will mention', *luaigh*+Verb+FutInd

*lua**itear***, 'it is mentioned', *luaigh*+Verb+PresInd+Auto

Roots like *dóigh* 'burn' and *luaigh* 'mention', when truncated, use a mixture of broad and slender suffixes. Broad f-suffixes and slender t-suffixes are used. Slenderising of *dó-* and *lua-* is necessary (i.e. by inserting an '*i*' at the end of the stem) before adding a slender t-suffix.

Table 11. Type 1 Verbal Suffixes

| | *Suffix* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Tense / Mood* | *Synthetic Form (includes pronoun)* | | | | | | | *Pron. reqd.* |
| | *1P Sg.* | *2P Sg.* | *3P Sg.* | *1P Pl.* | *2P Pl.* | *3P Pl.* | *Auto.* | *Default* |
| Present Indicative | aim | | | aimid | | | tar | ann |
| | im | | | imid | | | tear | eann |
| Past Indicative | no suffix | | | amar | | | adh | no suffix |
| | | | | eamar | | | eadh | |
| Future Indicative | | | | faimid | | | far | faidh |
| | | | | fimid | | | fear | fidh |
| Imperfect Indic. | ainn | tá | | aimis | | aidís | taí | adh |
| (Past Habitual) | inn | teá | | imis | | idís | tí | eadh |
| Conditional Mood | fainn | fá | | faimis | | faidís | faí | fadh |
| Past / Present / | finn | feá | | fimis | | fidís | fí | feadh |
| Future | | | | | | | | |
| Present | | | | aimid | | | tar | a |
| Subjunctive | | | | imid | | | tear | e |
| Past Subjunctive | ainn | tá | | aimis | | aidís | taí | adh |
| | inn | teá | | imis | | idís | tí | eadh |
| Imperative | aim | no suffix | | aimis | aigí | aidís | tar | adh |
| Present Only | im | | | imis | igí | idís | tear | |

A summary of the inflection rules of first conjugation stems are given in Table 12 below:

**Table 12. Inflection of 1st Conjugation Verbs**

| *Root Type* | *Example* | *Action* | *Lexicon Class* |
|---|---|---|---|
| Broad | *mol* 'praise' | append broad suffixes | V1-BR |
| Broad stems with long vowel ending in *-igh* | *cráigh* 'torment'<br>*dóigh* 'burn'<br>*buaigh* 'win' | remove *–igh* & append broad f-suffixes and slender t-suffixes | V1-BR-LV |
| Broad stems with short vowel ending in *-igh* | *guigh* 'pray'<br>*luigh* 'lie' | remove *–igh* & append Type 2 suffixes except for Future Indicative and Conditional | V1-SV |
| Slender | *bris* 'break' | append slender suffixes | V1-SL |
| Slender stems with long vowel ending in *-éigh* | *léigh* 'read'<br>*pléigh* 'discuss' | remove *–igh* & append slender f-suffixes and slender t-suffixes | V1-SL-LV |
| Slender (exceptions) | *siúil* 'walk'<br>*tionóil* 'convene' | broaden & append broad suffixes | V1-SL-EX |
| Slender ending in *-áil* | *sábháil* 'save' | broaden except for t-suffixes & append broad f-suffixes slender t-suffixes | V1-SL-LC |

*Second conjugation of regular verbs*

The second conjugation contains polysyllabic (mainly disyllabic) stems. In all but a few cases, the last syllable (*-aigh/-igh*) is removed before a suffix is appended, as in (57). Other verb endings which are commonly removed before a suffix is appended include *-ail/-il, -ain/-in, -air/-ir,* and *-is.*

(57) *ceannaigh* 'buy'

*ceann**óidh***, 'will buy, *ceannaigh*+Verb+FutInd

*ceann**óimid***, 'we will buy, *ceannaigh*+Verb+FutInd+1P+Pl

*ceann**ófar***, 'it is bought, *ceannaigh*+Verb+FutInd+Auto

The following is an example of the less common case, where the whole root is used, i.e. the final syllable is not removed (or syncopated). Note that the stem used is *freastal*, rather than *freastail*, the form usually given in dictionaries and grammar references. (Likewise, in Table 14, *taisteal* is used rather than *taistil*.)

(58) *freastal* 'attend'
*freastal***óidh**, 'will attend, *(verb, future indicative)*
*freastal***aítear**, 'is attended, *(verb, present indicative, autonomous)*

Table 13 lists the inflectional suffixes for second conjugation verbs. The information in this table is based on pages 101-102 of *New Irish Grammar* (Christian Brothers, 1988).
As before synthetic forms are shown where they exist and in all other cases the default column is used in conjunction with the appropriate pronoun, except where "no suffix" is indicated.

**Table 13. Type 2 Verbal Suffixes**

| | *Suffix* | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Tense / Mood* | *Synthetic Form (includes pronoun)* | | | | | | | *Pron. reqd.* |
| | *1P Sg.* | *2P Sg.* | *3P Sg.* | *1P Pl.* | *2P Pl.* | *3P Pl.* | *Auto.* | *Default* |
| Present Indicative | aím | | | aímid | | | itear | aíonn |
| | ím | | | ímid | | | aítear | íonn |
| Past Indicative | | | | aíomar | | | aíodh | no |
| | | | | íomar | | | íodh | suffix |
| Future Indicative | | | | óimid | | | ófar | óidh |
| | | | | eoimid | | | eofar | eoidh |
| Imperfect Indicative (Past Habitual) | aínn | aíteá | | aímis | | aídís | aítí | aíodh |
| | ínn | iteá | | ímis | | ídís | ití | íodh |
| Past/Present/Future/ Conditional | óinn | ófá | | óimis | | óidís | ófaí | ódh |
| | eoinn | eofá | | eoimis | | eoidís | eofaí | eodh |
| Present Subjunctive | | | | aímid | | | aítear | aí |
| | | | | ímid | | | itear | í |
| Past Subjunctive | aínn | aíteá | | aímis | | aídís | aítí | aíodh |
| | ínn | íteá | | ímis | | ídís | ití | íodh |
| Present Imperative | aím | aigh | | aímis | aígí | aídís | aítear | aíodh |
| | ím | igh | | ímis | ígí | ídís | itear | íodh |

A summary of the inflection rules of second conjugation stems are given below:

**Table 14. Inflection of 2$^{nd}$ Conjugation Verbs**

| *Root Type* | *Example* | *Action* | *Lexicon Class* |
|---|---|---|---|
| Broad stems ending in *–aigh* | *ceannaigh* 'buy'<br>*tosaigh* 'start' | remove *–aigh*<br>& append broad suffixes | V2-BR |
| Other broad stems | *lorg* 'search' | append broad suffixes | V2-BR |
| Slender stems treated as broad | *freastail* 'attend'<br>*taistil* 'travel' | append broad suffixes<br>Note that broad stem forms used in this implementation:<br>freastal, taisteal | V2-BR |
| Slender stems ending in *-igh* | *bailigh* 'gather'<br>*éirigh* 'rise' | remove *–igh*<br>& append slender suffixes | V2-SL |
| Slender stems syncopated | *imir* 'play'<br>*taitin* 'like' | syncopate<br>& append slender suffixes | V2-SL-sync |
| Broad stems syncopated | *iompair* 'transport'<br>*codail* 'sleep' | syncopate<br>& append broad suffixes | V2-BR-sync |

*Irregular Verbs*

The following are the eleven commonly used irregular verbs. There are no morphological rules employed for either the irregular or the irregular-defective verbs - inflected forms are listed in full in grammar references, e.g. Christian Brothers (1988, p109-116) or Mac Giolla Phádraig (1963, p28-33, p48-59).

- abair 'say'
- beir 'catch, give birth to'
- *bí* 'be'
- clois 'listen'
- déan 'do, make'
- faigh 'get'
- feic 'see'
- ith 'eat'

- tabhair 'give'
- *tar* 'come'
- téigh 'go'

*Defective verbs*

Defective verbs are verbs which are missing many inflected forms. The following are examples of defective verbs (Christian Brothers, 1988, p113-116; Bammesberger, 1983, p90):

- *is* 'to be'
- *ar* 'says'
- dar 'seems'
- dóbair 'almost happened'
- tharla 'happened'

The defective verb *is* 'to be' is the copula (i.e. a grammatical link between subject and predicate) and is both irregular and defective (Christian Brothers, 1988, p113 & p122).

The irregular verb *bí* 'to be' is used to describe a state or condition etc. (59), whereas *is* 'to be' is used for classification (60), identification (61), ownership (62) and emphasis (63) (Christian Brothers, 1988, p122).

(59) *Tá Máire tinn*, 'Máire is sick'
*Bhí slua ann*, 'There was a crowd there'
*Bí ag éisteacht*, 'Listen'

(60) *Is dochtúir í*, 'She is a doctor'

(61) *Is í Máire an dochtúir,* 'Máire is the doctor'

(62) *Is le Máire an mála,* 'The bag belongs to Máire'

(63) *Is anseo atá sí,* 'She is <u>here</u>'

**Nominal Inflection**

The inflectional morphology of nouns is not as straightforward as that of the regular verbs. Most inflections use some combination of initial mutation, final mutation and suffixation. As Ó Siadhail puts it, "the inflectional patterns of the noun are varied and complicated" (1989, p148).

Nouns are inherently masculine or feminine, and they are inflected for case, number and definiteness (Stenson, 1981, p28). Case and number are combined in one inflected form which usually involves final mutations to the stem in the singular, and final mutation and suffixation in the plural. Definiteness is usually indicated by initial mutation. An emphatic form of a noun is achieved through the addition of the appropriate emphatic suffix. Both the gender of the noun, and the broad or slender nature of the stem influence the inflected form of the noun.

Before describing the formation of case/number inflections some of the factors which influence nominal inflection are outlined:

- Gender
- Case (Common, Genitive, Vocative, Dative)
- Number
- Definiteness
- Emphasis.

Syntactic **gender** usually follows the semantic gender (Ó Dochartaigh, 1992, p62 ), although there are some notable exceptions such as *cailín* 'girl' (masculine) and *stail* 'stallion' (feminine). Also, the gender of a small number of nouns varies between the different dialects. In (64) and (61) (Ó Dónaill, 1977), the first gender is given as the standard and the second is listed as a variant.

(64) a. *an t-ainm* 'name' (masculine)
b. *an ainm* 'name', (feminine varient in Munster dialect)

(65) a. *loch* 'lake', (masculine)
b. *loch* 'lake', (feminine variant)

The **common case** described in this work is the common form defined in *New Irish Grammar* (Christian Brothers, 1988, p 26). This is described as corresponding "to the traditional nominative, accusative and dative cases to be found in many previous grammars" as shown in example (66) (Christian Brothers, 1988, p10). The common form is described in Modern Irish (Ó Siadhail, 1989, p107) as the "unmarked" form. There are a small number of nouns which have a separate dative form but the dative[8] case is no longer a productive case in general.

(66) *a Tá **fear** ag an doras*
Is a-man at the door
'A man is at the door'.
*b. Chonaic Seán an **fear***
Saw Seán the man
'Seán saw the man'
c. Thug sé bonn don **fhear**.
Gave he a-coin to-the man
'He gave a coin to the man'

The same form of the noun is used in all three cases above. In (66a), *fear* 'man' is the subject, in (66b), it is the direct object and in (66c), it is the indirect object. It is lenited in (66c) since it is governed by the compound preposition *don* 'to the'. According to The Christian Brothers in *New Irish Grammar* (1988, p10) lenition and eclipsis do not change the essential form.

The **genitive case** is used more frequently in Irish than in English (McGonagle, 1991, p3). As well as marking the possessive case, four other contexts in which it is used are given below:

- A noun that qualifies (follows) another noun takes the genitive case (possessive) (Mac Giolla Phádraig, 1963, p73).
  - *hata Mháire* 'Mary's hat': *Mháire* is the genitive case of *Máire*,
  - *mála scoile* 'school bag': *scoile* is the genitive case of *scoil* 'school'
  - *Tom Sheáin* 'Tom son of Seán': *Sheáin* is the genitive case of *Seán*.

- A noun following a verbal noun takes the genitive case (Christian Brothers, 1988, p128).
  - *ag dúnadh an dorais* 'closing the door': *dorais* is the genitive case of *doras* 'door'.
  - *ag moladh an cheoil* (praising the music): *cheoil* is the genitive case of *ceol* 'music'.
  - *ag imirt peile* 'playing football': *peile* is the genitive case of *peil* 'football'

- A noun following the simple prepositions *chun* 'to', *cois* 'by', *dála* 'by', *fearacht* 'like', *timpeall* 'around', and *trasna* 'across' takes the genitive case (Christian Brothers, 1988, p134).
  - *chun an tí* 'to the house': *tí* is the genitive case of *teach* 'house'
  - *cois na habhann* 'by the river': *abhann* is the genitive case of *abhainn* 'river'.
  - *dála an scéil* 'by the way' lit. 'regarding the story': *scéil* is genitive case of *scéal* 'story' or 'news'
  - *fearacht a dhearthár* 'like his brother': *dearthár* is genitive case of *deartháir* 'brother'
  - *timpeall na páirce* 'around the field': *páirce* is the genitive case of *páirc* 'field'
  - *trasna an bhóthair* 'across the road': *bóthair* is the genitive case of *bóthar* 'road'

- A noun following compound prepositions takes the genitive case (Christian Brothers, 1988, p138).
  - *ar chúl an toilg* 'behind the sofa': *toilg* is the genitive case of *tolg* 'sofa'
  - *ar feadh míosa* 'for a month': *míosa* is the genitive case of *mí* 'month'
  - *tar éis na hoibre* 'after the work': *oibre* is the genitive case of *obair* 'work'

- If there are uncertain quantities of a noun then the noun will be in the genitive case (partitive genitive).
  - *a lán airgid* 'a lot of money': *airgid* is the genitive case of *airgead* 'money'
  - *mórán ama* 'much time': *ama* is the genitive case of *am* 'time'
  - *do chuid fola* 'your blood': *fola* is the genitive case of *fuil* 'blood'

The **vocative case** of a noun is used whenever a noun is directly addressed in speech. This is particularly true when addressing a person as in (67), but is also true when addressing animals (68), or inanimate objects (69), as for example in children's' literature. In all cases, a vocative noun is preceded by the vocative particle '*a'* which causes the noun to be lenited, except for vowel-initial words which are not lenited.

(67) *a Mháire* 'Mary'
*a chailín* 'girl
*a ghrá* 'love'

(68) a *Bhéirín* 'Little Bear'
*a éan* 'bird'

(69) *a Ghealach* 'Moon'
*a chrann* 'tree'

The vocative singular takes the same form as the genitive singular for masculine nouns whose genitive singular is formed by slenderising (first declension nouns). For all other nouns, both masculine and feminine, the vocative takes the same form as the common singular.

The **dative case** takes the same form as the nominative and accusative cases in all but a few instances, and is therefore included in the common case by default. In Modern Irish it has ceased to be a morphologically productive case, i.e. a new noun will not have a separate dative form.

Where fossilised dative forms occur, e.g. (70) to (72), they are included (in this implementation) in the irregular nouns lexicon. In (70), *cionn* is a dative case of *ceann* 'head', in (71), *láimh* is the dative case of *lámh* 'hand' and in (72), *Éirinn* is the dative case of *Éireann* 'Ireland'.

(70) *ceann* 'head'
*os cionn* 'overhead'
*thar cionn* 'excellent'

(71) *lámh* 'hand'
*de láimh* 'by hand'

(72) *Éire* 'Ireland'
*in Éirinn* 'in Ireland'

A noun is either **singular** or **plural** in **number**. The common singular is the unmarked form of the noun. A noun is said to have a **strong plural** if the plural form is the same for all three grammatical cases. Conversely, if the plural form is different in the common and genitive case the noun is said to have a **weak plural**. This distinction influences noun-adjective agreement and will be dealt with later under Adjectives.

There are, in general, two methods of plural formation. Weak plurals are formed by mutation of the final consonant cluster, and possibly accompanied by stem ablaut or syncope. Strong plurals are formed by the addition of a suffix also possibly accompanied by ablaut or syncope (Ó Dochartaigh, 1992, p62). As is the case with Breton (Ternes, 1992, p414) there are a large number of plural suffixes and no morphological rules for predicting which plural suffix is used.

In many cases there are a number of possible plurals. In *Gramadach na Gaeilge* (Rannóg an Aistriúcháin, 1958, p ix) the choice of plurals available is mentioned and the reasons for the particular choices which were made. An example of the possible plurals for *capall* 'horse' and the form chosen as the standard form is given in (73).

(73) *capall* 'horse'
*capai**ll**, cai**ple**, capai**llí*** 'horses' (alternative plural forms)
*capai**ll*** 'horses' (official standard)

The unpredictable nature of plural formation from a phonological point of view is demonstrated in (74) and (75), where similar sounding words *bé* 'woman' and *gé* 'goose' form their plurals in different ways. (76) and (77) demonstrate the same point with the words *baile* 'home' and *béile* 'meal'.

(74) *bé* 'woman'
*bé**ithe*** women'

(75) *gé* 'goose'
*gé**anna*** 'geese'

(76) *baile* 'home'
*bail**te*** 'homes''

(77) *béile* 'meal'
*béil**í*** 'meals'

There are some preferences for certain plural suffixes which are dialect specific and often non-standard. The suffix *–(e)achal* is favoured over *–(e)acha* in Connaught for certain noun-classes, whereas the

Donegal dialect tends to use *–(e)annaí* rather than *–(e)anna* (Ó Dochartaigh, 1992, p31). Some examples are given in (78) and (79).

(78) *maidin* 'morning

*maidin**eacha*** 'mornings' (standard)

*maidin**eachaí*** 'mornings, (Connaught)

(79) *carr* 'car'

*carr**anna*** 'cars' (standard)

*carr**annaí*** 'cars' (Donegal, also Connaught)

In summary, most plurals are formed by the addition of a suffix, though a few are formed by varying the broad or slender quality of the stem. Some plurals involve both phenomena. Polysyllabic nouns with a final unstressed syllable may be subject to syncopation when a suffix is appended. Replacement of the final syllable is also a feature of some plural formations. This can occur in combination with suffixation. Finally, there are also a small number of nouns where the plural takes a form which is unrelated to the singular form (suppletion). Further examples of plural formation may be seen in Appendix D.

**Definiteness** relates to whether or not the noun is preceded by the definite article (there is no indefinite article in Irish). Nouns are inflected after articles depending on gender, case and number. (80) shows the indefinite form *páirc* 'field', the definite common singular *an pháirc*, and the definite genitive singular *na páirce*, of the feminine noun *páirc*.

(80) *páirc* 'field'

*an pháirc* 'the field'

*imeall na páirce* 'edge of the field'

Nouns may have an **emphatic (contrastive) form** after a possessive determiner (possessive adjective). It is formed by adding an emphatic suffix such as *–sa* or *–se* (also *–na, –ne, -san, -sean*), depending on whether the stem is broad or slender. Although this is possibly a derivational suffix, it is included here under inflectional morphology since it is fully productive and does not alter the part-of-speech category. (81) and (82) show broad and slender stems respectively with the appropriate emphatic suffixes.

(81) *mo theach* 'my house'

*mo theach**sa*** 'my house'

*a teach**sa*** 'her house'

*a theach**san*** 'his house'

(82) *mo pháirc* 'my field'

*mo pháirc****se*** 'my field'

*a páirc****se*** 'her field'

*a pháirc****sean*** 'his field'

*Formation of Common Singular*

The common case, singular, indefinite form of a noun is taken to be the root form. Common singular definite forms feature some initial mutations but no final mutations or suffixes, apart from the optional emphatic suffix. Set out below are the ways in which the common case of singular nouns are produced.

Common Singular: Initial Mutations

a) Masculine
   - i. Indefinite: *unmarked*
   - ii. Definite:
     - vowel-initial: *prefixed with 't-'*
     - words starting with *'sl', 'sn'* , *'sr'* or *'s'* + vowel: *unmarked*
     - other words: *unmarked*

b) Feminine
   - i. Indefinite: *unmarked*
   - ii. Definite:
     - vowel-initial: *unmarked*
     - words starting with *'sl', 'sn', 'sr'* or *'s'* + vowel: *prefixed with 't'*
     - other words: *lenited*

Common Singular: Final Mutations

a) Non-emphatic: *unmarked*

b) Emphatic:
   - i. 1st/2nd person: Suffix +*sa*, +*se*
   - ii. 3rd person:
     - Fem: Suffix +*sa*, +*se*
     - Masc: Suffix +*san*, +*sean*

**Fig 5. Formation of common singular of nouns**

*Formation of Genitive Singular*

The rules for the formation of the genitive singular of nouns are given below in Fig.6. The five categories of final mutation of the genitive singular are the basis of the five declensions of nouns in Irish (see Table 15 below). In reference grammars and many dictionaries, the declension number is listed to aid in the formation of the genitive singular.

**Table 15. Declensions of Nouns**

| *Declension* | *Genitive Singular* |
|---|---|
| 1 | Slenderise stem ending or<br>change ending, *-each/-ach* to *-í/-aí* or *-ach* to *-aigh* |
| 2 | Add suffix *-e* and slenderise if necessary or change *-each* is to *-í* |
| 3 | Add suffix *-a* and broaden if necessary |
| 4 | No change |
| 5 | If stem ends in a vowel, add suffix 'n' or 'd' and broaden if slender or<br>suffix *–(e)ach* or<br>syncopate plus suffix *–(e)ach* or<br>change *-ú/-aí* to *-aithe*. |

Genitive Singular: Initial Mutations

a) Masculine roots
   i. Indefinite: *unmarked*
   ii. Definite:
      - vowel-initial: *unmarked*
      - words starting with *sl, sn, sr* or *s* + vowel: *lenited*
      - other words: *lenited*

b) Feminine roots
   i. Indefinite: *unmarked*
   ii. Definite:
      - vowel-initial: *prefixed with 'h'*
      - words starting with *'sl', 'sn', 'sr'* or *'s'* + vowel: *unmarked*
      - other words: *unmarked*

Genitive Singular: Final Mutations

1st Declension: Slenderise stem or change *–each /-ach* to *–í /-aí* or *–ach* to *-aigh*

2nd Declension: Add suffix *-e* and slenderise if necessary or change *-each* is to *–í*

3rd Declension: Add suffix *-a* and broaden if necessary

4th Declension: No change

5th Declension: If stem ends in a vowel, add suffix *'n'* or *'d'* and broaden if slender.
Also other non-standard genitives.

**Fig 6. Formation of genitive singular of nouns**

*Formation of Vocative Singular*

The vocative singular of nouns are produced is as follows:

Vocative Singular: Initial Mutations

i. vowel-initial: *unmarked*
ii. words starting with *'sl'*, *'sn'*, *'sr'* or *'s'* + vowel: *lenited*
iii. other words: *lenited*

Vocative Singular: Final Mutations

1st Declension: follows genitive singular
2nd-5th Decls.: follows the common singular

**Fig 7. Formation of vocative singular of nouns**

*Formation of Plurals*

Plural formation involves initial mutation plus final mutation consisting of one or more of the processes summarised in Fig. 8.

Plurals: Initial Mutation

a) Indefinite: *unmarked*

b) Definite:

   i. Nominative:

      - vowel-initial: prefix h
      - words starting with *sl, sn, sr* or *s* + vowel: *unmarked*
      - other words: *unmarked*

   ii. Genitive:

      - vowel-initial: *eclipsed*
      - words starting with *sl, sn, sr* or *s* + vowel: *unmarked*
      - other words: *eclipsed*

   iii. Vocative

      - vowel-initial: *unmarked*
      - words starting with *'sl', 'sn', 'sr'* or *'s'* + vowel: *lenited*
      - other words: *lenited*

Plurals: Final Mutation

a) Suffixation

   i. +a
   ii. +acha
   iii. +aí
   iv. +anna
   v. +anta
   vi. +e
   vii. +eacha
   viii. +eanna
   ix. +eanta
   x. +í
   xi. +idí
   xii. +na
   xiii. +nna
   xiv. +ta
   xv. +te
   xvi. +tha
   xvii. +the

b) Slenderise

c) Broaden & Suffix +a

d) Syncopate & Suffix +acha

e) Syncopate & Slenderise & Suffix +e, +eacha

f) Replace stem ending

g) Replace stem ending & Suffix +the

h) Suppletion

**Fig 8. Formation of plurals of nouns**

### Adjectival Inflection

Descriptive adjectives behave in a similar manner to nouns in that they are inflected according to case, number and gender, by means of initial mutation, and final mutation and suffixation. They take their gender, number and case from the noun which they qualify. The genitive plural form depends also on whether the noun they are modifying has a weak or strong plural.

The number of adjectival declensions varies from three to eight depending on the grammar source consulted (Christian Brothers, 1988; Bráithre Críostaí, 1999; Ó Baoill and Ó Tuathail, 1992; Rannóg an Aistriúcháin, 1958).

Descriptive adjectives can be used in two ways (Christian Brothers, 1988, p59): predicatively or attributively. Predicatively an adjective qualifies a noun or pronoun indirectly, as in (83) to (84). Used in this way, the base form (positive degree) is not inflected to agree with case, number or gender of the noun. Further details may be found in New Irish Grammar (Christian Brothers, 1988, p 59).

(83) Is *breá an lá é.*
Is lovely the day it
'It is a lovely day'

(84) Tá *sé fuar.*
Is it cold
'It is cold'

Adjectives used attributively, i.e. directly qualifying a noun, are inflected to agree with the case, number and gender of the noun, as in the following examples. In (85), following the definite article *an* 'the' the feminine noun *bróg* 'shoe' causes lenition in the adjective *mór* 'big'. In (86), since the masculine noun *capall* 'horse' is not lenited after the definite article, neither is the adjective.

(85) *an bhróg* ***mhór***
the shoe big

(86) *an capall* ***mór***
the horse big

In the case of plurals the adjective is lenited following a masculine noun with a slender plural, as in (87).

(87) *na bróga* ***móra***
the shoes big

(88) na capaill ***mhóra***
the horses big

The relevant facts about adjectival inflection are summarised as follows:

Adjectives Qualifying Masculine Nouns

a) Common
   i. Initial Mutation
      - Lenite if the noun is lenited, otherwise no lenition

b) Genitive & Vocative
   i. Initial Mutation
      - Lenite
   ii. Final Mutation (one of the following)
      1. broad stem, slenderise
      2. broad stem, no change
      3. broad stem, change ending (*-(e)ach* becomes *–(a)igh*)
      4. slender stem, no change

Adjectives Qualifying Feminine Nouns

c) Common & Vocative
   i. Initial Mutation
      - Lenite

d) Genitive
   i. Initial Mutation
      - Lenite
   ii. Final Mutation (one of the following)
      1. broad stem, slenderise, suffix *-e*
      2. broad stem, change ending (*-(e)ach* becomes *–(a)í*)
      3. broad stem, change ending (*-och* removed), suffix *-thí*
      4. broad stem, change ending (*-ách* removed), suffix *+thaí*
      5. broad stem, no change
      6. slender stem, broaden, suffix *+a*
      7. slender stem, suffix *–e*
      8. syncopate and suffix –a/-e

**Fig 9. Formation of adjectives qualifying singular nouns**

### Adjectives Qualifying Common Plural Nouns

i. Initial Mutation

- noun slender final consonant, lenite
- noun final vowel or broad consonant, no change

ii. Final Mutation

- broad stem adj., suffix -*a*
- slender stem adj., broaden and suffix -*a*
- slender stem adj., suffix -*e*

### Adjectives Qualifying Genitive Plural Nouns

i. Initial Mutation

- no change

ii. Final Mutation

- noun strong plural, broad stem adj. suffix -*a*
- noun strong plural, slender stem adj. suffix -*e*
- noun strong plural, vowel-final adj., no change
- noun weak plural, all adjs. no change

### Adjectives Qualifying Vocative Plural Nouns

iii. Final Mutation

- broad stem adj., suffix -*a*
- slender stem adj., suffix -*e*

**Fig 10. Formation of adjectives qualifying plural nouns**

### Pronouns

The personal pronouns and their emphatic (contrastive) forms are listed in Table 16.

**Table 16. Pronouns**

| *Personal Pronoun* | | *Contrastive (emphatic) Personal Pronoun* |
|---|---|---|
| *mé* | 'me' | *mise* |
| *tú* | 'you' | *tusa* |
| *sí* | 'she' (as subject) | *sise* |
| *í* | 'she/her' | *ise* |
| *sé* | 'he' (as subject) | *seisean* |
| *é* | 'he' | *eisean* |
| *sinn* | 'we' | *sinne* |
| *sibh* | 'you' pl. | *sibhse* |
| *siad* | 'they' (as subject) | *siadsan* |
| *iad* | 'they' | *iadsan* |

### Articles

There are two forms of the definite article 'the', *an* and *na* for singular and plural respectively. The form *na* is also used with feminine singular nouns in the genitive case.
The full paradigm is given in Table 17.

**Table 17. Articles**

| *Article* | *Usage* |
|---|---|
| *an* | common case, singular |
| | genitive case singular, masculine |
| *na* | genitive case singular, feminine |
| | all plurals |

### Prepositional Pronouns

The inflection of prepositions for gender, number and person is a feature typical of the Celtic languages (Fife, 1993, p22). Eighteen inflected prepositions are detailed in various grammar sources (Christian Brothers, 1988; Mac Giolla Phádraig, 1963), as well as in *An Foclóir Póca* (An Roinn Oideachas, 1986a), and the simple prepositions from which they are derived are listed in Table 18. (89) and (90) show inflected forms of the prepositions *chuig* 'to' and *ar* 'on', which are created by combining the simple preposition and relevant pronoun.

(89) *chugam* (*chuig+mé*) 'to me'
*chugat* (*chuig+tú*) 'to you"
*chugaibh* (*chuig+sibh*) 'to you' (plural)

(90) *uirthi* (*ar+í*) 'on her, it'
*air* (*ar+é*) 'on him, it'
*orthu* (*ar+iad*) 'on them'

**Table 18. Simple Prepositions**

| *Preposition* | | | *Preposition* | | |
|---|---|---|---|---|---|
| *1.* | *ag* | at | *2.* | *idir* | between |
| *3.* | *ar* | on | *4.* | *ionsar* | to |
| *5.* | *as* | out of | *6.* | *le* | with |
| *7.* | *chun, chuig* | to | *8.* | *ó* | from |
| *9.* | *de* | from | *10.* | *roimh* | Before |
| *11.* | *do* | to | *12.* | *seach* | other than |
| *13.* | *faoi* | under | *14.* | *thar* | Over |
| *15.* | *fara* | as well as | *16.* | *trí* | Through |
| *17.* | *i* | in | *18.* | *um* | About |

## 2.6 Summary

In this chapter, some background information on Irish was given, and the fundamental characteristics of the language were outlined, including phonology, morphophonology and inflectional morphology.

# 3. Finite-State and Two-level Morphology

## *3.1 Introduction*

This chapter introduces the topic of computational morphology in general and focuses on two-level and finite-state morphology in particular.

## *3.2 Computational Morphology*

Computational morphology involves combining computational techniques with linguistic methodologies in order to create robust and efficient computer systems that can encode the morphology of a language and make it available to other NLP modules.

Knowledge about the morphology of a language is a vital component of most NLP systems. Morphemes do not combine freely (Karlsson and Karttunen, 1997, p97); only certain combinations and orderings are allowed. The morphotactics of a language describe all the possible morpheme combinations (and precludes invalid combinations) (Sproat, 1992, p125). Rules about the spelling changes that take place at the boundaries when morphemes are combined must also be included. Therefore, the essential ingredients in a computational morphology system are:

a) information about the constituent parts of words, i.e. **morphemes**
b) the rules for combining them, i.e. **morphotactics**
c) the effects of morpheme combinations and alternations, i.e. **morphographemics.**

The main approaches to computational morphology are affix-stripping, statistical methods, and finite-state and two-level morphology. The first two will be mentioned briefly and the third will be discussed in greater detail since it is the methodology used in this work.

An **affix-stripping** program analyses a word by removing known inflectional and derivational affixes. It assumes that the remainder is the stem, which it then looks up in a lexicon of stems. It can take a machine-readable dictionary as its starting point (Klavans, 1997, p672; Jurafsky and Martin, 2000, p87). These programs tend to be language-dependent (Karlsson, 1994, p2570). Affix-stripping was developed initially for English, where it works quite well, but it is not flexible enough to deal with complex non-concatenative inflections, i.e. stem modifications, found in other languages (Sproat, 1992, p xiii; Karlsson, 1994, p2570). Affix-stripping programs are sometimes known as 'stemmers'.

**Statistical methods** (unlike affix-stripping and two-level morphology), do not rely on a lexicon and pre-defined rules but seek to deduce the structure of words by analysing a corpus and looking for patterns in the texts and their relative frequencies. This methodology is used particularly in the areas of part-of-speech tagging and in speech recognition and synthesis. (Jurafsky and Martin, 2000, p10-13, Karlsson, 1994, p2572). In automatic part-of-speech tagging the objective is to assign grammatical part-of-speech information to the words in a corpus of text, and in particular to overcome the problem of ambiguity, i.e. deciding the most likely tag for homographs. Bi-gram and tri-gram analysis is commonly used to make decisions in ambiguous situations. A number of successful applications have been developed including the following: CLAWS (Garside, et al, 1987), TAGGIT (Greene and Rubin, 1971), and the Xerox tagger (Cutting *et al*, 1992). Automatic part-of-speech tagging may also be carried out using using rule-based methods, e.g. the Brill tagger (Brill, 1992). This tagger iteratively learns its tagging rules, and reports levels of accuracy comparable to those of statistically based taggers.

More recent developments include research by Sheremetyeva *et al*, (1998) and Goldsmith (2001). Sheremetyeva *et al* describe a system called Rapid Deployment Morphology, which uses 'quasi-roots' (stem-endings which identify inflectional paradigms). This methodology has been implemented for Russian and Serbo-Croatian. Goldsmith's system uses 'signatures' (morphological patterns) to aid unsupervised acquisition of morphology from a corpus of text.

**Finite-state (and two-level morphology)** is, however, the most common way of implementing morphology (Jurafsky and Martin, 2000, p5-6; Sproat, 1992, p153). It owes its success to several factors. Finite-state machines operate in a straightforward manner and are mathematically well understood (Koskenniemi, 1997, p99). Having a sound mathematical basis, there are number of useful set operations available to it. Finite-state transducers are inherently bi-directional – the same system works equally well in parsing or generation.

In practical terms, a number of efficient algorithms for implementing finite-state machines have been developed (Mohri, 1997) and there are rule-compilers which can automatically convert linguistic rules into finite-state transducers.

Two-level morphology is language-independent – successful implementations have been developed for many languages including Finnish, Russian, English, Spanish, Ancient Greek, Japanese, Basque and Arabic (Karlsson, 1994, p2571; Koskenniemi, 1997, p101; Sproat, 1992, p153). The rules and lexicons are separate from the processing mechanism, which works on many typologically different languages (Sproat, 1992, p153). A disadvantage of finite-state morphology compared to statistical methods is that it is time-consuming initially to set up the lexicon and rules, but this is often offset by greater accuracy.

## *3.3 Two-level and Finite-State Morphology*

Johnson (1972), in his doctoral thesis, examined the formal characteristics of Chomsky and Halle's (1968) system of generative phonology in which sequences of features are mapped on to other sequences using rules which define the surrounding context as well as the sequence to be transformed (Ritchie et al, 1992, p16; Gazdar, 1985, p598). Johnson identified three types of rule, which he called *iterative*, *simultaneous*, and *linear* and showed that they were equivalent in power to various types of finite-state transducer (Ritchie et al, 1992, p16).

Independently of Johnson's work, Kaplan and Kay (1981; 1994) proposed that phonological rules be implemented as FSTs and that the individual transducers be cascaded together in the style of ordered rules of generative phonology, where the output of one rule can be the input to another rule (Ritchie et al, 1992, p16; Sproat, 1992, p136). These transducers could be composed into one large transducer, with the same behaviour as the original phonological rules (Sproat, 1992, p139; Ritchie et al, 1992, p20; Gazdar, 1985, p598). At the time, this was computationally infeasible due to the large number of states in the intermediate stages (Sproat, 1992, p137-9). (Currently, with better FSA minimisation algorithms and improved computing power this is not such an issue).

Following on from Kaplan and Kay, Kimmo Koskenniemi (1983) defined a *two-level* model of morphology and phonology. Morphotactic information is encoded in the lexicon. Morphophonological and morphographemic alternations are encoded as two-level rules which are implemented as finite-state transducers. Koskenniemi observed that most rules do not feed one another, therefore each rule expresses a single fact which is independent of the other rules (Jurafsky and Martin, 2000, p105-6; Karlsson, 1994, p2571). This differs primarily from Kaplan and Kay's model in that the rules are unordered and individual phonological transducers are executed in parallel rather than in series (Gazdar, 1985, p599; Karlsson, 1994, p2571). All rules operate on just two levels; lexical and surface. This avoided the practical problems of a cascade of transducers, where one rule feeds into another, producing intermediate machines with an unwieldy number of states (Ritchie et al, 1992, p21; Sproat, 1992, p139).

**The lexicons** in the original Kimmo systems (Koskenniemi, 1983; Karttunen, 1983; Antworth, 1990) were implemented as linked annotated letter-trees (Sproat, 1992, p131; Karlsson and Karttunen, 1997, p97). Look-up was achieved by processing the (surface) input and traversing the tree segment by segment (letter by letter). The last segment of a morpheme was labelled as a lexical entry. The label also contained grammatical information pertaining to the morpheme and, where appropriate, a pointer to a continuation pattern (a sub-lexicon of suffixes). The root lexicon contains the free morphemes in the language and the pointers can operate in a recursive manner depending on the morphotactics of the language (Karlsson, 1994, p2570-1). See Fig. 11, which follows Sproat (1992, p128-131).

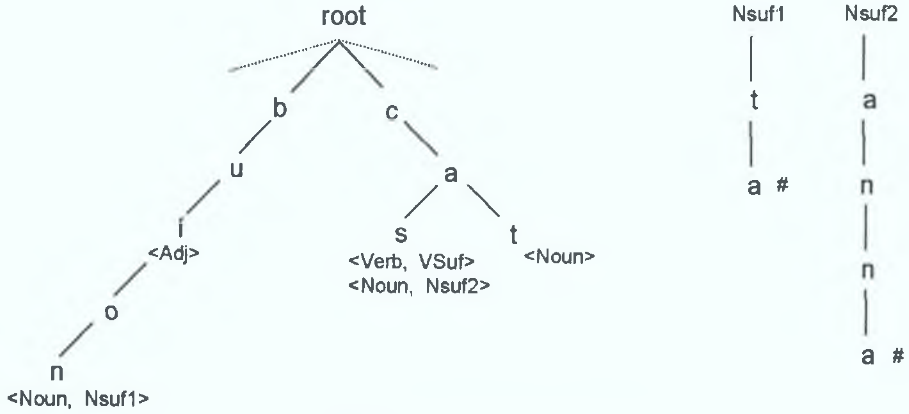


**Fig 11. Lexical letter-trees**

**Two-level rules** also differ from traditional (generative) rules by explicitly coding when they are obligatory and when they are optional, by using four different rule operators, <=>, =>, <= and /<=, as opposed to the single rewrite operator, ->, of generative phonology (Jurafsky and Martin, 2000, p108).

The idea of **correspondence pairs** (also called concrete pairs or feasible pairs) is central to two-level morphology. It is this which allows two-level rules to represent constraints on two levels. The symbols `a:b` mean that lexical `a` maps to surface `b`. By default, a symbol on the lexical (upper) level corresponds to the same symbol on the surface (lower) level. Other correspondence pairs can be defined in the lexicon explicitly, or they may be inferred from the rules (Sproat, 1992, p135). Where there is a choice of possible correspondences the most restrictive will apply. For example, in general, lexical 'y' maps to surface 'y', (`y:y`), but in certain restricted circumstances, lexical 'y' maps to surface 'i' , (`y:i`), e.g. the plural of certain English nouns such as 'city'.

The following is the basic syntax of Koskenniemi's rule formalism (Sproat, 1992, p145-6; Jurafsky and Martin, 2000, p108):

```
CP op LC _ RC
```

`CP` stands for correspondence pair, `LC` stands for left context, `RC` stands for right context and `op` stands for operator. The first three operators imply optionality and the fourth defines an obligatory rule.

1. `a:b =>  LC _ RC` (context restriction rule)
2. `a:b <=  LC _ RC` (surface coercion rule)
3. `a:b /<= LC _ RC` (exclusion rule)
4. `a:b <=> LC _ RC` (composite rule)

The context restriction rule (1) states that a *may* be realised as b, but *only* when it is preceded by the specified left context (LC) and followed by the specified right context (RC). The surface coercion rule (2) means that a is *always* realised as b in the given context. The exclusion rule (3) means that a is *never* realised as b in the specified contexts. The composite rule (4) is a combination of the context restriction (1) and the surface coercion (2) rules. It states that a must correspond to b in the given context and only in that particular context.

A simple example of a type (1) rule might be:

```
        y:i => ?+ _ e s
```

meaning y *may* be realised as i, but *only* when it is preceded by one or more symbols (?+) on the left and followed by 'es' on the right.

According to Clemenceau (1997, p83-4) a fundamental difference between two-level rules and the rewrite rules of Chomsky and Halle (1968) is that a two-level rule expresses a *relationship* between an underlying form and a surface form, whereas a rewrite rule *rewrites* the underlying form as the surface, and the underlying form is then no longer available to any other rules. Furthermore, a rewrite rule always goes from underlying to surface form and may not easily be reversed, whereas a two-level rule is a mapping which can be applied in either direction.

**Finite-state vs. Two-Level Morphology**

Sometimes a distinction is drawn between finite-state and two-level morphology. Two-level morphology (Koskenniemi, 1983) involves exactly two levels of representation: surface and underlying (lexical). These two levels are related by a set of simultaneous or parallel correspondence rules. By contrast, in finite-state morphology (Kaplan and Kay, 1994; Beesley and Karttunen, 2001) a complex morphological operation can be broken into a cascade of simpler mappings that operate in succession, resulting in a number of intermediate representations. Each simple mapping is a regular relation (a two-level mapping). The rule compiler can then use composition of regular relations to automatically compile a cascade of regular relations with n+1 representations into a single 2-level regular relation which directly relates level 1 with level n+1 of the cascade bi-directionally.

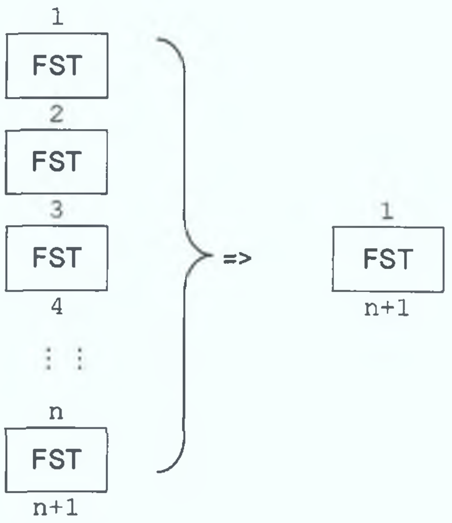


**Fig 12. Cascade of transducers composed as a two-level transducer.**

The mathematical properties of a cascade of relations, and of two-level rules, are well-understood; any *n*-level description can be composed into two levels (Karttunen *et al*, 1992). Sproat cites Schützenberger (1961) as having shown that any system of cascaded FSTs can be composed into a single equivalent FST (1992, p139). Parallel FSTs may also be intersected into a single transducer. It follows, therefore, that the two systems are mathematically equivalent. In principle, any system of cascaded FSTs implementing ordered rules which is composed into one transducer can be interpreted as a single FST operating "in parallel" (Sproat, 1992, p141). According to Karttunen *et al* (1992) there is no computational or theoretical reason to favour two-level descriptions.

However, finite-state based morphology has the advantage that the developer can think "procedurally", and adopt a modular design in terms of intermediate mappings, while at the same time the mathematics of of the composition of regular (intermediate) relations generates a resulting bi-directional FST that directly maps between surface and lexical representations. Traditional two-level systems (Koskenniemi, 1983), by contrast, are often harder to design for complex phenomena as each rule has to apply in parallel.

There are also some differences in the way in which linguistic rules are expressed. In the tradition of generative phonology there is a general tendency to list one (ideal) form of a morpheme in the lexicon and to represent the alternations in the rule component. The opposite is true in the case of two-level descriptions – often a number of allomorphs are listed in the lexicon thus requiring fewer rules. This is due to the fact that complex rule interactions, which are easy to state in terms of ordered rules are often more difficult to express in a two-level model. Because of the difficulties of encoding the complicated rule interactions found in ordered rule systems, two-level descriptions tend to be relatively shallow (Sproat, 1992, p141-3).

When the relationship between lexical and surface forms is complex it makes sense to decompose the relation into a sequence of simpler relations. In many cases this is not only easier, but linguistically justified (Karttunen *et al*, 1992). For efficient recognition and generation, a cascade of simpler relations can be composed into a single transducer.

**Lexicon Transducers**

Karttunen *et al* (1992) describe a further refinement to the finite-state methodology with the discovery that the lexicon itself can be implemented as a FST. The letter-tree approach described above has a number of shortcomings. Generation is less efficient than analysis. As the morphosyntactic information is held at the leaf nodes only, an incorrect path may be traversed and backtracked several times before the correct one is found. Also, the lexical information is incomplete. Allomorphs rather than morpheme categories tend to be encoded in the lexical form, (e.g. `+s` or `+es` rather than `+Plural`). Another disadvantage is that although annotated letter-trees are a type of finite-state network, they cannot be minimised since all the information is at the leaf nodes and removing branches would result in information loss (Karttunen *et al*, 1992).

The solution proposed by Karttunen *et al* (1992) is to implement the lexicon, as well as the rules, as FSTs. Morphosyntactic information is encoded on the branches rather than the leaves. This means that morphological categories are represented as part of the lexical form, backtracking is reduced and standard minimisation techniques can be applied. It has the added advantage that the lexicon can now be considered to be the first stage in a cascade of finite-state transducers. The lexicon and the rule transducers can be composed to produce a single two-level transducer for the language, which transduces directly between lexical and surface representations.

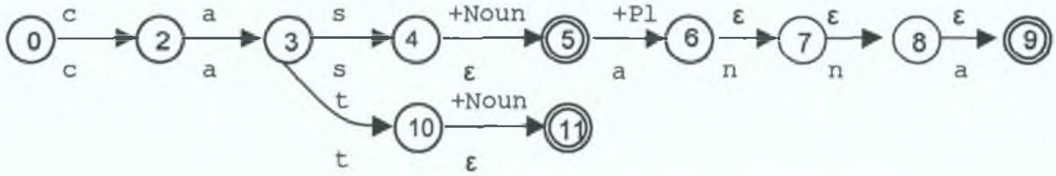


**Fig 13. Fragment of Lexicon Finite State Transducer**

## *3.4 Summary*

This chapter introduced the topic of computational morphology concentrating in particular on on two-level finite-state morphology.

# 4. Finite-State Technology

*"Finite-state to the rescue"*
(Kornai, 1999, p4)

## *4.1 Introduction*

This chapter deals with finite-state technology in detail. It covers finite-state automata, finite-state transducers, regular expressions, and the operations which can be performed on automata and transducers. The Xerox finite-state toolset is also introduced.

## *4.2 Background*

In 1936 the mathematician and logician Alan Turing (1936) described an abstract model of computation based on a machine reading from and writing to tapes (Hein, 1995, p698). This model is still valid for present day computers. The model describes a computing device in terms of a control unit containing rules which processes symbols taken from a finite list, the **alphabet**. The machine proceeds from an initial or **start state** by looking at symbols on the tape and deciding what action to take based on its current state, the symbol in question and inbuilt rules. This processing continues until the sequence of input symbols is finished and the machine reaches a **final state,** which may be an **accepting state** or a **rejecting state**. There are various sequences of symbols that can lead to an accepting state and each such sequence is known as a **word**. The set of all such words constitutes the **language** of the machine.

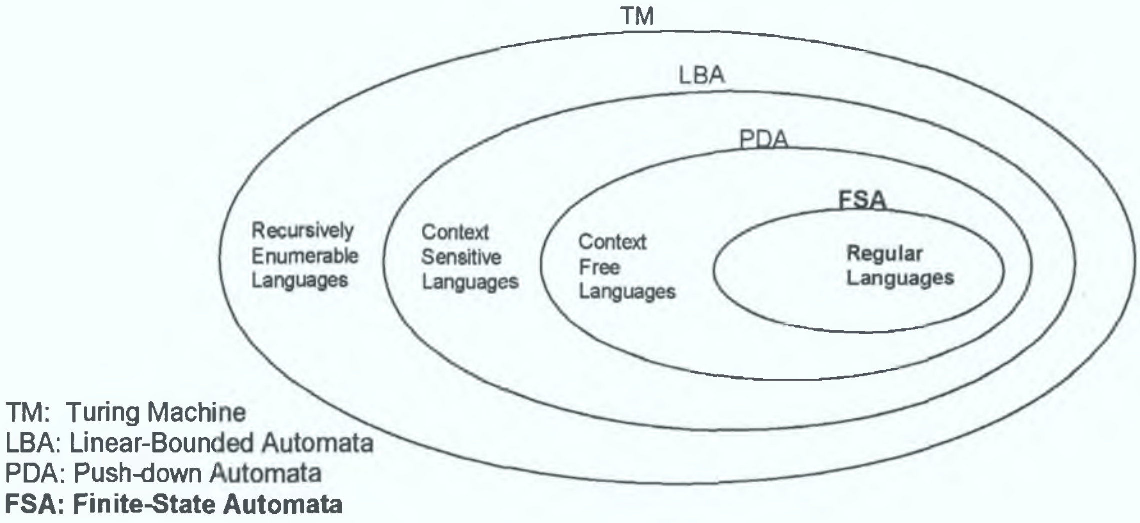

TM
LBA
PDA
FSA
Recursively Enumerable Languages
Context Sensitive Languages
Context Free Languages
Regular Languages
TM: Turing Machine
LBA: Linear-Bounded Automata
PDA: Push-down Automata
**FSA: Finite-State Automata**

**Fig 14. Formal Languages and Computing Machines**

A **Turing Machine** (TM) can read one symbol at a time and may overwrite that symbol. It can move forward or back one symbol, or can remain in the same position. This can be used to model decisionmaking and recursion. In the 1950's Chomsky worked on formal grammars, which bear a close relationship to abstract machines or automata (Hopcroft *et al*, 2001, p1). A TM is the most general model of computing and comes at the top of what is known as the Chomsky Hierarchy (Fig. 14), which defines three further sub-levels of machine (or automata) and their respective classes of language. Each successive level is a subset of the preceding levels, and the respective automata are more restricted in their functioning.

## *4.3 Finite State Automata*

Unlike the TM described earlier, a **Finite-State Automaton** (FSA) can move in one direction only (e.g. left to right) and can read the input looking at each symbol in turn. It cannot not write to the tape. The automaton, which is always in one of a finite set of states, moves to a new state based on the symbol just read and its current internal state. A FSA is also known as a Finite State Machine.

A FSA can model systems that at all times must be in one of a finite number of states. According to Hopcroft *et al.* (2001, p2-3) the purpose of a "state" is to remember the relevant portion of a system's history, and the advantage of having a finite number of states is that such a system can be implemented using limited resources.

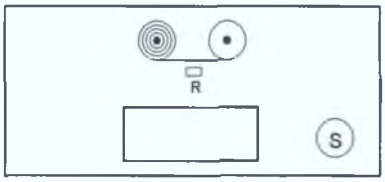


**Fig 15. FSA –start state**

Operating Instructions:

1. Load the input tape (note: the tape moves in one direction only).
2. Press the start button. ⓢ
3. If the tape contains a sequence of symbols which matches a sequence of transitions in the machine's internal network leading to a final state then the input will be accepted, otherwise it will be rejected.
4. Note: If a final state is reached before all the input symbols are processed, this will also cause the tape to be rejected.

R
Accepted
S
R
Rejected
S

**Fig 16. FSA –final state; input accepted or rejected**

**Fomal Specification of a Finite State Automaton.**

FSA = (Q, $\Sigma$, $\delta$, $q_0$, F) is the formal 5-tuple definition of a finite state automaton where

Q = Finite set of states.

$\Sigma$ = Finite set of symbols i.e. the alphabet.

$\delta$ = A transition function which has two arguments; a state and an input symbol, and which returns a new state e.g. $\delta(q_i,x) = q_j$

$q_0 \in Q$ = The start state, a special state which is a member of Q.

$F \subseteq Q$ = Set of final or accepting states which is a subset of Q.

The following transition network represents a language containing two words, 'cat' and 'can'. Final states are indicated by double circles.

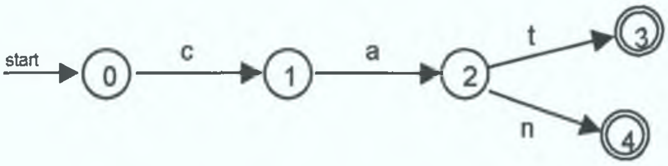


**Fig 17. Finite state network**

This transition network, and the language defined by it, can be encoded as a Finite State Automata using the following 5-tuple definition:

FSA = (Q, $\Sigma$, $\delta$, $q_0$, F)

Q = {0,1,2,3,4}

$\Sigma$ = {a,c,n,t}

$\delta$ = {(0,c,1), (1,a,2), (2,t,3), (2,n,4)}

$q_0$ = 0

F = {3,4}

The State Transition Table (below) shows the set of transitions from current state to next state depending on the symbol just read. For example if the machine is in state 0 and an 'a' is read it will fail, but if a 'c' is read it will proceed to state 1.

**Table 19. State Transition Table**

| *State* | *a* | *c* | *n* | *t* |
|---|---|---|---|---|
| 0 | fail | 1 | fail | fail |
| 1 | 2 | fail | fail | fail |
| 2 | fail | fail | 4 | 3 |
| 3: | fail | fail | fail | fail |
| 4: | fail | fail | fail | fail |

The sequences of symbols in the example network (Fig. 17) that lead to a final or accepting state are **c a t** and **c a n**. These valid sequences of symbols are called **words**.

The **language** of this automaton is the set of words it accepts:

Language = {cat, can}.

The FSA can therefore be seen as a **language acceptor** (Hein, 1995, p585). Equivalently, the same FSA can also be seen as a **language generator**. Starting in the start state the FSA will output the label of each transition it takes from the start state to a final state.

An important constraint on a language is that the alphabet is finite. Therefore, although a language can have an infinite number of strings, the strings must be drawn from a fixed and finite alphabet (Hopcroft *et al*, 2001, p31).

**Deterministic and non-Deterministic Automata**

If the transition function δ($q_i$,x), with current state $q_i$ and input *x*, uniquely determines the next state then the FSA is deterministic. If the transition function (or in this case: transition relation) permits several possible next states then the automaton is non-deterministic. The transition network in Fig. 18 is non-deterministic as δ($q_2$,n) can lead to states 4 and 6; therefore the transition function does not uniquely determine the next state. Every non-deterministic FSA can be transformed into an equivalent deterministic FSA that defines the same language (Hein, 1995, p588-621).

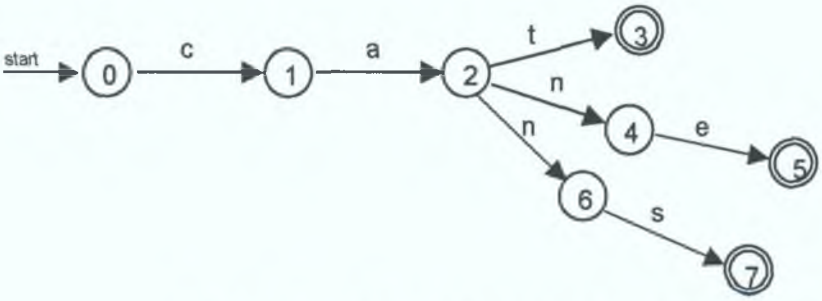


**Fig 18. Non-deterministic finite-state network**

## *4.4 Regular Expressions*

A language accepted by a FSA can always be encoded as a **regular expression**. A regular expression for the language {cat, can} is:

```
        c a [n | t]
```

meaning ‘c’ followed by ‘a’ followed by either ‘n’ or ‘t’.

Languages accepted by Finite State Automata are **Regular Languages**. Regular languages belong to a category of languages that are characterised by the fact that the symbol at a particular location in a string is dependent only on a bounded amount of information about the preceding symbols (Kaplan, 1997, p362). A regular language is described using regular expressions. Therefore regular expressions can be used to define Finite State Automata.

In computational terms any input text is thought of as a “string” of symbols. Regular expressions are a way of describing classes of strings. Regular expressions are used for pattern matching in word processors e.g. “Find walk*”, which will match ‘walk’, ‘walked’, ‘walking’ etc. or to list all text files in a directory e.g. “List *.txt”, which will list all files ending in ‘.txt’.

The notation for operations and operators of regular expressions was first defined by Kleene (1956) as a means of modelling neural networks in the brain (Jurafsky and Martin, 2000 p11). They have since been adopted and used extensively in the area of lexical analysis. The following are some examples of regular expressions:

(1) recogni[s|z]e — matches ‘recognise’ or ‘recognize’

(2) neighbo(u)r — matches ‘neighbour’ or ‘neighbor’ - the ‘u’ is optional

(3) re[?+]ing — matches all sequences starting with ‘re’, followed by one or more symbols and ending in ‘ing’

(4) use[?*] — matches all sequences starting with ‘use’ followed by zero or more symbols

Example (4) will match the ‘use’ in ‘reuse’ and example (2) will match with ‘neighbor’ in ‘neighborhood’. It is often necessary to anchor regular expressions to a part of a string, particularly the start and the end. (2) and (4) could be extended to (5) and (6).

(5) neighbo(u)r# — matches ‘neighbour’ or ‘neighbor’ - the ‘u’ is optional and # signifies the end of the string

(6) #use[?*] — matches sequences of symbols beginning with ‘use’ and followed by zero or more symbols

## *4.5 Finite-State Transducers*

A variation on the Finite State Automata is the Finite State Transducer (FST) (also known as a Mealy Machine). This type of machine may write to as well as read the tape.

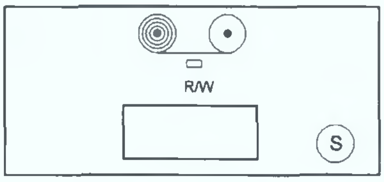


**Fig 19. Finite-state transducer**

The tape still moves in one direction only and one symbol at a time - it therefore accepts the same class of languages as the FSA (i.e. Regular Languages).

The advantage of the FST over the FSA is that in addition to accepting or rejecting the input as being part of the language, it can "translate" an input string into an output string and thereby output useful information regarding the input. A FSA defines a formal language by defining a set of strings, but a FST defines a relation between languages by defining relations between sets of strings (Jurafsky and Martin, 2000, p71).

For example, in Fig. 20 taking the path 0,1,2,3,5 through the network we can output `cat+Noun` from the upper level labels while we read `cat` on the lower level labels of the arcs.

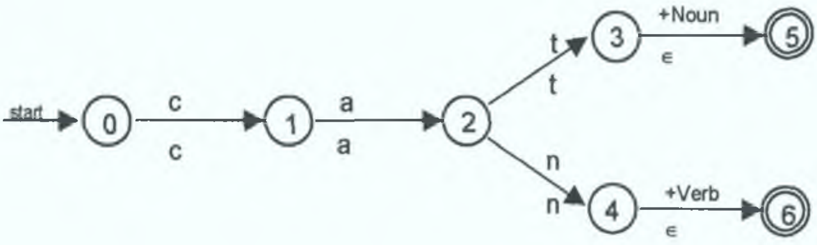


**Fig 20. FST network with upper and lower level labels on arcs**

A **regular relation** is a mapping between two regular languages. The alphabet of the machine in Fig. 20 is a set of regular relations between pairs of symbols of the form *upper:lower*. By default, every letter maps onto itself, therefore in set Σ of the following quintuple, 'a' is assumed to mean 'a:a'. 'ϵ', epsilon, is the symbol for the empty string. The empty string is used when an upper level string maps onto a lower level string which is not the same length (or vice versa).

The formal 5-tuple definition which encodes the finite state automaton in Fig. 20 is:

FST = $(Q, \Sigma, \delta, q_0, F)$

$Q = \{0, 1, 2, 3, 4, 5, 6\}$

$\Sigma = \{a{:}a, c{:}c, n{:}n, t{:}t, +\text{Noun}{:}\in, +\text{Verb}{:}\in\}$

$\delta = \{(0, c{:}c, 1), (1, a{:}a, ), (2, n{:}n, 4), (2, t{:}t, 3), (3, +\text{Noun}{:}\in, 5), (4, +\text{Verb}{:}\in, 6)\}$

$q_0 = 0$

$F = \{5, 6\}$

The FST is the encoding of a mapping (regular relation) between two languages, in the case at hand a lexical (upper) language and a surface (lower) language. A FST, like a FSA, is a language acceptor; it can accept (or reject) a string from either the lexical or surface. More importantly it can function as a **language generator**; if a string from the lexical language is input the corresponding surface string is output (generated) and vice versa. A transducer is a device which can take one type of input and produce a different type of output, therefore this FST, which can accept a lexical string and output a surface string (and vice versa), makes it a **lexical transducer**. The fact that it can work in either direction means that the lexical transducer is inherently **bi-directional**.

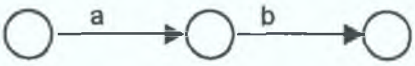
## *4.6 Finite-State Operations*

Words are the valid combinations of symbols in the alphabet, and a language is the set of valid combinations defined by the FSA. Since a FSA can represent a regular language (i.e. a set of words), many of the mathematical operations on sets such as **intersection, difference, concatenation, union** etc. may be applied to FSAs. Not all set operations apply to FSTs since a FST represents not just a regular language but a regular *relation* between languages. Union and concatenation may be applied to FSTs, whereas intersection and difference apply only to regular languages rather than regular relations. **Inversion**, **composition** and **projection** are operations which apply to regular relations. The following presentation follows Beesley and Karttunen (2001, p23-30,44-47)

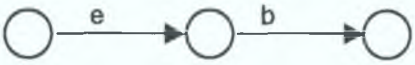
### Intersection (or conjunction): A & B

Intersection applies to regular languages only. The resulting network after intersecting two regular languages contains only the arcs common to both networks.

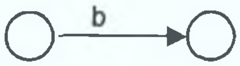
A



B



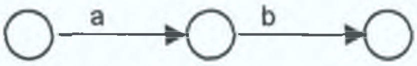
A & B



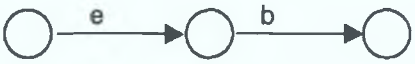
**Difference (subtraction): A - B**

When network B is subtracted from network A we are left with only the arcs in A which are not in B.

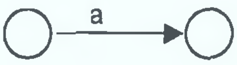
A



B



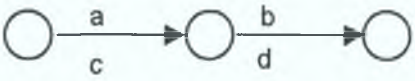
A - B



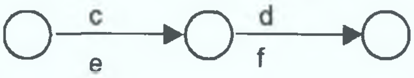
**Concatenation: AB**

Concatenation can be applied to both regular languages (FSAs) and regular relations (FSTs). The final state of the first network is joined to the start state of the second network.

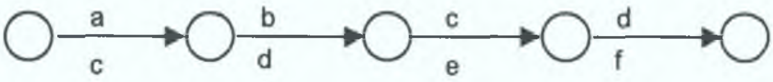
A



B



A B

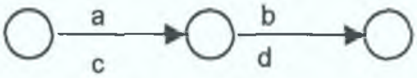


This operation is used in a lexical transducer, e.g. to concatenate a suffix to a stem.

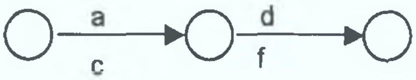
### Union A | B

Union applies to both regular languages and regular relations. The union of two networks contains all of the arcs of both networks.

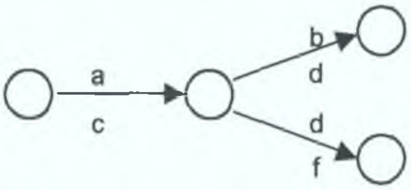
A



B



A | B

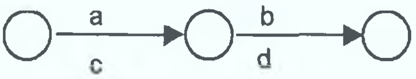


This operation is used to combine lexical FSTs.

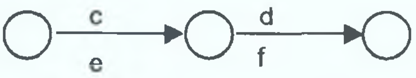
### Composition: A .o. B

Composition operates on regular relations only (FSTs). It composes together one language from each of the regular relations, i.e. the lower language of the first network with the upper language of the second network.

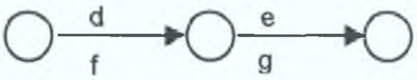
A



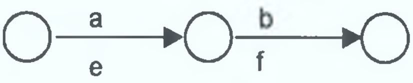
B



C

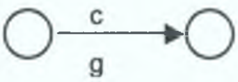


A .o. B



Composition can act as a filter, as in the following example: if there are not matching labels on the lower and upper languages those arcs will be absent from the resulting network.

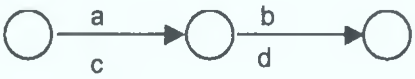
B .o. C



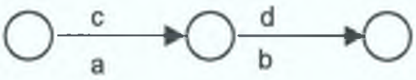
**Inversion: A.i**

Inversion of a network involves swapping the upper and lower level labels on the arcs. This allows the transducer network to act as a language acceptor or generator (Jurafsky and Martin, 2000, p73). It is possible to apply inversion to a regular language network (FSA), but the resulting network will remain the same.

A



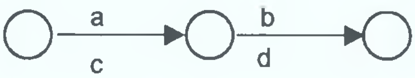
A.i



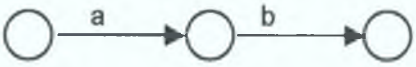
**Projection: A.u , A.l**

Projection is an operation which can be applied to regular relations, i.e. FSTs. Either the upper or lower side language can be selected (i.e. projected) from the regular relation.

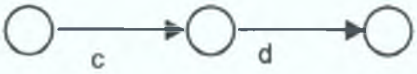
A



A.u



A.l



## *4.7 Finite-State Technology and Linguistics*

Finite-state theory resulted from the convergence of ideas about computing devices and regular languages, particularly as a result of Kleene's discovery that the languages recognised by deterministic finite state automata are exactly the class of regular languages (Hein, 1995, p586). Finite-state methods have been applied in the area of natural language processing as early as 1958 (Kornai, 1999, p3), and according to Mohri (1997, p355) their use can be justified both linguistically and computationally. Linguistic phenomena can be described in a natural manner, and finite-state machines are efficient in terms of both time and space (Mohri, 1997, p355).

Nevertheless, the popularity of finite-state models in linguistics went into decline in the 1950's and 1960's, due mainly to negative comments by Chomsky in *Syntactic Structures* (1957), where he suggests that finite-state models are not powerful enough to model grammatical structures (Crystal, 1997a, p152). After this, finite-state models received little attention throughout the late 50's and the 60's when most attention was focussed on transformational models (Kornai, 1999, p3).

Johnson's (1972) discovery that phonological rewrite rules could be modelled using finite state transducers went largely unnoticed at the time (Jurafsky and Martin, 2000, p88).

In the 1980's Kaplan and Kay at Xerox PARC also proposed that FST's be used for phonological rewrite rules (Jurafsky and Martin, 2000, p88; Sproat, 1992, p136) and produced the first algorithms for finite state computing. These were extended by Kimmo Koskenniemi (1983) who defined a Two-Level Morphology, proposing that FSTs operate in parallel rather than in sequence (Sproat, 1992, p139). Karttunen and others implemented the KIMMO system at Xerox. A PC version by Antworth (1990) was later made widely available.

During the 1990's the area of finite state computing for lexical analysis progressed and several implementations are now available. The following are listed by Kornai (1999, p2): XFST from Xerox, HTK from Entropic, Watson from Ribbit Software Systems, and FSM from AT&T/Bell Labs.

Currently finite-state technology is used with great success in morphological processing and according to Uszkoreit (1997, p340), finite-state parsers have been constructed which out-perform their competition in coverage and performance, and finite-state methods are also being applied in semantics and discourse modelling.

## *4.8 Xerox Finite-State Tools*

Xerox Research Centre Europe (XRCE) and the Palo Alto Research Center (PARC) have developed a set of finite-state tools which provide a means of implementing finite-state morphologies. The tools are independent of any one natural language and have been used to implement morphologies for many of the major European languages (English, Spanish, French, German etc.) as well as Arabic, Turkish, Japanese and others.

Only the aspects of the tools used in the current work are described here, but full details of the tools may be found in the book *Finite State Morphology: Xerox Tools and Techniques* (Beesley and Karttunen, 2001), or on the XRCE site at "http://www.xrce.xerox.com/research/mltt/fst".

**lexc: lexicon compiler**

*lexc* is the finite-state tool which has been developed by Xerox for defining two-level lexicons. It accepts a text file containing a user-defined lexicon encoded using to the following syntax:

```
Lexical-item       Continuation-class;
```

The lexical item is usually the unmarked form of the word (the root or headword given in a dictionary). In the context of this work the lexical item is the **stem** (the root in most cases) to which inflectional affixes are attached, i.e. a free morpheme. The continuation class can be a pointer to another lexicon or it can be the end-of-string marker '#'. Example (1) shows two entries for 'cat', one of which is followed by the end-of-string marker '#' and the second which points to the continuation class Noun-Pl, where the plural form of the word will be defined.

```
(1) cat             #;
    cat             Noun-Pl;
```

As this is a two-level morphology each lexical-item has two representations; a **lexical level** and a **surface level** representation. They are encoded in the lexicon in the format **lexical:surface**. Where there is no colon present in an entry, the lexical item is assumed to map to itself, as in (1), where 'cat' means 'cat:cat     #;'.

We make use of the two-level representation to encode valuable morphological information about the words as example (2) shows. The symbols to the left of the colon represent the lexical level, i.e. the morphological analysis 'cat+Noun+Sg', and the symbols to the right of the colon represent the surface form 'cat', i.e. the orthographic form of the word.

```
(2) cat+Noun+Sg:cat             #;
    cat+Noun+Pl:cat             Noun-Pl;
```

In finite-state transducer terms, the lexical level corresponds to the **upper level** of the FST and the surface level corresponds to the **lower level** of the FST as shown in Fig. 21.

$$\frac{\text{Upper Level}}{\text{Lower Level}} \begin{matrix}\leftrightarrow \\ \leftrightarrow\end{matrix} \frac{\text{Lexical Level}}{\text{Surface Level}} \text{ e.g. } \frac{\text{cat+Noun+Pl}}{\text{cats}}$$

**Fig 21. Two-level representation**

Fig. 22 shows a fragment of a lexicon using *lexc* syntax which illustrates how some English plurals might be handled. The lexical items (noun stems) i.e. 'cat', 'dog' etc. point to different continuation classes depending on the manner in which their plural is formed.

```
Multichar_Symbols
+Noun +Sg +Pl ^I

LEXICON Nouns
cat            Noun-Pl-S;
box            Noun-Pl-ES;
fry            Noun-Pl-IES;

LEXICON Noun-Pl-S
+Noun+Pl:s     #;
+Noun+Sg:0     #;

LEXICON Noun-Pl-ES
+Noun+Pl:es    #;
+Noun+Sg:0     #;

LEXICON Noun-Pl-IES
0:^I           Noun-Pl-ES;
```

**Fig 22. Fragment of English Noun Lexicon**

The tags, which are chosen by the lexicon developer (linguist) to describe the morphological features, e.g. `+Noun`, `+Pl` etc. (and the tags used for triggering replace rules e.g. '`^I`'), are all declared as **multi-character symbols** in the *lexc* syntax at the start of the text file.

By convention, the morphological features on the lexical level are distinguished from text by a leading '+' symbol and the mark-up tags triggering replace rules on the surface level are distinguished by a leading '^' symbol.

Unless otherwise specified, the first lexicon in the text file is the root lexicon. This root lexicon contains a list of stems each of which is associated with a continuation class (or an end-of-string marker). The first stem listed in Fig. 22, '`cat`', is associated with continuation class '`Noun-Pl-S`'. In the `Noun-Pl-S` lexicon, symbols are concatenated to both the *lexical* and *surface representations*. Lexical information

'+Noun+Pl' is added to the upper level and the symbol '*s*' is added to the lower level. This completes the entry as we now encounter the end-of-string marker; the outcome is shown in (3) below.

(3) cat**+Noun+Pl**:cats

The complete listing of word-forms and their lexical representations encoded in Fig. 22 is given Table 20.

**Table 20. Output of Sample Noun FST**

| *Lexical Level* | *Surface Level* |
|---|---|
| cat+Noun+Sg | cat |
| cat+Noun+Pl | cats |
| box+Noun+Sg | box |
| box+Noun+Pl | boxes |
| fry+Noun+Sg | fry |
| fry+Noun+Pl | fry^Ies |

Fig. 23 shows a network representation of the lexicon encoded in Fig. 22. In reality, the stem morphemes would be expanded and each of the constituent letters would be represented by an arc, e.g. 'cat' would require three labelled arcs.

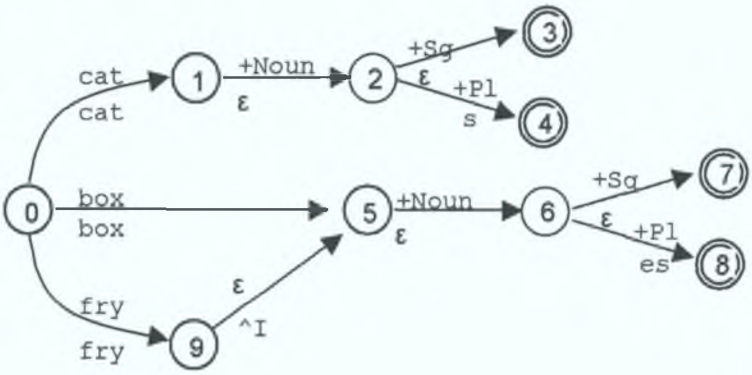


**Fig 23. Sample of Noun FST Network**

*Replace Rule Triggers in the Lexicon*

So far we have dealt with simple concatenative morphology, e.g. a plural suffix is concatenated to a stem. The following example shows how **replace rule triggers** are used for stem modification. In (4), in addition to adding the plural suffix '`-es`' to the stem '`fry`', the stem itself must also be modified. The '`y`' must be replaced by an '`i`' in the context of the plural suffix '`-es`'. This can not be carried out using concatenation in the lexicon.

(4) `fry+Noun+Pl:fry^Ies`

A **replace rule** must act upon the output of the lexicon to effect this change. Not every '`y`' must be replaced by an '`i`', therefore the appropriate strings are indicated in the lexicon by concatenating a tag ('`^I`' in this example) to the lower level string. This acts as a trigger for a replace rule which will make this change outside of the lexicon. (The manner in which replace rules operate will be discussed in detail the next section).

A trigger such as '`^I`' can be used to avoid the problem of a rule applying too liberally. For example, without the '`^I`' tag, in certain circumstances, a word such as '`yes`' could be transformed into '`ies`'.

All replace rules in this system are implemented in two phases:

- inflectional mark-up tags (triggers) are inserted in the lexicon surface strings
- replace rules encoded as regular expressions are constrained to fire only in the context of these triggers.

*Ambiguity*

Many words have more than one morphological analysis. The word 'fry' could also be listed in the verb stem lexicon. It is quite common for a surface form to have more than one morphological analysis, whereas generation will usually provide only one surface form (except perhaps in the case of variant plurals, e.g. the plural of lexicon: lexicons and lexica).

(5) `fry+Noun+Sg:fry`
`fry+Verb+PresInd+1P+Sg:fry`

(6) `fry+Noun+Pl:fries`
`fry+Verb+PresInd+3P+Sg:fries`

**xfst: Xerox finite-state tool**

*xfst* is a Xerox linguistic tool which accepts regular expressions and compiles them into finite-state networks. These networks can be saved and run independently. *xfst* provides many built-in functions which allow for easy manipulation of the finite-state networks. The following is a brief outline of some aspects of the Xerox regular expression notation and the *xfst* tool.

Following on from the more general operators listed in Section 4.4, Table 21 lists the notation used in replace rule regular expressions. Some of these operators (e.g. @->, [., .], .o. etc. ) are specific to the Xerox finite-state calculus. Many of these special operators are shorthand for more complicated regular expressions. They do not extend the power of the regular expression language, but they do make it easier to define lexical rules (Beesley and Karttunen, 2001, p58)

**Table 21. Xerox Regular Expression Operators**

| *Notation* | *Description* |
|---|---|
| a\|b | a or b |
| a b | a and b |
| \a | not a |
| .#. | start or end of string |
| ; | a semi-colon terminates a regular expression |
| (a) | optional a |
| a* | zero or more a |
| a+ | one or more a |
| [ ] | square brackets are used for grouping regular expressions |
| % | literalises the following symbol, e.g. %+ is just a character not a plus operator |
| .o. | composition, an ordered operation, i.e. a .o. b is not equal to b .o. a |
| -> | maps, e.g. a -> b, lexical a maps to surface b |
| @-> | maps the longest sequence found |
| \| \| | in the context of... (both left and right contexts match on the on upper level) |
| _ | position in the string of the symbol(s) to be mapped |
| ? | any symbol |
| a^<3 | less than three a's concatenated together |
| a^>4 | more than four a's concatenated together |
| [] | empty string |
| [. | dotted brackets treats an infinite number of empty strings as a single empty |
| .] | string, e.g. [..] -> h replaces one empty string with a h |
| ... | repeat matched symbols e.g. a+ @-> <...> will surround one or more a with angle brackets |

### Replace Rules

The format of the replace rule regular expression is as follows:

```
String -> Replacement-String || Left-context _ Right-context;
```

The left and right contexts are optional; either, both or none may be used, as shown in the following examples below.

In (7), lexical level (upper level) 'b' maps to a surface level (lower level) 'B' in any context.

```
(7)  b -> B;
(8)  b -> B || .#. _ ;
(9)  b -> B || _ .#. ;
(10) b -> B || a _ c ;
```

The symbol .#. anchors the search context to the start or end of a string, and the underscore '_' marks the position in the string of the symbols to be replaced. Therefore, in (8), only 'b' at the start of a string is mapped to 'B', and in (9), only 'b' at the end of a string is mapped to 'B'. In example (10), 'b' must be preceded by 'a' and followed by 'c' in order to be mapped to lower-level 'B'.

The replace rule required by example (4) above could be encoded using the following regular expressions. (Note that '%' literalises '^' in (12) and (13)).

```
(11) define Cons [b|c|d|f|g|h|j|k|l|m|n|p|q|r|s|t|v|w|x|y|z];
(12) y -> i || Cons+ _ %^I e s .#.;
(13) %^I -> [];
```

A regular expression can be assigned to a variable name using the **define** command. In (11), the user-defined variable 'Cons' represents the consonants listed within the square brackets. In example (12), an upper level 'y' maps to a lower level 'i' when preceded by one or more consonants, followed by the tag '^I', the symbols 'es', and the end-of-string marker '.#.'. In (13), '^I' is removed since it is no longer required, i.e. it is mapped to the empty string.

These regular expressions can be compiled into a replace rule transducer using *xfst*. This transducer may then be composed with the lexicon transducer. The ordering is important, since the lower-level of the first transducer must match the upper level of the second transducer, as shown in Fig. 24.

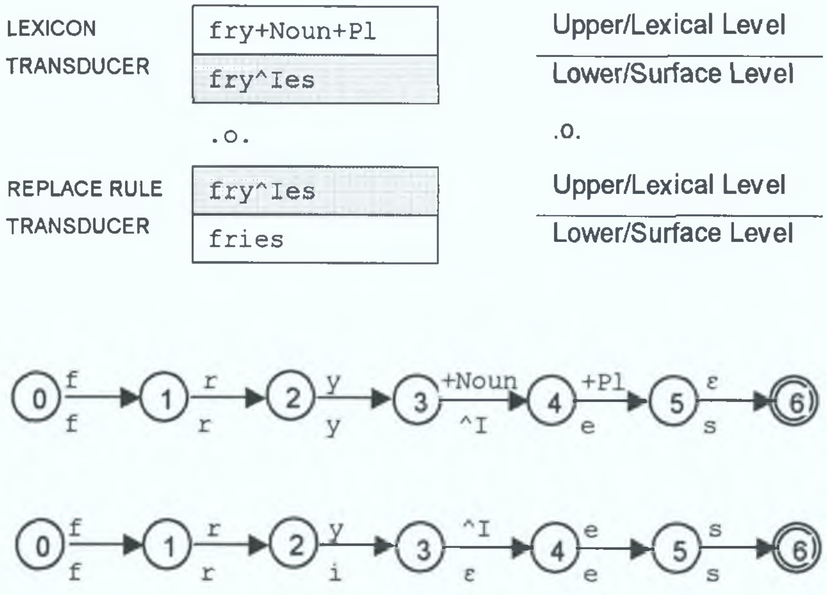


**Fig 24. Rule transducer and lexicon transducer *before* composition**

The composed transducer (Fig. 25) is bi-directional, and combines a single lexical and surface level without any intermediate levels. As Fig. 25 shows, if the lexical string 'fry+Noun+Pl' is input then the surface string 'fries' will be generated as output. Likewise, if the surface string 'fries' is input then the lexical string 'fry+Noun+Pl' will be output.

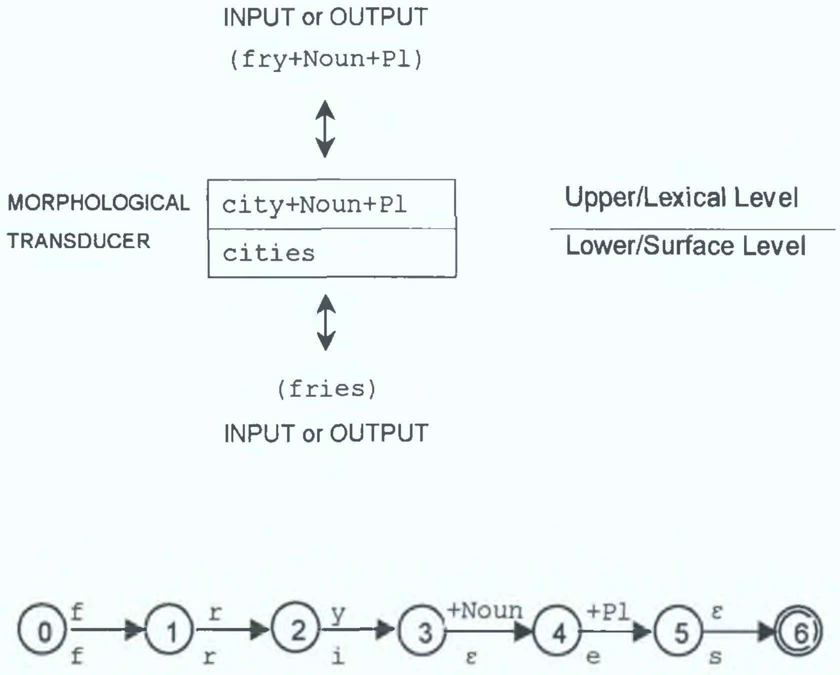


**Fig 25. Bi-directional Morphological Transducer after composition.**

## *4.9 Summary*

This chapter described finite-state automata, finite-state transducers, regular expressions and the operations which can be performed on automata and transducers. The Xerox finite-state toolset was also introduced.

# PART 2 Implementation

# 5. The Application of Finite-State Morphology to Irish

## *5.1 Introduction*

In this chapter, the implementation of a finite-state morphology of Irish is described. The scope of the work is outlined, and followed by a description of the implementation of the morphological phenomena and morphotactics of nouns, adjectives and verbs. The treatment of some commonly used non-inflected parts-of-speech is also discussed.

## *5.2 Scope of Implementation*

The present work describes the implementation of a morphological analyser and generator for Irish using finite-state morphology. The morphological rules of the language have been encoded enabling all of the inflected word-forms for nouns, adjectives and verbs to be generated from a list of stems. As the system is bi-directional, an inflected word form may also be analysed to determine its stem and morphosyntactic features. Syntax, the grouping of word-forms at the phrase or sentence level, is not addressed in this work.

A two-level morphology relates the surface representation of word-forms to their lexical representations as shown in Fig. 26. The surface level form is usually the orthographical form, i.e. the written form, although it could equally be a phonetic transcription. The lexical level form is a morphological description of the surface form. It consists of a concatenation of morphemes, usually a stem and one or more morphological description tags.

| Lexical Level | ubh+Noun+Fem+Com+Pl+Def |
|---|---|
| Surface Level | huibheacha |

**Fig 26. Two-level representation of an inflected noun**

The following examples are two-level representations of a noun, verb and adjective. (1) gives two-level representations of some more inflected forms of the noun *ubh* 'egg'. (2) shows some of the forms of the verb *déan* 'do' or 'make', and (3) shows the adjective *beag* 'small'.

(1) ubh+Noun+Fem+Com+Sg+Idf:**ubh**
ubh+Noun+Fem+Gen+Sg+Def:**huibhe**
ubh+Noun+Fem+Gen+Strong+Pl+Def:**n-uibheacha**

Through relating a surface form such as *(na) huibhe* 'of (the) egg' to its lexical root *ubh* 'egg' we can distinguish between surface forms which are mutated (*huibhe*), and surface forms such as *hata* 'of the hat' where the initial *h* is present in the lexical root *hata* 'hat'.

(2) `déan+Verb+FutInd:`**`déanfaidh`**
`déan+Verb+FutInd+Neg:`**`dhéanfaidh`**
`déan+Verb+FutInd+Q:`**`ndéanfaidh`**

(3) `beag+Adj+Fem+Com+Sg:`**`bheag`**
`beag+Adj+Masc+Com+Sg+LenYES:`**`bheag`**
`beag+Adj+Masc+Com+Sg+LenNO:`**`beag`**

Some adjectival tags contain extra information which may be required at a syntactic level. For example, after some compound prepositions a masculine noun may be either lenited or eclipsed; if it is lenited an accompanying adjective must also be lenited, but if the noun is eclipsed no change occurs to the start of the adjective (Christian Brothers, 1988, p62). The '`LenYES`' and '`LenNO`' tags in example (3) refer to the noun which the adjective is modifying, and can used to test for agreement in such cases.

Although rules for syntactically combining word-forms are not included in this work, in parsing text, the morphological analyses generated by this implementation could provide the necessary information to test for agreement between individual phrase elements. The following simple phrase (4) contains an article, noun and adjective.

(4) *na cait bheaga*
the cats small
'the small cats'

The morphological analyses of the words in (4) are as follows:

(5) `an+Art+Com+Pl+Def:`**`na`**
`cat+Noun+Masc+Com+Pl+Def:`**`cait`**
`beag+Adj+Com+Slender+Pl:`**`bheaga`**

From this analysis we can determine that there is agreement between the article *na* and the noun *cait,* both being common case, plural and definite. (The `+Def` tag on the article is for clarity and consistency only, since there is no indefinite article in Irish). There is also number agreement between the noun *cait* 'cats' and the adjective *bheaga* 'small'. The word *cait* is slender, as it ends in a slender consonant, indicated by the slender vowel *'i'* immediately preceding the last consonant. (This can be determined automatically by examining the orthography.) *Bheaga* is marked as the form used to modify slender plural nouns and therefore agrees with *cait* in this respect.

### Overview of Morphological Transducer

The morphotactics of the language, i.e. what stems and affixes can co-occur and in what order, are captured in the lexicon. Stems which require modification are marked up by adding inflectional mark-up tags (replace rule triggers) systematically introduced in the lexicon. These modifications are implemented using replace rules. Replace rules systematically consume their trigger mark-up.

The overall morphological transducer consists of several components, i.e. various lexicon transducers and replace rule transducers, as shown in Fig. 27. There is a separate lexicon for each word-class. In the case of verbs, nouns and adjectives, systematic stem modifications are encoded using replace rules.

In addition to the three main inflected parts of speech (verbs, nouns and adjectives) there are a small number of other inflected lexical classes, namely prepositional pronouns, personal pronouns, and the article. Their inflected forms are listed in full in the lexicon without the use of  rules. To enhance coverage of the language, many of the most frequently used items of the non-inflected classes such as determiners (other than the article), adverbs, prepositions (simple and compound), conjunctions, numerals, interjections, particles and abbreviations are also included in the lexicon in the same manner.

In the next sections, design and implementation of the morphological phenomena of nouns, adjectives and verbs is described in detail.

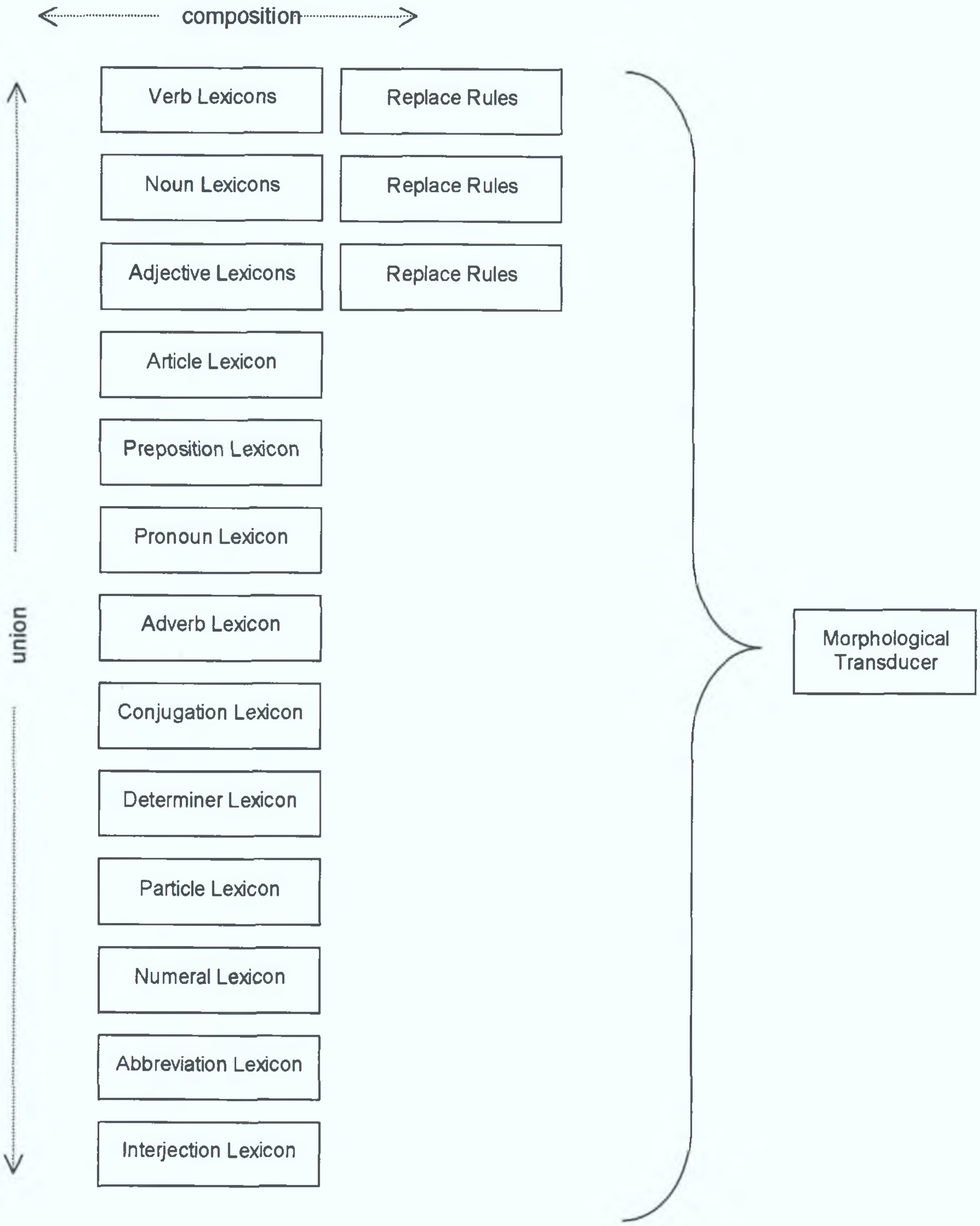


**Fig 27. Overview of Irish Morphological Transducer Architecture**

## *5.3 Design Issues*

The main objective in the design of the morphological transducer has been to make it as intuitive as possible for a lexicographer to add new items to the lexicon. Consideration has also been given to the fact that in order to achieve (near) full coverage of the language it is necessary that this task be automated to the greatest possible extent.

In order to support a modular design, a separate lexicon is used for each of the different lexical categories, i.e. the noun lexicon, the verb lexicon, the adjective lexicon. (Individual lexicon transducers can later be unioned into one overall morphological transducer using the *xfst* function *merge*.)

Within a lexicon, lexical items (stems) are assigned to separate classes depending on the inflectional suffixes and internal modifications they require. Each stem class has an associated continuation class where morphological tags and suffixes are concatenated to the stem. Internal modifications to stems cannot be implemented in the lexicon, but the fact that they are required is signalled lexically through the concatenation of special tags or replace rule triggers to the stem. The actual modifications are implemented using replace rules.

Replace rules are encoded as regular expressions, and each is applied to all (surface) strings defined in the lexicon. If the required conditions specified in the rule are present in a particular string, e.g. the specified context and trigger, then the replacement takes place. In order to generate the desired inflectional forms for a new word, a new stem must be added to the correct stem class. The stem then acquires the suffixes and replace rule triggers that have been defined for this stem class.

It is important to keep the number of lexical classes to a minimum in order to make the addition of new items as straightforward as possible. Once the classes of words are defined and documented, the task of adding new words can be carried out by a lexicographer who is not required to have any knowledge of *lexc* or programming in general (Beesley and Karttunen, 2001, p213).

The noun lexicon is particularly challenging in this respect as there is potentially a large number of classes. (In this implementation, 11 verb classes, 12 adjective classes and 50 noun classes have been identified). Complexity is consigned to replace rules rather than the lexicon whenever possible. For example, in nominal inflection there are no morphological rules for determining which plural suffix a noun stem takes, therefore, the distinction is made by using separate classes in the lexicon. On the other hand, although some plural suffixes have a broad and slender form, (to match the broad or slender nature of the stem), we do not need separate classes to distinguish between broad and slender stems. This can be determined phonologically by a replace rule and thus reduces the number of stem classes required.

In addition, if a word is regular for a majority of its inflectional forms but deviates from the paradigm in some minor way(s), it is included in the lexicon with the regular members rather than creating a special class for it. The incorrect forms which are generated are later removed from the finite-state network (by using composition to filter them out) and replaced with the correct forms. This is discussed in more detail in the Section 5.8 FST Manipulation.

In most cases, a stem class has a succession of continuation classes. These are arranged in order of generality, starting with the features specific to the particular stem class (e.g. a suffix or final mutation) and finishing with the continuation classes which are general to many stem classes (e.g. initial mutations). Each continuation class concatenates lexical and/or surface material to the end of the current string.

As is evident from Fig. 28, strings output from the lexicon (e.g. Fig. 28 line 5) are in an intermediate form, as they contain mark-up tags (triggers in the form `^Trigger`) signifying that further processing is required. The final forms are produced when all replace rules have been applied (e.g. Fig. 28 line 10).

A replace rule implements a single spelling change (mutation). The replace rules mostly relate to stems, but some are also used to model changes to affixes, e.g. in the case of allomorphs. In general, the rules deal with the triggers in the reverse order to which they were applied appended in the lexicon. General rules (e.g. initial mutations) are applied first, followed by the more specific rules. To demonstrate this process Fig. 28 shows one inflected form of the root *ceannaigh* 'buy' from the verb lexicon (see Figs. 31-33) and the results of applying the relevant rules (see Section 5.7).

| | LEXICON CONTINUATION CLASSES |
|---|---|
| 1. Lexicon Verbs | ceannaigh:ceannaigh |
| 2. Lexicon V2-BR | ceannaigh:ceannaigh^aigh |
| 3. Lexicon V2-BR-0 | ceannaigh+Verb+PresInd:ceannaigh^aigh^Verb |
| 4. Lexicon V2-BR-PresInd | ceannaigh+Verb+PresInd:ceannaigh^aigh^Verbaíonn |
| 5. Lexicon NegQ | ceannaigh+Verb+PresInd+Neg:ceannaigh^aigh^Verbaíonn^Sé |
| | REPLACE RULE CASCADE |
| 6. Replace Rule 1 | ceannaigh+Verb+PresInd+Neg:cheannaigh^aigh^Verbaíonn^Sé |
| 7. Replace Rule 4 | ceannaigh+Verb+PresInd+Neg:cheannaigh^aigh^Verbaíonn |
| 8. Replace Rule 15 | ceannaigh+Verb+PresInd+Neg:cheann^aigh^Verbaíonn |
| 9. Replace Rule 17 | ceannaigh+Verb+PresInd+Neg:cheann^Verbaíonn |
| 10. Replace Rule 32 | ceannaigh+Verb+PresInd+Neg:cheannaíonn |

**Fig 28. Inflected verb form using a cascade of replace rules**

In addition to keeping the number of classes to a minimum, this design uses the least number of mark-up tags. It dispenses with them as early as possible in order to simplify the strings, and consequently allow for less cumbersome replace rule regular expressions. Noun initial mutations use gender and case mark-up tags (`^M`, `^F`, `^C`, `^G` and `^V`) but since they are not used in any subsequent rule they are dispensed with at this point. The final mutation rules rely mainly on triggers, indicating the process required rather than the features of the stem itself. All tags, both morphological and replace rule triggers, are documented in a tag grammar to ensure consistency of naming and order of concatenation (Section 6.2 and Appendix B).

The noun and verb lexicons are organised using the traditional declensional and conjugational paradigms (Christian Brothers, 1988; Bráithre Críostaí, 1999), which describe the morphological and morphosyntactic features of these word-classes. This use of the declensional categories facilitates the re-use of existing lexical resources to populate a two-level lexicon, e.g. semi-automatic processing of machine readable dictionaries, most of which contain some grammatical information. An example of the type of grammatical and phonetic information available in *An Foclóir Póca* (An Roinn Oideachais, 1986a) is given in (6), where *f2* indicates "feminine noun, second declension".

(6) *lampróg* lampro:g f2 glow-worm; firefly

The lexicon is designed to be flexible in terms of dialect and standard language. The work reported in the present dissertation is based mainly on the standard grammar references (Christian Brothers, 1988; Bráithre Críostaí, 1999; Rannóg an Aistriúcháin, 1958), but includes tags for dialectal forms. Currently the three broad categories, `+CC` *Canúint Chonnachta*, 'Connaught Dialect', `+CD` *Canúint Dhún na nGall*, 'Donegal Dialect' and `+CM` *Canúint na Mumhan*, 'Munster Dialect' have been included for illustrative purposes. Forms which belong to a particular dialect may be included in the lexicon and tagged as such (see Appendix B). Forms which are not marked for dialect are assumed to be standard (common) forms. When forms are appropriately tagged, it is possible to extract various subsets of language from the final transducer, e.g. standard forms only, standard forms plus a particular dialect. The same principle could also be applied to historical forms, e.g. perhaps `sgéal+Hist+Noun` 'story' representing an older form of `scéal+Noun` 'story'.

In the following sections the lexicons and replace rules are described. Fig. 29 gives a more detailed view of how the components of the system are related.

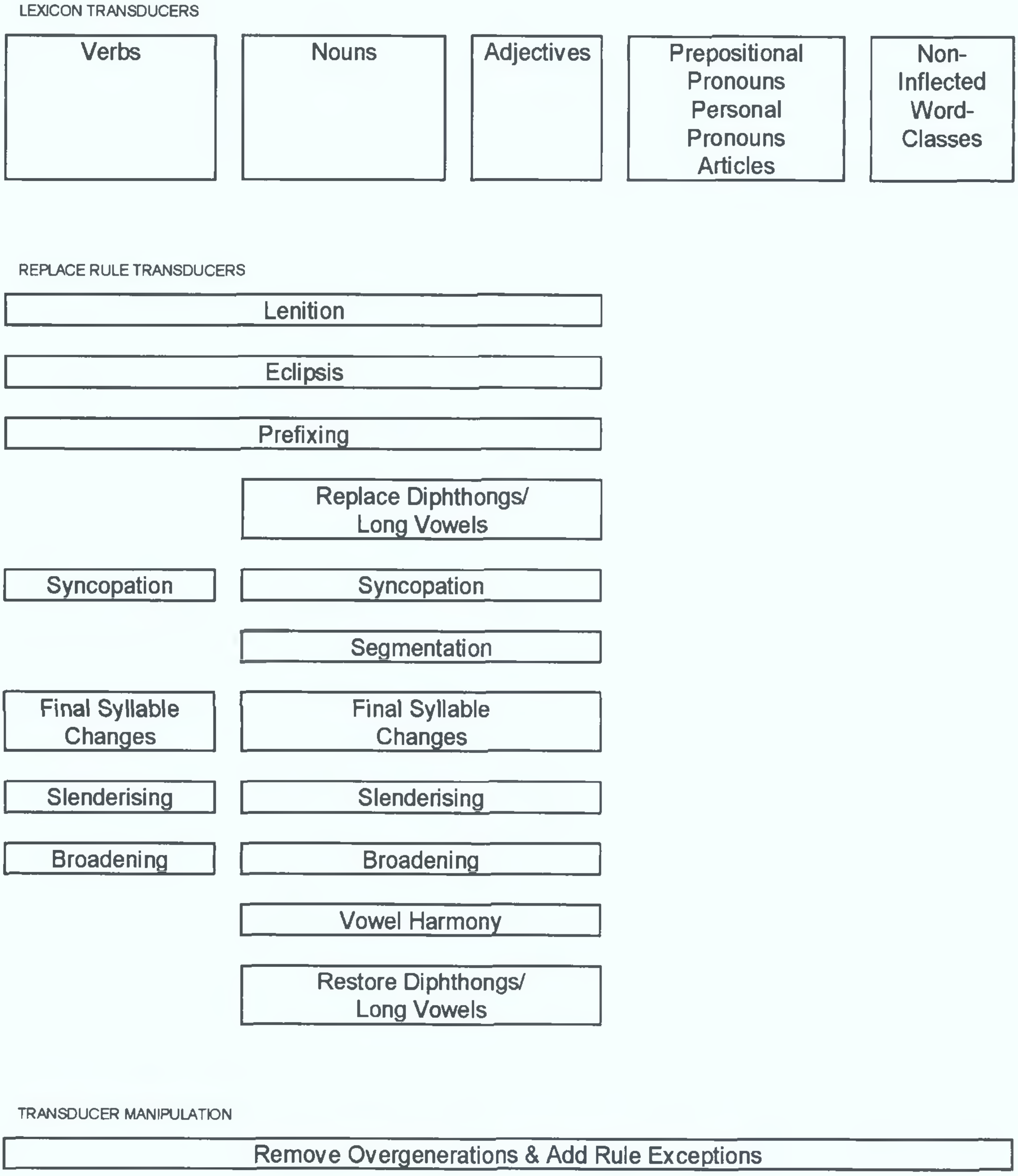


**Fig 29. Morphological Transducer Architecture**

## *5.4 Verb Lexicon*

This section describes the implementation of inflectional verb morphology in the lexicon. (Verbal nouns and verbal adjectives are not covered in the current implementation).

Verbs are encoded in three separate lexicons. The first lexicon contains monosyllabic stems and a small number of polysyllabic stems (first conjugation), the second contains polysyllabic stems (second conjugation), and the third contains irregular and defective verbs. Both conjugations include broad and slender stems. The second conjugation class also includes some polysyllabic stems, which undergo syncopation. In the third lexicon, all inflected forms are listed together with their morphosyntactic analysis, as shown in example (7) where *déarfaimid*, 'we will say' and *beimid*, 'we will be' are given.

(7) `abair+Verb+FutInd+1P+Pl:déarfaimid`
`bí+Verb+FutInd+1P+Pl:beimid`

The verbal root given in Irish dictionaries is usually the second person imperative indicative, e.g. Ó Dónaill (1977). This form is used as the stem in the present work in all but a small number of cases, where a different stem is chosen for processing convenience. For example, *freastal* rather than *freastail* 'attend' is used, likewise *taisteal* rather than *taistil*, 'travel' is used, as these forms require less morphological change in general.

(8) *freastail* 'attend' freastal+Verb+Imper+2P+Sg
*freastalaíonn* 'attend' freastal+Verb+PresInd
*freastalóimid* 'we will attend' freastal+Verb+FutInd+1P+Pl
(9) *taistil* 'travel' taisteal+Verb+Imper+2P+Sg
*taistealaíonn*, 'travel' teasteal+Verb+PresInd
*taistealóimid* 'we will travel' teasteal+Verb+FutInd+1P+Pl

The regular verbs are divided up in the lexicons as follows:

- Conjugation
    - Broad or Slender Stem Ending
        - Tense and Mood

Fig. 30 shows the components in the overall verb processing scheme. The final mutation components are applied to the regular verbs only (first and second conjugation). Although the irregular verbs have many suppletive forms they still conform to initial mutation rules, therefore initial mutation replace rules apply to regular and irregular lexicon output.

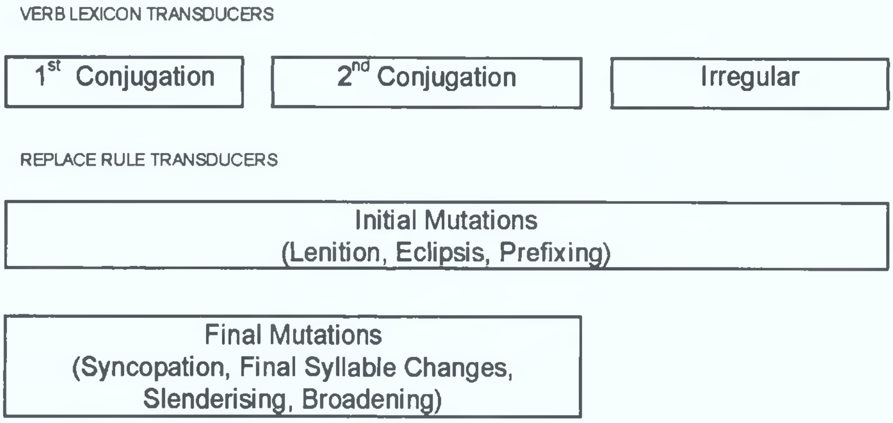


**Fig 30. Verb Transducer Architecture**

*Extract from Verb Lexicon*

Figs. 31 to 33 contains an extract from the 2nd Conjugation Verb Lexicon, which is representative of verb lexicons in general. A full listing of verb classes may be seen in Appendix G.

Initially, morphosyntactic tags and inflectional mark-up tags are declared as multi-character symbols at the start of the lexicon. The morphosyntactic tags used in these lexicons follow, where possible, the recommendations of Beesley and Karttunen (2001, Appendix C) and are closely related to the morphosyntactic features described in the PAROLE Morphosyntactical Tagset (Appendix C).

The `Verbs` lexicon is declared as the root lexicon. The `Verbs` lexicon contains several examples of both broad and slender polysyllabic stems, including some which are syncopated.
The stem ***ceannaigh*** 'buy', for example, points towards the continuation class `V2-BR`, which as the name indicates, contains verbs from the second conjugation which are broad. As this is a polysyllabic stem ending in ***–aigh***, the final syllable must be removed before the appropriate suffixes can be added; therefore the inflectional mark-up tag `^aigh` is appended to the surface form before moving on to the next continuation class, `V2-BR-0`.

In `V2-BR-0`, three actions are performed: the various tense/mood morphological tags, e.g. `+Verb+PresInd` are appended to the lexical level, an inflectional mark-up tag e.g. `^Verb`, is appended to the surface level, and there is a pointer to the appropriate next continuation class.

There is a choice of eight routes which the string arriving at lexicon `V2-BR-0` can take through the network at this stage. The first pointer is to `V2-BR-PresInd`, which is the class of present indicative suffixes. The first suffix `-aionn` is the default suffix which does not contain any person or number information; therefore no information is added to the lexical side (this suffix form must be used in conjunction with a pronoun). The next two suffixes `-aím` and `-aímid` are used for first person singular and plural, respectively, and the relevant information is appended to the lexical level, i.e. `+1P+Sg` and `+1P+Pl`. Lastly, the autonomous suffix `-aítear`, is given. This is the impersonal form, and is marked accordingly with `+Auto` on the lexical level. Each of these four forms point to continuation class `NegQ`. `NegQ` completes the inflected forms by appending the relevant tags for the positive, negative, interrogative, and negative-interrogative forms. The positive form is unchanged, otherwise a lenition mark-up tag, `^Sé`, is required to create the negative form, or an eclipsis mark-up tag, `^Urú`, is required to create the question form and negative-question form.

The manner in which these inflectional mark-up tags (replace rule triggers) are processed will be described in Section 5.7.

```
Multichar_Symbols
+Verb +1P +2P +3P +Auto +Sg +Pl
+PresInd +PastInd +FutInd +ImpInd +Cond +PresSubj +PastSubj
+Imper +Neg +Q

^Verb ^Sé ^Caol ^Lea ^LeaS ^VAdj ^Vnoun ^igh ^Fr ^Urú

LEXICON Root
        Verbs;

LEXICON Verbs

!STEM                   CONT. CLASS       GLOSS
ceannaigh               V2-BR;            ! buy
clúdaigh                V2-BR;            ! cover
freastal                V2-BR;            ! attend
bailigh                 V2-SL;            ! gather
cuimhnigh               V2-SL;            ! remember
bagair                  V2-BR-sync;       ! threaten
ceangail                V2-BR-sync;       ! tie
coigil                  V2-SL-sync;       ! save
eitil                   V2-SL-sync;       ! fly

LEXICON V2-BR                 ! 2nd. Conj. - broad
0:^aigh                       V2-BR-0;

LEXICON V2-SL                 ! 2nd. Conj. - slender
0:^Caol^igh                   V2-BR-0;

LEXICON V2-BR-sync            ! 2nd. Conj. - broad & syncop.
0:^Coim                       V2-BR-0;

LEXICON V2-SL-sync            ! 2nd. Conj. - slender & syncop.
0:^Caol^Coim                  V2-BR-0;

LEXICON V2-BR-0               ! 2nd. Conjugation - broad
+Verb+PresInd:^Verb           V2-BR-PresInd;
+Verb+PastInd:^Verb           V2-BR-PastInd;
+Verb+FutInd:^Verb            V2-BR-FutInd;
+Verb+ImpInd:^Verb            V2-BR-ImpInd;
+Verb+Cond:^Verb              V2-BR-Cond;
+Verb+PresSubj:^Verb          V2-BR-PresSubj;
+Verb+PastSubj:^Verb          V2-BR-PastSubj;
+Verb+Imper:^Verb             V2-BR-Imper;

LEXICON V2-BR-PresInd
0:aíonn           NegQ;
+1P+Sg:aím        NegQ;
+1P+Pl:aímid      NegQ;
+Auto:aítear      NegQ;

LEXICON V2-BR-PastInd
0:^Fr^Sé          NegQPast;
+1P+Pl:aíomar^Sé  NegQPast;
+Auto:aíodh       NegQAuto;
```

Fig 31. Extract of Irish 2nd. Conjugation Verb Lexicon – Part i

```
LEXICON V2-BR-FutInd
0:óidh            NegQ;
+1P+Pl:óimid      NegQ;
+Auto:ófar        NegQ;

LEXICON V2-BR-ImpInd
+1P+Sg:aínn^Sé          NegQLen;
+2P+Sg:aíteá^Sé         NegQLen;
+3P+Sg:aíodh^Sé         NegQLen;
+1P+Pl:aímis^Sé         NegQLen;
+2P+Pl:aíodh^Sé         NegQLen;
+3P+Pl:aídís^Sé         NegQLen;
+Auto:aítí^Sé           NegQLen;

LEXICON V2-BR-Cond
+1P+Sg:óinn^Sé          NegQLen;
+2P+Sg:ófá^Sé           NegQLen;
+3P+Sg:ódh^Sé           NegQLen;
+1P+Pl:óimis^Sé         NegQLen;
+2P+Pl:ódh^Sé           NegQLen;
+3P+Pl:óidís^Sé         NegQLen;
+Auto:ófaí^Sé           NegQLen;

LEXICON V2-BR-PresSubj
0:aí              NegSubj;
+1P+Pl:aímid      NegSubj;
+Auto:aítear      NegSubj;

LEXICON V2-BR-PastSubj
+1P+Sg:aínn       NegQLen;
+2P+Sg:aíteá      NegQLen;
+3P+Sg:aíodh      NegQLen;
+1P+Pl:aímis      NegQLen;
+2P+Pl:aíodh      NegQLen;
+3P+Pl:aídís      NegQLen;
+Auto:aítí        NegQLen;

LEXICON V2-BR-Imper
+1P+Sg:aím        NegImper;
+2P+Sg:^Fr        NegImper;
+3P+Sg:aíodh      NegImper;
+1P+Pl:aímis      NegImper;
+2P+Pl:aígí       NegImper;
+3P+Pl:aídís      NegImper;
+Auto:aítear      NegImper;

LEXICON NegQ
#;
+Neg:^Sé    #;
+NegQ:^Urú  #;
+Q:^Urú     #;
```

**Fig 32. Extract of Irish 2nd. Conjugation Verb Lexicon – Part ii**

```
LEXICON NegQPast
0:^Sé         #;
+Neg:^Sé      #;
+NegQ:^Sé     #;
+Q:^Sé        #;

LEXICON NegQSaor
#;
+Neg:0        #;
+NegQ:0       #;
+Q:0          #;

LEXICON NegQLen
0:^Sé         #;
+Neg:^Sé      #;
+NegQ:^Sé     #;
+Q:^Urú       #;

LEXICON NegImper
#;
+Neg:0        #;

LEXICON NegSubj
0:^Urú        #;
+Neg:^Sé      #;
```

Fig 33. Extract of Irish 2nd. Conjugation Verb Lexicon – Part iii

## *5.5 Noun Lexicon*

As with the verb lexicon, all concatenative noun morphology is handled in the lexicon including the insertion of inflectional mark-up tags which act as triggers for stem modification replace rules. As before, the lexicon surface level containing inflectional mark-up tags is, in fact, an **intermediate surface level**. These surface level strings are modified by a series of replace rule transducers as shown in Fig. 34 in order to achieve the desired surface level string.

Fig. 34 shows the components which are composed to create the noun FST. The noun and irregular noun lexicons, and the various replace rule components each constitute finite state transducers. The noun lexicons are merged (unioned) and then composed with the first replace rule transducer, the result of which is composed with the next rule transducer and so on.

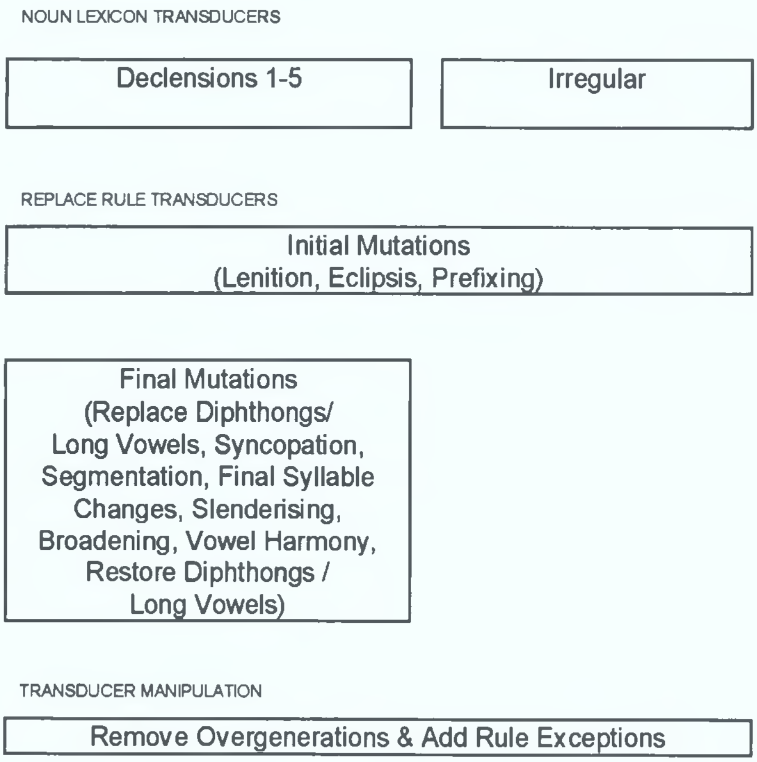

NOUN LEXICON TRANSDUCERS
Declensions 1-5
Irregular
REPLACE RULE TRANSDUCERS
Initial Mutations (Lenition, Eclipsis, Prefixing)
Final Mutations (Replace Diphthongs/ Long Vowels, Syncopation, Segmentation, Final Syllable Changes, Slenderising, Broadening, Vowel Harmony, Restore Diphthongs / Long Vowels)
TRANSDUCER MANIPULATION
Remove Overgenerations & Add Rule Exceptions

**Fig 34. Noun Transducer Architecture**

The noun lexicon is organised primarily according to noun declension. There are five declensions based on the formation of the genitive singular of nouns. Each declension is further sub-divided into masculine and feminine nouns (as initial mutation varies according to the gender of the noun). Within gender, nouns are further sub-categorised according to their method of plural formation.

- Declension
    - Gender
        - Plural formation

The sample of the Irish noun lexicon given in Figs. 35-37 contains lexical entries from seven classes covering three declensions, both genders, and a variety of plural formations. A full listing of noun classes may be seen in Appendix F.

Three of the stems listed, *cat*, 'cat', *fear* 'man', and *éan* 'bird' belong to continuation class `Nm1-1`. As the naming convention indicates, this class contains nouns which are masculine, form their genitive singular according to the first declension, and whose plurals are formed according to subcategory one.

`Nm1-1` contains four continuation classes; one of which handles the singular inflected forms, i.e. `Nm1-Singular,` and three which handle the plural form of the three cases; common, genitive and vocative. (Some lexical items such as *aonar* 'one person' have no plural form and therefore point directly to the `Nm1-Singular` rather than going through `Nm1-1`.)

The nouns in `Nm1-1` have weak plurals, i.e. the plural is not the same for all grammatical cases, and therefore each form proceeds to a different continuation class, i.e. `PL-CAOLÚ`, `PL-TADA` and `PL-(LEA)A`. Both `Nf2-6` and `Nf4-1` have strong plurals; therefore each of the three cases uses the same continuation class, e.g. `PL-(E)ANNA` or `PL-Í,` respectively.

The morphosyntactic tags for lexical category, gender, and case, e.g. `+Noun+Masc+Com`, are assigned in `Nm1-Singular` along with surface mark-up tags for gender and case, e.g. `^M^C` which will be used in the processing of initial mutations (Fig. 34). Strings in `Nm1-Singular` proceed to three continuation classes: `Com-sg`, `Gen-sg-D1`, and `Voc-sg-1`, where final mutation mark-up tags are appended for each of the three cases.

```
Multichar_Symbols
+Prop +Noun +Masc +Fem +Com +Gen +Voc +Sg +Pl +Def +Idf
+Strong +Weak +CC +CD +CM
^F ^M ^C ^G ^V ^Urú ^Sé ^hv ^tv ^ts ^Lea ^Caol ^Emph

LEXICON Root
        Nouns;

LEXICON Nouns

!STEM                   CONT. CLASS        GLOSS
aodhán                  Nm1-Prop;          ! Aodhán (name)
aonar                   Nm1-Singular;      ! one person
cat                     Nm1-1;             ! cat
fear                    Nm1-1;             ! man
éan                     Nm1-1;             ! bird
áit                     Nf2-6;             ! place
eibhlín                 Nf4-Prop;          ! Eibhlín (name)
saoirse                 Nf4-Singular;      ! freedom
réalta                  Nf4-1;             ! star

LEXICON Nm1-Prop
+Prop:0                 Nm1-Singular;

LEXICON Nm1-Singular
+Noun+Masc+Com:^M^C     Com-sg;
+Noun+Masc+Gen:^M^G     Gen-sg-D1;
+Noun+Masc+Voc:^M^V     Voc-sg-1;

LEXICON Nm1-1
Nm1-Singular;
+N+Masc+Com:^C          PL-CAOLÚ;
+N+Masc+Gen+Weak:^G     PL-TADA;
+N+Masc+Voc:^V          PL-(LEA)A;

LEXICON Nf2-Singular
+Noun+Fem+Com:^F^C      Com-sg;
+Noun+Fem+Gen:^F^G      Gen-sg-D2;
+Noun+Fem+Voc:^F^V      Voc-sg-0;

LEXICON Nf2-6
Nf2-Singular;
+Noun+Fem+Com:^C        PL-(E)ANNA;
+Noun+Fem+Gen+Strong:^G PL-(E)ANNA;
+Noun+Fem+Voc:^V        PL-(E)ANNA;

LEXICON Nf4-Prop
+Prop:0           Nf4-Singular;

LEXICON     Nf4-Singular
+Noun+Fem+Com:^F^C      Com-sg;
+Noun+Fem+Gen:^F^G      Gen-sg-D4;
+Noun+Fem+Voc:^F^V      Voc-sg-0;
```

Fig 35. Extract of Irish Noun Lexicon – Part i

```
LEXICON     Nf4-1
Nf4-Singular;
+Noun+Fem+Com:^C        PL-Í;
+Noun+Fem+Gen+Strong:^G PL-Í;
+Noun+Fem+Voc:^V        PL-Í;

LEXICON Com-sg
+Sg:0             Com-sg-initial;

LEXICON Gen-sg-D1
+Sg:^Caol         Gen-sg-initial;

LEXICON Gen-sg-D2
+Sg:^Caole        Gen-sg-initial;

LEXICON Gen-sg-D4
+Sg:0             Gen-sg-initial;

LEXICON Voc-sg-0
+Sg:^Sé           #;

LEXICON Voc-sg-1
+Sg:^Caol^Sé      #;

LEXICON PL-TADA
+Pl:0             Pl-initial;

LEXICON PL-CAOLÚ
+Pl:^Caol         Pl-initial;

LEXICON PL-(LEA)A
+Pl:^Leaa         Pl-initial;

LEXICON PL-(E)ANNA
+Pl:^LCanna       Pl-initial;

LEXICON PL-Í
+Pl:í             Pl-initial;

LEXICON Com-sg-initial
+Def:^Sé^tv^ts    #;
+Idf:0            Com-sg-emphasis;
+Len:^Sé          Com-sg-emphasis;
+Urú:^Urú         Com-sg-emphasis;

LEXICON Com-sg-emphasis
#;
+Emph+1P:s^Emph   #;
+Emph+2P:s^Emph   #;
+Emph+3P:s^Emphan #;
```

Fig 36. Extract of Irish Noun Lexicon – Part ii

```
LEXICON Gen-sg-initial
+Idf:0            #;
+Def:^Sé^hv^ts    #;

LEXICON Pl-initial
+Idf:0            Pl-emphasis;
+Def:^Sé^Urú^hv   #;

LEXICON Pl-emphasis
#;
+Emph+1P:n^Emph   #;
+Emph+2P:s^Emph   #;
+Emph+3P:s^Emph   #;
```

**Fig 37. Extract of Irish Noun Lexicon – Part iii**

The continuation class `Com-sg` adds the morphosyntactic tag `+Sg` (no final mutation triggers are required since this is the unmarked form) before proceeding to the initial mutations continuation class `Com-Sg-initial`. This class has two entries, one each for definite and indefinite nouns. In the case of definite nouns (i.e. nouns preceded by an article) the tag `+Def` is concatenated to the lexical representation, and initial mutation triggers are concatenated to the surface representation, i.e. mark-up tags `^Sé^tv^ts`. For any particular stem only one of these three mark-up tags will succeed depending on its phonological characteristics, i.e. if it begins with a vowel or '*s*' then '*t*' will be prefixed to it, otherwise it will be lenited. Indefinite nouns, i.e. those without an article, receive an `+Idf` tag and no initial mutation triggers are required.

There are, however, other grammatical conditions (apart from the presence of the definite article) which cause initial mutation in nouns. Therefore such strings also have general purpose lenition and eclipsis tags applied in `Com-Sg-initial`. (An inflected form never receives more than one initial mutation: lenition and eclipsis never co-occur).

Nouns, synthetic verb-forms, and prepositional pronouns all have emphatic forms. The emphatic form of a noun can be used after possessive determiners, such as *mo* 'my', *do* 'your' etc. This form is created by appending the appropriate emphatic suffix, e.g. *–se*, *-sa* etc., to the surface form. The indefinite, lenited, and eclipsed forms all point to the `Com-sg-emphasis` class where the string may terminate unchanged, e.g. by encountering the end-of-string marker '#', or receive a particular emphatic suffix and mark-up tag, e.g. `s^Emph` and related emphasis and person tag, e.g. `+Emph+1P`.

The continuation class `Gen-sg-D1` appends the tag `+Sg` to the lexical level and the mark-up tag '`^Caol`' to the surface level. This will create the required context for a slenderising replace rule to fire, as the genitive singular of first declension nouns is formed by slenderising. The next stop is the continuation class for initial mutations of genitive singular nouns, `Gen-sg-initial`.

The `Gen-sg-initial` class contains both definite and indefinite lexical forms, again depending on whether the noun is preceded by an article or not. A genitive singular without an article receives the `+Idf` lexical tag and no further mark-up. The definite form receives the `+Def` lexical tag and the relevant surface mark-up tags, i.e. `^Sé^hv^ts`.

The continuation class `Voc-sg-1` also adds the lexical tag `+Sg`, and mark-up tags `^Caol` and `^Sé`. The initial mutation (`^Sé`) is handled in this class since there is no distinction between definite and indefinite (unlike common and genitive cases).

Each of the three plural continuation classes associated with `Nm1-1`: `PL-CAOLÚ`, `PL-TADA` and `PL-(LEA)A`, append the tag `+Pl` on the lexical level and the appropriate inflectional mark-up tags on the surface level. Class `PL-CAOLÚ` appends the mark-up tag `^Caol` (slender) and `PL-(LEA)A` appends both a mark-up tag `^Lea` (broad) and a suffix '*–a*'. `PL-TADA` does not append any mark-up tags as this plural form does not require any final mutation.

In all cases, the plural continuation classes proceed to `Pl-initial` which appends the appropriate initial mutation lexical and surface mark-up, and emphatic lexical and surface mark-up.

**Lexicon Output**

The output for two stems from this lexicon extract is listed in full below. Table 22 lists the output of the lexicon transducer for the stem *áit* 'place'. This is followed by Fig. 38 showing a two-level network with lexical and surface levels, and state transitions. This particular stem has strong plurals: all end in *-anna* or *-eanna*.

The second example *cat* 'cat', (Table 23), is similar except for the fact that it has weak plurals. This is evident from Fig. 39, in that plural forms take three different paths.

Table 22. Lexicon output for stem *áit* 'place'

| *Lexical level: morphological tags* | *Surface level (including inflectional mark-up tags)* |
|---|---|
| `áit+Noun+Fem+Com+Pl+Def` | `áit^C^LCanna^Sé^Urú^hv` |
| `áit+Noun+Fem+Com+Pl+Idf` | `áit^C^LCanna` |
| `áit+Noun+Fem+Com+Sg+Def` | `áit^F^C^Sé^tv^ts` |
| `áit+Noun+Fem+Com+Sg+Idf` | `áit^F^C` |
| `áit+Noun+Fem+Com+Sg+Idf+Emph` | `áit^F^Cs^Emph` |
| `áit+Noun+Fem+Com+Sg+Idf+Len` | `áit^F^C^Sé` |
| `áit+Noun+Fem+Com+Sg+Idf+Len+Emph` | `áit^F^C^Sés^Emph` |
| `áit+Noun+Fem+Com+Sg+Idf+Urú` | `áit^F^C^Urú` |
| `áit+Noun+Fem+Com+Sg+Idf+Urú+Emph` | `áit^F^C^Urús^Emph` |
| `áit+Noun+Fem+Gen+Sg+Def` | `áit^F^G^Caole^Sé^hv^ts` |
| `áit+Noun+Fem+Gen+Sg+Idf` | `áit^F^G^Caole` |
| `áit+Noun+Fem+Gen+Strong+Pl+Def` | `áit^G^LCanna^Sé^Urú^hv` |
| `áit+Noun+Fem+Gen+Strong+Pl+Idf` | `áit^G^LCanna` |
| `áit+Noun+Fem+Voc+Pl+Def` | `áit^V^LCanna^Sé^Urú^hv` |
| `áit+Noun+Fem+Voc+Pl+Idf` | `áit^V^LCanna` |
| `áit+Noun+Fem+Voc+Sg` | `áit^F^V^Sé` |

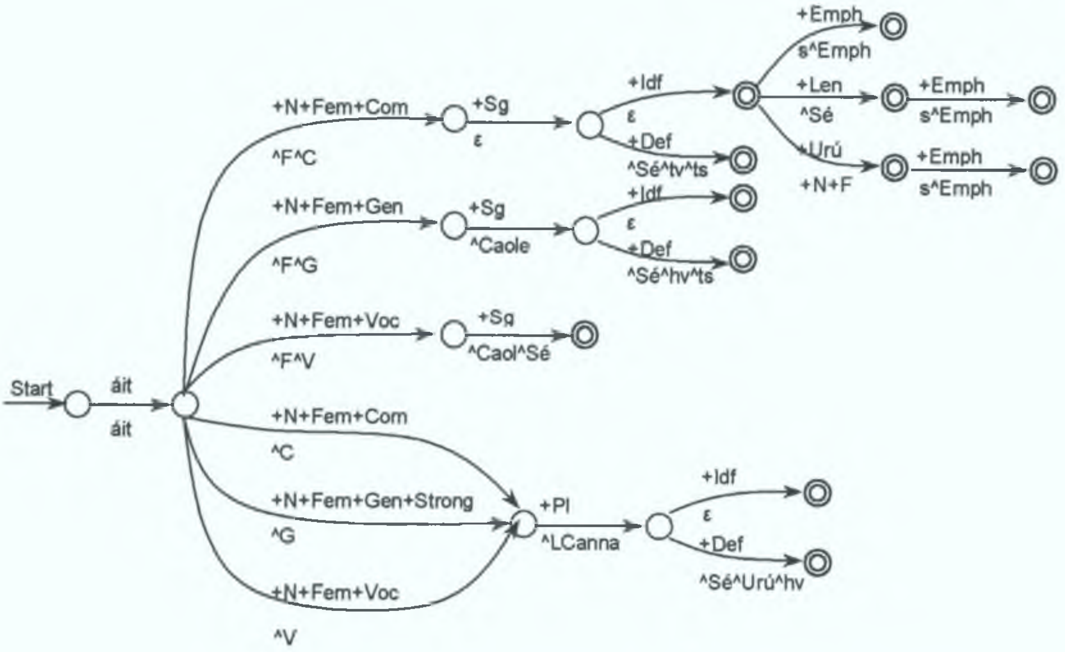


Fig 38. Two-level network for the stem *áit* 'place' (strong plural)

Table 23. Lexicon output for stem *cat* 'cat'

| *Lexical level: morphological tags* | *Surface level (including inflectional mark-up tags)* |
|---|---|
| `cat+Noun+Masc+Com+Pl+Def` | `cat^C^Caol^Sé^Urú^hv` |
| `cat+Noun+Masc+Com+Pl+Idf` | `cat^C^Caol` |
| `cat+Noun+Masc+Com+Sg+Def` | `cat^M^C^Sé^tv^ts` |
| `cat+Noun+Masc+Com+Sg+Idf` | `cat^M^C` |
| `cat+Noun+Masc+Com+Sg+Idf+Emph` | `cat^M^Cs^Emph` |
| `cat+Noun+Masc+Com+Sg+Idf+Len` | `cat^M^C^Sé` |
| `cat+Noun+Masc+Com+Sg+Idf+Len+Emph` | `cat^M^C^Sés^Emph` |
| `cat+Noun+Masc+Com+Sg+Idf+Urú` | `cat^M^C^Urú` |
| `cat+Noun+Masc+Com+Sg+Idf+Urú+Emph` | `cat^M^C^Urús^Emph` |
| `cat+Noun+Masc+Gen+Sg+Def` | `cat^M^G^Caol^Sé^hv^ts` |
| `cat+Noun+Masc+Gen+Sg+Idf` | `cat^M^G^Caol` |
| `cat+Noun+Masc+Gen+Weak+Pl+Def` | `cat^G^Sé^Urú^hv` |
| `cat+Noun+Masc+Gen+Weak+Pl+Idf` | `cat^G` |
| `cat+Noun+Masc+Voc+Pl+Def` | `cat^V^Leaa^Sé^Urú^hv` |
| `cat+Noun+Masc+Voc+Pl+Idf` | `cat^V^Leaa` |
| `cat+Noun+Masc+Voc+Sg` | `cat^M^V^Caol^Sé` |

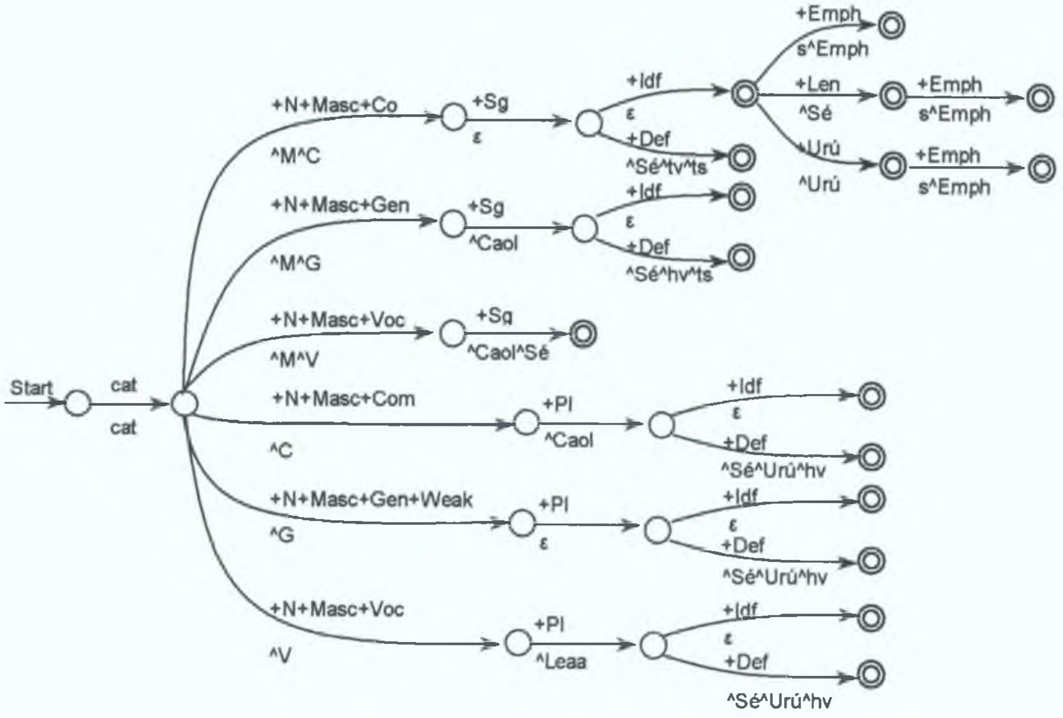


Fig 39. Two-level network for the stem *cat* 'cat' (weak plural)

## *5.6 Adjective Lexicon*

Adjective inflection employs the same processes as noun inflection but to a lesser degree[9].

As there does not appear to be a general consensus among the grammar reference works consulted (Christian Brothers, 1988; Bráithre Críostaí, 1999; Ó Baoill and Ó Tuathail, 1992; Rannóg an Aistriúcháin, 1958) on the declensions of adjectives, the following new scheme has been adopted for the purposes of this work. Four categories based on stem ending, i.e. slender stems, broad stems, stems ending in a vowel and stem endings requiring syncopation, are distinguished.

- Stem ending
    - Gender
        - Plural formation

Fig. 40 shows the various components used in the processing of Adjectives. The same replace rule modules are used for both nouns and adjectives (c.f. Fig. 34).

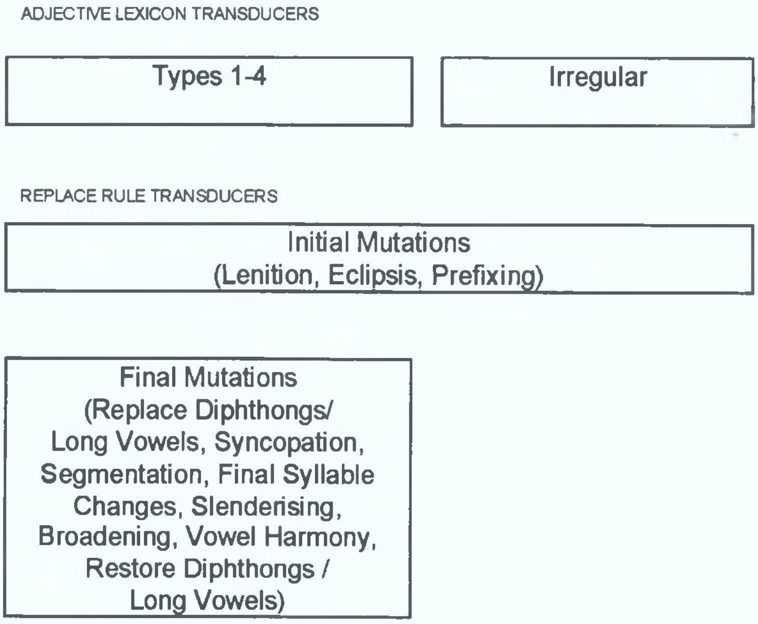


**Fig 40. Adjective Transducer Architecture**

Sample adjectives from each of the four categories of adjectives are shown in the extract (Fig. 41-43) from the Adjective Lexicon. A full listing of Adjective Classes may be seen in Appendix H.

Five sub-classes of broad stemmed adjective (Type 1) are distinguished in this implementation. Fig. 41 shows samples from the first and third sub-classes. For example, *fliuch* 'wet', belongs to class `Adj1-1` containing broad stemmed adjectives of sub-class 1 (ie. monosyllabic endings in –ll, -nn and –ch(t)). As adjectives take their gender from the noun which they qualify, there are six continuation classes; both masculine and feminine singular forms of each of the three grammatical cases.

There are five continuation classes for plural forms. Gender is not a distinguishing factor in plural forms, but common case plurals differ depending on whether the preceding noun is slender or not, i.e. if the noun is slender then the adjective must also be slender.

Likewise, genitive case plurals differ depending on whether the preceding noun has a strong plural or not. If the noun has weak plurals the adjective is not inflected; otherwise it is inflected according to its lexical class (`Adj1-1`), in this case by appending the suffix *–a.* Two adjectives from lexicon `Adj1-3` (broad stemmed adjectives ending in –(e)ach), are also given in Fig. 41.

Type 2 adjectives are slender stemmed. Two sub-classes are identified and examples of each are shown. Type 3 adjectives are those ending in a vowel. All belong to the same class and undergo very little inflection other than lenition of common case plurals when the noun they qualify is slender. Type 4 adjectives are those which undergo syncopation. Two samples, *ramhar* 'fat' and *folamh* 'empty', are given in Fig. 41.

```
Multichar_Symbols
+Adj +Pos +Comp +Sup +Masc +Fem +Com +Gen +Voc +Sg +Pl
+Weak +Strong +Slender +NotSlen +LenNO +LenYES

^Adj ^Sé ^Ath ^Caol ^Lea ^Coim

LEXICON Root
        Adjectives;

LEXICON Verbs

!STEM                   CONT. CLASS       GLOSS
fliuch                  Adj1-1;           ! wet
gann                    Adj1-1;           ! scarce

iontach                 Adj1-3;           ! wonderful
baolach                 Adj1-3;           ! dangerous

áitiúil                 Adj2-1;           ! local
bliantúil               Adj2-1;           ! annual, yearly

maith                   Adj2-2;           ! good

aibí                    Adj3-1;           ! ripe
buí                     Adj3-1;           ! yellow

ramhar                  Adj4-1;           ! fat
folamh                  Adj4-1;           ! empty


LEXICON Adj1-1
+Adj+Fem+Com:^Adj             Fem_com_voc_sg;
+Adj+Fem+Gen:^Adj             Fem_gen-D2A_sg;
+Adj+Fem+Voc:^Adj             Fem_com_voc_sg;

+Adj+Masc+Com:^Adj            Masc_com_sg;
+Adj+Masc+Gen:^Adj            Masc_gen-D4_voc_sg;
+Adj+Masc+Voc:^Adj            Masc_gen-D4_voc_sg;

+Adj+Com+NotSlen:^Adj         PL-A;
+Adj+Com+Slender:^Adj         PL-A-SLENDER;
+Adj+Gen+Weak:^Adj            PL-TADA;
+Adj+Gen+Strong:^Adj          PL-A;
+Adj+Voc:^Adj                 PL-A;

LEXICON Adj1-3
+Adj+Fem+Com:^Adj             Fem_com_voc_sg;
+Adj+Fem+Gen:^Adj             Fem_gen-D2B_sg;
+Adj+Fem+Voc:^Adj             Fem_com_voc_sg;
+Adj+Masc+Com:^Adj            Masc_com_sg;
+Adj+Masc+Gen:^Adj            Masc_gen-D1_voc_sg;
+Adj+Masc+Voc:^Adj            Masc_gen-D1_voc_sg;
```

**Fig 41. Extract of Irish Adjective Lexicon – Part i**

```
+Adj+Com+NotSlen:^Adj          PL-A;
+Adj+Com+Slender:^Adj          PL-A-SLENDER;
+Adj+Gen+Weak:^Adj             PL-TADA;
+Adj+Gen+Strong:^Adj           PL-A;
+Adj+Voc:^Adj                  PL-A;

LEXICON Adj2-1
+Adj+Fem+Com:^Adj              Fem_com_voc_sg;
+Adj+Fem+Gen:^Adj              Fem_gen-D3_sg;
+Adj+Fem+Voc:^Adj              Fem_com_voc_sg;
+Adj+Masc+Com:^Adj             Masc_com_sg;
+Adj+Masc+Gen:^Adj             Masc_gen-D4_voc_sg;
+Adj+Masc+Voc:^Adj             Masc_gen-D4_voc_sg;

+Adj+Com+NotSlen:^Adj^Lea      PL-A;
+Adj+Com+Slender:^Adj^Lea      PL-A-SLENDER;
+Adj+Gen+Weak:^Adj             PL-TADA;
+Adj+Gen+Strong:^Adj^Lea       PL-A;
+Adj+Voc:^Adj^Lea              PL-A;

LEXICON Adj2-2
+Adj+Fem+Com:^Adj              Fem_com_voc_sg;
+Adj+Fem+Gen:^Adj              Fem_gen-D2A_sg;
+Adj+Fem+Voc:^Adj              Fem_com_voc_sg;
+Adj+Masc+Com:^Adj             Masc_com_sg;
+Adj+Masc+Gen:^Adj             Masc_gen-D4_voc_sg;
+Adj+Masc+Voc:^Adj             Masc_gen-D4_voc_sg;

+Adj+Com+NotSlen:^Adj          PL-E;
+Adj+Com+Slender:^Adj          PL-E-SLENDER;
+Adj+Gen+Weak:^Adj             PL-TADA;
+Adj+Gen+Strong:^Adj           PL-E;
+Adj+Voc:^Adj                  PL-E;

LEXICON Adj3-1
+Adj+Fem+Com:^Adj              Fem_com_voc_sg;
+Adj+Fem+Gen:^Adj              Fem_gen-D4_sg;
+Adj+Fem+Voc:^Adj              Fem_com_voc_sg;
+Adj+Masc+Com:^Adj             Masc_com_sg;
+Adj+Masc+Gen:^Adj             Masc_gen-D4_voc_sg;
+Adj+Masc+Voc:^Adj             Masc_gen-D4_voc_sg;

+Adj+Com+NotSlen:^Adj          PL-TADA;
+Adj+Com+Slender:^Adj          PL-SLENDER;
+Adj+Gen+Weak:^Adj             PL-TADA;
+Adj+Gen+Strong:^Adj           PL-TADA;
+Adj+Voc:^Adj                  PL-TADA;
```

Fig 42. Extract of Irish Adjective Lexicon – Part ii

```
LEXICON Fem_com_voc_sg
+Sg:^Sé           #;

LEXICON Fem_gen-D2A_sg
+Sg:^Caole        #;

LEXICON Fem_gen-D2B_sg
+Sg:^Ath          #;

LEXICON Fem_gen-D2C_sg
+Sg:^Aththí       #;

LEXICON Fem_gen-D2D_sg
+Sg:^Aththaí      #;

LEXICON Fem_gen-D3_sg
+Sg:^Leaa         #;

LEXICON Fem_gen-D4_sg
+Sg:0             #;

LEXICON Masc_com_sg
+Sg+LenYES:^Sé    #;
+Sg+LenNO:0       #;

LEXICON Masc_gen-D1_voc_sg
+Sg:^Caol^Sé      #;

LEXICON Masc_gen-D4_voc_sg
+Sg:^Sé           #;

LEXICON PL-A
+Pl:a             #;

LEXICON PL-A-SLENDER
+Pl:a^Sé          #;

LEXICON PL-E
+Pl:e             #;

LEXICON PL-E-SLENDER
+Pl:e^Sé          #;

LEXICON PL-SLENDER
+Pl:^Sé           #;

LEXICON PL-TADA
+Pl:0       #;
```

Fig 43. Extract of Irish Adjective Lexicon – Part iii

## *5.7 Replace Rules*

### Overview

The lexicon transducer and replace rule transducers are composed together. The order of composition is important. The output of the lexicon transducer becomes the input to the first replace rule transducer (See Fig. 24). The output of the first rule transducer becomes the input to the second and so on for the remaining transducers. The surface form of the initial lexicon undergoes several stages of refinement before the final surface form is produced. The end result (after composition) is a single bi-directional 2-level transducer without any intermediate levels, which maps a morphosyntactically specified lexical form to a fully inflected surface form and vice versa.

Finite-state transducers (networks) can be manipulated using various mathematical operations such as concatenation, projection and subtraction. Unwanted strings (rule overgenerations) can be removed from the network and new correct strings can be added (Beesley and Karttunen, 2001, p250-252). This is represented in the Transducer Manipulation component of Fig. 29.

Irish inflectional morphology relies heavily on stem changes (initial and final mutations) to indicate different grammatical functions. In the lexicon, stem changes are flagged by concatenating tags (triggers) to the surface level string specifiying the type of change required (c.f. Tables 22 & 23 cat and áit output from lexicon). In addition to the suffix *–anna*, (10) contains triggers for initial mutation (`^Sé^Urú^hv`) and broad/slender harmonization between the stem and suffix (`^LC`). It also contains a `^C` tag indicating that it is a common noun, which is used by initial mutation rules. Some inflections consist only of stem changes without any affixation. (11) contains tags for slenderisation (`^Caol`) and initial mutation (`^Sé^Urú^hv`) only.

(10) `áit+Noun+Fem+Com+Pl+Def:` **`áit^C^LCanna^Sé^Urú^hv`**

(11) `cat+Noun+Masc+Com+Pl+Def:` **`cat^C^Caol^Sé^Urú^hv`**

Replace rules take the concatenation of (surface level) morphemes output from the lexicon, and output a modified concatenation of morphemes as will be demonstrated in the following sections (Jurafsky and Martin, 2000, p77).

The rules are encoded as regular expressions. Each regular expression is compiled into a finite state transducer (using the *xfst* tool). Each rule transducer specifies only the constraints necessary for that rule, allowing all other strings to pass through unchanged (Jurafsky and Martin, 2000, p78).

Individual FSTs are composed together to produce one larger FST. In general, the FSTs must be composed in the order described in this implementation, as the output of one transducer becomes input to another.

The replace rule regular expressions (REs) are described in the following order, which follows the flow of control in Fig. 29.

- Define variables

- Initial Mutations (nouns, adjectives and verbs)
  - Lenition (RE 1-4)
  - Eclipsis (RE 5-7)
  - Prefix vowel-initial and s-initial stems (RE 8-13)

- Final Mutations: Verbs
  - Final Syllable Changes (RE 14-17)
  - Syncopation (RE 18-21)
  - Slenderising (RE 22-27)
  - Broadening (RE 28-32)

- Final Mutations: Nouns and Adjectives
  - Replace Diphthongs and Long Vowels (RE 33-36)
  - Syncopation (RE 37-38)
  - Segment Stems (RE 39)
  - Tidy up (RE 40-45)
  - Final Syllable Changes (RE 46-61)
  - Check Broad/Slender Tags (RE 62-63)
  - Slenderising (RE 64-73)
  - Broadening (RE 74-80)
  - Check Orthographic Vowel Harmony (RE 81-86)
  - Restore Diphthongs and Long-Vowels (RE 87-90)

- Post-Processing
  - Filter out Overgenerations
  - Apply Corrections (i.e. rule exceptions)

There are just under 100 replace rules in the current implementation. Sample strings from the lexicon are chosen in order to illustrate the function of each rule. The sample strings are shown before and after the application of the rule. When these morphosyntactic rules have been applied to the strings, the triggering tags are eliminated from the strings by simple replace rules, which map the triggers to the empty string, e.g. `%^Sé -> [].`

### Variable Definitions

The following variables are defined for later use in replace rules.

```
define Vowel a|e|i|o|u|á|é|í|ó|ú|%^AO|%^IA|%^AE|%^UA;
define Cons  b|c|d|f|g|h|j|k|l|m|n|p|q|r|s|t|v|w|x|y|z;
define Nountag %^F|%^M|%^C|%^G|%^V;
```

### Initial Mutations

Initial mutation rules are common to verbs, nouns and adjectives.
The following tags (replace rule triggers) relating to initial mutations have already been inserted into the appropriate lower level strings in the lexicon:

- `^Sé` – triggers lenition replace rules
- `^Urú` – triggers eclipsis replace rules
- `^tv` – triggers *t* before vowel replace rules
- `^hv` – triggers *h* before vowel replace rules
- `^ts` – triggers *t* before *s* replace rules

The replacement rule regular expression format is repeated here for convenience (Beesley and Karttunen, 2001, p122-127). Left and right contexts are optional.

```
String -> Replacement String || Left-context _ Right-context;
```

*Lenition (Séimhiú) Regular Expressions (^Sé)*

Lenition takes place in stems which have the `^Sé` tag, begin with a lenitable consonant (see Table 5) and contain a verb, noun, or adjective tag. Verbs and adjectives are identified by `^Verb` and `^Adj` tags respectively. Nouns are identified by gender and case tags. Lenition applies to a) common nouns (`^C`) both masculine (`^M`) and feminine (`^F`), b) masculine genitive (`^G`) nouns and c) vocative nouns (`^V`).

There are three replace rules which implement lenition and a fourth replace rule which removes the lenition trigger, `^Sé`, when it is no longer required. RE 1 is a general lenition rule and RE 2 is a rule specifically for `s`-initial stems, which has a more restricted right context. RE 3 applies to verbs only and it inserts `d'` before past, imperfect, and conditional inflections of vowel-initial or `f`-initial verb stems, which are already lenited. RE 4 eliminates the lenition trigger.

(12) shows surface strings before lenition replace rules (RE 1 to 4) are applied.

```
(12) alt+Noun+Masc+Com+Sg+Def:alt^M^C^Sé^tv^ts
     cathaoir+Noun+Fem+Com+Sg+Def:cathaoir^F^C^Sé^tv^ts
     saol+Noun+Masc+Com+Sg+Def:saol^M^C^Sé^tv^ts
```

**RE 1.** `[..] -> h || .#. [m|b|c|d|f|g|p|t] _ ?+`
`[[[%^F|%^M] %^C]|[%^M %^G]|[%^V]|[%^Adj]|[^Verb]] ?* %^Sé;`

**RE 2.** `[..] -> h || .#. s _ [Vowel|l|n|r] ?*`
`[[[%^F%|^M] %^C]|[%^M %^G]|[%^V]|[%^Adj]|[^Verb]] ?* %^Sé;`

**RE 3.** `[..] -> d ' || .#. _  [Vowel|f h] ?+ %^Verb ?* %^Sé`

**RE 4.** `%^Sé -> [];`

In RE 1 `[..] -> h` inserts a symbol 'h' after the first consonant of any stem starting with a lenitable consonant, followed by one or more symbols, followed by one of the listed tag combinations, followed by zero or more symbols (an affix), followed by the lenition tag `^Sé`. `[..]` limits the regular expression's power to one 'h' insertion. '`[ ] -> h`' could map an infinite number of empty strings to h's as there are an infinite number of empty strings, taking up no space, between the lenitable consonant and the next symbol (Beesley and Karttunen, 2001, p161). '`%`' is used to literalise symbols which have special functions: in this case '^' is literalised.

Example (13) shows the result after RE 1 is applied to (12). *c* is mapped to *ch* in *cathaoir* 'chair' but no replacement takes place in *alt* 'paragraph' or *saol* 'life'.

(13) alt+Noun+Masc+Com+Sg+Def:alt^M^C^Sé^tv^ts
cathaoir+Noun+Fem+Com+Sg+Def:**ch**athaoir^F^C^Sé^tv^ts
saol+Noun+Masc+Com+Sg+Def:saol^M^C^Sé^tv^ts

RE 3 maps *s* to *sh* in *saol* 'life' as shown below.

(14) alt+Noun+Masc+Com+Sg+Def:alt^M^C^Sé^tv^ts
cathaoir+Noun+Fem+Com+Sg+Def:**ch**athaoir^F^C^Sé^tv^ts
saol+Noun+Masc+Com+Sg+Def:**sh**aol^M^C^Sé^tv^ts

RE 4 removes the '^Sé' symbol by mapping it to the empty string.

(15) alt+Noun+Masc+Com+Sg+Def:alt^M^C^tv^ts
cathaoir+Noun+Fem+Com+Sg+Def:**ch**athaoir^F^C^tv^ts
saol+Noun+Masc+Com+Sg+Def:**sh**aol^M^C^tv^ts

*Eclipsis (^Urú)*

Eclipsis takes place in stems which have the ^Urú tag, begin with an eclipsable consonant (see Table 6) and contain a verb, noun or adjective tag. Eclipsis applies to genitive plurals of nouns, nouns following simple prepositions, and several verb tenses. In RE 5, the eclipsable consonants are mapped to their appropriate eclipsed pairs. As this set of replacements share the same left and right contexts they are listed, to the left of the double-pipe symbol '||' and separated by commas.

Vowels are eclipsed by prefixing '*n-*' to the vowel as in RE 6. RE 7 eliminates the '`^Urú`' tag from the surface string. RE 5 to 7 will be applied to genitive plural feminine stems *cathaoir* 'chair' and *adharc* 'horn' as in (16) and the result is given in (17).

(16) `cathaoir+Noun+Fem+Gen+Strong+Pl+Def:cathaoir^G^Coim^LCacha^Urú^hv`
`adharc+Noun+Fem+Gen+Weak+Pl+Def:adharc^G^Urú^hv`

**RE 5.**

```
b -> m b ,
c -> g c ,
d -> n d ,
f -> b h f ,
g -> n g ,
p -> b p ,
t -> d t || .#. _ ?+ [[%^G]| [[%^M | %^F] %^C]|%^Verb|%^Adj] ?* %^Urú
```

**RE 6.**

```
 [..] -> n %- || .#. _ Vowel ?*
[[%^G]| [[%^M | %^F] %^C]|[%^Verb]] ?* %^Urú
```

**RE 7.** `%^Urú -> []`

(17) cathaoir+Noun+Fem+Gen+Strong+Pl+Def:**gc**athaoir^G^Coim^Lcacha^hv
adharc+Noun+Fem+Gen+Weak+Pl+Def:**n-a**dharc^G^hv

*Prefixing of vowel-initial and s-initial stems (^tv, ^hv, ^ts)*

In RE 8, the '`^tv`' tag triggers prefixing of '*t-*' to vowel-initial common case masculine definite nouns (a definite markup tag is not required at this point as '`^tv`' is only inserted into the appropriate noun strings in the lexicon).

(18) `alt+Noun+Masc+Com+Sg+Def:alt^M^C^tv^ts`
`cathaoir+Noun+Fem+Com+Sg+Def:chathaoir^F^C^tv^ts`
`saol+Noun+Masc+Com+Sg+Def:shaol^M^C^tv^ts`

**RE 8.** `[..] -> t %- || .#. _ Vowel ?* %^M %^C ?* %^tv`

**RE 9.** `%^tv -> [];`

(19) shows the result of applying RE 8 and RE 9 to the strings in (18). RE 8 prefixes '*t-*' to *alt*, and RE 9 eliminates the '`^tv`' tag from all three strings.

(19) alt+Noun+Masc+Com+Sg+Def:**t-**alt^M^C^ts

cathaoir+Noun+Fem+Com+Sg+Def:chathaoir^F^C^ts

saol+Noun+Masc+Com+Sg+Def:shaol^M^C^ts

*h* is prefixed to a vowel-initial noun following the article *na* 'the', which is used before common case plural nouns and genitive singular feminine nouns. In the example below (20), the '^hv' tag triggers the prefixing of 'h' to a vowel-initial feminine noun, *adharc* 'horn'.

(20) adharc+Noun+Fem+Gen+Sg+Def:adharc^F^G^Caole^hv^ts

**RE 10.** [..] -> h || .#. _ Vowel ?*
[[%^F %^G] | [%^C]|[%^Verb]] ?* %^hv

**RE 11.** %^hv -> [];

In (21) '*h*' is prefixed to the vowel and the '^hv' tag is eliminated.

(21) adharc+Noun+Fem+Gen+Sg+Def:**h**adharc^F^G^Caole^ts

RE 12 in the context of the '^ts' tag prefixes '*t*' to s-initial feminine, common case nouns and masculine, genitive case nouns. This rule creates no mappings in the sample input strings as the only s-initial string *saol* 'life' is common case and masculine. RE 13, however, eliminates the ^ts marker from the surface strings as shown in (23) .

(22)alt+Noun+Masc+Com+Sg+Def:**t-**alt^M^C

cathaoir+Noun+Fem+Com+Sg+Def:chathaoir^F^C

saol+Noun+Masc+Com+Sg+Def:shaol^M^C

**RE 12.** [..] -> t || .#. _ s [Vowel|l|n|r|h] ?*
[[%^F %^C]|[%^M %^G]] ?* %^ts

**RE 13.** %^ts -> [];

(23) alt+Noun+Masc+Com+Sg+Def:**t-**alt^M^C

cathaoir+Noun+Fem+Com+Sg+Def:chathaoir^F^C

saol+Noun+Masc+Com+Sg+Def:shaol^M^C

(24) shows two strings at the end of the initial mutation rules. The first string requires no further processing. The second string still contains some triggers: '^Coim' (*coimriú*) for syncopation and '^LC' (*leathan/caol*) for broad or slender harmony checking. These will be dealt with in the next section on final mutations.

(24) adharc+Noun+Fem+Gen+Weak+Pl+Def:n-adharc
cathaoir+Noun+Fem+Gen+Strong+Pl+Def:gcathaoir**^Coim^LC**acha

There are two sets of replace rules for final mutations, one for verbs and one for nouns and adjectives, as the replacements and contexts are different. Final mutations for verbs are relatively straightforward and are dealt with first, followed by the more complex nominal final mutations.

**Final Mutations - Verbs**

The replace rules which deal with final mutations in verbs use the following tags:

- `^igh/^aigh` – Final syllable deletion
- `^Coim` – (*coimriú*) syncopation
- `^Caol` – (*caolú*) slenderising
- `^Lea/^LC` – (*leathnú*) broadening

*Final Syllable Deletion (^igh, ^aigh).*

All 2nd conjugation verbs ending in *-(a)igh* lose the last syllable before verbal suffixes are appended, e.g. *bailigh* 'gather'. Some 1st conjugation verbs (monosyllabic roots), such as *brúigh* 'press' have two syllables phonetically and so the ending *-igh* is also removed from these stems before verbal suffixes are attached.

```
(25) brúigh+Verb+PresInd+1P+Sg:brúigh^igh^Verbim
     bailigh+Verb+PresInd+1P+Sg:bailigh^Caol^igh^Verbaím
     bailigh+Verb+Imper+2P+Sg:bailigh^Caol^igh^Verb^Fr
```

**RE 14.** `i g h -> [] || _ (%^Caol) %^igh %^Verb [[\%^Fr]|.#.]`

**RE 15.** `a i g h -> [] || _ %^aigh %^Verb [\%^Fr]`

**RE 16.** `%^igh -> []`

**RE 17.** `%^aigh -> []`

```
(26) brúigh+Verb+PresInd+1P+Sg:brú^Verbim
     bailigh+Verb+PresInd+1P+Sg:bail^Caol^Verbaím
     bailigh+Verb+Imper+2P+Sg:bailigh^Caol^Verb^Fr
```

There are some inflections which do not add any suffix (past indicative and imperative). Therefore these strings should not lose their ending. Such strings are lexically flagged with the `^Fr` (*fréamh* 'root') mark-up tag to signify that they keep the root form. RE 14 includes the context `[\%^Fr]`, which reads as "`not ^Fr`", meaning eliminate *igh* where the `^Verb` tag is not followed by `^Fr`.

*Syncopation (^Coim)*

Syncopation is implemented by mapping short vowels in the final syllable to the empty string when followed by one or two symbols (consonants) and the syncopation tag '`^Coim`'.
Second conjugation verbs, i.e. polysyllabic stems, which end in (a)il, (a)in, (a)ir, and (a)is are syncopated (with some exceptions) in all tenses except the past indicative and the imperative, second person singular (the root form of verbs). Those tenses which do not involve syncopation are marked with the `^Fr` (*fréamh*, 'root') tag in the lexicon. (27) shows the verb *freagair* 'answer' before RE 18 and RE 19.

(27)`freagair+Verb+FutInd:freagair^Coim^Verbóidh`
`freagair+Verb+PastInd:d'fhreagair^Coim^Verb^Fr`

**RE 18.** `a i -> [] || _ [l|n|r|s] %^Coim %^Verb [\%^Fr]`

**RE 19.** `i   -> [] || _ [l|n|r|s] %^Caol %^Coim %^Verb [\%^Fr]`

In (28) the future indicative string is syncopated whereas the past indicative is not, as RE 18 and RE 19 apply only in the context of strings which do not have the `^Fr` tag.

(28)`freagair+Verb+FutInd:freagr^Coim^Verbóidh`
`freagair+Verb+PastInd:d'fhreagair^Coim^Verb^Fr`

**RE 20.** `%^Fr -> []`

**RE 21.** `%^Coim -> []`

*Slenderising (^Caol)*

For all verb suffixes the broad allomorph is appended as standard in the lexicon. These suffixes will be slenderised where necessary, i.e. in the context of verb stems with the `^Caol` mark-up tag. Where the relevant verbal suffixes start with either *f* or *t*, these characters are used to locate the start of the suffix (c.f. Table 11).

In the following example, broad suffixes are slenderised to match the slender 1st conjugation stem *bain* 'extract' or 'take', by removing the broad vowel *a* when it is followed by a slender vowel.

(29) `bain+Verb+FutInd+Neg:bhain^Caol^Verbf`**`ai`**`dh`
`bain+Verb+ImpInd+Auto:bhain^Caol^Verbt`**`aí`**

**RE 22.** `a -> [] || %^Caol %^Verb [f|t]* _ i Cons+`

**RE 23.** `a -> [] || %^Caol %^Verb [f|t]* _ í Cons* .#.`

(30) bain+Verb+FutInd+Neg:bhain^Caol^Verbf**i**dh

bain+Verb+ImpInd+Auto:bhain^Caol^Verbt**í**

In the following case, the suffixes are slenderised through the insertion of the slender vowel *e* before a broad vowel.

(31) bain+Verb+ImpInd+2P+Sg:bhain^Caol^Verbt**á**

bain+Verb+PastInd+1P+Pl:bhain^Caol^Verb**a**mar

**RE 24.** [..] -> e || %^Caol %^Verb [f|t]* _ á .#.

**RE 25.** [..] -> e || %^Caol %^Verb [f|t]* _ a Cons+

(32) bain+Verb+ImpInd+2P+Sg:bhain^Caol^Verbt**eá**

bain+Verb+PastInd+1P+Pl:bhain^Caol^Verb**ea**mar

In (33) the suffix attaching to 2nd conjugation stems, such as *aithin* 'recognise', is slenderised by replacing *ó* with *eo*.

(33) aithin+Verb+Cond+1P+Pl:d'aithn^Caol^Verb**ó**imis

**RE 26.** ó -> e o || %^Caol %^Verb [f|t]* _ ?+

(34) aithin+Verb+Cond+1P+Pl:d'aithn^Caol^Verb**eo**imis

**RE 27.** %^Caol -> []

*Broadening (^Lea, ^LC)*

Next the ^Lea and ^LC mark-up tags are processed.

Broadening of slender words usually involves removing '*i*' but can also require substitution of vowels, e.g. *io* becomes *ea*, or the insertion of a broad vowel, e.g. *í* becomes *ío*. There are some slender verb stems which are broadened when verbal suffixes are added. These are marked with the tag ^Lea (*leathnú* 'broaden'). There are also some slender verb stems which are broadened except when the verbal suffix starts with *t*. These are marked with the tag ^LC (*leathan/caol* 'broad/slender').

In RE 28, the ^LC tag is eliminated in strings where there is either no suffix, or the suffix starts with *t*. Where the ^LC tag remains, the stem is broadened in RE 29. (35) and (36) show the stems *sábháil* 'save' and *taispeáin* 'show' before and after RE 28.

(35) sábháil+Verb+ImpInd+1P+Sg:shábháil**^LC**^Verbainn

sábháil+Verb+ImpInd+2P+Sg:shábháil**^LC**^Verbtá

taispeáin+Verb+Cond+1P+Sg:thaispeáin^Lea^Verbfainn

**RE 28.** `%^LC -> [] || _ %^Verb [t|.#.]`

(36) sábháil+Verb+ImpInd+1P+Sg:shábháil^LC^Verbainn
sábháil+Verb+ImpInd+2P+Sg:shábháil^Verbtá
taispeáin+Verb+Cond+1P+Sg:thaispeáin^Lea^Verbfainn

In (37) stems containing ^Lea or ^LC are broadened through the removal of the final *i*.

**RE 29.** `i -> [] || ?+ _ Cons [%^LC|%^Lea] %^Verb`

(37) sábháil+Verb+ImpInd+1P+Sg:shábhál^LC^Verbainn
sábháil+Verb+ImpInd+2P+Sg:shábháil^Verbtá
taispeáin+Verb+Cond+1P+Sg:thaispeán^Lea^Verbfainn

As there are no more verb replace rules the remaining mark-up tags are removed.

**RE 30.** `%^LC -> []`

**RE 31.** `%^Lea -> []`

**RE 32.** `%^Verb -> []`

**Final Mutations – Nouns and Adjectives**

The replace rules which deal with final mutations in nouns and adjectives use the following tags:

- ^Coim – (*coimriú*) syncopation
- ^Ath (*athrú*) – Final syllable deletion
- ^Caol – (*caolú*) slenderising
- ^Lea – (*leathnú*) broadening
- ^LC – (*leathan/caol*) broad/slender harmony.

Before applying replace rules for initial and final mutation of stems, some processing of diphthongs and long vowels is carried out.

*Replace Diphthongs and Long-Vowels (^IA, ^UA, ^AE, ^AO)*

Final mutation replace rules require some pre-processing of diphthongs and certain vowel combinations in order to allow treatment of these vowel combinations as single units. The diphthongs *'ua'* and *'ia'* are replaced by the multi-character symbols '^UA' and '^IA' respectively. For example, the genitive case of *cluas* 'ear' is *cluaise* 'of the ear', which is formed by suffixing '*-e*' and slenderising the stem. Slenderisation usually requires inserting an '*i*' after the broad vowels. In *cluas*, the diphthong *'ua'* must be slenderised as *'uai'* and not each vowel individually as in *'uiai'*, to give the correct genitive form *cluaise* 'of the ear'.

Some stems are syncopated, i.e. an un-stressed final syllable loses its vowel(s) when a suffix is added. Certain short vowel combinations are stressed like long vowels, e.g. the '*ao*' in *cathaoir* sounds like /i:/ and '*ae*' (sometimes) sounds like /e:/ as in *Gael* 'Irish person'. These long vowels are replaced by `^AO` and `^AE` respectively, which blocks syncopation.

In (38) and (39) the plural forms of *cathair* 'city' and *cathaoir* 'chair' demonstrate the effect of the long vowel combination in the plural. The plural form of *cathair* is *cathr+acha* 'cities' - the *'ai'* is deleted (syncopated) when the suffix *+acha* is added. The long vowels *ao* /i:/ resist syncopation, e.g. *cath**ao**ir* becomes *cath**aoi**r+e+acha* 'chairs' - the *'aoi'* is not deleted when the suffixes *+e+acha* are attached. (The *+e* is required to slenderise the suffix as the stem ends in a slender consonant, i.e. the final consonant *'r'* is preceded by a slender vowel *'i'*).

(38) *cath**ai**r* 'city'
*cathracha 'cities'*

(39) *cath**aoi**r* 'chair'
*cath**aoi**reacha* 'chairs'

RE 33 to 36 replace the specified vowel combinations with a multi-character symbol. (`^IA` is implicitly a multi-character symbol in *xfst* syntax when the symbols are concatenated together – it does not have to be explicitly defined as in *lexc* syntax.)

(40) `cathaoir+Noun+Fem+Gen+Strong+Pl+Def:gcathaoir^Coim^LCacha`

**RE 33.** `i a -> %^IA;`

**RE 34.** `u a -> %^UA;`

**RE 35.** `a e -> %^AE;`

**RE 36.** `a o -> %^AO;`

(41) `cathaoir+Noun+Fem+Gen+Strong+Pl+Def:gcath`**`^AO`**`ir^Coim^LCacha`

*Syncopation (^Coim)*

In certain polysyllabic stems the vowels in the final syllable are removed when a suffix is added. The expression `?^<3` denotes less than three symbols and `Nountag` is declared at the start as a shorthand for the various noun tags, i.e `^M`, `^F`, `^C`, `^G` and `^V`.

(42) cathaoir+Noun+Fem+Gen+Strong+Pl+Def:gcath^AO**i**r^Coim^LCacha
cathair+Noun+Fem+Gen+Strong+Pl+Def:gcath**ai**r^Coim^LCacha
daingean+Adj+Voc+Pl:daing**ea**n^Adj^Coim^Caole

**RE 37.** `[a|e|i|o|u] -> [] || Cons (a|e) _ ?^<3 [Nountag+ | %^Adj] %^Coim`

(43) is the result of applying RE 37 to example (42) above. The `^AO` tag in the first string blocks the dropping of any vowels, whereas in the second string the *a* and *i* have been eliminated (syncopated). The context `(a|e)` is required for cases where more than one vowel is removed, e.g. nouns ***cathaoir*** 'chair', ***cathair*** 'city', or adjective ***daingean*** 'fortified'.

(43) cathaoir+Noun+Fem+Gen+Strong+Pl+Def:gcath^AO**i**r^Coim^LCacha
cathair+Noun+Fem+Gen+Strong+Pl+Def:gcathr^Coim^LCacha
daingean+Adj+Voc+Pl:daingn^Adj^Coim^Caole

The `^Coim` marker is now mapped to the empty string.

**RE 38.** `%^Coim -> [];`

*Stem Segmentation (^X)*

An `^X` tag is inserted before the first vowel of the final syllable of every stem in the noun and adjective lexicons as shown in the following regular expression. The `^X` tag is used as a left context, and the mark-up tag acts as a right context thus restricting the application of final mutations to the final syllable of the string only.

**RE 39.** `Vowel+ @-> %^X... || _ Cons* [Nountag |%^Adj];`

RE 39 identifies the final syllable of every noun and adjective stem in order to confine final mutation replacements to the final syllable. (It is not necessary to mark the final syllable of verbs as verbal inflection, in general, tends to use allomorphic suffixes and limited final mutation ).

Vowels, in the specified contexts, are prefixed by an '`^X`' tag. '`Vowel+`' means a string of one or more vowels and the '`@->`' operator matches the longest string of vowels. '`%^X...`' inserts '`^X`' before the matched string of one or more vowels, which is denoted by '`...`'. The context in this instance states that the vowel(s) must be followed by zero or more consonants, and one of the tags listed in the square brackets. As one of these tags always follows a stem (in a well-formed string) they are used to anchor the search to the last syllable of the stem.

In 44 stem segmentation is illustrated using two noun stems: *alt* 'paragraph' and *cathaoir* 'chair', and one adjective stem *maith* 'good'.

```
(44) alt+Noun+Masc+Com+Sg+Idf:alt^M^C
     cathaoir+Noun+Fem+Com+Sg+Idf:cath^AOir^F^C
     maith+Adj+Pos:maith^Adj
```

The effect of applying RE 39 to the above examples is shown in (45).

```
(45) alt+Noun+Masc+Com+Sg+Idf:^Xalt^M^C
     cathaoir+Noun+Fem+Com+Sg+Idf:cath^Xaoir^F^C
     maith+Adj+Pos:m^Xaith^Adj
```

The final syllable is identified by anchoring the right context to the end of the string by specifying tags which have been inserted in the lexicon. Provided the output of the lexicon transducer is validated against a (lower level) tag grammar (see Fig. 50), we can be confident that this regular expression will always correctly identify the final segment of the string.

After segmentation, the gender tags (`^F, ^M`) and case tags (`^C, ^G, ^V`) are also removed, as they are not required by subsequent rules.

**RE 40.** `%^M -> [];`

**RE 41.** `%^F -> [];`

**RE 42.** `%^C -> [];`

**RE 43.** `%^G -> [];`

**RE 44.** `%^V -> [];`

**RE 45.** `%^Adj -> [];`

*Final Syllable Replacements (^Ath)*

The following replace rules use the `^Ath` (*athrú* 'change') mark-up tag and are a mixed bag of changes mainly concerning the final stem consonants at the juncture of stem and suffix morphemes. The more regular changes to vowels are described under slenderising and broadening in the sections following on from this.

There are a number of ways in which many of these replacements could be implemented. In most cases the particular choice taken was based on what was most effective in the terms of the overall inflectional patterns of the stems in question and in other cases it was simply a matter of personal choice.

The change from *'eadh'* to *'í'*, e.g. *geimhreadh* 'winter' to *geimhrí* 'winters', in the plural, is unusual in that it takes place in two stages. The ordering of the rules is important as one feeds into the other.

(46) soitheach+Noun+Masc+Com+Pl+Def:soith^X**each**^Ath  
geimhreadh+Noun+Masc+Com+Pl+Def:geimhr^X**eadh**^Ath

**RE 46.** c h -> í ,  
d h -> í || %^X (e) a _ %^Ath

(47) soitheach+Noun+Masc+Com+Pl+Def:soith^X**eaí**^Ath  
geimhreadh+Noun+Masc+Com+Pl+Def:geimhr^X**eaí**^Ath

**RE 47.** e a í -> í || _ %^Ath;

(48) soitheach+Noun+Masc+Com+Pl+Def:soith^X**í**^Ath  
geimhreadh+Noun+Masc+Com+Pl+Def:geimhr^X**í**^Ath

The following rules using the ^Ath take place in one stage and are illustrated, as before, through the use of sample strings from the lexicon.

RE 48 when applied, for example, to *finné* 'witness' changes the *é* to *éi* before the suffix *-the*. In order to keep the lexicon simple (i.e. a single class for *–the* suffixes only, rather than including a special class for *–ithe* suffixes also) the inflected form is generated in the lexicon as *finnéthe* and corrected here by means of a replace rule which in effect inserts the *i* required by the correct spelling *finnéithe* 'witnesses'.

(49) finné+Noun+Masc+Com+Pl+Idf:finn^X**é**^Aththe

**RE 48.** é -> é i || %^X _ %^Ath;

(50) finné+Noun+Masc+Com+Pl+Idf:finn^X**éi**^Aththe

RE 49 deals with a similar phenomenon where the final *í*, e.g. in *ainmhí* 'animal' changes to *i* with the addition of the plural suffix *–the*, resulting in *ainmhithe* 'animals'.

(51) ainmhí+Noun+Masc+Gen+Strong+Pl+Idf:ainmh^X**í**^Aththe

**RE 49.** í -> i || _ %^Ath;

(52) ainmhí+Noun+Masc+Gen+Strong+Pl+Idf:ainmh^X**i**^Aththe

RE 50 replaces the final *ú* with *ui*, as in *tarlú* 'happening', (53), when the plural suffix *–the* is added.

(53) tarlú+Noun+Masc+Com+Pl+Def:tarl^X**ú**^Aththe

**RE 50.** ú -> u i || _ %^Ath t

(54) tarlú+Noun+Masc+Com+Pl+Def:tarl^X**ui**^Aththe

RE 51 demonstrates another non-standard final formation. In example (55), the final *ú* is removed when the *–aithe* suffix is appended to the genitive singular form of the noun.

(55) tarlú+Noun+Masc+Gen+Sg+Def:tharl^X**ú**^Athaithe

**RE 51.** ú -> [] || _ %^Ath a

(56)tarlú+Noun+Masc+Gen+Sg+Def:tharl^X^Athaithe

In RE 52, a similar change occurs; where the final *aí* is removed, e.g. in *conaí* 'home', when the *–aithe* suffix is appended to the genitive singular form.

(57)cónaí+Noun+Masc+Gen+Sg+Def:chón^X**aí**^Athaithe

**RE 52.** a í -> [] || _ %^Ath a

(58)cónaí+Noun+Masc+Gen+Sg+Def:chón^X^Athaithe

RE 53 demonstrates the replacement of the final two consonants, *bh,* in *leanbh* 'child' with *aí*, in the plural form. This could also be described as removing the final two consonants and adding the suffix *aí*.

(59)leanbh+Noun+Masc+Com+Pl+Idf:l^Xean**bh**^Ath

**RE 53.** b h -> a í || _ %^Ath

(60)leanbh+Noun+Masc+Com+Pl+Idf:l^Xean**aí**^Ath

In RE 54, two final consonants, *dh*, are also replaced, this time by *t*, in the noun *iarraidh* 'attempt', when the plural suffix *–aí* is added. As *–aí* is a broad suffix and the stem is slender (*iarrait-*), the ^Lea (*leathnú* 'broadening') tag has been inserted in the lexicon. Processing of this tag is be described in a later section.

(61)iarraidh+Noun+Fem+Com+Pl+Def:hiarr^Xai**dh**^Ath^Leaaí

**RE 54.** d h -> t || i _ %^Ath %^Lea

(62)iarraidh+Noun+Fem+Com+Pl+Def:hiarr^Xai**t**^Ath^Leaaí

In (64), the final *t* is dropped when the plural suffix *–a* is added, e.g. *tiomáint* 'driving' in the genitive case.

(63) tiomáint+Noun+Fem+Gen+Sg+Idf:tiom^Xái**nt**^Ath^Leaa

**RE 55.** t -> [] || n _ %^Ath;

(64) tiomáint+Noun+Fem+Gen+Sg+Idf:tiom^Xái**n**^Ath^Leaa

In the following example, *tagairt* 'reference' becomes *tagartha* 'of the reference' in the genitive singular, i.e. *h* is inserted after the final *t*.

(65) tagairt+Noun+Fem+Gen+Sg+Idf:tag^Xai**rt**^Ath^Leaa

**RE 56.** [..] -> h || r t _ %^Ath;

(66) tagairt+Noun+Fem+Gen+Sg+Idf:tag^Xai**rth**^Ath^Leaa

After RE 57, the final *e* is removed, e.g. *buille* 'blow' and *míle* 'mile' or 'thousand' when a plural suffix is appended. (In effect the final *e* is replaced by a plural suffix in this type of noun).

(67) buille+Noun+Masc+Com+Pl+Idf:buill^X**e**^Athi
míle+Noun+Masc+Com+Pl+Idf:míl^X**e**^Athte

**RE 57.** e -> [ ] || %^X _ %^Ath

(68) buille+Noun+Masc+Com+Pl+Idf:buill^X^Athí
míle+Noun+Masc+Com+Pl+Idf:míl^X^Athte

The next rule, although a vowel change, is included here since it does not fall into the category broadening or slenderising, i.e. *io* becomes *ea*. This change, which is quite irregular, is restricted by a right context which only allows this replacement to take place when the *io* is followed by *c* or *s* (these were the only examples found during implementation). The examples *crios* 'belt' and *sioc* 'frost' are given below.

(69)crios+Noun+Masc+Gen+Sg+Idf:cr^X**io**s^Ath^Leaa
sioc+Noun+Masc+Gen+Sg+Def:ts^X**io**c^Ath^Leaa

**RE 58.** i o -> e a || _ [c|s] %^Ath

(70)crios+Noun+Masc+Gen+Sg+Idf:cr^X**ea**s^Ath^Leaa
sioc+Noun+Masc+Gen+Sg+Def:ts^X**ea**c^Ath^Leaa

The following are changes which apply primarily to adjectives. The final syllable *-íoch* of *buíoch* 'thankful' becomes *-íthí*, and *[(e)á,ó,eo,ua]ch* becomes *thaí* in various other adjectives, e.g. *gnách* 'usual' becomes *gnáthaí* when qualifying a feminine noun in the genitive singular.

(71) buíoch+Adj+Fem+Gen+Sg:b^Xuí**och**^Aththí
gnách+Adj+Fem+Gen+Sg:gn^Xá**ch**^Aththaí

**RE 59.** o c h -> [] || %^X (?) í _ %^Ath;

**RE 60.** c h -> [] || %^X (e) [á|ó|o|%^UA] _ %^Ath;

(72) buíoch+Adj+Fem+Gen+Sg:b^Xuí^Aththí
gnách+Adj+Fem+Gen+Sg:gn^Xá^Aththaí

The ^Ath tag is removed and we move on to vowel changes in the next section.

**RE 61.** %^Ath -> []

*Broad and Slender Tag Checking (^Caol, ^Lea)*

In the lexicon, all noun stems which should be slender (e.g. masculine noun, genitive case, first declension) are marked with a ^Caol (slender) tag regardless of whether they are already slender or not. The string is then checked to see if the stem is already slender (if the last vowel is an *i* or *í*) and if so the ^Caol tag is removed.

**RE 62.** %^Caol -> [] || [i|í] Cons* _ ;

The check on the ^Lea (*leathan* 'broad') tag operates in a similar manner by checking that the last vowel is a vowel other than *i* or *í*.

**RE 63.** %^Lea -> [] || [a|o|u|á|ó|ú|%^UA|%^AO|%^IA] Cons* _ ;

*Slenderising (^Caol)*

There are several ways in which slenderising can occur in nouns. Sometimes similar looking words are inflected in different ways. These stems are assigned to different classes in the lexicon and mark-up triggers are used to provide the essential differences in context needed to differentiate between them.
The ^Caol (*caolú* 'slenderising') tag is processed by making changes to the vowels based on the local context in which they are found. The following regular expressions dealing with the various ways of slenderising nouns are illustrated by means of examples from the lexicon.

In the following example, *marcach* 'rider' and *misneach* 'courage' are slenderised. The ending *-ach/-each* becomes *–aigh/-igh*, and as the final consonants also change, this is carried out in two stages. Firstly, the final consonants *ch* are replaced by *gh* and later the actual slenderising takes place under the *i*-insertion rule (RE 68).

(73) marcach+Noun+Masc+Com+Pl+Idf:marc^Xa**ch**^Caol

misneach+Noun+Masc+Gen+Sg+Idf:misn^Xea**ch**^Caol

**RE 64.** c h -> g h || [a|ú] _ %^Caol;

(74) marcach+Noun+Masc+Com+Pl+Idf:marc^Xa**gh**^Caol

misneach+Noun+Masc+Gen+Sg+Idf:misn^Xea**gh**^Caol

Nouns such as *nead* 'nest', (and *beach* 'bee', *sceach* 'bush') are slenderised by replacing *ea* with *ei* in the context of an *–e* suffix. Adjectives such as *daingean* 'tight' are not mutated in this way and so a further constraint has been added; the stem must end in either *ch* or *n*.

(75)nead+Noun+Fem+Gen+Sg+Idf:n^X**ea**d^Caole

daingean+Adj+Masc+Gen+Sg:dhaing^X**ea**n^Adj^Caol

**RE 65.** a -> i || %^X e _ [[c h] | d ] %^Caol e

(76)nead+Noun+Fem+Gen+Sg+Idf:n^X**ei**d^Caole

daingean+Adj+Masc+Gen+Sg:dhaing**ea**n^Caol

In the following types of nouns and adjectives, *ea* is slenderised as *i*. The examples given below are *fear* 'man', *misneach* 'courage' and *beag* 'small'.

(77)fear+Noun+Masc+Gen+Sg+Idf:f^X**ea**r^Caol

misneach+Noun+Masc+Gen+Sg+Idf:misn^X**ea**gh^Caol

beag+Adj+Masc+Gen+Sg:bh^X**ea**g^Caol

**RE 66.** e a -> i || %^X _ Cons+ %^Caol

(78)fear+Noun+Masc+Gen+Sg+Idf:f^X**i**r^Caol

misneach+Noun+Masc+Gen+Sg+Idf:misn^Xigh^Caol

beag+Adj+Masc+Gen+Sg:bh**i**g^Caol

The following noun and adjective are slenderised, *éan* 'bird' and *séipéal* 'church', through replacing *éa* with *éi*.

(79)éan+Noun+Masc+Gen+Sg+Idf:^X**éa**n^Caol
séipéal+Noun+Masc+Gen+Sg+Idf:séip^X**éa**l^Caol

**RE 67.** a -> i || %^X é _ Cons+ %^Caol;

(80) éan+Noun+Masc+Gen+Sg+Idf:^X**éi**n^Caol
séipéal+Noun+Masc+Gen+Sg+Idf:séip^X**éi**l^Caol

The most usual way of slenderising is to insert *i* after the vowels in the final syllable. The following nouns are all slenderised in this way; *seol* ‘sail’, *fuinneog* ‘window’, *saol* ‘life’, *naomh* ‘saint’, *leabhar* ‘book’, *ard* ‘high’, *bád* ‘boat’, *bán* ‘white’, *cnoc* ‘hill’, *toll* ‘hollow’, *glór* ‘voice’, *mór* ‘big’, *bun* ‘base’, *rún* ‘secret’ *casúr,* ‘hammer’ and *cluas* ‘ear’.

(81)seol+Noun+Masc+Gen+Sg+Idf:s^X**eo**l^Caol
fuinneog+Noun+Fem+Gen+Sg+Idf:fuinn^X**eo**g^Caole
saol+Noun+Masc+Gen+Sg+Idf:s^X^AOl^Caol
naomh+Noun+Masc+Gen+Sg+Idf:n^X**^AO**mh^Caol
leabhar+Noun+Masc+Gen+Sg+Idf:leabh^X**a**r^Caol
ard+Adj+Masc+Gen+Sg:^X**a**rd^Caol
bád+Noun+Masc+Gen+Sg+Idf:b^Xád^Caol
bán+Adj+Masc+Gen+Sg:bh^X**á**n^Caol
cnoc+Noun+Masc+Gen+Sg+Idf:cn^X**o**c^Caol
toll+Adj+Fem+Gen+Sg:t^X**o**ll^Caole
bocht+Adj+Masc+Gen+Sg:bh^X**o**cht^Caol
glór+Noun+Masc+Gen+Sg+Idf:gl^X**ó**r^Caol
mór+Adj+Masc+Gen+Sg:mh^X**ó**r^Caol
bun+Noun+Masc+Gen+Sg+Idf:b^X**u**n^Caol
rún+Noun+Masc+Gen+Sg+Idf:r^X**ú**n^Caol
casúr+Noun+Masc+Gen+Sg+Idf:cas^X**ú**r^Caol
cluas+Noun+Masc+Gen+Sg+Idf:cl^X**^UA**s^Caol

**RE 68.** [..] -> i || %^X (e) [%^AO|a|á|o|ó|u|ú|%^UA] _ Cons+ %^Caol

(82)seol+Noun+Masc+Gen+Sg+Idf:s^X**eoi**l^Caol
fuinneog+Noun+Fem+Gen+Sg+Idf:fuinn^X**eoi**g^Caole
saol+Noun+Masc+Gen+Sg+Idf:s^X**^AOi**l^Caol
naomh+Noun+Masc+Gen+Sg+Idf:n^X**^AOi**mh^Caol
leabhar+Noun+Masc+Gen+Sg+Idf:leabh^X**ai**r^Caol
ard+Adj+Masc+Gen+Sg:^X**ai**rd^Caol
bád+Noun+Masc+Gen+Sg+Idf:b^X**ái**d^Caol
bán+Adj+Masc+Gen+Sg:bh^X**ái**n^Caol

cnoc+Noun+Masc+Gen+Sg+Idf:cn^X**oi**c^Caol
toll+Adj+Fem+Gen+Sg:t^X**oi**ll^Caole
glór+Noun+Masc+Gen+Sg+Idf:gl^X**ói**r^Caol
bun+Noun+Masc+Gen+Sg+Idf:b^X**ui**n^Caol
rún+Noun+Masc+Gen+Sg+Idf:r^X**úi**n^Caol
casúr+Noun+Masc+Gen+Sg+Idf:cas^X**úi**r^Caol
cluas+Noun+Masc+Gen+Sg+Idf:cl^X**^UAi**s^Caol

The following nouns, *síol* 'seed', *suíomh* 'position' and *fionn* 'fair haired person' are slenderised by deleting the *o* in *ío* and *io*.

(83)síol+Noun+Masc+Gen+Sg+Idf:s^X**ío**l^Caol
suíomh+Noun+Masc+Gen+Sg+Idf:s^Xu**ío**mh^Caol
fionn+Noun+Masc+Gen+Sg+Idf:f^X**io**nn^Caol

**RE 69.** o -> [] || %^X (u) [i|í] _ Cons+ %^Caol

(84)síol+Noun+Masc+Gen+Sg+Idf:s^X**í**l^Caol
suíomh+Noun+Masc+Gen+Sg+Idf:s^Xu**í**mh^Caol
fionn+Noun+Masc+Gen+Sg+Idf:f^X**i**nn^Caol

In the following nouns, *grian* 'sun' and *sliabh* 'mountain', *ia* is replaced by *éi*.

(85)grían+Noun+Fem+Gen+Sg+Idf:gr^X**ia**n^Caole
sliabh+Noun+Masc+Gen+Sg+Idf:sl^X**ia**bh^Caole

**RE 70.** i a -> é i || %^X _ Cons+ %^Caol (t) e;

(86)grian+Noun+Fem+Gen+Sg+Idf:gr^X**éi**n^Caole
sliabh+Noun+Masc+Gen+Sg+Idf:sl^X**éi**bh^Caole

The noun *rian* 'track', is slenderised by inserting *i* after the vowels *ia*.

(87)rian+Noun+Masc+Gen+Sg+Idf:r^X**ia**n^Caol

**RE 71.** [..] -> i || %^X i a _ Cons+ %^Caol .#.

(88)rian+Noun+Masc+Gen+Sg+Idf:r^X**iai**n^Caol

Alternatively the noun *scian* 'knife', is slenderised by deleting the *a* in *ia*.

(89) scian+Noun+Fem+Gen+Strong+Pl+Idf:sc^X**ia**n^Caole

**RE 72.** `a -> [] || %^X i _ Cons+ %^Caol e;`

(90)scian+Noun+Fem+Gen+Strong+Pl+Idf:sc^X**i**n^Caole

Finally, the `^Caol` tag is mapped to the empty string.

**RE 73.** `%^Caol -> [];`

*Broadening (^Lea)*

In the following examples, *greim* 'grip' is broadened to become *greama* and *spéir* 'sky' becomes *spéartha*. In both cases *i* (preceded by *e* or *é*) is replaced by *a* in the context of a broadening tag `^Lea`.

(91)greim+Noun+Masc+Gen+Sg+Idf:gr^X**ei**m^Ath^Leaa
spéir+Noun+Fem+Gen+Strong+Pl+Idf:sp^X**éi**r^Leatha

**RE 74.** `i -> a || [e|é] _ Cons+ %^Lea`

(92)greim+Noun+Masc+Gen+Sg+Idf:gr^X**ea**m^Ath^Leaa
spéir+Noun+Fem+Gen+Strong+Pl+Idf:sp^X**éa**r^Leatha

In the following examples *binn* 'point' is broadened, by replacing *i* with *ea*, to become *beanna* and likewise *crith* 'shake', becomes *creatha*.

(93)binn+Noun+Fem+Com+Pl+Idf:b^X**i**nn^Leaa
crith+Noun+Masc+Com+Pl+Idf:cr^X**i**th^Lea^LCanna

**RE 75.** `i -> e a || [b|c|m|r] (h) %^X _ Cons+ %^Lea`

(94)binn+Noun+Fem+Com+Pl+Idf:b^X**ea**nn^Leaa
crith+Noun+Masc+Com+Pl+Idf:cr^X**ea**th^Lea^LCanna

The most usual way of broadening is to remove *i* when it follows another vowel or diphthong. The following are a range of examples of nouns which are broadened in this manner; *stair* 'history', *gáir* 'shout', *droim* 'back, *bádóir* 'boatman', *súil* 'eye', *traein* 'train', *aoir* 'satire', *bliain* 'year' and *buairt* 'worry'.

(95)stair+Noun+Fem+Gen+Strong+Pl+Idf:st^X**ai**r^Leatha
gáir+Noun+Fem+Gen+Strong+Pl+Idf:g^X**ái**r^Leatha
droim+Noun+Masc+Gen+Strong+Pl+Idf:dr^X**oi**m^Lea^Lcanna
bádóir+Noun+Masc+Gen+Sg+Idf:bád^X**ói**r^Leaa
súil+Noun+Fem+Gen+Weak+Pl+Idf:s^X**úi**l^Lea
traein+Noun+Fem+Gen+Sg+Idf:tr^X**^AEi**n^Leaach
aoir+Noun+Fem+Gen+Strong+Pl+Idf:^X**^AOi**r^Leatha

bliain+Noun+Fem+Gen+Strong+Pl:bl^X**iai**n^Leata
buairt+Noun+Fem+Gen+Strong+Pl+Idf:b^X**^UAi**rt^Ath^Leaaí

**RE 76.** i -> [] || [a|á|o|ó|ú|%^AE|%^AO|%^IA|%^UA] _ Cons+ %^Lea;

(96)stair+Noun+Fem+Gen+Strong+Pl+Idf:st^X**ar**^Leatha
gáir+Noun+Fem+Gen+Strong+Pl+Idf:g^X**ár**^Leatha
droim+Noun+Masc+Gen+Strong+Pl+Idf:dr^X**om**^Lea^Lcanna
bádóir+Noun+Masc+Gen+Sg+Idf:bád^X**ó**r^Leaa
súil+Noun+Fem+Gen+Weak+Pl+Idf:s^X**ú**l^Lea
traein+Noun+Fem+Gen+Sg+Idf:tr^X**^AEn**^Leaach
aoir+Noun+Fem+Gen+Strong+Pl+Idf:^X**^AO**r^Leatha
bliain+Noun+Fem+Gen+Strong+Pl:bl^X**ia**n^Leata
buairt+Noun+Fem+Gen+Strong+Pl+Idf:b^X**^UA**rt^Ath^Leaaí

In the following examples *cuid* 'part' and *fuil* 'blood' are broadened by replacing the *ui* with an *o*.

(97)cuid+Noun+Fem+Gen+Strong+Pl+Idf:c^X**uid**^Lea^LCanna
fuil+Noun+Fem+Gen+Sg+Idf:f^X**uil**^Leaa

**RE 77.** u i -> o || _ Cons+ %^Lea;

(98)cuid+Noun+Fem+Gen+Strong+Pl+Idf:c^X**o**d^Lea^LCanna
fuil+Noun+Fem+Gen+Sg+Idf:f^X**o**l^Leaa

In the following examples *feadaíl* 'whistle' and *tír* 'country' are broadened by the insertion of *o* after the final slender vowel.

(99)feadail+Noun+Fem+Gen+Sg+Idf:fead^X**aí**l^Leaa
tír+Noun+Fem+Com+Pl+Def:t^X**í**r^Leatha

**RE 78.** [..] -> o || (a) í _ Cons+ %^Lea;

(100) feadaíl+Noun+Fem+Gen+Sg+Idf:fead^X**aío**l^Leaa
tír+Noun+Fem+Com+Pl+Def:t^X**ío**r^Leatha

The following two inflectional mark-up tags are now eliminated.

**RE 79.** %^Lea -> [];

**RE 80.** %^X -> [];

*Check Orthographic Vowel Harmony (^LC)*

Suffixes and stems must match with respect to broadness or slenderness. Vowel harmony is ensured by checking if any adjustments need to be made between stem and suffix. All verbal suffixes have a broad and slender allomorph, e.g. *-faidh/-fidh*. A broad verbal suffix is appended as standard in the lexicon and marked with the `^Caol` (slender) tag where appropriate. Slenderisation rules then apply to these strings as already outlined.

In the case of nouns, quite often, the stem changes to accommodate a suffix. There are, however, some broad and slender allomorphs such as *-acha/-eacha* (101) and *-anna/-eanna* (102). Again, the broad suffixes, e.g. *-acha* or *–anna*, are appended as standard in the lexicon, preceded by the `^LC` tag.

(101) *nead*
*nead**acha*** 'nests'
*stoirm*
*stoirm**eacha*** 'storms'

(102) *carr* 'car'
*carr**anna*** 'cars'
*áit* 'place'
*áit**eanna*** 'places'

In RE 81 the `^LC` marker is replaced by *e* in cases where the final stem syllable is slender, thus slenderising the suffix. This check is carried out after stem modifications have been made, i.e. after `^Coim`, `^Caol` and `^Lea` tags have been processed.

(103) `nead+Noun+Fem+Gen+Strong+Pl+Idf:nead`**`^LC`**`acha`
`stoirm+Noun+Fem+Gen+Strong+Pl+Idf:stoirm`**`^LC`**`acha`
`carr+Noun+Masc+Gen+Strong+Pl+Idf:carr`**`^LC`**`anna`
`áit+Noun+Fem+Gen+Strong+Pl+Idf:áit`**`^LC`**`anna`

**RE 81.** `%^LC -> e || [i|í] Cons* _`

(104) `nead+Noun+Fem+Gen+Strong+Pl+Idf:nead`**`^LC`**`acha`
`stoirm+Noun+Fem+Gen+Strong+Pl+Idf:stoirm`**`e`**`acha`
`carr+Noun+Masc+Gen+Strong+Pl+Idf:carr`**`^LC`**`anna`
`áit+Noun+Fem+Gen+Strong+Pl+Idf:áit`**`e`**`anna`

**RE 82.** `%^LC -> []`

Emphatic suffixes also have allomorphs (*-se/-sa*). The *s* is appended in the lexicon as standard. RE 83-84 completes the suffix through the addition of the appropriate vowel, i.e. *e* if the stem is slender and *a* if the stem is broad.

**RE 83.** `[..] -> e || [e|é|i|í] Cons* s _ %^Emph`

**RE 84.** `[..] -> a || [a|á|o|ó|u|ú] Cons* s _ %^Emph`

Stems ending in *s* are hyphenated when an emphatic suffix *–sa/-se* is added, e.g. *mo chos-sa* 'my foot'.

**RE 85.** `[..] -> %- ||  s _ s (a|e) %^Emph`

**RE 86.** `%^Emph -> []`

*Restore Diphthongs and Long-vowels*

Diphthongs are now reinstated.

**RE 87.** `%^AE -> a e;`

**RE 88.** `%^IA -> i a;`

**RE 89.** `%^AO -> a o;`

**RE 90.** `%^UA -> u a;`

## 5.8 FST Manipulation

The lexicon and rule transducers are composed and unioned together to create one large morphological transducer. This transducer network may be manipulated by using composition and union, to filter out unwanted paths and to add in new paths, as described in the section on using composition for filtering in (Beesley and Karttunen, 2001, p358-361).

The following is an example of a script which is input to the *xfst* tool in order to manipulate the network in the required manner, i.e. remove unwanted paths and add new paths.

```
clear stack

! (1) REMOVE VOCATIVE PLURAL INDEFINITE FORMS

read regex ~$[%+Voc %+Pl %+Idf] .o. @"noun.fst";
save stack noun.fst

! (2) REMOVE INCORRECT STRINGS AND ADD CORRECT REPLACEMENTS
clear stack

define Bad    {tonn} %+Noun %+Fem %+Gen %+Pl      ! (d)tonnta
             |{crios} %+Noun %+Masc ?+ %+Pl       ! (g)creasanna
             |{ceann} %+Noun %+Masc %+Voc %+Sg    ! chinn
             |{pobal} %+Noun %+Masc %+Voc %+Sg    ! phobail
             |{stór} %+Noun %+Masc %+Voc %+Sg     ! stóir
             ;

define Good   [{tonn} %+Noun %+Fem %+Gen %+Pl %+Def]:{dtonn}
             |[{tonn} %+Noun %+Fem %+Gen %+Pl %+Idf]:{tonn}
             |[{crios} %+Noun %+Masc %+Nom %+Pl %+Def]:{criosanna}
             |[{crios} %+Noun %+Masc %+Nom %+Pl %+Idf]:{criosanna}
             |[{crios} %+Noun %+Masc %+Gen %+Pl %+Def]:{gcriosanna}
             |[{crios} %+Noun %+Masc %+Gen %+Pl %+Idf]:{criosanna}
             |[{crios} %+Noun %+Masc %+Voc %+Pl %+Def]:{chriosanna}
             |[{ceann} %+Noun %+Masc %+Voc %+Sg]:{cheann}
             |[{pobal} %+Noun %+Masc %+Voc %+Sg]:{phobal}
             |[{stór} %+Noun %+Masc %+Voc %+Sg]:{stór}
             ;

! REMOVE THE BAD STRINGS FROM THE EXISTING NETWORK
read regex ~$[Bad] .o. @"noun.fst";

! CREATE A NEW NETWORK OF GOOD REPLACEMENT STRINGS
read regex Good;

! COMBINE THE TWO NETWORKS
union net
save stack noun.fst
```

**Fig 44. Updating the morphological transducer**

As stated earlier, it is sometimes more convenient to allow some unwanted strings be generated and later removed than to create complicated exceptions to rules designed to prevent their generation in the first place. The code in Fig. 44 (1), removes the indefinite form of vocative plural nouns, i.e. all strings which contain the tags `+Voc +Pl +Idf` from the network. This is achieved by specifying the complement ( `~S`) of the set of strings containing `+Voc +Pl +Idf`, i.e. the set of strings not containing the string `+Voc +Pl +Idf`. This is composed with the transducer network, saved as "noun.fst", and thus all strings containing this sequence of tags are filtered out (Beesley and Karttunen, 2001, p360).

Some stems follow a general paradigm except for some minor exceptions. It is also more convenient to treat them as a regular members of their particular word-class and make the necessary adjustment(s) to specific forms at the end (Beesley and Karttunen, 2001, p367).
Table 24 shows some inflected forms which are generated according to general rules for genitive singular (slenderise and add *–e*) and genitive plural (eclipse and add *–ta*) for this class of nouns. The genitive plural of *tonn* 'wave', is an exception to the rule in that the *–ta* suffix is not added. It is, however, placed in the same class as the other nouns listed causing a genitive plural *(d)tonnta* to be generated. In the final morphological transducer this is replaced by *(d)tonn.*

**Table 24. Rule exception example**

| *2nd Declension Nouns*<br>*Feminine, Strong Plural* | *Genitive Sg* | *Genitive Pl.* |
|---|---|---|
| *buíon* 'group, band' | na **buíne** | na **mbuíonta** |
| *grian* 'sun' | na **gréine** | na **ngrianta** |
| *mian* 'wish, desire' | na **méine** | na **mianta** |
| *pian* 'pain' | na **péine** | na **bpianta** |
| *tonn* 'wave' | na **toinne** | na **dtonn** |

In Fig. 44 (2), a number of incorrectly generated strings are defined using regular expressions and are assigned to a variable named "Bad". The use of curly brackets is a shorthand for spacing out the individual characters in a stem, i.e. {tonn} is equivalent to "t o n n". Again, the complement of these bad strings is composed with the transducer network, which in effect filters out the bad strings. The new transducer is resaved. A network of good strings is then added to the new transducer using the union operator.

## *5.9 Closed Inflected Word-Classes*

Lexicons have been coded for a number of closed inflected word classes, such as conjugated prepositions (prepositional pronouns), personal pronouns, and the article.

### Prepositional Pronouns

Prepositional pronouns are simple prepositions which are conjugated for person and number by combining them with the personal pronouns *mé* 'me', *tú* 'you', *sé* 'him', *sí* 'her', *sinn* 'us', *sibh* 'you pl.', *siad* 'them'. i.e. *ag* 'at' is combined with the pronouns to produce *agam, agat, aige, aici, againn, agaibh* and *acu* respectively. Because of the limited number of such prepositions (eighteen) all inflected forms are simply listed together with their morphosyntactic description. A sample from the preposition lexicon is given Fig. 45-46.

```
Multichar_Symbols
+Prep +Simp +Comp +Emph +Cpx +Rel +Poss +1P +2P +3P +Fem +Masc +Sg +Pl

LEXICON Prepositions

! SIMPLE PREPOSITIONS
!
a+Prep+Simp:a                               #;  ! a chlog, a dhíth
ag+Prep+Simp:ag                             #;  ! at
ar+Prep+Simp:ar                             #;  ! on
as+Prep+Simp:as                             #;  ! out
chuig+Prep+Simp:chuig                       #;  ! to

! CONJUGATED PREPOSITIONS (prepositional pronouns)

ag+Prep+Simp:ag                             #;                 ! at
ag+Prep+Comp+1P+Sg:agam                     Emphasis-br;       ! at me
ag+Prep+Comp+2P+Sg:agat                     Emphasis-br;       ! at you
ag+Prep+Comp+3P+Sg+Masc:aige                Emphasis-sl-3P;    ! at him
ag+Prep+Comp+3P+Sg+Fem:aici                 Emphasis-sl-3P;    ! at her
ag+Prep+Comp+1P+Pl:againn                   Emphasis1P;        ! at us
ag+Prep+Comp+2P+Pl:agaibh                   Emphasis-sl;       ! at you
ag+Prep+Comp+3P+Pl:acu                      Emphasis-br-3P;    ! at them

... etc.

! COMPOUND PREPOSITIONS
!
<a r %  c h ú l %+Prep:0 %+Comp:0>          #;    ! behind
<a r %  f e a d h %+Prep:0 %+Comp:0>        #;    ! during
<a r %  f u d %+Prep:0 %+Comp:0>            #;    ! throughout
<a r %  n ó s %+Prep:0 %+Comp:0>            #;    ! in the manner of
```

**Fig 45. Extract of Preposition Lexicon, part i**

```
! COMPLEX COMPOUND PREPOSITIONS
!
i+Prep+Cpx:sa                       #;    ! in
i+Prep+Cpx:san                      #;    ! in the (singular)
i+Prep+Cpx:sna                      #;    ! in the (plural)

! COMPLEX RELATIVE COMPOUND PREPOSITIONS
!
ar+Prep+Cpx+Rel:arna          #;  ! on his/her/their/its having been
faoi+Prep+Cpx+Rel:faoina      #;  ! under his/her/their/its
le+Prep+Cpx+Rel:lena          #;  ! with his/her/their/its

LEXICON Emphasis-br
#;                                        ! no emphasis
+Emph:sa          #;                      ! -sa

LEXICON Emphasis-sl
#;                                        ! no emphasis
+Emph:se          #;                      ! -se

LEXICON Emphasis-br-3P
#;                                        ! no emphasis
+Emph:san         #;                      ! -san

LEXICON Emphasis-sl-3P
#;                                        ! no emphasis
+Emph:sean        #;                      ! -sean

LEXICON Emphasis-1P
#;                                        ! no emphasis
+Emph:e           #;                      ! -e
```

**Fig 46. Extract of Preposition Lexicon, part ii**

Emphatic forms of conjugated prepositions are created through appending the appropriate suffix. The suffix is modified by replace rules ensure broad/slender harmony with the stem.

As well as simple and conjugated prepositions, the preposition lexicon also contains compound prepositions, complex compound prepositions and complex relative compound prepositions. Unlike the latter two (complex) types of preposition, compound prepositions consist of two words and are therefore treated differently to all other lexical items encountered to date.

As already mentioned, the *lexc* format allows us to use a shorthand format upper:lower to specify the upper and lower level representations of strings, without having to explicitly state individual mappings. For example cat+Pl:cats interpreted as c:c, a:a, t:t, +Pl:0. Because there are two parts to the compound prepositions we cannot use this shorthand. However, by surrounding the items with angle brackets we can revert to the longer notation. Therefore, `<a r %  c h ú l %+Prep:0 %+Comp:0>` is equivalent to `<a:a r:r %_:%_ c:c h:h ú:ú l:l %+Prep:0 %+Comp:0>`, where the underscore is used to denote a space character.

### Pronouns

The following is an extract from the pronoun lexicon. The pronouns *sí*, *sé* and *siad* are used only when the pronoun follows the verb in subject position as in (105).

(105) *Chuaigh sí amach* 'She went out'
*Téigh gan í* 'Go without her'

```
Multichar_Symbols

+Pron +Pers +Emph +Ref +Idf +1P +2P +3P +Fem +Masc +Sg +Pl +VerbSubj

LEXICON Pronouns

mé+Pron+Pers+1P+Sg:mé                     #;    ! me
tú+Pron+Pers+2P+Sg:tú                     #;    ! you
sí+Pron+Pers+3P+Sg+Fem+VerbSubj:sí        #;    ! she
í+Pron+Pers+3P+Sg+Fem:í                   #;    ! she/her
sé+Pron+Pers+3P+Sg+Masc+VerbSubj:sé       #;    ! he
é+Pron+Pers+3P+Sg+Masc:é                  #;    ! he
sinn+Pron+Pers+1P+Pl:sinn                 #;    ! we
sibh+Pron+Pers+2P+Pl:sibh                 #;    ! you pl.
siad+Pron+Pers+3P+Pl+VerbSubj:siad        #;    ! they
iad+Pron+Pers+3P+Pl:iad                   #;    ! they

! Emphatic/Contrastive

mise+Pron+Pers+1P+Sg+Emph:mise                  #;    ! me/myself
tusa+Pron+Pers+2P+Sg+Emph:tusa                  #;    ! you/yourself
sise+Pron+Pers+3P+Sg+Fem+VerbSubj+Emph:sise     #;    ! she
ise+Pron+Pers+3P+Sg+Fem+Emph:ise                #;    ! she/herself
seisean+Pron+Pers+3P+Sg+Masc+VerbSubj+Emph:seisean    #;    ! he
eisean+Pron+Pers+3P+Sg+Masc+Emph:eisean               #;    ! he
sinne+Pron+Pers+1P+Pl+VerbSubj+Emph:sinne             #;    ! we
sibhse+Pron+Pers+2P+Pl+Emph:sibhse                    #;    ! you pl.
siadsan+Pron+Pers+3P+Pl+VerbSubj+Emph:siadsan         #;    ! they/them
iadsan+Pron+Pers+3P+Pl+Emph:iadsan                    #;    ! they

! Reflexive

féin+Pron+Ref:féin                              #;    ! self

! Indefinite

ceachtar+Pron+Idf:ceachtar                      #;    ! anyone
neachtar+Pron+Idf:neachtar                      #;    ! not anyone
```

**Fig 47. Extract of Pronoun Lexicon**

### Articles

The article lexicon is reproduced in Fig. 48.

```
Multichar_Symbols

+Art +Def +Masc +Fem +Com +Gen +Sg +Pl

LEXICON Articles

an+Art+Com+Sg+Def:an          #;
an+Art+Gen+Sg+Def+Masc:an     #;
na+Art+Gen+Sg+Def+Fem:na      #;

na+Art+Com+Pl+Def:na          #;
na+Art+Gen+Pl+Def:na          #;
```

**Fig 48. Article Lexicon**

## *5.10 Functional Word-Classes*

A number of non-inflected word-classes are also included in the lexicon since the number of lexical items is small, but their frequency of usage is high. The following word-classes, consisting mainly of function words, are encoded:

- Determiners (possessive pronouns/adjectives, demonstratives and interrogative pronouns)
- Adverbs
- Prepositions (simple and compound)
- Conjunctions
- Numerals
- Interjections
- Particles.

The stems for functional and closed inflected items have been extracted from a machine-readable form of the pocket dictionary, *An Foclóir Póca* (An Roinn Oideachais, 1986a)

*Determiners*

This class includes possessive determiners (also known as possessive pronouns or possessive adjectives) (91), demonstrative determiners (92) and interrogative determiners (93):

(91) *mo* 'my'

*do* 'your' (singular)

*a* 'his', her' or 'their'

*ár* 'our'

*bhur* 'your' (plural)

(92) *seo* 'this'

*sin* 'that'

(93) *cad* 'what'

*cé* 'who'

*Adverbs*

Many adjectives are used as adverbs following the particle *go* as in example (94):

(94) *Sin obair* ***chrua***

That work **hard**

'That is **hard** work'

*Tá sí ag obair* ***go crua***

Is she working **hard**

'She is working **hard**'

*Conjunction*

Conjunctions are classified as either co-ordinate (95) or sub-ordinate (96):

(95) *agus* 'and'

(96) *ach* 'but'

*Numerals*

Numerals are classified as cardinal (97), ordinal (98), personal (99) and adjectival (100):

(97) *aon* 'one'

*dó* 'two'

*trí* 'three'

(98) *céad* ‘first’
*dara* ‘second’
*tríú* ‘third’

(99) *duine* ‘one person’
*beirt* ‘two people’
*triúr* ‘three people’

(100) *dhá lámh* ‘two hands’

*Interjection*

The following are some examples of interjections:

(101) *á* ‘ah!’
*faraor* ’unfortunately’
*abú* ’victory to’

*Particles*

There are many categories of particle including verbal (102), adverbial (103), vocative, continuative, numeral, degree, patronymic (104) used in surnames, relative etc. :

(102) *ní* negative verbal particle

(103) *go* adverbial particle

(104) *uí* patronymic particle

### *5.11 Summary*

This chapter described the implementation of the morphological phenomena and morphotactics of nouns, adjectives, and verbs. The treatment of some commonly used non-inflected parts-of-speech was also discussed.

# 6. Testing and Evaluation

## 6.1 Introduction

This chapter deals with maintaining accuracy in the system and assessing language coverage. Some areas of further work are suggested.

## 6.2 Testing and Development

During the development of the system, testing focussed on two main areas; the integrity of the rules and the well-formedness of the lexical, intermediate, and surface representations.

### Rule integrity

It is important that confidence in the integrity and accuracy of the system is maintained as it is changed and enhanced. During development, there is a great danger that parts of the system which are tested and working, will be disrupted when adding new rules or when fixing a problem. Therefore, it is crucial that regression testing is carried out consistently from the start in order to avoid this problem.

The *xfst* tool has several very useful features, documented in (Beesley and Karttunen, 2001, p393), which enable rigorous and consistent testing. Using projection and subtraction, the network of old surface forms can be subtracted from the network of new surface forms giving the list of new word-forms which have just been added. By examining this list one can check that only correct word-forms have been generated. Conversely, subtracting the new network from the old network gives the list of words-forms which have been lost. This may be as intended, or it could signal a problem. If these checks are performed after each change to the system, any unintentional effects can be quickly spotted and the problem can rectified before continuing.

### Well-formedness

It is important for both analysis and generation that the lexical tags are consistent in naming and in the proper order of concatenation. It would be undesirable, for example, for some nouns to be specified as `nounstem+Noun+Com+Sg` and others as `nounstem+Noun+Sg+Com`. It is also necessary that the tags in the intermediate surface levels conform to a specified standard in order to ensure that replace rules will fire as intended.

For example, in the case of nouns, there are mandatory tags to describe lexical class, gender, case, number, and definiteness (see Appendices A and B) as well as a number of optional tags. Some tags are drawn from lists of mutually exclusive tags, i.e. gender must be either `+Fem` or `+Masc`. The tags must also appear in a specified order to provide a consistent interface to other systems. These requirements may be checked using a tag-grammar written as a regular expression (Beesley and Karttunen, 2001, p387).

The regular expression, given below in Fig. 49, states that all lexical (upper) level strings in the noun lexicon transducer must conform to the following grammar:

- *one or more symbols from the alphabet*
- followed by an optional dialect tag
- followed by an optional proper noun tag
- followed by the noun tag
- followed by a gender tag
- followed by a case tag
- followed by a number tag (including strong/weak tag for genitive plurals)
- followed a tag for definiteness
- followed by an optional initial mutation tag
- followed by an optional emphatic-form tag.

The exclamation mark indicates that any text on the same line which follow it is treated as a comment not part of the regular expression. Round brackets indicate optionality. Not all strings will contain a tag from the list of options enclosed in round brackets. Where the tags are enclosed in square brackets one of the options must appear in the string.

```
[á|é|í|ó|ú|a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|z]+
[
  (%+CC|%+CD|%+CM)              ! dialect: canúint Connachta/
                                ! Dún na nGall/Mumhan
  (%+Prop)                      ! proper noun
   %+Noun                       ! noun marker
   [%+Fem|%+Masc]               ! gender
   [
    [[%+Com|%+Dat] [%+Sg|%+Pl|%+Num] [%+Def|%+Idf] ]|
    [%+Gen [%+Sg | [[%+Strong|%+Weak] %+Pl] ] [%+Def|%+Idf] ] |
    [%+Voc [%+Sg|%+Pl] (%+Def)]
   ]
  (%+Len|%+Urú)
  (%+Emph)
];
```

**Fig 49. Lexical tag grammar for Nouns**

By taking the upper level of the noun transducer (using projection) and subtracting the lexical grammar network defined in Fig. 48, we are left with all strings not conforming to the grammar. Any strings in the resulting network signify a problem which requires attention. (The morphological tag grammar for the entire morphological transducer is given in Appendix B.)

The same type of test can be carried out to check that intermediate surface forms are well-formed. It is essential when composing rule transducers with the lexicon that the lower level of the lexicon transducer matches the upper level of the first replace rule transducer.

The surface level of all strings in the noun lexicon transducer should conform to the following grammar:

- one or more symbols from the alphabet i.e. a stem
- optionally followed by gender tag (gender is not required for plural formation)
- followed by case tag
- optionally followed by a syncopate or change tag (final mutations)
- optionally followed by a slenderise or broaden tag (final mutations)
- optionally followed by a check vowel harmony marker (required by allophonic suffixes)
- zero or more symbols from the alphabet, i.e. a suffix
- optionally followed by one of four sets of initial mutation markers.

```
[á|é|í|ó|ú|a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|z|%^X]+
 [
  (%^M|%^F)                    ! masc, fem
  [%^C|%^G|%^V]                ! com., gen. or voc.
  (%^Coim|%^Ath)               ! ^Coim (coimriú) syncopate
                               ! ^Ath (athrú) change/replace
  (%^Caol|%^Lea)               ! ^Caol (caolú) slenderise
                               ! ^Lea (leathnú) broaden
  (%^LC)                       ! ^LC (leathan-caol) br./slen harmony
  (á|é|í|ó|ú|a|c|d|e|g|h|i|n|o|t|u)*     ! e.g. anna, í etc
  ((%^Sé %^tv %^ts) |                    ! used in common singular
   (%^Sé %^hv %^ts) |                    ! used in genitive singular
   (%^Sé) |                              ! used in vocative singular
   (%^Sé %^Urú %^hv)                     ! used in plurals
  )
 ];
```

**Fig 50. Inflectional mark-up tag grammar ( lower level) for Noun Lexicon**

By taking the lower level of the noun transducer (using projection) and subtracting the (intermediate) surface tag grammar network defined in Fig. 50, we are left with all strings not conforming to the grammar. As with the lexical tag test, any strings in the resulting network signify a problem which requires attention.

## 6.3 Assessing Language Coverage

In the first phase of this work, all of the morphological phenomena of Irish inflectional morphology were implemented. Examples from standard reference grammars (Christian Brothers, 1988; Bráithre Críostaí, 1999) were used. The focus of the reference grammars is to illustrate all aspects of grammar as comprehensively as possible. The examples used do not necessarily reflect the relative frequency of use of those phenomena. It is possible, therefore, that the lexicon, at the end of the implementation phase, may contain obscure words and lack commonly used words.

The aims of the second phase of the work are therefore to assess the language coverage of the lexicon in terms of frequency of use, and to augment the lexicon where necessary to ensure that the most frequently used words in the language are included. The resources used to evaluate language coverage are *Corpas Náisiúnta na Gaeilge (*Institiúid Teangeolaíochta Éireann), a corpus of contemporary Irish text, and *An Foclóir Póca* (ITÉ/An Gúm), a machine-readable dictionary containing 15,000 headwords.

At the end of phase one, the test lexicons contained the following number of stems:

**Table 25. Lexicon: End of Phase1**

| | |
|---|---:|
| Nouns | 275 |
| Verbs | 120 |
| Adjectives | 85 |
| Others | 30 |
| Total | 510 |

From these stems the system generated over 5000 unique surface forms and over 14,000 morphological descriptions (a surface form can have several analyses).

According to Crystal (1997b, p87), if the unique word types in a sufficiently large sample of text (in any language) are listed in order of decreasing frequency, certain statistical predictions can be made. The 15 most frequently used words in the text account for about 25% of the text, the 100 most frequently used word account for about 60% of the text etc.

**Table 26. Text Coverage Predictions**

| *Most freq. used words* | *% of text (Crystal)* | *% of text (CNG)* |
|---|---|---|
| 15 | 25.0 | 25.0 |
| 100 | 60.0 | 45.0 |
| 1000 | 85.0 | 67.0 |
| 4000 | 97.5 | 80.0 |

A list of the most frequently used words in the *Corpas Náisiúnta na Gaeilge* (*CNG*) was computed using Wordsmith Tools. The corpus consists of over 14 million words (tokens) of running text and approximately 280,000 word-types. The word-types were listed in descending order of frequency[10] and the percentages were calculated as shown in Table 26.

In accordance with the predictions, the results show that the 15 most frequently used words in the Irish corpus represent 25% of the text. The results for 100, 1000 and 4000 most frequently used words in Irish show a lower coverage than predicted by Crystal. The 15 most frequently used forms in Irish are mainly non-inflected function words (See Appendix J), which may explain the agreement at this level. A more direct comparison could be made by using a lemmatised frequency list for Irish.

If the transducer contained the 1000 most frequently used (lemmatised) words in the language then we would expect a recognition rate of about 85% in a random corpus of the language. In order to assess the coverage of the system, developed by the end of phase one, a list of the 1000 most frequent word types from *Corpas Náisiúnta na Gaeilge (CNG)* were used. The lexicon, because of its relatively small size and its focus on forms rather than frequencies, contained just over one third of the 1,000 most frequently used words in Irish. The recognition rate for the lexicon at the end of Phase 1 is 37% (Table 27).

**Table 27. Coverage Analysis – Phase 1**

| *1000 most freq. used words* | *Nouns* | *Verbs* | *Adjs.* | *Other* | *Total* |
|---|---|---|---|---|---|
| Recognised by Transducer | 159 | 72 | 25 | 114 | 370 |
| Not recognised by Trans. | 448 | 76 | 52 | 54 | 630 |
| Total | 607 | 148 | 77 | 168 | 1000 |

The remaining 63% of the 1000 word-forms was analysed and just over half were contained in *An Foclóir Póca* as shown in Table 28, meaning that they coincided with headwords in the dictionary. Of the remaining forms the greatest proportion (69%) were nouns.

**Table 28. Analysis of forms not recognised by transducer – Phase 1**

| *Not recognised by Trans.* | *Nouns* | *Verbs* | *Adjs.* | *Other* | *Total* |
|---|---|---|---|---|---|
| Contained in *An Foclóir Póca* (headwords) | 233 | 9 | 31 | 45 | 318 |
| Not in *An Foclóir Póca* (inflected forms/abbreviations, proper nouns) | 215 | 67 | 21 | 9 | 312 |
| Total | 448 | 76 | 52 | 54 | 630 |

All word-forms belonging to non-inflected word classes contained in *An Foclóir Póca* were added to the relevant lexicons. This is a straightforward task as there is just one lexicon for each non-inflected lexical class. The stems corresponding to the non-recognised inflected items were also added. This is a more complex task, however, as the correct sub-lexicon (continuation class) within the appropriate lexicon must be identified (see Section 6.4).

After manual addition of all stems relating to the 1000 most frequently used word forms, the lexicon at the end of Phase 2, is as shown in Table 29. From these stems the system generates over 10,000 unique surface forms and over 30,000 morphological descriptions.

**Table 29. Lexicon: End of Phase2**

| | |
|---|---|
| Nouns | 651 |
| Verbs | 141 |
| Adjectives | 130 |
| Others | 245 |
| Total | 1167 |

Based on the percentages for Irish (CNG) given in Table 27, we can estimate that the recognition rate for the lexicon at the end of Phase 2 is 67%.

## 6.4 Lexicon Maintenance

For a morphological transducer to be of practical use it must be clear how new stems should be added. For the inflected parts of speech (nouns, verbs, and adjectives) the correct lexicon class must be selected in order to generate the correct inflections. The lexicon classes are listed in Appendices F, G and H. Adding verb roots is relatively straightforward; in most cases the relevant conjugation can be determined automatically based on the number of syllables. The broad or slender nature of the root can also be determined automatically.

Currently in order to add new noun stems manually to the noun lexicon the following steps are required:

a) locate the appropriate declension table

b) go to the feminine or masculine section as appropriate

c) go to weak or strong plural section as appropriate

d) locate the correct plural formation category

e) if a change of syllable is involved then check the relevant table to check that this particular change is included.

In order to do this the following information about the stems is required:

- *gender*
- declension number <u>or</u> how its genitive singular is formed
- how the plurals are formed, and whether they are the same for all cases (i.e. a strong plural).

Machine-readable dictionaries usually include gender and plural-formation information and in some cases declensional information. This information could be used to populate the lexicon semi-automatically.

## 6.5 Summary

In this chapter, the issues of maintaining accuracy in the system and assessing language coverage were covered. Lexicon maintenance is also discussed.

# 7. Conclusion

In this work, I have designed and implemented a lexicalised, bi-directional finite-state based inflectional morphology of Irish. The use of two-level morphology and finite-state technology is both theoretically and technologically attractive. From a theoretical point of view the morphological features of Irish are well suited to finite-state techniques. The concatenative nature of Irish affixation can be implemented without difficulty using finite-state transducers. Stem mutations are all influenced by their immediate locale (there are no long distance dependencies within words), and may be expressed effectively as regular expressions and ultimately finite-state transducers.

Although the task of assigning the correct lexical categories to stems is labour intensive, this can in part be automated and the explicit encoding of morphosyntactic information makes this a flexible and re-usable resource. This morphological transducer can be used in the following areas of NLP:

- Generation of inflected forms for spelling checking
- Word form analyser in a language parsing application
- Word form generator in a language generation application
- Corpus part-of-speech tagging and lemmatization
- Text-to-speech synthesis
- As an interface to a machine-readable dictionary
- As a stemmer for Information Retrieval (document retrieval) and Extraction

and it provides a basis for progressing to finite-state syntactic chunking of Irish.

There is much scope for further work in this area. Initially, I plan to (semi-automatically) convert existing machine-readable dictionaries to *lexc* format in order to increase the language coverage of the morphological transducer. As this work deals only with inflectional morphology, there is a need to extend the system to also include derivational morphology.

A morphological guesser (Beesley and Karttunen, 2001, p452) could be implemented for stems not contained in the lexicon. This would define forms which are phonologically possible.

This finite-state morphology could be of benefit in dealing with the following issues:

- Dialectal forms or variants
- Multiple plural forms
- Historical forms
- Standardisation – changes in grammar and/or orthographic rules.

Further investigation of phonologically and etymologically conditioned alternations could lead to a reduction in the number of sub-lexicons particularly in the case of nouns.

It would also be interesting to investigate the applicability or otherwise of this model of mutation handling to the other Celtic languages such as Welsh, Scottish Gaelic or Breton.

Finally, finite-state techniques could be used for light parsing or ‘chunking’ of texts in noun phrases, verb phrases, and prepositional phrases.

## Glossary of Terms

*Allomorph*: a variation in the form of a morpheme which does not affect its meaning or function (Crystal, 1997a, p15)

*Alternation*: (alternant) the relationship between alternative forms of a linguistic unit, e.g. a morpheme alternant is another term for allomorph.

*Attenuation*: see Syncopation.

*Automaton*: a general term for a device that mechanically processes an input string with the aim of deciding whether it belongs to some set of strings, or of producing an output string (Illingworth, 1986, p25)

*Broad consonant*: in Irish phonetics, a consonant immediately preceding or following a broad (back, velar) vowel in the same word (Pei, 1966, p32,27).

*Broadening*: a form of inflexion whereby a slender consonant is made broad, denoted by vowel-changes preceding the consonant, e.g. by removing the letter *i* from before the consonant (Christian Brothers, 1988, p9). See Velarisation.

*Caolú*: see Slenderisation.

*Coimriú*: see Syncopation.

*Clitic*: a grammar term used to describe a form which looks like a word but which cannot stand alone as a normal utterance being structurally dependant on a neighbouring word, e.g. "the". Proclitics are dependant on the following word and enclitics are dependant on the preceding word (Crystal, 1997a, p64). Clitics may be attached to the form that they are dependant on, e.g. wasn't or *d'fhéach* 'looked'.

*Derivation*: in morphology, a term used to refer to one of the two main categories or processes of word formation, the other being inflection. The result of a derivational process is a new word, unlike inflection which results in a new form of the same word (Crystal, 1997a, p111).

*Depalatalisation*: see Slenderisation.

*Determiner*: items which co-occur with nouns to express a range of semantic contrasts such as quantity and number (Crystal, 1997a, p112).

*Eclipsis*: in grammars of Celtic languages; a term for certain phonetical changes, especially nasalisation, of the initial phonemes of words when they directly follow certain words or flectional forms. (Pei, 1966, p78)

*Etymology*: the study of the origins and history of the form and meaning of words (Crystal, 1997a, p140).

*Final Mutation*: mutation of the final syllable of stems in certain grammatical contexts (Ó Siadhail, 1989, p134).

*Finite-State*: having a finite number of states.

*Fricative*: in phonetics, a classification of consonant forms based on the manner of articulation, i.e. sounds made when two organs come so close together that there is audible friction (Crystal, 1997a, p159).

*Harmony*: A phonological process which causes segments of a particular class to agree in the specification of some phonological features across a certain domain, often a word (Sproat, 1992, p246). In Irish there is harmony in respect of broadness and slenderness between stems and affixes, as stated in the grammatical rule, *Leathan le leathan agus caol le caol* 'Broad with broad and slender with slender'.

*Inflection*: (inflexion) in morphology, a term used to refer to one of the two main categories or processes of word formation, the other being derivation. Inflectional affixes signal grammatical relationships (Crystal, 1997a, p195).

*Initial Mutation*: in Celtic languages, mutation of the initial phoneme of words in certain grammatical contexts.

*Intervocalic*: in phonetics, it refers to a consonant sound between two vowels (Crystal, 1997a, p201).

*Leathnú*: see Broadening.

*Lemma*: in lexicology, the item at the beginning of a dictionary entry, i.e. the headword (Crystal, 1997a, p217). Also, the combination of surface form and underlying analysis (Karlsson & Karttunen, 1997, p96).

*Lenition*: a weakening of the overall strength of a sound. This usually involves a change from a stop to a fricative, from a fricative to an approximant, from a voiceless to a voiced sound, or a sound being reduced to zero (Crystal, 1997a, p 218). In Celtic languages, lenition is a phonetic change to a consonant between two vowels (Pei, 1966, p145).

*Morpheme*: a minimal distinctive unit of grammar, an abstract unit, realised in speech by actual units known as morphs (Crystal, 1997a, p248).

*Morphographemics*: Orthographic representation of morphophonology (Karlsson, 1994, p2570).

*Morphology*: the branch of grammar which deals with the structure or form of words (Crystal, 1997a, p249).

*Morphophonology*: the study of the phonological factors which affect the appearance of morphemes and correspondingly the grammatical factors which affect the appearance of phonemes (Crystal, 1997a, p250).

*Morphosyntax*: syntactic features represented by morphological means, e.g. through the use of bound morphemes in inflectional morphology (Bussmann, 1996, p316).

*Morphotactics*: the study of the arrangement of morphemes in linear sequence (Crystal, 1997a, p249).

*Nasalisation*: in phonetics, a term used to describe sounds (both vowels and consonants) which are made when the soft palate is lowered allowing air to escape audibly through the nose (Crystal, 1997a, p254).

*Orthography*: (spelling) the conventional representation in writing of the spoken word (Pei, 1966, p255).

*Palatalisation*: in phonetics a term used to describe sounds which are made when the front of the tongue touches or approaches the hard palate. Sometimes used in relation to vowels but more commonly in relation to consonants (Crystal, 1997a, p275). See Slenderisation.

*Regular Expression*: a language for specifying text search strings (Jurafsky and Martin, 2000, p22).

*Replace Rule*: formulism for describing a morphological or phonological change which takes place in a particular context.

*Root*: the base form of a word which cannot be further analysed, and from which stems and words may be derived by affixation. (Sproat, 1992, p249).

*Séimhiú*: see Lenition.

*Slender consonant*: in Irish phonetics, a consonant immediately preceding or following a slender (front, palatal) vowel in the same word (Pei, 1966, p32,27).

*Slenderisation*: a form of inflexion whereby a broad consonant is made slender, denoted by vowel-changes preceding the consonant, e.g. by inserting the letter *i* before the consonant (Christian Brothers, 1988, p9). See Palatalisation.

*Stem*: the part of a word to which inflectional suffixes are attached. A stem may consist of a single root morpheme, or two root morphemes (as in a compound) or a root morpheme plus a derivational affix (Crystal, 1997a, p362).

*Strong plural*: A noun in Irish which has the same plural form in the common and genitive cases is said to have a strong plural.

*String*: in computation, a sequence of symbols.

*Suppletive*: in morphology a term which is used to show a relationship between morphemes which have different roots (Crystal, 1997a, p372).

*Syncopation*: the omission of a short unaccented vowel (or vowels) form the final syllable of a polysyllabic stem when lengthened by an inflexion beginning with a vowel (Christian Brothers, 1988, p9).

*Transducer*: in formal language theory, any automaton that produces an output (Illingworth, 1986, p391).

*Urú*: see Eclipsis.

*Velarisation*: in phonetics a term used to describe sounds involving the movement of the back of the tongue towards the soft palate (velum) (Crystal, 1997a, p409). See Broadening.

*Weak plurals*: A noun in Irish which has different plural forms in the common and genitive cases is said to have weak plurals.

# References

An Roinn Oideachais, 1986a. *Foclóir Póca English-Irish/Irish-English Dictionary.* Baile Átha Cliath: An Gúm.

An Roinn Oideachais, 1986b. *Foclóir Póca English-Irish/Irish-English Dictionary Learner's Cassette.* Baile Átha Cliath: An Gúm/ITÉ.

An Roinn Oideachais agus Eolaíochta, 2001. *Tuarascáil Staitistiúil 1999/2000 Statistical Report.* Baile Átha Cliath: Oifig an tSoláthair.

Andersen, H., 1987. Introduction: Sandhi. *In:* ed. H. Andersen, *Sandhi Phenomena in the Languages of Europe.* Mouton de Gruyter. pp. 1-8.

Annunciata le Muire, An tSiúr and Ó Huallacháin, C. L., An tAthair, 1966. *Bunchúrsa Foghraíochta. [A Basic Course in Phonetics].* Baile Átha Cliath: Oifig an tSoláthair.

Antworth, E. L., 1990. *PC-KIMMO: A Two-level Processor for Morphological Analysis.* Dallas, Texas: Summer Institute of Linguistics.

Baayen, R. H., 2001. *Word Frequency Distributions.* Kluwer Academic Publishers.

Bammesberger, A., 1983. *An Outline of Modern Irish Grammar.* Heidelberg: Winter.

Bauer, L., 1983. *English Word-Formation.* Cambridge University Press.

Bauer, L., 1988. *Introducing Linguistic Morphology.* Edinburgh University Press.

Beesley, K. and Karttunen, L., 4 Oct 2001 (draft). *Finite State Morphology: Xerox Tools and Techniques.* Cambridge University Press (forthcoming).

Bloomfield, L., 1935. *Language.* London: Allen and Unwin.

Bráithre Críostaí, 1999. *Graiméar Gaeilge na mBráithre Críostaí. [The Christian Brothers' Irish Grammar].* 2nd. ed. Baile Átha Cliath: An Gúm.

Brill, E., 1992. A simple rule-based part of speech tagger. *In: Proceedings of Third Conference on Applied Natural Language Processing.* Trento, Italy.

Bunreacht na hÉireann - The Constitution of Ireland, 1937.

Bussmann, H., 1996. *Routledge Dictionary of Language and Linguistics.* London, New York: Routledge.

Campbell, G. L., 2000. Irish. *In: Compendium of the Worlds Languages.* (2nd. ed.). Routledge.

Central Statistics Office, 1998. *Census 1996 - Principal Socio-economic Results.* Dublin: Stationery Office.

Chomsky, N. and Halle, M., 1968. *The Sound Patterns of English.* New York: Harper and Row.

Christian Brothers, 1988. *New Irish Grammar.* Dublin: Fallons.

Clemenceau, D., 1997. Finite-State Morphology: Inflections and Derivations in a Single Framework Using Dictionaries and Rules. *In:* eds. E. Roche and Y. Schabes, *Finite-State Language Processing.* MIT Press. pp. 67-98.

Coates, R., 1994. Morphophonemics. *In:* ed. R. E. Asher, *The Encyclopedia of Language and Linguistics.* Pergamon. pp. 2602-2612.

Crystal, D., 1997a. *A Dictionary of Linguistics and Phonetics.* (4th. ed.) Blackwell.

Crystal, D., 1997b. *The Cambridge Encyclopedia of Language.* (2nd. ed.) Cambridge University Press.

Cutting, D., Kupiec, J., Pedersen, J. and Sibun, P., 1992. A practical part-of-speech tagger. *In:*

*Proceedings of Third Conference on Applied Natural Language Processing*. Trento, Italy.

Davies, J., 1993. *The Welsh Language*. University of Wales Press.

Davis, T., 1914. *In:* ed. A. Griffith, *Thomas Davis: the thinker and teacher*. Dublin: M. H. Gill and Son. pp. 54-5.

Dobrovolsky, M., 1997. Interfaces. *In:* eds. W. O' Grady, M. Dobrovolsky and F. Katamba, *Contemporary Linguistics - An Introduction*. London; New York: Longman. pp. 245-267.

Fife, J., 1993. Introduction. *In:* eds. M. J. Ball and J. Fife, *The Celtic Languages*. London; New York: Routledge. pp. 3-25.

Garside, R., Leech, G. and Samson, G., 1987. *The Computational Analysis of English: A Corpus-based Approach*. London: Longman.

Gazdar, G., 1985. Review article: Finite state morphology. *In: Linguistics,* 23 (4), 597-607.

Goldsmith, J., 2001. Unsupervised Learning of Morphology of a Natural Language. *In: Computational Linguistics,* 27 (2), 153-197.

Greene, B. and Rubin, G., 1971. *Automatic Grammatical tagging of English*. Dept of Linguistics, Brown University.

Hein, J. L., 1995. *Discrete Structures, Logic and Computability*. Sudbury, MA: Jones and Bartlett.

Hopcroft, J. E., Motwani, R. and Ullman, J. D., 2001. *Introduction to Automata Theory, Languages, and Computation*. (2nd. ed.) Reading, MA: Addison-Wesley.

Hughes, A. J., 2001. Advancing the Language: Irish in the Twenty-First Century. *In: New Hibernia Review / Iris Éireannacha Nua,* 5 (1), 101-126.

Illingworth, V., 1986. *Dictionary of Computing*. Oxford University Press.

Johnson, C. D., 1972. *Formal Aspects of Phonological Description*. Mouton.

Jurafsky, D. and Martin, J. H., 2000. *Speech and Language Processing*. Upper Saddle River, N.J: Prentice Hall.

Kaplan, R. M., 1997. Finite State Technology. *In:* eds. G. Varile and A. Zampolli, *Survey of the State of the Art in Human Language Technology*. Cambridge Unniversity Press. pp. 361-365.

Kaplan, R. M. and Kay, M., 1981. Phonological rules and finite-state transducers. *Paper presented at ACL/Linguistics Society of America*. New York.

Kaplan, R. M. and Kay, M., 1994. Regular models of phonological rule systems. *In: Computational Linguistics,* 20 (3), 331-378.

Karlsson, F., 1994. Computational Morphology. *In:* ed. R. E. Asher, *The Encyclopedia of Language and Linguistics*. Pergamon.

Karlsson, F. and Karttunen, L., 1997. Sub-Sentential Processing. *In:* eds. G. Varile and A. Zampolli, *Survey of the State of the Art in Human Language Technology*. Cambridge University Press. pp. 96-100.

Karttunen, L., 1983. KIMMO: a general morphological processor. *In: Texas Linguistic Forum,* 22, 165-186.

Karttunen, L., Kaplan, R. M. and Zaenen, A., 1992. Two-Level Morphology with Composition. *In: Proceedings of Coling-92: International Conference on Computational Linguistics*. Nantes, France.

Klavans, J., 1997. Computational Linguistics. *In:* eds. W. O' Grady, M. Dobrovolsky and F. Katamba, *Contemporary Linguistics - An Introduction.* London, New York: Longman. pp. 372-415.

Kleene, S. C., 1956. Representation of events in nerve nets and finite automata. *In:* eds. C. E. Shannon and J. McCarthy, *Automata Studies*. Princeton University Press. pp. 3-42.

Kornai, A., 1999. Introduction. *In:* ed. A. Kornai, *Extended Finite State Models of Language.* Cambridge University Press. pp. 1-5.

Koskenniemi, K., 1983. *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production.* PhD. thesis. University of Helsinki.

Koskenniemi, K., 1997. Representations and Finite-State Components in Natural Language. *In:* eds. E. Roche and Y. Schabes, *Finite-State Language Processing.* MIT Press. pp. 99-116.

Lewis, G., 2000. *Turkish Grammar.* Oxford University Press.

Mac Giolla Phádraig, B., 1963. *Réchúrsa Gramadal.* (3rd. ed.) Baile Átha Cliath: Longmans, Brún agus Ó Nualláin.

McGonagle, N., 1991. *Irish Grammar - A Basic Handbook.* Conamara: Clo Iar-Chonnachta.

Mohri, M., 1997. On the Use of Sequential Transducers in Natural Language Processing. *In:* eds. E. Roche and Y. Schabes, *Finite-State Language Processing.* MIT Press. pp. 353-382.

Murray, R. W., 1997. Historical Linguistics: the study of language change. *In:* eds. W. O' Grady, M. Dobrovolsky and F. Katamba, *Contemporary Linguistics - An Introduction.* London; New York: Longman. pp. 313-371.

Ó Baoill, D. and Ó Riagáin, P., 1990. Reform of the Orthography, Grammar and Vocabulary of Irish. *In:* eds. I. Fodor and C. Hagge, *Language Reform - History and Future. (Vol. V).* Hamburg: Helmut Buske. pp. 173-195.

Ó Baoill, D. P. and Ó Tuathail, É. 1992. *Úrchúrsa Gaeilge.* Baile Átha Cliath: Institiúid Teangeolaíochta Éireann.

Ó Cuív, B., 1987. Sandhi phenomena in Irish. *In:* ed. H. Andersen, *Sandhi Phenomena in the Languages of Europe.* Mouton de Gruyter. pp. 395-414.

Ó Dochartaigh, C., 1992. The Irish Language. *In:* ed. D. MacAulay, *The Celtic Languages.* Cambridge(England); New York: Cambridge University Press. pp. 11-99.

Ó Dónaill, N., 1977. *Foclóir Gaeilge-Béarla. [Irish-English Dictionary].* Baile Átha Cliath: An Gúm.

O' Grady, W. and de Guzman, V. P., 1997. Morphology: the analysis of word structure. *In:* eds. W. O' Grady, M. Dobrovolsky and F. Katamba, *Contemporary Linguistics - An Introduction.* London; New York: Longman. pp. 132-180.

Ó Searcaigh, C., 1997. *Out in the Open.* Indreabhán: Cló Iar-Chonnachta.

Ó Siadhail, M., 1988. *Learning Irish.* Yale University Press.

Ó Siadhail, M., 1989. *Modern Irish.* Cambridge University Press.

Palmer, L. R., 1972. *Descriptive and Comparative Linguistics: a critical introduction.* London: Faber.

Pei, M., 1966. *Glossary of Linguistic Terminology.* Columbia University Press.

Rannóg an Aistriúcháin, 1958. *Gramadach na Gaeilge agus Litriú na Gaeilge: An Caighdeán Oifigiúil. [Irish Grammar and Irish Spelling: The Official Standard].* Baile Átha Cliath: Oifig an tSoláthair.

Ritchie, G. D., Russell, G. J., Black, A. W. and Pulman, S. G., 1992. *Computational Morphology: Practical Mechanisms for the English Lexicon.* Cambridge, Mass.: MIT Press.

Roche, E. and Schabes, Y., 1997. Introduction. *In:* eds. E. Roche and Y. Schabes, *Finite-State Language Processing.* MIT Press.

Ruhlen, M., 1987. *A Guide to the World's Languages.* Stanford University Press.

Russell, P., 1995. *An Introduction to the Celtic Languages.* London; New York: Longman.

Schützenberger, M., 1961. A remark on finite transducers. *In: Information and Control,* 4, 185-196.

Sheremetyeva, S., Jin, W. and Nirenburg, S., 1998. Rapid Deployment Morphology. *In: Machine Translation,* 13, 239-268.

Sproat, R., 1992. *Morphology and Computation.* Cambridge, Mass: MIT Press.

Steinbergs, A., 1997. The classification of languages. *In:* eds. W. O' Grady, M. Dobrovolsky and F. Katamba, *Contemporary Linguistics - An Introduction.* London; New York: Longman. pp. 372-415.

Stenson, N., 1981. *Studies in Irish Syntax.* Tübingen: Gunter Narr.

Stephens, J., 1993. Breton. *In:* eds. M. J. Ball and J. Fife, *The Celtic Languages.* New York: Routledge. pp. 349-409.

Ternes, E., 1987. A grammatical hierarchy of joining. *In:* ed. H. Andersen, *Sandhi Phenomena in the Languages of Europe.* Mouton de Gruyter. pp. 11-21.

Ternes, E., 1992. The Breton Language. *In:* ed. D. MacAulay, *The Celtic Languages.* Cambridge(England); New York: Cambridge University Press. pp. 371-452.

Thomas, A. R., 1992. The Welsh Language. *In:* ed. D. MacAulay, *The Celtic Languages.* Cambridge(England); New York: Cambridge University Press. pp. 251-345.

Turing, A., 1936. On computable numbers, with an application to the Entscheidungsproblem. *In: Proc. London Math. Soc,* 2 (42), 230-265.

Uszkoreit, H., 1997. Mathematical Methods - Overview. *In:* eds. G. Varile and A. Zampolli, *Survey of the State of the Art in Human Language Technology.* Cambridge Unniversity Press. pp. 337-342.

**Software**

Xerox Finite-State Tools (tools: lexc, xfst, twolc; operating system: Linux/Solaris). For details contact: Xerox Research Centre Europe, Attn: Licensing of Finite-State Programming Languages, 6 chemin de Maupertuis, 38240 Meylan, France. See also: http://www.xrce.xerox.com (Accessed 1/7/2002)

WordSmith Tools (tools: Concord, Wordlist; operating system: Windows) For details see: http://www.oup.com/elt/global/isbn/6890/ (Accessed 1/7/2002)

# Index

[1] Although spelled differently these forms sound the same.

[2] There are some defective lexemes which lack certain forms of the general paradigm, e.g. the defective verb *inquit* in Latin (Sproat, 1992, p24)

[3] He *hoovered* the flat, *jeyes-fluided* the bins, *harpicked* the loo, *vimmed* the bath. Then he *flashed* the floor, *windowlened* the windows and *eau-de-cologned* the beds.

[4] Sandhi comes from Sanskrit meaning 'putting together' or 'joining' (Ternes, 1987).

[5] Ten percent of the population are recorded as using the language on a daily basis in the 1996 census. By excluding the school-going population (those over 3 years of age and under nineteen) on the pessimistic basis that their use of the language is confined to the classroom only, the figure is two per cent. The exact percentage lies somewhere in between.

[6] 108 Irish medium schools (7507 students) in Gaeltacht regions and 112 Irish medium schools (19,491 students) outside of Gaeltacht regions (An Roinn Oideachais agus Eolaíochta, 2001, p27). These number 220 of the 3172 primary schools in the country.

[7] These verbal particles can be thought of as verb clitics since they never appear independently of the verb and are not stressesd (Stenson, 1981, p32; O' Grady and de Guzman, 1996, p140).

[8] Some new formations of strong plurals in Munster dialect are based on dative formations, e.g. *na fearaibh* 'the men' rather than *na fir* 'the men'.

[9] An adjective agrees with the gender, number and case of the noun it modifies, but "this is done by means of very few changes in form, and not for all adjectives" (Stenson 1981, p30). Stenson also cites Green (1966, p35) as saying that adjective inflection appears to be on the way out (1981, p30 & p161).

[10] According to Baayen (2001, p1-12), relative and mean frequencies change systematically as the sample size increases, due to the fact that a) words do not occur randomly in texts, and b) lexical frequency distributions contain large numbers of low-probability words.

# Appendix A

## Morphological Feature Tags

The following tables contain the morphosyntactic tags used in the Irish two-level morphology.

**Table 0. General Morphosyntactic Tags**

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +CM | canúint na Mumhan, Munster dialect |
| +CC | canúint Chonnachta, Connaught dialect |
| +CD | canúint Dhún na nGall, Donegal dialect |

**Table 1. Noun Morphosyntactic Tags**

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Noun | noun |
| +Fem | feminine |
| +Masc | masculine |
| +Com | common case (nominative/accusative/most datives) |
| +Gen | genitive case |
| +Voc | vocative case |
| +Dat | dative case (in irregular noun lexicon) |
| +Prop | proper noun |
| +Sg | singular in number |
| +Pl | plural in number |
| +Def | definite noun e.g.. preceded by an article |
| +Idf | indefinite noun e.g. not preceded by an article |
| +Strong | strong plural (same plural for common, gen. and voc. cases) |
| +Weak | weak plural (different com, gen, voc plurals) |
| +Emph | emphasis: *ár dteach**sa*** 'our house', *ár bpáirc**se*** 'our field' |
| +Sé | lenition after simple prep. eg *ar t**h**ír* 'on land' |
| +Urú | eclipsis after compound prep eg *ar an **g**cat* 'on the cat' |
| +Subst | substantives, words acting like nouns but not declined |
| +Poss | possessive form with vowels, e.g. *m'aois* 'my age', *d'aois* 'your age' |

**Table 2. Verb Morphosyntactic Tags**

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Verb | verb |
| +1P +2P +3P | first, second and third person |
| +Auto | autonomous form |
| +Sg +Pl | singular and plural |
| +PresInd | present Indicative |
| +PastInd | past Indicative |
| +PastIndDep | past Indicative dependant form (irregular verbs) |
| +FutInd | future indicative |
| +ImpInd | imperfect indicative |
| +Cond | conditional |
| +PresSubj | present subjunctive |
| +PastSubj | past subjunctive |
| +Imper | imperative |
| +VerbalNoun | verbal noun |
| +VerbalAdj | verbal adjective |
| +Neg | negative form |
| +Cop | copula |
| +Q | interrogative form |
| +NegQ | negative interrogative form |
| +Rel | relative |

**Table 3. Adjective Morphosyntactic Tags**

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Adj | adjective |
| +Pos | positive |
| +Comp | comparative |
| +Sup | superlative |
| +Masc | masculine gender |
| +Fem | feminine gender |
| +Com | common case |
| +Gen | genitive case |
| +Voc | vocative case |
| +Sg | singular |
| +Pl | plural |
| +Strong | an adj. qualifying a strong plural noun will also have the same plural form in all cases |
| +Weak | an adj. qualifying  a weak plural noun, in the gen.case, is not inflected |
| +Slender | adj qualifying a plural noun ending in a slender consonant |
| +NotSlen | adj. qualifying a plural noun ending in a broad consonant or a vowel |
| +LenYES +LenNO | a masc noun after prepositions (e.g. *ag an* 'at the', *ar an* 'on the', *as an* 'out of the' etc.), is either lenited or eclipsed according to preference/dialect. If it is lenited then the adj is likewise lenited. If it is eclipsed then the adj has no initial mutation (New Irish Grammar, Christian Brothers, 1988). |
| +Cop | with copula, e.g. *b'fhearr* 'would prefer' |

Table 4. Pronoun Morphosyntactic Tags

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Pron | pronoun |
| +Pers | personal |
| +Emph | emphatic (contrastive) form of personal pronoun |
| +Ref | reflexive |
| +Idf | indefinite |
| +1P +2P +3P | first, second or third person |
| +Fem | feminine gender |
| +Masc | masculine gender |
| +Sg +Pl | singular or plural in number |
| +VerbSubj | pronoun as verb subject, e.g. *Chuaigh sí amach* 'She went out' |

Table 5. Determiner Morphosyntactic Tags

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Det | determiner |
| +Dem | demonstrative |
| +Poss | possessive |
| +Q | interrogative |
| +1P +2P +3P | first, sceond or third person |
| +Fem | feminine gender |
| +Masc | masculine gender |
| +Sg +Pl | singular or plural in number |

Table 6. Article Morphosyntactic Tags

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Art | article |
| +Def | definite |
| +Masc | masculine gender |
| +Fem | feminine gender |
| +Com | common case |
| +Gen | genitive case |
| +Sg | singular |
| +Pl | plural |

Table 7. Adverb Morphosyntactic Tags

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Adv | adverb |
| +Gen | general, e.g. *go tapaidh*, quickly |
| +Deg | degree, e.g. *sách tapaidh*, 'fairly quickly' |
| +Phr | phrasal, e.g. *dul amach* 'going out' |
| +Q | interrogative, e.g. *cá bhfuil sé* 'where is it/he' |
| +Rel | relative, e.g. *nuair a tharla sé* 'when it happened' |

Table 8. Preposition Morphosyntactic Tags

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Prep | preposition |
| +Simp | simple |
| +Comp | compound |
| +Emph | emphatic form of prep pronoun |
| +Cpx | complex compound |
| +Rel | complex relative compound |
| +Poss | possessive, e.g. *ina* 'in his' *inár* 'in our' |
| +1P +2P +3P | first, sceond or third person |
| +Fem | feminine gender |
| +Masc | masculine gender |
| +Sg +Pl | singular or plural in number |

Table 9. Conjunction Morphosyntactic Tags

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Conj | conjunction |
| +Coord | co-ordinate |
| +Subord | subordinate |

**Table 10. Numeral Morphosyntactic Tags**

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Num | numeral |
| +Card | cardinal, e.g. *aon dó trí...* 'one, two three ...' |
| +Ord | ordinal, e.g. *céad dara tríú...* 'first, second, third ...' |
| +Pers | personal, e.g. *duine, beirt, triúr…* 'one person, two people, three people ...' |
| +Adj | adjectival, e.g. *mo* ***dhá*** *lámh* 'my two hands', *an* ***chéad dhá*** *theach…*'the first two houses' |

**Table 11. Interjection Morphosyntactic Tags**

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Itj | interjection, e.g. á 'aah', *faraor* 'unfortunately' |

**Table 12. Particle Morphosyntactic Tags**

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Part | particle |
| +Verb | verbal particle |
| +Voc | vocative particle, e.g. ***a*** *Mháire* 'Mary!' |
| +Neg | negative, e.g. ***ní*** *raibh* 'was not' |
| +Cont | continuative, e.g. ***ag*** *rith* 'running' |
| +ConP | continuative Passive, e.g. ***á*** *dhéanamh* 'being done' |
| +Q | interrogative verbal particle, e.g. ***an*** *raibh* 'was?' |
| +Adv | adverbial, e.g. ***go*** *holc* ' badly' |
| +Num | numeral, e.g. ***a*** *haon* 'one' |
| +Deg | degree, e.g. ***níos*** *fearr* 'better' |
| +Pat | patronym, e.g. ***Ó*** *Beirn*, ***Ní*** *Bheirn*, ***Uí*** Bheirn |
| +DV | defective Verb, e.g. ***arsa*** *Seán* 'said Seán' |
| +Inf | infinitive, e.g. *litir* ***a*** *scríobh* 'to write a letter' ***le*** *feiceáil* 'to be seen' |
| +Rel | relative, e.g. *an lá* ***ar*** *tháinig sé* 'the day that he came' |
| +Subj | subjunctive, e.g. ***go*** *raibh maith agat* 'thank you' |
| +Imp | imperative, e.g. ***ná*** *déan*, 'don't do it' |
| +Conj | conjunction, e.g. ***sular*** *tháinig sé* 'before he came' |
| +Prep | preposition with verbal noun phrase |
| +PastIrreg | irregular past tense verbal particle, e.g. ***an*** *raibh sé* 'was he?'' |
| +PastReg | regular past tense verbal particle, e.g. ***ar*** *chuala sé* 'did he hear' |
| +Fut | future tense, e.g. ***an*** *mbeidh tú ann*? 'will you be there?' |
| +Pres | present tense, e.g. ***an*** *bhfuil tú ann?* 'are you there' |
| +Cond | conditional, e.g. ***má*** *bhíonn tú ann* 'if you would be there' |

**Table 13. Abbreviation Morphosyntactic Tags**

| *Morphosyntactic Tag* | *Description* |
|---|---|
| +Abr | abbreviation, e.g. ***lch.*** *(leathanach)* 'page' |

# Appendix B

## *Morphological Tag Grammar*

The following is the lexical tag grammar for the morphological transducer.

```
[á|é|í|ó|ú|a|b|c|d|e|f|g|h|i|j|k|l|m|n|o|p|q|r|s|t|u|v|w|x|y|
 z|%^X|%]+
(%+CC|%+CD|%+CM)
[
  (%+Prop)
  %+Noun [%+Fem|%+Masc]
          [[[%+Com|%+Dat] [%+Sg|%+Pl|%+Num] [%+Def|%+Idf] ]|
           [%+Gen [%+Sg | [[%+Strong|%+Weak] %+Pl] ] [%+Def|%+Idf] ]
|
         [%+Voc [%+Sg|%+Pl] (%+Def)]
          ]
        (%+Sé|%+Urú)
        (%+Emph) (%+Poss)
| %+Subst (%+Len|%+Urú)

|
  %+Adj [[%+Fem  [%+Com|%+Gen|%+Voc] %+Sg] |
         [%+Masc [%+Gen|%+Voc] %+Sg] |
         [%+Masc %+Com %+Sg [%+LenYES|%+LenNO]]|
         [%+Gen [%+Strong|%+Weak] %+Pl] |
             [%+Com [%+Slender|%+NotSlen] %+Pl]|
             [%+Voc %+Pl]|
         [%+Pos|%+Comp|%+Sup]
         ]
|
  %+Art [%+Gen|%+Com] [%+Sg|%+Pl] %+Def (%+Fem|%+Masc)
|
  %+Verb
[%+FutInd|%+ImpInd|%+PastInd|%+PastIndDep|%+PresInd|%+PresSubj|%+Past
Subj
         |%+Imper|%+Cond]  ([[%+1P|%+2P|%+3P] [%+Sg|%+Pl]] | %+Auto)
(%+Neg|%+Q|%+NegQ)

| %+Cop [%+Pres|%+Past|%+Cond] (%+Neg|%+Q|%+NegQ|%+Rel) (%+Pron)(%+V)
| %+VerbalAdj
| %+VerbalNoun

| %+Prep [ %+Simp|
          [%+Pron [%+1P|%+2P|%+3P] [%+Sg|%+Pl] (%+Fem|%+Masc)]|
        [%+Cpx (%+Poss|%+Rel)] |
           %+Comp
         ]
         (%+Emph)

| %+Pron [[%+Pers [%+1P|%+2P|%+3P] [%+Sg|%+Pl] (%+Fem|%+Masc)
(%+VerbSubj) (%+Emph)] |
          [%+Idf|%+Ref]
         ]

| %+Adv [%+Gen | %+Deg | %+Phr | %+Q | %+Rel | %+Dir]
| %+Part
        [%+Pat|%+Verb| %+Adv|%+Voc| %+Deg|%+DV|%+Num|%+Prep|%+Cont]
      (%+Comp | %+Sup)
        (%+Neg) (%+Q)
        (%+PastIrreg|%+Conj|%+Fut|%+Pres |%+Cond|%+PastReg|%+Subj |
%+Imp| %+Rel)

| %+Num [%+Card | %+Ord | %+Adj]

| %+Conj [%+Coord|%+Subord]

| %+Det [[%+Dem|%+Q] |
         [%+Poss [%+1P|%+2P|%+3P] [%+Sg|%+Pl] (%+Fem|%+Masc)]
        ]

| %+Itj

| %+Abr
];
```

# Appendix C

## *Parole Morphosyntactical Tagset*

The PAROLE tagset was used for part-of-speech tagging of harmonised corpora in the EU funded LE-PAROLE Project (1996-1998). 14 European languages are tagged with this common tagset (Belgian and Swiss French, Catalan, French, Danish, English, French, Finnish, German, Greek, Irish, Italian, Portuguese, Spanish and Swedish).

Some additional codes (specific to Irish) which were added to the tagset during manual checking are underlined. Some language specific features were also added during checking and these are marked with an asterisk.

# PAROLE COMMON MORPHOSYNTACTICAL TAGSET

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 1. NOUN | | | | | | | | |
| 1. | 2. Type | 3. Gender | 4. Number | 5. Case | 6. Sem-Gender | *7. Contrast | *8. Usage | |
| N | c = common | f = fem | s = sing. | n = nom. | n/a | n = unmked | v = as verb | |
| | p = proper | m = mas | p = pl. | g = gen. | | y = marked | c = as conj. | |
| | s = substantive | | | v = voc. | | | a = as adverb | |
| | | | | d = dative | | | | |
| 2. VERB | | | | | | | | |
| 1. | 2. Type | 3. Mood | 4. Tense | 5. Person | 6. Number | 7. Gender | *8. Dependency | *9. Neg |
| V | m = main | i = indic. | p = pres. | 1 = first | s = sing | n/a | i = independant | (Tá & Is ) |
| | t = tá | s = subj. | s = past | 2 = sec. | p = pl. | | d = dependant | n = neg |
| | i = is | m = imper | h = past hab | 3 = third | | | | a = affirm. |
| | | c = cond. | f = future | 0 = free | | | | |
| | | n = infinitive | g = pres. hab | | | | | |
| | | p = participle | | | | | | |
| | | a = adjectival | | | | | | |
| | | r = relative | | | | | | |
| 3. ADJECTIVE | | | | | | | | |
| 1. | 2. Type | 3. Degree | 4. Gender | 5. Number | 6. Case | | | |
| A | q = qualificator | p = positive | f = fem. | s = sing | n = nom. | | | |
| | i = with verb 'is' | c = comparative | m = masc. | p = pl. | g = gen. | | | |
| | | s = superlative | | | v = voc. | | | |
| | | d = degree | | | | | | |
| 4. PRONOUN | | | | | | | | |
| 1. | 2. Type | 3. Person | 4. Gender | 5. Number | 6. Case | 7. Posessor | | |
| P | p = personal | 1 = first | f = fem. | s = sing. | n/a | n/a | | |
| | c = contrastive | 2 = sec. | m = masc. | p = pl. | | | | |

| | | |
|---|---|---|
| | x = reflexive | 3 = third |
| | o=indefinite | 0 = null |
| 5. DETERMINER | | |
| **1.** | **2. Type** | **3. Person** |
| D | d = demonstrative | 1 = first |
| | p = possessive | 2 = sec. |
| | q = interrogative | 3 = third |
| 6. ARTICLE | | |
| **1.** | **2. Type** | **3. Gender** |
| T | d = definite | f = fem. |
| | | m = masc. |
| 7. ADVERB | | |
| **1.** | **2. Type** | **3. Degree** |
| R | g = general | b = base |
| | d= degree | c = comparative |
| | p = phrasal | s = superlative |
| | q = interrogative | |
| | r = relative | |
| 8. ADPOSITION | | |
| **1.** | **2. Type** | **3. Formation** |
| S | p = preposition | s = simple |
| | | c = compound |
| | | e = emphat. comp. |
| | | x = complex comp. |
| | | r = cmplx. rel. cmp. |

| **4. Gender** | **5. Number** | **6. Case** | **7. Posessor** |
|---|---|---|---|
| f = fem. | s = sing | n/a | n/a |
| m = masc. | p = pl. | | |

| **4. Number** | **5. Case** |
|---|---|
| s = sing | n = nom. |
| p = pl. | g = gen. |

| **4. Function** | **5. Wh-ness** |
|---|---|
| m = mod? | n/a |
| s = spe? | |
| v = verbal | |

| **4. Gender** | **5. Number** | _*6 Person_ |
|---|---|---|
| f = fem. | s = sing | 0 = null |
| m = masc. | p = pl. | 1 = first |
| n = null | n = null | 2 = sec. |
| | | 3 = third |

**9. CONJUNCTION**

| **1. POS** | **2. Type** |
|---|---|
| C | c = coordinate |
| | s = subordinative |

**10. NUMERALS**

| **1. POS** | **2. Type** |
|---|---|
| M | c = cardinal |
| | o = ordinal |
| | p = personal |

**11. INTERJECTION**

| **1. POS** |
|---|
| I |

**12. UNIQUE MEMBERSHIP CLASS**

| **1. POS** | **2. Particle Type** |
|---|---|
| U | i = infinitive |
| | n = negative |
| | c = continuative |
| | c(p)= con. passive |
| | q = interrogative |
| | a = adverbial |
| | r = relative |
| | |
| | b = defective verb |

| **3. Ctype** | **4. Coord-pos** |
|---|---|
| n/a | n/a |

| **3. Gender** | **4. Number** | **5. Case** |
|---|---|---|
| n/a | n/a | n/a |

**3. B-Function**

q = quotative
i = impressionistic
a = almost

| | |
|---|---|
| | v = vocative |
| | m = numeral |
| | d = degree |
| | p = patronym |
| | o = other |

**13. RESIDUALS**

| **1. PoS** | **2 Type** |
|---|---|
| X | f = foreign |
| | s = symbol |
| | t = toponym |
| | a = acronym |
| | b = abbreviation |
| | n = number |

**14. PUNCTUATION**

| **1. PoS** | 2. Type |
|---|---|
| F | e = sentence final |
| | i = sen. Internal |
| | a = quote/par init. |
| | z = quote/par fin. |
| | b = hyphen/ underscore/ dash |
| | u = ! |
| | q = ? |
| | x = apostrophe |

**15. ABBREVIATION**

| **1. PoS** |
|---|
| Y |

# Appendix D

## Noun Plural Formation

**Table 1. Nouns – Plurals formed by suffixation**

| *Suffix* | *Singular* | *Plural* | *Lexicon class* |
|---|---|---|---|
| -a | *bróg* 'shoe' | *bróga* 'shoes' | PL-(LEA)A |
| -acha | *craobh* 'branch' | *craobhacha* 'branches' | PL-(E)ACHA |
| -aí | *cumhacht* 'power' | *cumhachtaí* 'powers' | PL-AÍ |
| | *rud* 'thing' | *rudaí* 'things' | |
| -anna | *bus* 'bus' | *busanna* 'buses' | PL-(E)ANNA |
| -e | *súil* 'eye' | *súile* 'eyes' | PL-(CAOL)E |
| -eacha | *cathaoir* 'chair' | *cathaoireacha* 'chairs' | PL-(E)ACHA |
| -eanna | *áit* 'place' | *áiteanna* 'places' | PL-(E)ANNA |
| -í | *cailín* 'girl' | *cailíní* 'girls' | PL-Í |
| -idí | *caoga* 'fifty' | *caogaidí* 'fifties' | PL-IDÍ |
| -na | *comharsa* 'neighbour' | *comharsana* 'neighbours' | PL-NA |
| -nna | *cnó* 'nut' | *cnónna* 'nuts' | PL-NNA |
| -ta | *grian* 'sun' | *grianta* 'suns' | PL-TA |
| -te | *coill* 'wood' | *coillte* 'woods' | PL-TE |
| -tha | *glór* 'voice' | *glórtha* 'voices' | PL-(LEA)THA |
| -the | *rí* 'king' | *ríthe* 'kings' | PL-THE |

**Table 2. Nouns: Plurals formed by slenderisation**

| *Process* | *Singular* | *Plural* | *Lexicon class* |
|---|---|---|---|
| slenderise | *cat* 'cat' | *cait* 'cats' | PL-CAOLÚ |
| | *inneall* 'engine' | *innill* 'engines' | |

Table 3. Nouns – plurals formed by broadening and suffixation

| *Process* | *Singular* | *Plural* | *Lexicon class* |
|---|---|---|---|
| suffix -a and broaden if slender | *deoir* 'drop '<br>*binn* 'peak' | *deora* 'drops '<br>*beanna* 'peaks' | PL-(LEA)A |
| | *tír* 'country' | *tíortha* 'countries' | PL-(LEA)THA |

Table 4. Nouns – plural formed by syncopation and suffixation

| *Process* | *Singular* | *Plural* | *Lexicon class* |
|---|---|---|---|
| suffix -acha and syncopate | *cathair* 'city' | *cathracha* 'cities' | PL-(E)ACHA |

Table 5. Nouns – plurals formed by syncopation, slenderisation and suffixation

| *Process* | *Singular* | *Plural* | *Lexicon class* |
|---|---|---|---|
| suffix -e syncopate slenderise | *doras* 'door' | *doirse* 'doors' | PL-(CAOL)E |
| | *tobar* 'well' | *toibreacha* 'wells' | PL-(CAOL)EACHA |

**Table 6. Nouns – plural by replacing stem ending**

| *Process* | *Singular* | *Plural* | *Lexicon class* |
|---|---|---|---|
| replace -ach with –aí | *bealach* 'way' | *bealaí* 'ways' | PL-ATHRÚ |
| replace -each with –í | *soitheach* 'vessel' | *soithí* 'vessels' | PL-ATHRÚ |
| replace -adh with –aí | *samhradh* 'summer' | *samhraí* 'summers' | PL-ATHRÚ |
| replace –eadh with -í | *geimhreadh* 'winter' | *geimhrí* 'winters' | PL-ATHRÚ |

**Table 7. Nouns – plural formed by suffixation and replacement**

| *Process* | *Singular* | *Plural* | *Lexicon class* |
|---|---|---|---|
| suffix -the and replace -í with -i | *ainmhí* 'animal' | *ainmhithe* 'animals' | PL-THE |

**Table 8. Nouns – plural formed by suppletion**

| *Comment* | *Singular* | *Plural* | *Lexicon class* |
|---|---|---|---|
| suppletion (phonologically unrelated form) | *bean* 'woman' | *mná* 'women' | irregular |

# Appendix E

## Replace Rule Triggers

The following inflectional mark-up tags are appended to surface level strings in the lexicon.

**Table 1. Verb Replace Rule Triggers**

| *Inflectional Mark-up Tag*<br>*Surface (lower) level* | *Description and usage* |
|---|---|
| ^Verb | **verb** marker |
| ^Sé | **séimhiú** (lit. softening) 'lenition' - h added after certain initial consonants i.e. b c d f g m p s t |
| ^Urú | **urú** 'eclipsis': a consonant placed before word initial letter (b c d f g p t) e.g.'m' before 'b' - gen. pl. of *balla* 'wall' is *mballaí* 'of the walls' |
| ^igh | **remove** –igh ending from root form (slender roots) |
| ^aigh | **remove** –aigh ending from root form (broad roots) |
| ^Fr | **root form** is used, i.e. no ending is removed |
| ^Caol | **caolú** 'slenderise' (also called attenuation) usually inserts '*i*' after a final broad vowel (a o u á ó ú are broad vowels) |
| ^Lea | **leathnú** 'broadening' usually deletes final '*i*' |
| ^Coim | **coimriú** 'syncopation': the final unstressed vowel(s) is/are dropped |
| ^Ath | **athrú** 'change': in certain plurals the ending changes |
| ^LC | **leathan-caol**: check harmony of broad and slender vowels |
| ^VA | **verbal adjective** marker |
| ^VN | **verbal noun** marker |

**Table 2. Noun and Adjective Replace Rule Triggers**

| *Inflectional Mark-up Tag Surface (lower) level* | *Description and usage* |
|---|---|
| ^M / ^F | **masculine / feminine gender**<br>initial mutation of singular nouns is gender specific |
| ^C / ^G / ^V | **common / genitive / vocative**:<br>initial mutation of plural nouns is case specific |
| ^Sé | **séimhiú** (lit. softening) 'lenition' - h added after certain initial consonants i.e. b c d f g m p s t |
| ^Urú | **urú** 'eclipsis': a consonant placed before initial letter (b c d f g p t) e.g.'m' before 'b' - gen. pl. of *balla* 'wall' is *mballaí* 'of the walls' |
| ^tv | **'t-' before a vowel** e.g. éan 'bird' masc. sg. With def. art.: an t-éan 'the bird' |
| ^hv | **'h' before a vowel** e.g. *éan*: masc. pl. *na héin* 'the birds' |
| ^ts | **'t' before 's'** e.g. *sagart* 'priest' masc. sg. gen *teach an tsagairt* – 'the priest's house' |
| ^Caol | **caolú** 'slenderise' (also called attenuation) usually inserts '*i*' after a final broad vowel (a o u á ó ú are broad vowels) e.g. *an cat* 'the cat' in pl. form becomes *na cait* 'the cats' |
| ^Lea | **leathnú** 'broadening' usually deletes final '*i*' e.g. *súil* 'eye' in gen. becomes *radharc na súl* 'eyesight' (sight of the eyes) |
| ^Coim | **coimriú** 'syncopation': the final unstressed vowel(s) is/are dropped e.g. *saghas* 'type' becomes *saghs+anna* 'types' |
| ^Ath | **athrú** 'change': in certain plurals the ending changes e.g. *gealach* 'moon' in gen. becomes *gealaí* 'of the moon' |
| ^LC | **leathan-caol**: check harmony of broad and slender vowels |
| ^Emph | **empatic forms:** mo theach**sa** 'my house', do mháthair**se** 'your mother' |
| ^X | **syllable boundary**: this tag is inserted before the last vowel (or vowels) in stem, i.e. it marks the boundary between between the onset and peak of the final syllable.<br>e.g. *leabh^Xar* 'book' |
| ^AE / ^AO | **long vowels** 'ae' and 'ao' are collapsed into one portmanteau (multi-character) symbol e.g. the slender form of *–ae-* is *–aei-* not *–aiei-*. |
| ^IA / ^UA | **diphthongs** 'ia' and 'ua' are similarly collapsed into one symbol |
| ^Poss | **possessive marker** on vowel-initial nouns such as *m'aois* 'my age' |

# Appendix F

## Noun Lexicon Classes

- Noun continuation classes for the 5 declensions
- *Final Syllable Changes*
- Plural Continuation Classes

**Table 1.1. First Declension Stems: Weak Plurals (all masc.)**

Genitive Singular: Slenderise

| *Weak Plural Type* | | | *Continuation Class* | *Example* |
|---|---|---|---|---|
| *Common* | *Genitive* | *Vocative* | | |
| Slenderise | No change | +a | Nm1-1 | *cat: cait, cat, a chata* |
| +a | No change | +a | Nm1-2 | *úll: úlla, úll, a úlla* |

**Table 1.2. First Declension Stems: Strong Plurals (all masc.)**

Genitive Singular: Slenderise

| *Strong Plural Types* | *Continuation Class* | *Example* |
|---|---|---|
| +ta | Nm1-3 | *gaol: gaolta* |
| +tha | Nm1-4 | *glór: glórtha* |
| Change final syllable | Nm1-5 | *bealach: bealaí* |
| +anna | Nm1-6 | *carr: carranna* |
| +í | Nm1-7 | *cogadh: cogaí* |
| Syncopate +anna | Nm1-8 | *saghas: saghsanna* |
| Syncopate, slenderise +e | Nm1-9 | *bóthar: bóithre* |
| Syncopate, slenderise +eacha | Nm1-10 | *tobar: toibreacha* |
| +(e)acha | Nm1-11 | *cineál: cineálacha* |

**Table 2.1. Second Declension Stems – Weak Plurals**

Genitive Singular: Slenderise if necessary, suffix –e

| *Weak Plural Type* | | | *Continuation Class* | *Example* |
|---|---|---|---|---|
| *Common Pl* | *Gen Pl* | *Voc Pl* | | |
| Feminine | | | | |
| +a | No change | +a | Nf2-1 | *bróg: bróga, bróg, a bhróga* |
| Broaden +a | Broaden | Broaden +a | Nf2-2 | *deor: deora, deor, a dheora* |
| Slenderise +e | Broaden | Slenderise +e | Nf2-3 | *súil: súile, súl, a shúile* |
| Broaden +a | No change | Broaden +a | Nf2-4 | *gealach: gealacha* |

**Table 2.2. Second Declension Stems – Strong Plurals**

Genitive Singular: Slenderise if necessary, suffix -e

| *Strong Plural Types* | *Continuation Class* | *Example* |
|---|---|---|
| Feminine | | |
| +te | Nf2-5 | *coill: coillte* |
| +(e)anna | Nf2-6 | *fadhb: fadhbanna, caint : cainteanna* |
| +í | Nf2-7 | *eaglais: eaglaisí* |
| +(e)acha | Nf2-8 | *craobh: craobhacha, carraig : carraigeacha* |
| Broaden +tha | Nf2-9 | *spéir: spéartha* |
| +ta | Nf2-10 | *grian: grianta* |
| +(e)anta | Nf2-11 | *uair: uaireanta* |
| Masculine | | |
| +te | Nm2-1 | *sliabh: sléibhte* |

**Table 3. Third Declension stems – All Strong Plurals**

Genitive Singular: Broaden if necessary, suffix -a

| *Plural Type* | *Continuation Class* | *Example* |
|---|---|---|
| Feminine | | |
| +aí | Nf3-1 | *scoláireacht: scoláireacht**aí*** |
| Broaden, change +aí | Nf3-2 | *buairt: buart**haí*** |
| Broaden +anna | Nf3-3 | *cuid: cod**anna*** |
| +í | Nf3-4 | *tagairt: tagairt**í*** |
| Masculine | | |
| +í | Nm3-1 | *bádóir: bádóir**í*** |
| Broaden +anna | Nm3-2 | *am: am**anna*** |
| +aí | Nm3-3 | *rás: rás**aí*** |
| Broaden +anna | Nm3-4 | *droim: drom**anna*** |
| +acha | Nm3-5 | *anam: anam**acha*** |
| Broaden +a | Nm3-6 | *flaith: flath**a*** |
| +ta | Nm3-7 | *gleann: gleann**ta*** |

**Table 4. Fourth Declension stems – All Strong Plurals**

Genitive Singular: No change

| *Plural Type* | *Continuation Class* | *Example* |
|---|---|---|
| Feminine | | |
| +í | Nf4-1 | *bearna: bearna**í*** |
| Change | Nf4-2 | *aicme: aicm**í*** |
| Change | Nf4-3 | *féile: féil**te*** |
| Change, +the | Nf4-4 | *bé: bé**ithe*** |
| +(e)acha | Nf4-5 | *céadaoin: céadaoin**eacha*** |
| +(e) anna | Nf4-5 | *gé: gé**anna*** |
| +(e) anta | Nf4-6 | *oíche: oíche**anta*** |
| Masculine | | |
| +í | Nm4-1 | *coinín: coinín**í*** |
| Change | Nm4-2 | *ailtire: ailtir**í*** |
| Change | Nm4-3 | *baile: bail**te*** |
| Change, +the | Nm4-4 | *ainmhí: ainmh**ithe**, croí: croí**the*** |
| +nna | Nm4-6 | *tae: tae**nna*** |
| +(e)anna | Nm4-7 | *bus: bus**anna*** |

**Table 5. Fifth Declension stems**

Genitive Singular: (1) Broaden +suffix, (2) Syncopate +suffix or (3) Append suffix
Note: irregular nouns are included in this declension where possible)

| *Plural Type* | *Continuation Class* | *Example* |
|---|---|---|
| Feminine | | |
| Broaden +acha | Nf5-1 | *beoir: beor**acha*** |
| Syncopate +(e)acha | Nf5-2 | *cathair: cathr**acha*** |
| +na | Nf5-3 | *comharsa: comharsa**na*** |
| Masculine | | |
| +idí | Nm5-1 | *fiche: fich**idí*** |
| Syncopate +(e)acha | Nm5-2 | *athair: aithr**eacha*** |
| Change +the | Nm5-3 | *athrú: athr**uithe**, cónal: cón**aithe*** |

**Table 6. Final Syllable Changes**

| *Slender Vowel(s)* | *Broadened Vowel(s)* | *Example* |
|---|---|---|
| é | é i | *finné: finn**éi**+the* |
| í | i | *ainmh**í** : ainmh**i**+the* |
| ch | í | *beala**ch**: beala**í*** |
| dh | í | *marga**dh**: marga**í*** |
| eadh | í | *geimhr**eadh**: geimhr**í*** |
| each | í | *soith**each**: soith**í*** |
| ba | p | *lea**ba**: lea**p**+acha* |
| och | - | *buí**och**: buí+thí* |
| ch | th | *gná**ch**: gná**th**+aí* |
| t | - | *tiomáin**t**: tiomán+aí* |

Table 7. Noun-Plural Continuation Classes

| *Lexicon Class* | *Description* | *Suffix* | *Process* | *Example* |
|---|---|---|---|---|
| PL-AÍ | Add suffix ‘aí’ | -aí | | |
| PL-ATHRÚ | Change (replace) last syllable | | change | |
| PL-CAOL)E | Add suffix ‘e’ and slenderise stem if broad | -e | (slenderise) | |
| PL-(CAOL)EACHA | Add suffix ‘eacha’ and slenderise stem if broad | -eacha | (slenderise) | |
| PL-(CAOLÚ | Slenderise stem | | slenderise | |
| PL-(E)ACHA | Add suffix ‘acha’ if stem is broad OR add suffix 'eacha' if stem is slender | -acha<br>-eacha | | |
| PL-(E)ANNA | Add suffix ‘anna’ (broad) OR ‘eanna’ (slender) | -anna<br>-eanna | | |
| PL-Í | Add suffix ‘í’ | -í | | |
| PL-IDÍ | Add suffix ‘idí’ | -idí | | |
| PL-(LEA)A | Add suffix ‘a’ and broaden stem if slender | -a | (broaden) | |
| PL-(LEA)THA | Add suffix ‘tha’ and broaden stem if slender | -tha | (broaden) | |
| PL-LEATHNÚ | Broaden stem | | broaden | |
| PL-NA | Add suffix ‘na’ | -na | | |
| PL-NNA | Add suffix ‘nna’ | -nna | | |
| PL-TA | Add suffix ‘ta | -ta | | |
| PL-TADA | No change | | | |
| PL-TE | Add suffix ‘te’ | -te | | |
| PL-THE | Add suffix ‘the’ | -the | | |

# Appendix G

## Verb Lexicon Classes

**Table 1. 1st Conjugation Verbs**

| *Lexicon Class* | *Description* | *Example* | *Action* |
|---|---|---|---|
| V1-BR | Broad | *mol* 'praise' | append broad suffixes |
| V1-BR-LV | Broad stems with long vowel ending in *-igh* | *cráigh*, dóigh,<br>*buaigh* 'win' | remove *–igh*<br>& append broad f-suffixes and slender t-suffixes |
| V1-SV | Broad stems with short vowel ending in *-igh* | *guigh* 'pray'<br>*luigh* 'lie' | remove *–igh*<br>& append Type 2 suffixes except for Future Indicative |
| V1-SL | Slender | *bris* 'break' | append slender suffixes |
| V1-SL-LV | Slender stems with long vowel ending in *-éigh* | *léigh* 'read'<br>*pléigh* 'discuss' | remove *–igh*<br>& append slender f-suffixes and slender t-suffixes |
| V1-SL-EX | Slender (exceptions) | *siúil* 'walk'<br>*gearáin* 'complain' | broaden<br>& append broad suffixes |
| V1-SL-LC | Slender ending in *-áil* | *sábháil* 'save' | broaden except for t-suffixes<br>& append broad f-suffixes<br>slender t-suffixes |

**Table 2. 2nd Conjugation Verbs**

| *Lexicon Class* | *Description* | *Example* | *Action* |
|---|---|---|---|
| V2-BR | Broad stems ending in *–aigh* | *ceannaigh* 'buy'<br>*tosaigh* 'start' | remove *–aigh*<br>& append broad suffixes |
| | (plus some exceptions) | *freastal* 'attend'<br>*taisteal* 'travel'<br>*lorg* 'search' | |
| V2-SL | Slender stems ending in *-igh* | *bailigh* 'gather'<br>*éirigh* 'rise' | remove *–igh*<br>& append slender suffixes |
| V2-SL-sync | Slender stems syncopated | *imir* 'play'<br>*taitin* 'like' | syncopate<br>& append slender suffixes |
| V2-BR-sync | Broad stems syncopated | *iompair* 'carry'<br>*codail* 'sleep' | syncopate<br>& append broad suffixes |

# Appendix H

## Adjective Lexicon Classes

**Table 1. Adjectives ending in Broad Consonant**

| *Continuation Class* | *Description* | *Example* | *Singular* | *Plural* |
|---|---|---|---|---|
| Adj1-1 | monosyllabic ending in -ll -nn -ch(t) (except dall, donn, bocht) & some others | *mall* 'slow'<br>*gann* 'scarce'<br>*nocht* 'naked' | Feminine<br>Gen: slenderise & suffix -e<br>Voc: no change<br>Masculine<br>Gen & Voc: no change | suffix -a |
| Adj1-2 | monosyllabic others & polysyllabic ending in -(mh)ar | *bán* 'white'<br>*dall* 'blind'<br>*donn* 'brown'<br>*bocht* 'poor'<br>*bríomhar* 'lively'<br>*lúfar* 'athletic' | Feminine<br>Gen: slenderise & suffix -e<br>Voc: no change<br>Masculine<br>Gen & Voc: slenderise | suffix -a |
| Adj1-3 | polysyllabic ending in -ach/-each | *iontach* 'wonderful'<br>*bídeach* 'tiny' | Feminine<br>Gen: each->í,ach->aí<br>Voc: no change<br>Masculine<br>Gen & Voc: each->igh, ach ->aigh | suffix -a |
| Adj1-4 | polysyllabic ending in -íoch | *buíoch* 'thankful'<br>*eolaíoch* 'scientific' | Feminine<br>Gen: och -> thí<br>Voc: no change<br>Masculine<br>Gen & Voc: no change | suffix -a |
| Adj1-5 | polysyllabic ending in long vowel +íoch | *gnách* 'usual'<br>*sóch* 'happy'<br>*gleoch* 'noisy' | Feminine<br>Gen: ch -> thaí<br>Voc: no change<br>Masculine<br>Gen & Voc: no change | suffix -a |

**Table 2. Adjectives ending in Slender Consonant**

| *Continuation Class* | *Description* | *Example* | *Singular* | *Plural* |
|---|---|---|---|---|
| Adj2-1 | polysyllabic<br>ending –(i)úil | *áitiúil* 'local'<br>*cosúil* 'like | Feminine<br>Gen: broaden & suffix -a<br>Voc: no change<br>Masculine<br>Gen & Voc: no change | broaden & suffix -a |
| Adj2-2 | all others | *maith* 'good'<br>*glic* 'clever'<br>*fiáin* 'wild'<br>*séimh* 'gentle' | Feminine<br>Gen: suffix -e<br>Voc: no change<br>Masculine<br>Gen & Voc: no change | suffix -e |

**Table 3. Adjectives ending in a Vowel**

| *Continuation Class* | *Description* | *Example* | *Singular* | *Plural* |
|---|---|---|---|---|
| Adj3-1 | all<br>(except a few irregulars, e.g. breá, te | *aibí* 'ripe'<br>*crua* 'hard'<br>*lofa* 'rotten'<br>*cliste* 'clever'<br>*tanaí* 'thin' | No change | |

**Table 4. Adjectives which are Syncopated**

| *Continuation Class* | *Description* | *Example* | *Singular (final mutations)* | *Plural* |
|---|---|---|---|---|
| Adj4-1 | polysyllabic syncopated gen. sg. fem. & plural | *ramhar* 'fat'<br>*folamh* 'empty' | Feminine<br>Gen: syncopate, slenderise & suffix -e<br>Voc: no change<br>Masculine<br>Gen & Voc: slenderise | syncopate, suffix -a |
| Adj4-2 | polysyllabic syncopated plural | *bodhar* 'deaf'<br>*tirim* 'dry' | Feminine<br>Gen: slenderise & suffix -e<br>Voc: no change<br>Masculine<br>Gen & Voc: slenderise | syncopate, suffix -a |
| Adj4-3 | polysyllabic syncopated gen. sg. fem. & plural | *daingean* 'tight'<br>*saibhir* 'rich' | Feminine<br>Gen: syncopate, slenderise & suffix -e<br>Voc: no change<br>Masculine<br>Gen & Voc: slenderise | syncopate, slenderise, suffix -e |
| Adj4-4 | polysyllabic syncopated, gen. sg. fem. & plural | *socair* 'calm'<br>*deacair* 'difficult' | Feminine<br>Gen: syncopate & suffix -a<br>Voc: no change<br>Masculine<br>Gen & Voc: no change | syncopate, suffix -a |

# Appendix I

## Test Lexicon

This appendix contains a complete list of stems in the current version of the morphological transducer for the folllowing Lexical Classes:

- Nouns
- Verbs
- Adjectives
- Pronouns
- Determiners
- Articles
- Adverbs
- Prepositions
- Conjunctions
- Numerals
- Interjections
- Particles
- Abreviations

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| ab | Nm3-3 | |
| abhainn | Nf5-4 | river |
| ábhar | Nm1-1 | |
| achar | Nm1-1 | |
| acht | Nm3-2 | |
| adharc | Nf2-1 | horn |
| ae | Nm4-6 | liver |
| aer | Nm1-1 | |
| aghaidh | Nf2-6 | |
| aice | Nf4-2 | |
| aicme | Nf4-2 | class |
| aigne | Nf4-SG | |
| ailtire | Nm4-2 | architect |
| aimsir | Nf2-7 | weather |
| áine | Nf4-Prop | |
| ainm | Nm4-7 | name |
| ainmhí | Nm4-4 | animal |
| ainneoin | Subst | |
| aire | Nf4-SG | |
| aire | Nm4-2 | |
| áireamh | Nm1-SG | |
| airgead | Nm1-1 | money |
| airgeadas | Nm1-SG | |
| áirithe | Nf4-SG | |
| airteagal | Nm1-1 | |
| ais | Nf2-6 | |
| aiste | Nf4-2 | essay |
| aisteoireacht | Nf3-1 | acting |
| áit | Nf2-6 | place |
| aithne | Nf4-SG | |
| ál | Nm1-3 | clutch, brood |
| alt | Nm1-1 | |
| am | Nm3-2 | time |
| amharc | Nm1-1 | |
| amhrán | Nm1-1 | |
| amhras | Nm1-SG | |
| anam | Nm3-5 | soul |
| anna | Nf4-Prop | |
| antaine | Nm4-Prop | |
| aodh | Nm1-Prop | |
| aodhán | Nm1-Prop | |
| aoine | Nf4-3 | Friday |
| aoir | Nf2-9 | satire |
| aois | Nf2-6 | |
| aon | Nm1-1 | |
| aon | Nm1-3 | one, ace |
| aonar | Nm1-SG | |
| aonghas | Nm3-Prop | |
| ár | Nm1-SG | |
| arm | Nm1-1 | |
| asailín | Nm4-1 | small ass |
| áth | Nm3-2 | |
| athair | Nm5-2 | father |
| athrú | Nm5-3 | change |
| bá | Nf4-8 | bay |
| bád | Nm1-1 | boat |
| bádín | Nm4-1 | small boat |
| bádóir | Nm3-1 | boat person |
| bagairt | Nf3-4 | |
| baile | Nm4-3 | town |
| baint | Nf2-SG | |
| bairille | Nm4-2 | barrel |
| báisín | Nm4-1 | basin |
| báisteach | Nf2-4 | rain |
| ball | Nm1-1 | |
| balla | Nm4-1 | wall |
| barr | Nm1-1 | |
| bás | Nm1-6 | |

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| bata | Nm4-1 | stick |
| bé | Nf4-4 | maiden |
| beagán | Nm1-1 | |
| beairic | Nf2-7 | barrack |
| béal | Nm1-1 | mouth |
| bealach | Nm1-5 | way, road |
| bealtaine | Nf4-2 | |
| beannacht | Nf3-1 | blessing |
| béarla | Nm4-SG | |
| bearna | Nf4-1 | gap |
| bearnard | Nm1-Prop | |
| beatha | Nf4-1 | life |
| béile | Nm4-2 | meal |
| beirt | Nf2-6 | |
| beo | Nm4-1 | |
| beoir | Nf5-1 | beer |
| bia | Nm4-6 | food |
| binn | Nf2-2 | peak |
| bláth | Nm3-2 | flower |
| bláthnaid | Nf4-Prop | |
| bliain | Nf3-5 | year |
| blúire | Nm4-2 | bit eg of food |
| bogha | Nm4-6 | bough |
| bonn | Nm1-1 | |
| bord | Nm1-1 | table |
| bóthar | Nm1-9 | road |
| brath | Nm1-SG | |
| bráthair | Nm5-2 | brother (clerical) |
| breandán | Nm1-Prop | |
| breis | Nf2-6 | |
| brí | Nf4-6 | meaning |
| bríd | Nf2-Prop | |
| brídin | Nf4-Prop | |
| bríste | Nm4-2 | trousers |
| bróg | Nf2-1 | shoe |
| brú | Nm4-6 | hostel |
| bruas | Nm1-2 | (thick) lip |
| bua | Nm4-6 | victory |
| buairt | Nf3-2 | sorrow (worry) |
| buille | Nm4-2 | a blow |
| buíon | Nf2-10 | group, band |
| bun | Nm1-6 | |
| bus | Nm4-7 | bus |
| cabhail | Nf5-2 | body, trunk |
| cabhair | Nf5-2 | help |
| cág | Nm1-2 | jackdaw |
| cailín | Nm4-1 | girl |
| cailleach | Nf2-4 | hag, witch |
| cáin | Nf5-1 | tax |
| caint | Nf2-6 | talk |
| cáit | Nf4-Prop | |
| caitheamh | Nm1-SG | |
| caitlín | Nf4-Prop | |
| caitríona | Nf4-Prop | |
| caoga | Nm5-1 | fifty |
| caoi | Nf4-6 | way, manner |
| caoilfhionn | Nf4-Prop | |
| caoimhe | Nf4-Prop | |
| capaillín | Nm4-1 | pony (small horse) |
| carr | Nm1-6 | car |
| carraig | Nf2-8 | a rock |
| carráiste | Nm4-2 | carriage |
| cás | Nm1-6 | |
| cat | Nm1-1 | cat |
| cathair | Nf5-2 | city |
| cathal | Nm1-Prop | |
| cathaoir | Nf5-2 | chair |
| ceacht | Nm3-2 | lesson |
| cead | Nm3-2 | |
| céad | Nm1-3 | |
| ceann | Nm1-1 | head, one |
| céanna | Nm4-SG | |
| ceannaí | Nm4-4 | buyer (merchant) |
| ceannas | Nm1-SG | authority, headship |
| ceantar | Nm1-1 | |
| ceart | Nm1-2 | right |
| ceathrú | Nf5-3 | quarter |
| céile | Nm4-2 | |
| ceirnín | Nm4-1 | record (as in music) |
| ceirt | Nf2-8 | a rag |
| ceist | Nf2-6 | |
| ceo | Nm4-6 | fog |
| ceol | Nm1-3 | |
| ciara | Nf4-Prop | |
| ciarán | Nm1-Prop | |
| cineál | Nm1-11 | type, kind |
| cinniúint | Nf3-4 | fate |
| ciontú | Nm4-4 | conviction |
| císte | Nm4-2 | cake |
| cith | Nm3-4 | shower |
| clann | Nf2-1 | family |
| clár | Nm1-1 | |
| cleite | Nm4-2 | feather |
| cliabh | Nm1-1 | |
| cliath | Nf2-2 | |
| cló | Nm4-6 | print |
| clog | Nm1-1 | |
| club | Nm4-7 | club |
| cluiche | Nm4-2 | game |
| cnaipe | Nm4-2 | button |
| cnó | Nm4-6 | Nut |
| cnoc | Nm1-1 | hill |
| cogadh | Nm1-5 | |
| cogadh | Nm1-7 | |
| coill | Nf2-5 | wood |
| coinín | Nm4-1 | rabbit |
| coinne | Nf4-2 | |
| coinneáil | Nf3-SG | |
| coiste | Nm4-2 | committee |
| cóiste | Nm4-2 | coach |
| coláiste | Nm4-2 | college |

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| colm | Nm1-Prop | |
| colmán | Nm1-Prop | |
| comhair | Subst | |
| comhairle | Nf4-2 | advice |
| comhar | Nm1-SG | |
| comharsa | Nf5-3 | Neighbour |
| cónaí | Nm5-3 | home |
| conaire | Nm1-Prop | |
| contae | Nm4-4 | county |
| cor | Nm1-2 | |
| cór | Nm1-1 | |
| córas | Nm1-1 | |
| cos | Nf2-1 | foot |
| cothrom | Nm1-SG | |
| cráin | Nf5-1 | sow pig |
| crann | Nm1-1 | tree |
| craobh | Nf2-8 | branch |
| crios | Nm3-4 | belt |
| críostóir | Nm3-Prop | |
| crith | Nm3-4 | a shiver, shake |
| croí | Nm4-4 | heart |
| cú | Nm4-6 | hound |
| cuairt | Nf2-6 | |
| cuid | Nf3-3 | a share, part |
| cúis | Nf2-6 | |
| cúl | Nm1-1 | |
| cuma | Nf4-1 | |
| cumhacht | Nf3-1 | power |
| cuntas | Nm1-1 | |
| cúpla | Nm4-1 | |
| cur | Nm1-SG | |
| cúram | Nm1-7 | |
| cúrsa | Nm4-1 | |
| dalta | Nm4-1 | child, pupil |
| dán | Nm1-3 | |
| dáta | Nm4-1 | |
| dath | Nm3-2 | |
| dé | Nf4-4 | god |
| déag | Subst | |
| déaglán | Nm1-Prop | |
| déanaí | Nf4-SG | |
| deara | Subst | |
| dearcadh | Nm1-SG | |
| déardaoin | Nm4-SG | Thursday |
| dearg | Nm1-2 | |
| dearmad | Nm1-1 | |
| deartháir | Nm5-2 | brother |
| deas | Subst | |
| deireadh | Nm1-5 | |
| deis | Nf2-6 | |
| deo | Subst | |
| deoir | Nf2-2 | tear |
| diabhal | Nm1-1 | |
| diaidh | Subst | |
| diarmaid | Nm3-Prop | |
| diarmuid | Nm3-Prop | |
| díbirt | Nf3-SG | |
| dílleachta | Nm4-1 | orphen |
| díol | Nm3-SG | |
| dlí | Nm4-4 | law |
| dó | Nm4-6 | a burn |
| dochar | Nm1-SG | |
| dóigh | Nf2-6 | |
| domhan | Nm1-1 | |
| domhnach | Nm1-5 | Sunday |
| dónall | Nm1-Prop | |
| donncha | Nm1-Prop | |
| doras | Nm1-9 | door |
| dráma | Nm4-1 | |
| draoi | Nm4-4 | druid, wizard |
| dream | Nm3-2 | |
| dréimre | Nm4-2 | ladder |
| droim | Nm3-4 | back |
| dubh | Nm1-SG | |
| dúil | Nf2-3 | desire |
| dul | Nm3-SG | |
| dún | Nm1-3 | |
| éadach | Nm1-5 | cloth |
| éadan | Nm1-1 | |
| eagla | Nf4-1 | fear |
| eaglais | Nf2-7 | church (org) |
| eala | Nf4-1 | swam |
| éan | Nm1-1 | bird |
| earra | Nm4-1 | article eg for sale |
| easpa | Nf4-1 | lack |
| eibhlín | Nf4-Prop | |
| éigean | Nm1-SG | |
| eilís | Nf4-Prop | |
| éineacht | Subst | |
| éis | Subst | |
| éisteacht | Nf3-SG | |
| eithne | Nf4-Prop | |
| eoin | Nm4-Prop | |
| eolas | Nm1-SG | |
| fad | Nm1-1 | length |
| faide | Nf4-SG | length |
| fáilte | Nf4-2 | |
| fanacht | Nm3-SG | |
| farraige | Nf4-2 | |
| fás | Nm1-1 | growth |
| fáth | Nm3-2 | |
| féachaint | Nf3-4 | |
| feadh | Nm3-SG | |
| fear | Nm1-1 | man |
| fearghas | Nm3-Prop | |
| feidhm | Nf2-6 | |
| féidir | Subst | |
| féile | Nf4-3 | feast, festival |
| féirín | Nm4-1 | small gift |
| fia | Nm4-6 | deer |
| fiche | Nm5-1 | twenty |
| file | Nm4-2 | poet |
| finné | Nm4-4 | witness |
| fionn | Nm1-2 | fair, white |

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| fíor | Nf2-8 | figure/diagram |
| fios | Nm3-4 | knowledge |
| fiú | Subst | |
| flaith | Nm3-6 | prince |
| fleá | Nf4-8 | feast |
| focal | Nm1-1 | word |
| fogha | Nm4-6 | an attack |
| fogha | Nm4-6 | attach |
| fómhar | Nm1-1 | |
| fonn | Nm1-1 | |
| forbairt | Nf3-4 | |
| freagra | Nm4-1 | |
| fréamh | Nf2-8 | root |
| fud | Subst | |
| fuil | Nf3-3 | blood |
| fuinneog | Nf2-1 | window |
| gá | Nm4-SG | |
| gabháil | Nf3-6 | catch, seizure etc.. |
| gael | Nm1-1 | |
| gaeltacht | Nf3-1 | |
| gaillimh | Nf2-Prop | |
| gáir | Nf2-9 | a shout |
| gairdín | Nm4-1 | garden (small field?) |
| gaol | Nm1-3 | relationship, relative |
| garáiste | Nm4-2 | garage |
| garda | Nm4-1 | guard |
| garraí | Nm4-4 | field or garden |
| gasóigín | Nm4-1 | young shoot, boy scout |
| gé | Nf4-6 | goose |
| gealach | Nf2-4 | moon |
| geall | Nm1-3 | bet |
| gearmáin | Nf2-Prop | |
| geimhreadh | Nm1-5 | winter |
| girseach | Nf2-4 | child |
| glan | Nm1-SG | |
| glao | Nf4-8 | call |
| glas | Nm1-1 | |
| gleann | Nm3-7 | glen |
| gloine | Nf4-2 | glass |
| glór | Nm1-4 | voice |
| glúin | Nf2-3 | knee or generation |
| gnáth | Nm1-2 | |
| gnó | Nm4-4 | business |
| grá | Nm4-SG | |
| gráinne | Nf4-Prop | |
| greim | Nm3-4 | grip |
| grian | Nf2-10 | sun |
| grua | Nf4-8 | cheek |
| grúpa | Nm4-1 | group |
| hata | Nm4-1 | hat |
| iall | Nf2-8 | leash, shoelace |
| iarlaith | Nm3-Prop | |

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| iarracht | Nf3-1 | |
| iarraidh | Nf3-2 | request |
| íde | Nf4-Prop | |
| im | Nm2-1 | butter |
| imeacht | Nm3-3 | |
| imirt | Nf3-SG | |
| imní | Nf4-SG | |
| iníon | Nf2-8 | |
| inneall | Nm1-1 | engine |
| insint | Nf2-7 | |
| intinn | Nf2-7 | |
| íobairt | Nf3-4 | |
| íol | Nm1-2 | idol |
| iomaí | Nf4-4 | couch, bed |
| iomhá | Nf4-8 | image |
| iomlán | Nm1-1 | |
| iompar | Nm1-SG | |
| iompú | Nm4-4 | turning |
| ionaclú | Nm4-4 | innoculation |
| ionad | Nm1-1 | |
| iontas | Nm1-1 | |
| ispín | Nm4-1 | sausage (?) |
| iúil | Nm4-7 | July |
| iúl | Nm1-SG | |
| labhairt | Nf3-SG | |
| labhrás | Nm1-Prop | |
| laghad | Nm4-SG | |
| laghdú | Nm4-4 | decrease |
| láir | Nf5-1 | mare |
| láithreach | Nm1-SG | |
| lámh | Nf2-1 | hand |
| lán | Nm1-1 | |
| lao | Nm4-6 | calf |
| laoi | Nf4-4 | Narrative poem |
| lár | Nm1-1 | |
| lasair | Nf5-2 | flame |
| láthair | Nf5-4 | place, site |
| leabhar | Nm1-1 | book |
| leagan | Nm1-11 | version |
| leanbh | Nm1-5 | child |
| léaráid | Nf2-7 | |
| leas | Nm3-SG | |
| leasú | Nm5-3 | ammendment |
| leath | Nf2-1 | |
| leibhéal | Nm1-1 | |
| léine | Nf4-3 | shirt |
| leith | Nf2-6 | |
| leithéid | Nf2-7 | |
| liam | Nm4-Prop | |
| ligean | Nm1-SG | |
| líne | Nf4-3 | |
| líon | Nm1-SG | |
| liosta | Nm4-1 | list |
| litir | Nf5-2 | letter |
| litríocht | Nf3-1 | |
| liú | Nm4-6 | shout |
| loch | Nm3-2 | lake |

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| long | Nf2-1 | ship |
| lorcán | Nm1-Prop | |
| lorga | Nf4-1 | shin |
| luach | Nm3-2 | |
| luan | Nm1-3 | Monday |
| lucht | Nm3-2 | |
| má | Nf4-8 | plain |
| mac | Nm1-1 | |
| machaire | Nm4-2 | plain |
| maidin | Nf2-8 | |
| máire | Nf4-Prop | |
| máiréad | Nf4-Prop | |
| máirín | Nf4-Prop | |
| máirt | Nf4-6 | Tuesday |
| máirtín | Nm4-Prop | |
| malairt | Nf2-7 | |
| marbh | Nm1-1 | |
| marc | Nm1-6 | mark |
| marcach | Nm1-1 | jockey |
| margadh | Nm1-5 | market |
| más | Nm1-2 | |
| máthair | Nf5-4 | mother |
| meá | Nf4-8 | weight |
| méabh | Nf2-Prop | |
| méadú | Nm5-3 | increase |
| meán | Nm1-1 | |
| meas | Nm3-SG | |
| measc | Nm4-SG | jumble, confusion |
| méid | Nm4-SG | |
| meitheamh | Nm1-1 | |
| mian | Nf2-10 | |
| mícheál | Nm1-Prop | |
| míle | Nm4-3 | mile |
| mír | Nf2-6 | |
| misneach | Nm1-SG | courage |
| modh | Nm3-2 | |
| monarcha | Nf5-3 | factory |
| moncaí | Nm4-4 | monkey |
| mórán | Nm1-SG | |
| múinteoir | Nm3-1 | |
| muintir | Nf2-8 | |
| múr | Nm1-4 | wall, shower |
| naomh | Nm1-1 | saint |
| nathair | Nf5-2 | snake |
| nead | Nf2-8 | nest |
| neart | Nm1-SG | |
| ní | Nm4-4 | wash |
| nia | Nm4-6 | Nephew |
| nioclás | Nm1-Prop | |
| nócha | Nm5-1 | Ninety |
| nóiméad | Nm1-1 | |
| nóirín | Nf4-Prop | |
| nollaig | Nf3-4 | Christmas |
| nóra | Nf4-Prop | |
| nós | Nm1-6 | |
| nóta | Nm4-1 | Note |
| nua | Nm4-SG | |
| obair | Nf5-2 | work |
| ochtó | Nm5-1 | eighty |
| óg | Nm1-2 | |
| oibrí | Nm4-4 | worker |
| oíche | Nf4-7 | night |
| oideachas | Nm1-SG | |
| oifig | Nf2-7 | |
| óige | Nf4-SG | |
| oileán | Nm1-1 | |
| oiread | Subst | |
| ól | Nm1-SG | |
| ollscoil | Nf2-6 | |
| oscailt | Nf2-7 | |
| pá | Nm4-6 | pay |
| pádraig | Nm4-Prop | |
| paidrín | Nm4-1 | prayer (rosary) |
| páipéar | Nm1-1 | paper |
| páirtí | Nm4-4 | party |
| paiste | Nm4-2 | patch |
| páiste | Nm4-2 | child |
| pas | Nm4-7 | passport |
| pasáiste | Nm4-2 | passage |
| peaca | Nm4-1 | sin |
| peadar | Nm1-Prop | |
| pearsa | Nf5-3 | person |
| pian | Nf2-10 | pain |
| pictiúr | Nm1-1 | |
| pilib | Nm4-Prop | |
| pionta | Nm4-1 | pint |
| píopa | Nm4-1 | pipe |
| píosa | Nm4-1 | |
| plé | Nf4-SG | discussion |
| pobal | Nm1-1 | public |
| pointe | Nm4-2 | point |
| pól | Nm1-Prop | |
| post | Nm1-1 | post |
| praghas | Nm1-8 | price |
| proinsias | Nm4-Prop | |
| rabhadh | Nm1-5 | |
| radharc | Nm1-1 | |
| raidió | Nm4-6 | |
| rás | Nm3-3 | a race |
| réalta | Nf4-1 | star |
| réir | Nf2-SG | |
| réiteach | Nm1-1 | solution |
| rí | Nm4-4 | king |
| riail | Nf5-1 | rule |
| rialtas | Nm1-1 | |
| rian | Nm1-3 | mark, track |
| ribín | Nm4-1 | ribbon |
| rogha | Nf4-8 | bay |
| roinnt | Nf2-9 | division, portion |
| róisín | Nf4-Prop | |
| ros | Nm3-6 | |
| ruairí | Nm4-Prop | |
| rud | Nm3-3 | thing |

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| rún | Nm1-1 | secret, love |
| sagart | Nm1-1 | priest |
| saghas | Nm1-8 | type, kind |
| saighdiúir | Nm3-1 | soldier |
| sail | Nf2-6 | timber beam, cudgel |
| samhradh | Nm1-5 | summer |
| sampla | Nm4-1 | |
| saoi | Nm4-4 | master, expert |
| saoire | Nf4-SG | |
| saoirse | Nf4-SG | |
| saol | Nm1-3 | life |
| saor | Nm1-1 | |
| saothar | Nm1-1 | work |
| satharn | Nm1-1 | Saturday |
| scannán | Nm1-1 | |
| scéal | Nm1-3 | |
| scéim | Nf2-6 | |
| scoil | Nf2-6 | |
| scoláireacht | Nf3-1 | scholarship |
| scolóigín | Nm4-1 | servant, pupil, hard-working youth |
| scór | Nm1-4 | twenty |
| scríbhneoir | Nm3-1 | writer |
| scrín | Nf2-5 | shrine |
| scriosán | Nm1-1 | eraser |
| seachtain | Nf2-7 | |
| seachtó | Nm5-1 | seventy |
| séamas | Nm1-Prop | |
| seán | Nm1-Prop | |
| seans | Nm4-7 | chance |
| seasamh | Nm1-SG | |
| seasca | Nm5-1 | sixty |
| séimhiú | Nm4-4 | lenition |
| séipéal | Nm1-1 | church |
| seirbhís | Nf2-7 | service |
| seol | Nm1-3 | sail |
| seomra | Nm4-1 | |
| sicín | Nm4-1 | chick |
| síle | Nf4-Prop | |
| sinéad | Nf4-Prop | |
| siobhán | Nf4-Prop | |
| sioc | Nm3-4 | frost |
| síol | Nm1-3 | seed |
| siopa | Nm4-1 | shop |
| slabhra | Nm4-1 | chain |
| sláinte | Nf4-2 | health |
| slán | Nm1-2 | |
| sleá | Nf4-8 | spear |
| sleamhnú | Nm5-3 | slide |
| slí | Nf4-4 | way, road |
| sliabh | Nm2-1 | mountain |
| slinn | Nf2-5 | a slate |
| sloinne | Nm4-3 | surname |
| snáithe | Nf4-6 | thread syn. snáth |
| snáth | Nm3-2 | thread |

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| sochraid | Nf2-7 | funeral |
| soitheach | Nm1-5 | vessal |
| solas | Nm1-9 | light |
| son | Subst | |
| sórt | Nm1-1 | |
| spás | Nm1-6 | space |
| spéir | Nf2-9 | sky |
| spéis | Nf2-SG | |
| spiaire | Nm4-2 | spy |
| splanc | Nf2-8 | spark |
| sráid | Nf2-6 | |
| sraith | Nf2-6 | |
| stad | Nm4-7 | stop |
| staidéar | Nm1-SG | |
| staighre | Nm4-2 | stairs |
| stailc | Nf2-6 | strike |
| stair | Nf2-9 | history |
| stáisiún | Nm1-1 | |
| stát | Nm1-1 | |
| stoirm | Nf2-8 | storm |
| stop | Nm4-7 | stop |
| stór | Nm1-4 | store, treasure |
| stua | Nm4-6 | |
| sú | Nm4-6 | juice |
| súil | Nf2-3' | eye |
| suim | Nf2-6 | |
| suíomh | Nm1-1 | situation, site |
| suipéar | Nm1-1 | supper |
| tábla | Nf4-1 | table (e.g.of contents) |
| tadhg | Nm1-Prop | |
| tae | Nm4-6 | tea |
| tagairt | Nf3-4 | reference |
| taibhse | Nf4-2 | ghost |
| táille | Nf4-2 | fee |
| táilliúir | Nm3-1 | tailor |
| taithí | Nf4-SG | |
| tamall | Nm1-1 | |
| taobh | Nm1-6 | |
| tarlú | Nm5-3 | happen |
| teanga | Nf4-8 | |
| teideal | Nm1-1 | |
| teilifís | Nf2-SG | |
| tiarna | Nm4-1 | lord |
| timpeall | Nm1-1 | |
| tincéir | Nm3-1 | tinker |
| tine | Nf4-3 | fire |
| tiomáint | Nf3-4 | driving |
| tionchar | Nm1-1 | |
| tionónta | Nm4-1 | tennant |
| tír | Nf2-9 | country |
| titim | Nf2-SG | |
| tlú | Nm4-6 | tongs |
| tobar | Nm1-10 | well |
| togha | Nm4-6 | choice, variety |
| toil | Nf3-3 | will (desire) |
| toisc | Nf2-12 | purpose, etc.. |
| toitín | Nm4-1 | cigarette |

| Noun Stem | Lexicon Classes | Gloss |
|---|---|---|
| | | (small smoke) |
| tom | Nm1-1 | |
| tomás | Nm1-Prop | |
| tonn | Nf2-10 | wave (as in sea) |
| toradh | Nm1-5 | fruit, result |
| trá | Nf4-8 | beach |
| trácht | Nm3-SG | |
| traein | Nf5-1 | train |
| tram | Nm4-7 | tram |
| tráth | Nm3-2 | |
| tráthnóna | Nm4-1 | afternoon |
| treasa | Nf4-Prop | |
| tréimhse | Nf4-2 | |
| treo | Nf4-6 | |
| treoir | Nf5-1 | |
| tríocha | Nm5-1 | thirty |
| tríona | Nf4-Prop | |
| triúr | Nm1-1 | |
| trua | Nf4-8 | sorrow |
| tua | Nf4-8 | axe |
| tuairim | Nf2-7 | |
| tuairisc | Nf2-7 | an account |
| tuilleadh | Nm1-SG | |
| tuiscint | Nf3-SG | |
| turas | Nm1-1 | |
| tús | Nm1-SG | |
| uachtarán | Nm1-1 | |
| uair | Nf2-11 | hour, time |
| uasal | Nm1-9 | noble person |
| údar | Nm1-1 | |
| údarás | Nm1-1 | |
| uimhir | Nf5-2 | Number |
| uisce | Nm4-2 | water |
| úll | Nm1-2 | apple |
| ullmhú | Nm4-4 | preparation |
| úna | Nf4-Prop | |
| úr | Nm1-SG | |
| urú | Nm4-4 | eclipse |
| úsáid | Nf2-7 | |
| veidhlín | Nm4-1 | violin (loanword) |

| Noun Stems (irregular) | Lexicon Class | Gloss |
|---|---|---|
| tom | Nm1-1 | |
| tomás | Nm1-Prop | |
| tonn | Nf2-10 | wave (as in sea) |
| toradh | Nm1-5 | fruit, result |
| trá | Nf4-8 | beach |
| trácht | Nm3-SG | |
| traein | Nf5-1 | train |
| tram | Nm4-7 | tram |
| tráth | Nm3-2 | |
| tráthnóna | Nm4-1 | afternoon |
| treasa | Nf4-Prop | |
| tréimhse | Nf4-2 | |
| treo | Nf4-6 | |
| treoir | Nf5-1 | |
| tríocha | Nm5-1 | thirty |
| tríona | Nf4-Prop | |
| triúr | Nm1-1 | |
| trua | Nf4-8 | sorrow |
| tua | Nf4-8 | axe |
| tuairim | Nf2-7 | |
| tuairisc | Nf2-7 | an account |
| tuilleadh | Nm1-SG | |
| tuiscint | Nf3-SG | |
| turas | Nm1-1 | |
| tús | Nm1-SG | |
| uachtarán | Nm1-1 | |
| uair | Nf2-11 | hour, time |
| uasal | Nm1-9 | noble person |
| údar | Nm1-1 | |
| údarás | Nm1-1 | |
| uimhir | Nf5-2 | Number |
| uisce | Nm4-2 | water |
| úll | Nm1-2 | apple |
| ullmhú | Nm4-4 | preparation |
| úna | Nf4-Prop | |
| úr | Nm1-SG | |
| urú | Nm4-4 | eclipse |
| úsáid | Nf2-7 | |
| veidhlín | Nm4-1 | violin (loanword) |

| Noun Stems (irregular) | Lexicon Class | Gloss |
|---|---|---|
| bean | N-Irreg | |
| bith | N-Irreg | |
| bó | N-Irreg | |
| breatain+Prop | N-Irreg | |
| comhar | N-Irreg | |
| conall+Prop | N-Irreg | |
| conradh | N-Irreg | |
| dada | N-Irreg | |
| deirfiúr | N-Irreg | |
| dia | N-Irreg | |
| duine | N-Irreg | |
| éire+Prop | N-Irreg | |
| éireannach | N-Irreg | |

| Noun Stems (irregular) | Lexicon Class | Gloss |
|---|---|---|
| bean | N-Irreg | |
| bith | N-Irreg | |
| bó | N-Irreg | |
| breatain+Prop | N-Irreg | |
| comhar | N-Irreg | |
| conall+Prop | N-Irreg | |
| conradh | N-Irreg | |
| dada | N-Irreg | |
| deirfiúr | N-Irreg | |
| dia | N-Irreg | |
| duine | N-Irreg | |
| éire+Prop | N-Irreg | |
| éireannach | N-Irreg | |

| Noun Stems (irregular) | Lexicon Class | Gloss |
|---|---|---|
| eoraip+Prop | N-Irreg | |
| eorpach | N-Irreg | |
| feirste+Prop | N-Irreg | |
| foireann | N-Irreg | |
| gaeilge+Prop | N-Irreg | |
| gaolainn+CM+Prop | N-Irreg | |
| lá | N-Irreg | |
| leaba | N-Irreg | |
| life+Prop | N-Irreg | |
| londain+Prop | N-Irreg | |
| mac | N-Irreg | |
| méid | N-Irreg | |
| meiriceá+Prop | N-Irreg | |
| mí | N-Irreg | |
| sasana+Prop | N-Irreg | |
| scrúdú | N-Irreg | |
| smaoineamh | N-Irreg | |
| tada | N-Irreg | |
| talamh | N-Irreg | |
| talamh | N-Irreg | |
| teach | N-Irreg | |
| uladh+Prop | N-Irreg | |

| Verb Stem | Lexicon Class | Gloss |
|---|---|---|
| abair | VI1 | say |
| achainigh | V2-SL | request |
| agair | V2-BR-sync | |
| airigh | V2-SL | |
| áirigh | V2-SL | |
| aithin | V2-SL-sync | |
| aithris | V2-SL | |
| áitigh | V2-SL | |
| aontaigh | V2-BR | agree |
| arsa | Verb | |
| bagair | V2-BR-sync | |
| bailigh | V2-SL | |
| bain | V1-SL | |
| beannaigh | V2-BR | |
| beir | VI2 | catch, give birth to |
| bí | VI11 | be |
| bligh | V1-SV | |
| breathnaigh | V2-BR | look |
| bris | V1-SL | |
| brúigh | V1-BR-LV | |
| buaigh | V1-BR-LV | |
| buail | V1-SL | beat |
| bunaigh | V2-BR | found |
| caith | V1-SL | |
| can | V1-BR | sing |
| cas | V1-BR | turn |
| ceangail | V2-BR-sync | |
| ceannaigh | V2-BR | |
| ceap | V1-BR | think, invent |
| ciallaigh | V2-BR | mean,sense |
| cigil | V2-SL-sync | |
| clóigh | V1-BR-LV | |
| clois | VI3 | hear |
| clúdaigh | V2-BR | |
| codail | V2-BR-sync | |
| cogair | V2-BR-sync | |
| coigil | V2-SL-sync | |
| cónaigh | V2-BR | |
| corraigh | V2-BR | |
| cosain | V2-BR-sync | |
| cráigh | V1-BR-LV | |
| cruinnigh | V2-SL | |
| cuardaigh | V2-BR | |
| cuimhnigh | V2-SL | |
| cuir | V1-SL | |
| cúitigh | V2-SL | |
| cumhdaigh | V2-BR | |
| dathaigh | V2-BR | |
| déan | VI9 | do, make |
| díbir | V2-SL-sync | |
| dóigh | V1-BR-LV | |
| dtí | Verb | |
| dúisigh | V2-SL | |
| éiligh | V2-SL | |
| éirigh | V2-SL | |
| éist | V1-SL | |
| eitigh | V2-SL | |
| eitil | V2-SL-sync | |
| fág | V1-BR | |
| faigh | VI10 | get |
| fan | V1-BR | |
| féach | V1-BR | |
| féad | V1-BR | |
| feic | VI8 | see |
| fiafraigh | V2-BR | |
| figh | V1-SV | |
| fógair | V2-BR-sync | |
| foghlaim | V2-SL | |
| foilsigh | V2-SL | |
| freagair | V2-BR-sync | |
| freastal | V2-BR | |
| fuaigh | V1-BR-LV | |
| fulaing | V2-SL | |
| gabh | V1-BR | |
| gearáin | V1-SL-X | |
| gearr | V1-BR | cut |
| glac | V1-BR | |
| gortaigh | V2-BR | |
| guigh | V1-SV | |
| iarr | V1-BR | |
| imigh | V1-SV | go |
| imigh | V2-SL | |
| imir | V2-SL-sync | |
| inis | V2-SL-sync | |
| íoc | V1-BR | |
| iomair | V2-BR-sync | |
| iompair | V2-BR-sync | |
| ionsaigh | V2-BR | |
| is | VI12 | copula |
| ith | VI4 | eat |
| labhair | V2-BR-sync | |
| lag | V1-BR | |
| leag | V1-BR | knock, lay out |
| lean | V1-BR | |
| leanas | Verb | |
| léigh | V1-SL-LV | |
| léirigh | V2-SL | |
| lig | V1-SL | |
| loisc | V1-SL | |
| loit | V1-SL | |
| lorg | V2-BR | |
| luaigh | V1-BR-LV | |
| lúb | V1-BR | |
| luigh | V1-SV | |
| mair | V1-SL | live, last |
| mol | V1-BR | |
| múscail | V2-BR-sync | |
| nigh | V1-SV | |
| oscail | V2-BR-sync | |
| pléigh | V1-SL-LV | |
| réigh | V1-SL-LV | |
| réitigh | V2-SL | |
| rith | V1-SL | |
| roinn | V1-SL | |
| sábháil | V1-SL-LC | |
| sásaigh | V2-BR | |
| scaip | V1-SL | |

| | | |
|---|---|---|
| scríobh | V1-BR | |
| scrúdaigh | V2-BR | |
| seas | V1-BR | stand |
| seol | V1-BR | sail, launch |
| siúil | V1-SL-X | |
| smaoinigh | V2-SL | |
| socraigh | V2-BR | |
| spréigh | V1-SL-LV | |
| suigh | V1-SV | |
| tabhair | VI5 | give |
| taispeáin | V1-SL-X | |
| taisteal | V2-BR | |
| taitin | V2-SL-sync | |
| tar | VI6 | come |
| tarlaigh | V2-BR | happen |
| tarraing | V2-SL | |
| teastaigh | V2-BR | |
| téigh | V1-SL-LV | |
| téigh | VI7 | go |
| tíolaic | V1-SL-X | |
| tionóil | V1-SL-X | |
| tit | V1-SL | fall |
| tóg | V1-BR | |
| tosaigh | V2-BR | |
| tuig | V1-SL | understand |
| tuirling | V2-SL | |
| úsáid | V1-SL | |

| Adjective Stem | Lexicon Class | Gloss |
|---|---|---|
| achrannach | Adj1-3; | troublesome |
| aibí | Adj3-1; | ripe |
| aisteach | Adj1-3; | strange |
| áitiúil | Adj2-1; | local |
| amháin | Adj; | one |
| aoibhinn | Adj4-3; | pleasant |
| ard | Adj1-2; | tall |
| bacach | Adj1-3; | lame |
| bán | Adj1-2; | white |
| banúil | Adj2-1; | womanly, ladylike |
| baolach | Adj1-3; | dangerous |
| beag | Adj1-2; | small |
| bídeach | Adj1-3; | tiny |
| bliantiúil | Adj2-1; | yearly, annual |
| bocht | Adj1-2; | poor |
| bodhar | Adj4-2; | deaf |
| bríomhar | Adj1-2; | lively |
| buí | Adj3-1; | yellow |
| buíoch | Adj1-4; | thankful, grateful |
| cainteach | Adj1-3; | talkative |
| cinnte | Adj3-1; | certain |
| ciúin | Adj2-2; | quiet |
| cliste | Adj3-1; | clever |
| crua | Adj3-1; | hard |
| daibhir | Adj4-3; | poor |
| daingean | Adj4-3; | firm, tight |
| dall | Adj1-2; | blind |
| deacair | Adj4-4; | difficult |
| deimhin | Adj4-3; | certain |
| dílis | Adj4-3; | loyal |
| díomách | Adj1-5; | dissappointed |
| diomaíoch | Adj1-4; | ungrateful |
| díomuach | Adj1-5; | defeated |
| díreach | Adj1-3; | straight, direct |
| domhain | Adj4-3; | deep |
| donn | Adj1-2; | brown |
| dorcha | Adj3-1; | dark |
| eile | Adj; | other |
| eolaíoch | Adj1-4; | scientific |
| faillíoch | Adj1-4; | negligent |
| faíoch | Adj1-4; | copious |
| fearúil | Adj2-1; | manly |
| fiáin | Adj2-2; | wild |
| fionn | Adj1-1; | fair (hair) |
| fiosrach | Adj1-3; | inquisitive |
| fliuch | Adj1-1; | wet |
| folamh | Adj4-1; | empty |
| fuíoch | Adj1-4; | copious |
| gach | Adj; | every |
| gaelach | Adj1-3; | Irish |
| gann | Adj1-1; | scarce |
| gleoch | Adj1-5; | noisy |
| glic | Adj2-2; | clever |
| gnách | Adj1-5; | usual |
| imníoch | Adj1-4; | anxious |
| iontach | Adj1-3; | wonderful |
| íseal | Adj4-3; | low |
| lách | Adj1-5; | generous |
| lagbhríoch | Adj1-4; | weak, languid |
| láidir | Adj4-3; | strong |
| lofa | Adj3-1; | rotton |
| lúfar | Adj1-2; | athletic |
| magúil | Adj2-1; | jokingly |
| maith | Adj2-2; | good |
| mall | Adj1-1; | slow |
| meirgeach | Adj1-3; | rusty |
| milis | Adj4-3; | sweet |
| mín | Adj2-2; | fine, smooth |
| mór | Adj1-2; | big |
| nocht | Adj1-1; | naked |
| ramhar | Adj4-1; | fat |
| réidh | Adj2-2; | easy, smooth,ready |
| rua | Adj3-1; | red (hair) |
| sách | Adj1-5; | full |
| saibhir | Adj4-3; | rich |
| salach | Adj1-3; | dirty |
| sealadach | Adj1-3; | temporary |
| séimh | Adj2-2; | gentle |
| socair | Adj4-4; | calm |
| sóch | Adj1-5; | happy |
| sona | Adj3-1; | happy |
| sonaí | Adj3-1; | lucky |
| spleách | Adj1-5; | dependant |
| stairiúil | Adj2-1; | historic |
| sultmhar | Adj1-2; | jolly |
| taithíoch | Adj1-4; | accustomed to, familiar with |
| tanaí | Adj3-1; | thin |
| tapa | Adj3-1; | fast |
| tinn | Adj2-2; | sick |
| tirim | Adj4-2; | dry |
| toll | Adj1-1; | hollow |
| tromchroíoch | Adj1-4; | heavy-hearted or indigestible food |
| uaigneach | Adj1-3; | lonely |
| uasal | Adj4-3; | noble |
| umhal | Adj4-2; | humble |

| Pronoun | Lexicon Class | Gloss |
|---|---|---|
| mé | Pron; | me |
| tú | Pron; | you |
| sí | Pron; | she |
| í | Pron; | she/her |
| sé | Pron; | he |
| é | Pron; | he |
| sinn | Pron; | we |
| muid | Pron; | we |
| sibh | Pron; | you pl |
| siad | Pron; | they |
| iad | Pron; | they |
| mise | Pron; | !me contrastive |
| tusa | Pron; | you contrastive |
| sise | Pron; | she |
| ise | Pron; | she/herself |
| seisean | Pron; | he |
| eisean | Pron; | he |
| sinne | Pron; | we |
| muide | Pron; | we |
| sibhse | Pron; | you pl. (also ye or yous) |
| siadsan | Pron; | they/them |
| iadsan | Pron; | they |
| féin | Pron; | self |
| ceachtar | Pron; | any |
| neachtar | Pron; | not any |
| pé | Pron; | whoever |
| té | Pron; | person |
| cibé | Pron; | who/whichever |
| ea | Pron; | it |

| Determiner | Lexicon Class | Gloss |
|---|---|---|
| mo | Det; | my |
| do | Det; | your |
| a | Det; | her |
| a | Det; | his |
| ár | Det; | our |
| bhur | Det; | your |
| a | Det; | their |
| á | Det; | it |
| á | Det; | it |
| á | Det; | it |
| seo | Det; | this |
| sin | Det; | that |
| siúd | Det; | those |
| cad | Det; | what |
| céard | Det; | what |
| cé | Det; | who |
| cathain | Det; | when |
| cén | Det; | which |

| Article | Lexicon Class | Gloss |
|---|---|---|
| an | Art; | the |
| na | Art; | the |

| Adverb | Lexicon Class | Gloss |
|---|---|---|
| thart | Adv; | |
| suas | Adv; | |
| síos | Adv; | |
| anuas | Adv; | |
| aníos | Adv; | |
| thuas | Adv; | |
| thíos | Adv; | |
| thall | Adv; | |
| anall | Adv; | |
| anonn | Adv; | |
| abhus | Adv; | |
| thuaidh | Adv; | |
| theas | Adv; | |
| thoir | Adv; | |
| thiar | Adv; | |
| soir | Adv; | |
| siar | Adv; | |
| anoir | Adv; | |
| aniar | Adv; | |
| aneas | Adv; | |
| aduaidh | Adv; | |
| ann | Adv; | |
| anseo | Adv; | |
| ansin | Adv; | |
| abhaile | Adv; | |
| amach | Adv; | |
| isteach | Adv; | |
| amuigh | Adv; | |
| istigh | Adv; | |
| lasmuigh | Adv; | |
| laistigh | Adv; | |
| amú | Adv; | |
| anois | Adv; | |
| arís | Adv; | |
| annamh | Adv; | |
| cheana | Adv; | |
| choíche | Adv; | |
| déanach | Adv; | |
| fós | Adv; | |
| feasta | Adv; | |
| minic | Adv; | |
| riamh | Adv; | |
| uaireanta | Adv; | |
| inniu | Adv; | |
| inné | Adv; | |
| amárach | Adv; | |
| anocht | Adv; | |
| aréir | Adv; | |
| anuraidh | Adv; | |
| amhlaidh | Adv; | |
| beagnach | Adv; | |
| dáiríre | Adv; | |
| fosta | Adv; | |
| freisin | Adv; | |
| áfach | Adv; | |
| thuaidh | Adv; | |
| theas | Adv; | |
| thoir | Adv; | |

| Adverb | Lexicon Class | Gloss |
|---|---|---|
| thiar | Adv; | |
| ní | Adv; | |
| nach | Adv; | |
| ná | Adv; | |
| níor | Adv; | |
| nár | Adv; | |
| arú | Adv; | |
| conas | Adv; | |
| cá | Adv; | |
| cá | Adv; | |
| cathain | Adv; | |
| fadó | Adv; | |

| Preposition | Lexicon Class | Gloss |
|---|---|---|
| a | Prep; | a chlog, a dhíth |
| ag | Prep; | at |
| ar | Prep; | on |
| as | Prep; | out |
| chuig | Prep; | to |
| chun | Prep; | |
| cois | Prep; | |
| dála | Prep; | |
| de | Prep; | from |
| do | Prep; | to |
| faoi | Prep; | |
| fara | Prep; | as well as |
| fearacht | Prep; | |
| gan | Prep; | |
| go | Prep; | |
| gur | Prep; | |
| gurb | Prep; | |
| gurbh | Prep; | |
| i | Prep; | in |
| idir | Prep; | between |
| ionsar | Prep; | towards |
| is | Prep; | bliain is an t-am seo |
| le | Prep; | with |
| mar | Prep; | like, as |
| ó | Prep; | from |
| os | Prep; | |
| roimh | Prep; | from |
| seachas | Prep; | |
| seach | Prep; | other than |
| thar | Prep; | over |
| timpeall | Prep; | |
| trasna | Prep; | |
| trí | Prep; | through |
| um | Prep; | |
| agam | Prep; | at me |
| agat | Prep; | at you |
| aige | Prep; | at him |
| aici | Prep; | at her |
| againn | Prep; | at us |
| agaibh | Prep; | at you |
| acu | Prep; | at them |

| Preposition | Lexicon Class | Gloss |
|---|---|---|
| orm | Prep; | on me |
| ort | Prep; | on you |
| air | Prep; | on him |
| uirthi | Prep; | on her |
| orainn | Prep; | on us |
| oraibh | Prep; | on you |
| orthu | Prep; | on them |
| asam | Prep; | out of me |
| asat | Prep; | out of you |
| as | Prep; | out of him |
| aisti | Prep; | out of her |
| asainn | Prep; | out of us |
| asaibh | Prep; | out of you |
| astu | Prep; | out of them |
| chugam | Prep; | towards me |
| chugat | Prep; | towards you |
| chuige | Prep; | towards him |
| chuici | Prep; | towards her |
| chugainn | Prep; | towards us |
| chugaibh | Prep; | towards you |
| chucu | Prep; | towards them |
| díom | Prep; | from me |
| díot | Prep; | from you |
| de | Prep; | from him |
| di | Prep; | from her |
| dínn | Prep; | from us |
| díbh | Prep; | from you |
| díobh | Prep; | from them |
| dom | Prep; | to me |
| duit | Prep; | to you |
| dó | Prep; | to him |
| | Prep; | to her |
| dúinn | Prep; | to us |
| daoibh | Prep; | to you |
| dóibh | Prep; | to them |
| fúm | Prep; | under me |
| fút | Prep; | under you |
| faoi | Prep; | under him |
| fúithi | Prep; | under her |
| fúinn | Prep; | under us |
| fúibh | Prep; | under you |
| fúthu | Prep; | under them |
| faram | Prep; | as well as me |
| farat | Prep; | as well as you |
| fairis | Prep; | as well as her |
| farae | Prep; | as well as him |
| farainn | Prep; | as well as us |
| faraibh | Prep; | as well as you |
| faru | Prep; | as well as them |
| ionam | Prep; | in me |
| ionat | Prep; | in you |
| ann | Prep; | in him |
| inti | Prep; | in her |
| ionainn | Prep; | in us |
| ionaibh | Prep; | in you |
| iontu | Prep; | in them |
| | Prep; | between us |

| Preposition | Lexicon Class | Gloss |
|---|---|---|
| eadraibh | Prep; | between you |
| eatarthu | Prep; | between them |
| ionsorm | Prep; | towards me |
| ionsort | Prep; | to you |
| ionsair | Prep; | to him |
| ionsuirthi | Prep; | to her |
| ionsorainn | Prep; | to us |
| ionsoraibh | Prep; | to you |
| ionsorthu | Prep; | to them |
| liom | Prep; | with me |
| leat | Prep; | with you |
| leis | Prep; | with his |
| léi | Prep; | with her |
| linn | Prep; | with us |
| libh | Prep; | with you |
| leo | Prep; | with them |
| uaim | Prep; | from me |
| uait | Prep; | from you |
| uaidh | Prep; | from him |
| uaithi | Prep; | from her |
| uainn | Prep; | from us |
| uaibh | Prep; | from you |
| uathu | Prep; | from them |
| romham | Prep; | from me |
| romhat | Prep; | from you |
| roimhe | Prep; | from him |
| roimpi | Prep; | from her |
| romhainn | Prep; | from us |
| romhaibh | Prep; | from you |
| rompu | Prep; | from them |
| seacham | Prep; | other than me |
| seachad | Prep; | other than you |
| seacha | Prep; | other than him/her |
| seachainn | Prep; | other than us |
| seachaibh | Prep; | other than you |
| seacha | Prep; | other than them |
| tharam | Prep; | over me |
| tharat | Prep; | over you |
| thairis | Prep; | over him |
| thairsti | Prep; | over her |
| tharainn | Prep; | over us |
| tharaibh | Prep; | over you |
| tharstu | Prep; | over them |
| tríom | Prep; | through me |
| tríot | Prep; | through you |
| tríd | Prep; | through him |
| tríthí | Prep; | through her |
| trínn | Prep; | through us |
| tríbh | Prep; | through you |
| tríothu | Prep; | through them |
| umam | Prep; | about me |
| umat | Prep; | about you |
| uime | Prep; | about him |
| uimpi | Prep; | about her |
| umainn | Prep; | about us |
| umaibh | Prep; | about you |

| Preposition | Lexicon Class | Gloss |
|---|---|---|
| umpu | Prep; | about them |
| sa | Prep; | ! |
| san | Prep; | |
| sna | Prep; | |
| ó | Prep; | |
| ón | Prep; | from the |
| i | Prep; | in |
| in | Prep; | in |
| ina | Prep; | ina lámh |
| inár | Prep; | |
| inar | Prep; | |
| lena | Prep; | |
| ins | Prep; | |
| den | Prep; | from the |
| don | Prep; | to the |
| faoin | Prep; | |
| ar chúl | Prep; | |
| ar feadh | Prep; | |
| ar fud | Prep; | |
| ar nós | Prep; | |
| ar son | Prep; | |
| de bharr | Prep; | |
| de réir | Prep; | |
| faoi choinne | Prep; | |
| faoi dhéin | Prep; | |
| go ceann | Prep; | |
| i dteannta | Prep; | |
| i gcaitheamh | Prep; | |
| i gceann | Prep; | |
| i gcóir | Prep; | |
| i ndiaidh | Prep; | |
| i rith | Prep; | |
| in aghaidh | Prep; | |
| in aice | Prep; | |
| in áit | Prep; | |
| in éadan | Prep; | |
| le cois | Prep; | |
| le hais | Prep; | |
| os cionn | Prep; | |
| os comhair | Prep; | |
| tar éis | Prep; | |
| ama | Prep; | |
| faoina | Prep; | |
| lena | Prep; | |
| ina | Prep; | ina |
| óna | Prep; | |
| dár | Prep; | |
| maidir | Prep; | |

| Conjunction | Lexicon Class | Gloss |
|---|---|---|
| mar | Conj; | |
| ach | Conj; | |
| agus | Conj; | |
| arae | Conj; | |
| cé | Conj; | |

| Conjunction | Lexicon Class | Gloss |
|---|---|---|
| dá | Conj; | |
| go | Conj; | |
| gur | Conj; | |
| má | Conj; | |
| mura | Conj; | |
| murach | Conj; | |
| ná | Conj; | |
| ná | Conj; | |
| nach | Conj; | |
| nár | Conj; | |
| nó | Conj; | |
| nuair | Conj; | |
| óir | Conj; | |
| ráite | Conj; | |
| sula | Conj; | |

| Numeral | Lexicon Class | Gloss |
|---|---|---|
| aon | Num; | |
| dó | Num; | |
| trí | Num; | |
| ceathair | Num; | |
| cúig | Num; | |
| sé | Num; | |
| seacht | Num; | |
| ocht | Num; | |
| naoi | Num; | |
| deich | Num; | |
| dó | Num; | |
| dó | Num; | |
| ceathair | Num; | |
| céad | Num; | |
| dó | Num; | |
| trí | Num; | |
| aon | Num; | |
| dó | Num; | |
| trí | Num; | |
| ceathair | Num; | |
| cúig | Num; | |
| sé | Num; | |
| seacht | Num; | |
| ocht | Num; | |
| naoi | Num; | |
| deich | Num; | |

| Interjection | Lexicon Class | Gloss |
|---|---|---|
| á | Itj; | |
| abú | Itj; | |
| áiméan | Itj; | |
| ambaiste | Itj; | |
| bhuel | Itj; | |
| faraor | Itj; | |
| hurá | Itj; | |
| leoga | Itj; | |
| monuar | Itj; | |
| muise | Itj; | |
| ó | Itj; | |
| och | Itj; | |
| ochón | Itj; | |
| oró | Itj; | |

| Particle | Lexicon Class | Gloss |
|---|---|---|
| a | Part; | a Shíle |
| a | Part; | a haon, a dó |
| a | Part; | |
| a | Part; | uisce a ól (in vn phrase) |
| a | Part; | a géire a leabhair sé |
| ar | Part; | ar sise, ar seisean |
| arsa | Part; | arsa Síle |
| dar | Part; | dar léi impressionistic |
| dóbair | Part; | dóbair dó (almost) |
| ag | Part; | ag rith |
| go | Part; | go tobann, go deo, go holc |
| i | Part; | i gcónaí |
| ní | Part; | e.g. Ní Ghráda |
| uí | Part; | e.g. Uí Ghráda |
| ó | Part; | e.g. Ó Gráda |
| mac | Part; | e.g. Mac Griana |
| de | Part; | e.g. de Burgo |
| la | Part; | e.g. la Fontaine |
| le | Part; | e.g. le Clézio |
| a | Part; | an cailín a dól an deoch |
| ar | Part; | gach ar cheannaigh sé |
| an | Part; | an raibh tú ann |
| an | Part; | an mbeidh tú ann? an ólfaidh tú é |
| an | Part; | an bhfuil ? |
| an | Part; | an mbeadh sé réidh ? |
| cha | Part; | |
| chan | Part; | |
| nach | Part; | |
| ní | Part; | |
| go | Part; | |
| ar | Part; | ar chríochnaigh tú é ? |
| char | Part; | |
| go | Part; | |
| nár | Part; | |
| níor | Part; | |

| Particle | Lexicon Class | Gloss |
|---|---|---|
| ná | Part; | ná déan |
| mura | Part; | |
| mura | Part; | |
| má | Part; | |
| dá | Part; | |
| sula | Part; | |
| sular | Part; | |
| níos | Part; | |
| is | Part; | |
| chomh | Part; | |

| Abbreviation | Lexicon Class | Gloss |
|---|---|---|
| rté | Abr; | RTÉ |
| lch | Abr; | leathanach |
| uimh | Abr; | uimhir |
| dr | Abr | dochtúir |

# Appendix J

## 1000 Most Frequently Used Word-types in Corpus Náisiúnta na Gaeilge (ITÉ)

Corpus Náisiúnta na Gaeilge, a corpus of contemporary Irish texts, currently has approximately 14,800,000 tokens (280,000 types)

This list of the 1000 most frequently used words in the corpus was created using WordSmith Tools (Developer: Mike Smith, Distributor: Oxford University Press)

**N**: frequency rank order, e.g. 1 = most frequently found word
**Freq**: number of occurences in the corpus
**%**: percentage of corpus (freq./no. of tokens)*100, e.g. (671,353/14,800,000)*100 = 4.54

| N | Word | Freq. | % | N | Word | Freq. | % |
|---|---|---|---|---|---|---|---|
| 1. | a | 671,353 | 4.54 | 60. | aige | 24,847 | 0.17 |
| 2. | an | 642,192 | 4.35 | 61. | gach | 23,482 | 0.16 |
| 3. | ar | 365,885 | 2.48 | 62. | mór | 22,999 | 0.16 |
| 4. | agus | 338,380 | 2.29 | 63. | chur | 22,725 | 0.15 |
| 5. | na | 296,528 | 2.01 | 64. | á | 22,180 | 0.15 |
| 6. | go | 220,802 | 1.49 | 65. | ón | 21,692 | 0.15 |
| 7. | i | 203,829 | 1.38 | 66. | anois | 21,235 | 0.14 |
| 8. | ag | 199,736 | 1.35 | 67. | chomh | 21,210 | 0.14 |
| 9. | le | 156,032 | 1.06 | 68. | rud | 21,107 | 0.14 |
| 10. | is | 155,267 | 1.05 | 69. | chuid | 20,838 | 0.14 |
| 11. | sé | 138,461 | 0.94 | 70. | isteach | 20,234 | 0.14 |
| 12. | bhí | 124,349 | 0.84 | 71. | maith | 20,119 | 0.14 |
| 13. | é | 103,315 | 0.70 | 72. | dhéanamh | 19,659 | 0.13 |
| 14. | sin | 100,978 | 0.68 | 73. | lá | 19,558 | 0.13 |
| 15. | tá | 86,722 | 0.59 | 74. | níos | 18,726 | 0.13 |
| 16. | de | 85,893 | 0.58 | 75. | amháin | 18,428 | 0.12 |
| 17. | sa | 84,581 | 0.57 | 76. | níl | 18,231 | 0.12 |
| 18. | ach | 83,996 | 0.57 | 77. | daoine | 17,828 | 0.12 |
| 19. | mar | 82,116 | 0.56 | 78. | gaeilge | 17,778 | 0.12 |
| 20. | seo | 80,497 | 0.54 | 79. | chéile | 17,293 | 0.12 |
| 21. | ní | 78,407 | 0.53 | 80. | bith | 16,783 | 0.11 |
| 22. | ó | 76,518 | 0.52 | 81. | dtí | 16,340 | 0.11 |
| 23. | leis | 74,217 | 0.50 | 82. | cé | 15,883 | 0.11 |
| 24. | in | 73,402 | 0.50 | 83. | dul | 15,864 | 0.11 |
| 25. | raibh | 62,929 | 0.43 | 84. | liom | 15,758 | 0.11 |
| 26. | nó | 60,308 | 0.41 | 85. | dó | 15,739 | 0.11 |
| 27. | ina | 57,193 | 0.39 | 86. | níor | 15,717 | 0.11 |
| 28. | do | 55,569 | 0.38 | 87. | má | 15,564 | 0.11 |
| 29. | atá | 54,801 | 0.37 | 88. | áit | 15,468 | 0.10 |
| 30. | féin | 52,243 | 0.35 | 89. | chuir | 15,272 | 0.10 |
| 31. | bhfuil | 49,247 | 0.33 | 90. | idir | 15,224 | 0.10 |
| 32. | nach | 47,900 | 0.32 | 91. | mbeadh | 15,157 | 0.10 |
| 33. | ann | 47,252 | 0.32 | 92. | bheadh | 14,906 | 0.10 |
| 34. | mé | 45,655 | 0.31 | 93. | lena | 14,757 | 0.10 |
| 35. | as | 44,736 | 0.30 | 94. | orthu | 14,572 | 0.10 |
| 36. | faoi | 43,871 | 0.30 | 95. | dúirt | 14,547 | 0.10 |
| 37. | gur | 42,582 | 0.29 | 96. | dhá | 14,491 | 0.10 |
| 38. | eile | 42,448 | 0.29 | 97. | sna | 14,451 | 0.10 |
| 39. | sí | 41,535 | 0.28 | 98. | tháinig | 14,332 | 0.10 |
| 40. | ná | 39,952 | 0.27 | 99. | trí | 14,308 | 0.10 |
| 41. | chun | 38,822 | 0.26 | 100. | faoin | 14,250 | 0.10 |
| 42. | aon | 38,211 | 0.26 | 101. | agam | 14,089 | 0.10 |
| 43. | dá | 38,135 | 0.26 | 102. | cén | 13,962 | 0.09 |
| 44. | siad | 36,058 | 0.24 | 103. | ansin | 13,774 | 0.09 |
| 45. | den | 36,040 | 0.24 | 104. | thug | 13,606 | 0.09 |
| 46. | ba | 35,662 | 0.24 | 105. | féidir | 13,494 | 0.09 |
| 47. | nuair | 34,250 | 0.23 | 106. | mó | 13,420 | 0.09 |
| 48. | air | 32,584 | 0.22 | 107. | fáil | 13,289 | 0.09 |
| 49. | iad | 32,444 | 0.22 | 108. | fear | 12,971 | 0.09 |
| 50. | bheith | 30,175 | 0.20 | 109. | arsa | 12,895 | 0.09 |
| 51. | amach | 29,041 | 0.20 | 110. | leo | 12,684 | 0.09 |
| 52. | san | 27,667 | 0.19 | 111. | chéad | 12,667 | 0.09 |
| 53. | duine | 27,576 | 0.19 | 112. | beidh | 12,611 | 0.09 |
| 54. | acu | 27,303 | 0.18 | 113. | fad | 12,449 | 0.08 |
| 55. | don | 26,689 | 0.18 | 114. | bhíonn | 12,352 | 0.08 |
| 56. | tú | 26,342 | 0.18 | 115. | síos | 12,090 | 0.08 |
| 57. | gan | 25,873 | 0.18 | 116. | beag | 12,001 | 0.08 |
| 58. | mo | 25,868 | 0.18 | 117. | cad | 11,895 | 0.08 |
| 59. | í | 25,498 | 0.17 | 118. | mac | 11,839 | 0.08 |

| N | Word | Freq. | % | N | Word | Freq. | % |
|---|---|---|---|---|---|---|---|
| 119. | ceann | 11,826 | 0.08 | 178. | fós | 7,730 | 0.05 |
| 120. | cuid | 11,734 | 0.08 | 179. | bhaile | 7,678 | 0.05 |
| 121. | arís | 11,678 | 0.08 | 180. | léi | 7,646 | 0.05 |
| 122. | cur | 11,631 | 0.08 | 181. | dom | 7,527 | 0.05 |
| 123. | scéal | 11,573 | 0.08 | 182. | iarraidh | 7,471 | 0.05 |
| 124. | thabhairt | 11,378 | 0.08 | 183. | suas | 7,434 | 0.05 |
| 125. | am | 11,140 | 0.08 | 184. | thart | 7,417 | 0.05 |
| 126. | bíonn | 11,074 | 0.07 | 185. | fháil | 7,355 | 0.05 |
| 127. | nár | 10,976 | 0.07 | 186. | ceart | 7,254 | 0.05 |
| 128. | deir | 10,894 | 0.07 | 187. | maidir | 7,139 | 0.05 |
| 129. | rá | 10,810 | 0.07 | 188. | héireann | 7,120 | 0.05 |
| 130. | nua | 10,669 | 0.07 | 189. | di | 7,088 | 0.05 |
| 131. | réir | 10,600 | 0.07 | 190. | teanga | 7,065 | 0.05 |
| 132. | tar | 10,588 | 0.07 | 191. | teach | 7,058 | 0.05 |
| 133. | teacht | 10,485 | 0.07 | 192. | bheidh | 7,013 | 0.05 |
| 134. | éis | 10,422 | 0.07 | 193. | láthair | 7,002 | 0.05 |
| 135. | deireadh | 10,207 | 0.07 | 194. | b'fhéidir | 6,902 | 0.05 |
| 136. | rinne | 10,105 | 0.07 | 195. | os | 6,896 | 0.05 |
| 137. | leat | 10,055 | 0.07 | 196. | bliana | 6,816 | 0.05 |
| 138. | átha | 10,041 | 0.07 | 197. | dar | 6,816 | 0.05 |
| 139. | uirthi | 10,022 | 0.07 | 198. | déanamh | 6,806 | 0.05 |
| 140. | fhios | 10,021 | 0.07 | 199. | fuair | 6,805 | 0.05 |
| 141. | bhliain | 9,748 | 0.07 | 200. | úsáid | 6,804 | 0.05 |
| 142. | linn | 9,629 | 0.07 | 201. | gcuid | 6,700 | 0.05 |
| 143. | agat | 9,567 | 0.06 | 202. | uisce | 6,665 | 0.05 |
| 144. | uair | 9,566 | 0.06 | 203. | tír | 6,449 | 0.04 |
| 145. | taobh | 9,564 | 0.06 | 204. | anuas | 6,447 | 0.04 |
| 146. | aghaidh | 9,558 | 0.06 | 205. | chuig | 6,430 | 0.04 |
| 147. | obair | 9,535 | 0.06 | 206. | cheann | 6,429 | 0.04 |
| 148. | cliath | 9,527 | 0.06 | 207. | gcónaí | 6,417 | 0.04 |
| 149. | fíor | 9,434 | 0.06 | 208. | seisean | 6,403 | 0.04 |
| 150. | saol | 9,337 | 0.06 | 209. | siúl | 6,401 | 0.04 |
| 151. | anseo | 9,301 | 0.06 | 210. | lch | 6,400 | 0.04 |
| 152. | lucht | 9,261 | 0.06 | 211. | cuireadh | 6,399 | 0.04 |
| 153. | orm | 9,093 | 0.06 | 212. | dé | 6,394 | 0.04 |
| 154. | scríobh | 9,065 | 0.06 | 213. | áirithe | 6,382 | 0.04 |
| 155. | leor | 9,060 | 0.06 | 214. | mbíonn | 6,381 | 0.04 |
| 156. | againn | 9,001 | 0.06 | 215. | seán | 6,376 | 0.04 |
| 157. | bhíodh | 8,901 | 0.06 | 216. | ghaeilge | 6,357 | 0.04 |
| 158. | leith | 8,757 | 0.06 | 217. | roimh | 6,348 | 0.04 |
| 159. | mhaith | 8,757 | 0.06 | 218. | muid | 6,329 | 0.04 |
| 160. | céanna | 8,651 | 0.06 | 219. | ais | 6,310 | 0.04 |
| 161. | oíche | 8,519 | 0.06 | 220. | más | 6,258 | 0.04 |
| 162. | dóibh | 8,473 | 0.06 | 221. | ort | 6,240 | 0.04 |
| 163. | aici | 8,453 | 0.06 | 222. | ár | 6,204 | 0.04 |
| 164. | thar | 8,446 | 0.06 | 223. | méid | 6,175 | 0.04 |
| 165. | mhór | 8,445 | 0.06 | 224. | bhain | 6,136 | 0.04 |
| 166. | leabhar | 8,360 | 0.06 | 225. | déanta | 6,097 | 0.04 |
| 167. | baile | 8,317 | 0.06 | 226. | roinnt | 6,047 | 0.04 |
| 168. | fada | 8,310 | 0.06 | 227. | fiú | 5,993 | 0.04 |
| 169. | riamh | 8,244 | 0.06 | 228. | minic | 5,956 | 0.04 |
| 170. | chuaigh | 8,205 | 0.06 | 229. | ndiaidh | 5,930 | 0.04 |
| 171. | mbeidh | 8,116 | 0.05 | 230. | dhuine | 5,905 | 0.04 |
| 172. | freisin | 8,076 | 0.05 | 231. | dhiaidh | 5,902 | 0.04 |
| 173. | bliain | 8,009 | 0.05 | 232. | fearr | 5,891 | 0.04 |
| 174. | mise | 7,952 | 0.05 | 233. | siar | 5,867 | 0.04 |
| 175. | léir | 7,890 | 0.05 | 234. | oiread | 5,856 | 0.04 |
| 176. | éigin | 7,743 | 0.05 | 235. | leanas | 5,834 | 0.04 |
| 177. | ea | 7,733 | 0.05 | 236. | uí | 5,798 | 0.04 |

| N | Word | Freq. | % | N | Word | Freq. | % |
|---|---|---|---|---|---|---|---|
| 237. | éirinn | 5,779 | 0.04 | 296. | dóigh | 4,338 | 0.03 |
| 238. | feadh | 5,650 | 0.04 | 297. | eolas | 4,336 | 0.03 |
| 239. | duit | 5,574 | 0.04 | 298. | rialtas | 4,311 | 0.03 |
| 240. | alt | 5,495 | 0.04 | 299. | gcás | 4,304 | 0.03 |
| 241. | dara | 5,488 | 0.04 | 300. | máire | 4,264 | 0.03 |
| 242. | lán | 5,482 | 0.04 | 301. | lae | 4,224 | 0.03 |
| 243. | thaobh | 5,470 | 0.04 | 302. | bíodh | 4,218 | 0.03 |
| 244. | cinn | 5,469 | 0.04 | 303. | siúd | 4,196 | 0.03 |
| 245. | measc | 5,426 | 0.04 | 304. | tíre | 4,181 | 0.03 |
| 246. | súil | 5,409 | 0.04 | 305. | chonaic | 4,180 | 0.03 |
| 247. | gceist | 5,397 | 0.04 | 306. | thosaigh | 4,173 | 0.03 |
| 248. | uile | 5,382 | 0.04 | 307. | cuir | 4,167 | 0.03 |
| 249. | díreach | 5,345 | 0.04 | 308. | líon | 4,149 | 0.03 |
| 250. | óg | 5,339 | 0.04 | 309. | díobh | 4,145 | 0.03 |
| 251. | cinnte | 5,274 | 0.04 | 310. | háirithe | 4,136 | 0.03 |
| 252. | shin | 5,253 | 0.04 | 311. | sásta | 4,095 | 0.03 |
| 253. | nós | 5,204 | 0.04 | 312. | iontu | 4,066 | 0.03 |
| 254. | bhfad | 5,196 | 0.04 | 313. | éagsúla | 4,065 | 0.03 |
| 255. | oibre | 5,192 | 0.04 | 314. | bheag | 4,061 | 0.03 |
| 256. | faigh | 5,191 | 0.04 | 315. | haghaidh | 4,027 | 0.03 |
| 257. | dúinn | 5,185 | 0.04 | 316. | roinn | 4,020 | 0.03 |
| 258. | istigh | 5,171 | 0.03 | 317. | áfach | 4,006 | 0.03 |
| 259. | céard | 5,086 | 0.03 | 318. | náisiúnta | 4,006 | 0.03 |
| 260. | tharla | 5,061 | 0.03 | 319. | líne | 3,985 | 0.03 |
| 261. | cúpla | 5,041 | 0.03 | 320. | mura | 3,983 | 0.03 |
| 262. | bharr | 5,027 | 0.03 | 321. | d'éirigh | 3,961 | 0.03 |
| 263. | timpeall | 4,983 | 0.03 | 322. | éirí | 3,958 | 0.03 |
| 264. | gurb | 4,960 | 0.03 | 323. | chéanna | 3,952 | 0.03 |
| 265. | scoil | 4,958 | 0.03 | 324. | thíos | 3,906 | 0.03 |
| 266. | caint | 4,955 | 0.03 | 325. | airgid | 3,895 | 0.03 |
| 267. | amuigh | 4,954 | 0.03 | 326. | cosúil | 3,894 | 0.03 |
| 268. | fáth | 4,861 | 0.03 | 327. | dhaoine | 3,840 | 0.03 |
| 269. | lár | 4,843 | 0.03 | 328. | rí | 3,833 | 0.03 |
| 270. | ainm | 4,840 | 0.03 | 329. | airde | 3,826 | 0.03 |
| 271. | mbaile | 4,813 | 0.03 | 330. | fud | 3,826 | 0.03 |
| 272. | ábhar | 4,766 | 0.03 | 331. | focal | 3,814 | 0.03 |
| 273. | cheart | 4,749 | 0.03 | 332. | chóir | 3,812 | 0.03 |
| 274. | ndóigh | 4,736 | 0.03 | 333. | thiar | 3,807 | 0.03 |
| 275. | uaidh | 4,721 | 0.03 | 334. | ionad | 3,798 | 0.03 |
| 276. | bhean | 4,718 | 0.03 | 335. | mná | 3,796 | 0.03 |
| 277. | mórán | 4,710 | 0.03 | 336. | talamh | 3,781 | 0.03 |
| 278. | bhaint | 4,699 | 0.03 | 337. | dia | 3,780 | 0.03 |
| 279. | bean | 4,695 | 0.03 | 338. | doras | 3,778 | 0.03 |
| 280. | leithéid | 4,687 | 0.03 | 339. | chuala | 3,777 | 0.03 |
| 281. | ab | 4,667 | 0.03 | 340. | gaeltachta | 3,770 | 0.03 |
| 282. | rith | 4,628 | 0.03 | 341. | gurbh | 3,755 | 0.03 |
| 283. | chuige | 4,619 | 0.03 | 342. | bás | 3,744 | 0.03 |
| 284. | tríd | 4,601 | 0.03 | 343. | cúrsaí | 3,740 | 0.03 |
| 285. | tí | 4,595 | 0.03 | 344. | aire | 3,736 | 0.03 |
| 286. | inniu | 4,570 | 0.03 | 345. | féach | 3,732 | 0.03 |
| 287. | chaith | 4,539 | 0.03 | 346. | chaoi | 3,715 | 0.03 |
| 288. | lámh | 4,522 | 0.03 | 347. | abhaile | 3,684 | 0.02 |
| 289. | baint | 4,511 | 0.03 | 348. | luach | 3,678 | 0.02 |
| 290. | gá | 4,491 | 0.03 | 349. | mbliana | 3,673 | 0.02 |
| 291. | cá | 4,438 | 0.03 | 350. | shampla | 3,669 | 0.02 |
| 292. | cheana | 4,430 | 0.03 | 351. | iomlán | 3,639 | 0.02 |
| 293. | roimhe | 4,426 | 0.03 | 352. | stáit | 3,638 | 0.02 |
| 294. | uimh | 4,370 | 0.03 | 353. | úd | 3,613 | 0.02 |
| 295. | caite | 4,349 | 0.03 | 354. | cuma | 3,609 | 0.02 |

| N | Word | Freq. | % |
|---|---|---|---|
| 355. | mí | 3,582 | 0.02 |
| 356. | leagan | 3,572 | 0.02 |
| 357. | bhaineann | 3,555 | 0.02 |
| 358. | déan | 3,537 | 0.02 |
| 359. | muintir | 3,521 | 0.02 |
| 360. | arna | 3,520 | 0.02 |
| 361. | maidin | 3,517 | 0.02 |
| 362. | mhuintir | 3,462 | 0.02 |
| 363. | tabhairt | 3,461 | 0.02 |
| 364. | seachas | 3,455 | 0.02 |
| 365. | imeacht | 3,424 | 0.02 |
| 366. | curtha | 3,415 | 0.02 |
| 367. | bun | 3,413 | 0.02 |
| 368. | thuas | 3,371 | 0.02 |
| 369. | bia | 3,369 | 0.02 |
| 370. | domhain | 3,340 | 0.02 |
| 371. | pádraig | 3,332 | 0.02 |
| 372. | móra | 3,330 | 0.02 |
| 373. | mháthair | 3,309 | 0.02 |
| 374. | beo | 3,297 | 0.02 |
| 375. | rudaí | 3,289 | 0.02 |
| 376. | conas | 3,281 | 0.02 |
| 377. | caithfidh | 3,274 | 0.02 |
| 378. | blianta | 3,257 | 0.02 |
| 379. | fir | 3,240 | 0.02 |
| 380. | cionn | 3,239 | 0.02 |
| 381. | tomás | 3,205 | 0.02 |
| 382. | chuile | 3,192 | 0.02 |
| 383. | breá | 3,190 | 0.02 |
| 384. | déag | 3,187 | 0.02 |
| 385. | tugadh | 3,182 | 0.02 |
| 386. | cúig | 3,164 | 0.02 |
| 387. | amhlaidh | 3,142 | 0.02 |
| 388. | feasta | 3,127 | 0.02 |
| 389. | oideachais | 3,120 | 0.02 |
| 390. | ard | 3,104 | 0.02 |
| 391. | té | 3,094 | 0.02 |
| 392. | mhic | 3,092 | 0.02 |
| 393. | bhealach | 3,086 | 0.02 |
| 394. | aimsir | 3,082 | 0.02 |
| 395. | tamall | 3,081 | 0.02 |
| 396. | feirste | 3,078 | 0.02 |
| 397. | athair | 3,075 | 0.02 |
| 398. | d'fhág | 3,075 | 0.02 |
| 399. | faoina | 3,057 | 0.02 |
| 400. | gheall | 3,055 | 0.02 |
| 401. | bheirt | 3,054 | 0.02 |
| 402. | phobail | 3,048 | 0.02 |
| 403. | tús | 3,040 | 0.02 |
| 404. | thuaidh | 3,034 | 0.02 |
| 405. | inti | 3,033 | 0.02 |
| 406. | mba | 3,030 | 0.02 |
| 407. | airgead | 3,025 | 0.02 |
| 408. | domhan | 3,024 | 0.02 |
| 409. | labhairt | 3,021 | 0.02 |
| 410. | sibh | 3,014 | 0.02 |
| 411. | orainn | 3,011 | 0.02 |
| 412. | chor | 2,997 | 0.02 |
| 413. | seomra | 2,994 | 0.02 |
| 414. | shaol | 2,986 | 0.02 |
| 415. | thú | 2,985 | 0.02 |
| 416. | sise | 2,974 | 0.02 |
| 417. | chroí | 2,959 | 0.02 |
| 418. | níorbh | 2,959 | 0.02 |
| 419. | athrú | 2,952 | 0.02 |
| 420. | chaitheamh | 2,937 | 0.02 |
| 421. | fanacht | 2,931 | 0.02 |
| 422. | beirt | 2,927 | 0.02 |
| 423. | phobal | 2,906 | 0.02 |
| 424. | seans | 2,893 | 0.02 |
| 425. | acht | 2,878 | 0.02 |
| 426. | dream | 2,871 | 0.02 |
| 427. | dtaobh | 2,866 | 0.02 |
| 428. | tráth | 2,864 | 0.02 |
| 429. | deara | 2,857 | 0.02 |
| 430. | bhfeidhm | 2,843 | 0.02 |
| 431. | sórt | 2,812 | 0.02 |
| 432. | láidir | 2,810 | 0.02 |
| 433. | ceisteanna | 2,807 | 0.02 |
| 434. | seachtaine | 2,805 | 0.02 |
| 435. | dtús | 2,803 | 0.02 |
| 436. | léamh | 2,799 | 0.02 |
| 437. | feiceáil | 2,793 | 0.02 |
| 438. | son | 2,789 | 0.02 |
| 439. | clár | 2,788 | 0.02 |
| 440. | leabhair | 2,776 | 0.02 |
| 441. | tuairim | 2,770 | 0.02 |
| 442. | tosaigh | 2,769 | 0.02 |
| 443. | liam | 2,764 | 0.02 |
| 444. | ball | 2,763 | 0.02 |
| 445. | scéalta | 2,763 | 0.02 |
| 446. | dubh | 2,759 | 0.02 |
| 447. | sagart | 2,755 | 0.02 |
| 448. | heorpa | 2,751 | 0.02 |
| 449. | toisc | 2,750 | 0.02 |
| 450. | laghad | 2,748 | 0.02 |
| 451. | léiriú | 2,744 | 0.02 |
| 452. | beaga | 2,741 | 0.02 |
| 453. | bán | 2,732 | 0.02 |
| 454. | breathnú | 2,731 | 0.02 |
| 455. | iarracht | 2,728 | 0.02 |
| 456. | d'imigh | 2,726 | 0.02 |
| 457. | fios | 2,723 | 0.02 |
| 458. | tráthnóna | 2,718 | 0.02 |
| 459. | raidió | 2,713 | 0.02 |
| 460. | éireann | 2,711 | 0.02 |
| 461. | aithne | 2,705 | 0.02 |
| 462. | gabháil | 2,705 | 0.02 |
| 463. | tusa | 2,699 | 0.02 |
| 464. | ceol | 2,698 | 0.02 |
| 465. | óga | 2,678 | 0.02 |
| 466. | luath | 2,673 | 0.02 |
| 467. | mbun | 2,671 | 0.02 |
| 468. | dtreo | 2,669 | 0.02 |
| 469. | cead | 2,667 | 0.02 |
| 470. | mhí | 2,667 | 0.02 |
| 471. | tréimhse | 2,665 | 0.02 |
| 472. | ama | 2,662 | 0.02 |

| N | Word | Freq. | % |
|---|---|---|---|
| 473. | slán | 2,645 | 0.02 |
| 474. | thuig | 2,630 | 0.02 |
| 475. | ceithre | 2,629 | 0.02 |
| 476. | páistí | 2,624 | 0.02 |
| 477. | thóg | 2,620 | 0.02 |
| 478. | réiteach | 2,618 | 0.02 |
| 479. | dócha | 2,617 | 0.02 |
| 480. | bhéal | 2,615 | 0.02 |
| 481. | lorg | 2,613 | 0.02 |
| 482. | sula | 2,603 | 0.02 |
| 483. | cuireann | 2,600 | 0.02 |
| 484. | leath | 2,599 | 0.02 |
| 485. | bhuail | 2,591 | 0.02 |
| 486. | bord | 2,591 | 0.02 |
| 487. | scoile | 2,563 | 0.02 |
| 488. | rinneadh | 2,543 | 0.02 |
| 489. | réidh | 2,539 | 0.02 |
| 490. | ionann | 2,538 | 0.02 |
| 491. | dlí | 2,536 | 0.02 |
| 492. | fóill | 2,531 | 0.02 |
| 493. | béal | 2,527 | 0.02 |
| 494. | chineál | 2,525 | 0.02 |
| 495. | óna | 2,513 | 0.02 |
| 496. | deo | 2,506 | 0.02 |
| 497. | gcéad | 2,506 | 0.02 |
| 498. | féachaint | 2,503 | 0.02 |
| 499. | nollaig | 2,483 | 0.02 |
| 500. | súile | 2,480 | 0.02 |
| 501. | fómhair | 2,473 | 0.02 |
| 502. | céad | 2,472 | 0.02 |
| 503. | saothar | 2,463 | 0.02 |
| 504. | ngach | 2,452 | 0.02 |
| 505. | cibé | 2,450 | 0.02 |
| 506. | bhrí | 2,444 | 0.02 |
| 507. | bealach | 2,442 | 0.02 |
| 508. | stát | 2,436 | 0.02 |
| 509. | sampla | 2,431 | 0.02 |
| 510. | haon | 2,424 | 0.02 |
| 511. | um | 2,423 | 0.02 |
| 512. | chugainn | 2,416 | 0.02 |
| 513. | fhear | 2,415 | 0.02 |
| 514. | lú | 2,409 | 0.02 |
| 515. | míle | 2,409 | 0.02 |
| 516. | ceist | 2,407 | 0.02 |
| 517. | dhó | 2,407 | 0.02 |
| 518. | rogha | 2,404 | 0.02 |
| 519. | tabhair | 2,399 | 0.02 |
| 520. | labhair | 2,394 | 0.02 |
| 521. | béarla | 2,392 | 0.02 |
| 522. | oifig | 2,388 | 0.02 |
| 523. | deis | 2,386 | 0.02 |
| 524. | mhéad | 2,378 | 0.02 |
| 525. | eolais | 2,375 | 0.02 |
| 526. | cheist | 2,374 | 0.02 |
| 527. | iúl | 2,372 | 0.02 |
| 528. | ré | 2,369 | 0.02 |
| 529. | loch | 2,368 | 0.02 |
| 530. | d'fhéadfadh | 2,366 | 0.02 |
| 531. | déanaí | 2,361 | 0.02 |
| 532. | sinn | 2,356 | 0.02 |
| 533. | stair | 2,354 | 0.02 |
| 534. | thall | 2,354 | 0.02 |
| 535. | tseachtain | 2,353 | 0.02 |
| 536. | léinn | 2,351 | 0.02 |
| 537. | neart | 2,351 | 0.02 |
| 538. | fhéin | 2,349 | 0.02 |
| 539. | dán | 2,348 | 0.02 |
| 540. | aird | 2,339 | 0.02 |
| 541. | gceann | 2,329 | 0.02 |
| 542. | mbíodh | 2,324 | 0.02 |
| 543. | gcomhairle | 2,303 | 0.02 |
| 544. | imithe | 2,293 | 0.02 |
| 545. | gáire | 2,292 | 0.02 |
| 546. | gnó | 2,292 | 0.02 |
| 547. | pobal | 2,290 | 0.02 |
| 548. | casadh | 2,282 | 0.02 |
| 549. | aer | 2,281 | 0.02 |
| 550. | litir | 2,281 | 0.02 |
| 551. | chontae | 2,275 | 0.02 |
| 552. | t-am | 2,263 | 0.02 |
| 553. | uimhir | 2,255 | 0.02 |
| 554. | cois | 2,251 | 0.02 |
| 555. | cineál | 2,247 | 0.02 |
| 556. | uilig | 2,247 | 0.02 |
| 557. | cheap | 2,239 | 0.02 |
| 558. | cainte | 2,238 | 0.02 |
| 559. | inis | 2,237 | 0.02 |
| 560. | amhras | 2,228 | 0.02 |
| 561. | dhéanann | 2,227 | 0.02 |
| 562. | shíl | 2,227 | 0.02 |
| 563. | tine | 2,227 | 0.02 |
| 564. | chuma | 2,222 | 0.02 |
| 565. | oileán | 2,220 | 0.02 |
| 566. | mbéal | 2,218 | 0.02 |
| 567. | aois | 2,217 | 0.02 |
| 568. | lig | 2,214 | 0.01 |
| 569. | tugann | 2,212 | 0.01 |
| 570. | iontach | 2,199 | 0.01 |
| 571. | dár | 2,197 | 0.01 |
| 572. | leaba | 2,187 | 0.01 |
| 573. | luí | 2,187 | 0.01 |
| 574. | fheiceáil | 2,183 | 0.01 |
| 575. | taire | 2,176 | 0.01 |
| 576. | bhfuair | 2,175 | 0.01 |
| 577. | seasamh | 2,173 | 0.01 |
| 578. | mír | 2,172 | 0.01 |
| 579. | seacht | 2,163 | 0.01 |
| 580. | thit | 2,160 | 0.01 |
| 581. | bóthar | 2,159 | 0.01 |
| 582. | deich | 2,151 | 0.01 |
| 583. | aice | 2,147 | 0.01 |
| 584. | bunaithe | 2,147 | 0.01 |
| 585. | lean | 2,144 | 0.01 |
| 586. | thugtar | 2,138 | 0.01 |
| 587. | sráid | 2,135 | 0.01 |
| 588. | glacadh | 2,134 | 0.01 |
| 589. | suim | 2,133 | 0.01 |
| 590. | cailín | 2,127 | 0.01 |

| N | Word | Freq. | % |
|---|---|---|---|
| 591. | eagla | 2,113 | 0.01 |
| 592. | ngaeilge | 2,100 | 0.01 |
| 593. | meán | 2,099 | 0.01 |
| 594. | aontaithe | 2,097 | 0.01 |
| 595. | chathair | 2,087 | 0.01 |
| 596. | thugann | 2,085 | 0.01 |
| 597. | ainneoin | 2,084 | 0.01 |
| 598. | choinneáil | 2,083 | 0.01 |
| 599. | chuireann | 2,081 | 0.01 |
| 600. | bocht | 2,072 | 0.01 |
| 601. | fás | 2,064 | 0.01 |
| 602. | tríú | 2,062 | 0.01 |
| 603. | fáilte | 2,057 | 0.01 |
| 604. | ndaoine | 2,055 | 0.01 |
| 605. | mhéid | 2,053 | 0.01 |
| 606. | áireamh | 2,050 | 0.01 |
| 607. | rialtais | 2,043 | 0.01 |
| 608. | úsáidtear | 2,041 | 0.01 |
| 609. | deas | 2,036 | 0.01 |
| 610. | léiríonn | 2,027 | 0.01 |
| 611. | saor | 2,017 | 0.01 |
| 612. | trasna | 2,017 | 0.01 |
| 613. | tugtha | 2,014 | 0.01 |
| 614. | nóiméad | 2,012 | 0.01 |
| 615. | tsaoil | 2,012 | 0.01 |
| 616. | cuairt | 2,011 | 0.01 |
| 617. | barr | 2,008 | 0.01 |
| 618. | mhac | 2,000 | 0.01 |
| 619. | mhó | 2,000 | 0.01 |
| 620. | croí | 1,995 | 0.01 |
| 621. | píosa | 1,992 | 0.01 |
| 622. | éisteacht | 1,990 | 0.01 |
| 623. | sea | 1,984 | 0.01 |
| 624. | grá | 1,973 | 0.01 |
| 625. | leibhéal | 1,967 | 0.01 |
| 626. | smaoineamh | 1,965 | 0.01 |
| 627. | rún | 1,955 | 0.01 |
| 628. | aníos | 1,952 | 0.01 |
| 629. | comhairle | 1,950 | 0.01 |
| 630. | deirtear | 1,944 | 0.01 |
| 631. | éirigh | 1,940 | 0.01 |
| 632. | tagann | 1,938 | 0.01 |
| 633. | treoir | 1,937 | 0.01 |
| 634. | deacair | 1,931 | 0.01 |
| 635. | thógáil | 1,931 | 0.01 |
| 636. | chúrsaí | 1,925 | 0.01 |
| 637. | uathu | 1,925 | 0.01 |
| 638. | rua | 1,920 | 0.01 |
| 639. | cathrach | 1,917 | 0.01 |
| 640. | fiche | 1,914 | 0.01 |
| 641. | caitheamh | 1,912 | 0.01 |
| 642. | dheireadh | 1,909 | 0.01 |
| 643. | trácht | 1,903 | 0.01 |
| 644. | file | 1,897 | 0.01 |
| 645. | meas | 1,897 | 0.01 |
| 646. | tada | 1,894 | 0.01 |
| 647. | thagann | 1,881 | 0.01 |
| 648. | triúr | 1,880 | 0.01 |
| 649. | hé | 1,877 | 0.01 |
| 650. | deimhin | 1,876 | 0.01 |
| 651. | tagairt | 1,874 | 0.01 |
| 652. | speisialta | 1,867 | 0.01 |
| 653. | talún | 1,866 | 0.01 |
| 654. | bí | 1,848 | 0.01 |
| 655. | córas | 1,848 | 0.01 |
| 656. | táim | 1,846 | 0.01 |
| 657. | ól | 1,845 | 0.01 |
| 658. | micheál | 1,842 | 0.01 |
| 659. | greim | 1,838 | 0.01 |
| 660. | leas | 1,838 | 0.01 |
| 661. | fágtha | 1,835 | 0.01 |
| 662. | leasú | 1,835 | 0.01 |
| 663. | bhuel | 1,830 | 0.01 |
| 664. | lámha | 1,827 | 0.01 |
| 665. | trá | 1,826 | 0.01 |
| 666. | uaim | 1,826 | 0.01 |
| 667. | nárbh | 1,825 | 0.01 |
| 668. | ghlac | 1,814 | 0.01 |
| 669. | naomh | 1,814 | 0.01 |
| 670. | rialta | 1,810 | 0.01 |
| 671. | poiblí | 1,808 | 0.01 |
| 672. | tugtar | 1,807 | 0.01 |
| 673. | peadar | 1,804 | 0.01 |
| 674. | thuilleadh | 1,804 | 0.01 |
| 675. | radharc | 1,803 | 0.01 |
| 676. | chomhairle | 1,798 | 0.01 |
| 677. | chás | 1,797 | 0.01 |
| 678. | solas | 1,791 | 0.01 |
| 679. | brú | 1,788 | 0.01 |
| 680. | tithe | 1,787 | 0.01 |
| 681. | bheas | 1,785 | 0.01 |
| 682. | dáta | 1,781 | 0.01 |
| 683. | cuirtear | 1,779 | 0.01 |
| 684. | nithe | 1,779 | 0.01 |
| 685. | scoileanna | 1,778 | 0.01 |
| 686. | plé | 1,776 | 0.01 |
| 687. | fonn | 1,772 | 0.01 |
| 688. | haois | 1,771 | 0.01 |
| 689. | dtír | 1,769 | 0.01 |
| 690. | londain | 1,767 | 0.01 |
| 691. | fosta | 1,765 | 0.01 |
| 692. | bua | 1,760 | 0.01 |
| 693. | dheas | 1,758 | 0.01 |
| 694. | titim | 1,758 | 0.01 |
| 695. | uaireanta | 1,758 | 0.01 |
| 696. | éinne | 1,757 | 0.01 |
| 697. | fhada | 1,753 | 0.01 |
| 698. | breataine | 1,750 | 0.01 |
| 699. | dála | 1,749 | 0.01 |
| 700. | gcoinne | 1,746 | 0.01 |
| 701. | brath | 1,745 | 0.01 |
| 702. | focail | 1,745 | 0.01 |
| 703. | ollscoil | 1,744 | 0.01 |
| 704. | úr | 1,744 | 0.01 |
| 705. | sasana | 1,738 | 0.01 |
| 706. | laethanta | 1,731 | 0.01 |
| 707. | tom | 1,731 | 0.01 |
| 708. | fá | 1,730 | 0.01 |

| N | Word | Freq. | % |
|---|---|---|---|
| 709. | saoire | 1,730 | 0.01 |
| 710. | comhair | 1,727 | 0.01 |
| 711. | idirnáisiúnta | 1,727 | 0.01 |
| 712. | suí | 1,723 | 0.01 |
| 713. | scéil | 1,720 | 0.01 |
| 714. | amhrán | 1,717 | 0.01 |
| 715. | toradh | 1,713 | 0.01 |
| 716. | údarás | 1,713 | 0.01 |
| 717. | dath | 1,702 | 0.01 |
| 718. | ligean | 1,701 | 0.01 |
| 719. | tógadh | 1,700 | 0.01 |
| 720. | óir | 1,697 | 0.01 |
| 721. | rugadh | 1,693 | 0.01 |
| 722. | slí | 1,690 | 0.01 |
| 723. | láimh | 1,688 | 0.01 |
| 724. | fhearr | 1,687 | 0.01 |
| 725. | meiriceá | 1,686 | 0.01 |
| 726. | sláinte | 1,686 | 0.01 |
| 727. | feidhm | 1,682 | 0.01 |
| 728. | ceoil | 1,679 | 0.01 |
| 729. | freagra | 1,676 | 0.01 |
| 730. | modh | 1,676 | 0.01 |
| 731. | tuairisc | 1,676 | 0.01 |
| 732. | scannán | 1,673 | 0.01 |
| 733. | litríocht | 1,671 | 0.01 |
| 734. | dhéanfadh | 1,666 | 0.01 |
| 735. | shasana | 1,665 | 0.01 |
| 736. | airteagal | 1,663 | 0.01 |
| 737. | spéis | 1,663 | 0.01 |
| 738. | mbaineann | 1,661 | 0.01 |
| 739. | achar | 1,660 | 0.01 |
| 740. | bhfaca | 1,658 | 0.01 |
| 741. | dhul | 1,658 | 0.01 |
| 742. | hoíche | 1,657 | 0.01 |
| 743. | cás | 1,655 | 0.01 |
| 744. | bealtaine | 1,652 | 0.01 |
| 745. | gearr | 1,648 | 0.01 |
| 746. | ghnáth | 1,648 | 0.01 |
| 747. | gréine | 1,646 | 0.01 |
| 748. | eaglais | 1,644 | 0.01 |
| 749. | máthair | 1,643 | 0.01 |
| 750. | tullleadh | 1,638 | 0.01 |
| 751. | chúis | 1,637 | 0.01 |
| 752. | cuntas | 1,635 | 0.01 |
| 753. | farraige | 1,635 | 0.01 |
| 754. | freastal | 1,634 | 0.01 |
| 755. | turas | 1,632 | 0.01 |
| 756. | gceart | 1,629 | 0.01 |
| 757. | post | 1,627 | 0.01 |
| 758. | agamsa | 1,625 | 0.01 |
| 759. | éireannach | 1,624 | 0.01 |
| 760. | mhóir | 1,621 | 0.01 |
| 761. | breis | 1,619 | 0.01 |
| 762. | liomsa | 1,619 | 0.01 |
| 763. | pháirtí | 1,616 | 0.01 |
| 764. | pé | 1,615 | 0.01 |
| 765. | dearg | 1,612 | 0.01 |
| 766. | bhun | 1,604 | 0.01 |
| 767. | buí | 1,602 | 0.01 |
| 768. | seachtain | 1,601 | 0.01 |
| 769. | máirtín | 1,600 | 0.01 |
| 770. | tagtha | 1,600 | 0.01 |
| 771. | chlár | 1,593 | 0.01 |
| 772. | d'aon | 1,591 | 0.01 |
| 773. | chonradh | 1,590 | 0.01 |
| 774. | theacht | 1,590 | 0.01 |
| 775. | im | 1,588 | 0.01 |
| 776. | áitiúil | 1,587 | 0.01 |
| 777. | bhféadfadh | 1,587 | 0.01 |
| 778. | adeir | 1,586 | 0.01 |
| 779. | laistigh | 1,585 | 0.01 |
| 780. | leanbh | 1,583 | 0.01 |
| 781. | glan | 1,581 | 0.01 |
| 782. | uait | 1,581 | 0.01 |
| 783. | chaint | 1,580 | 0.01 |
| 784. | iúil | 1,576 | 0.01 |
| 785. | chuirfeadh | 1,574 | 0.01 |
| 786. | theastaigh | 1,574 | 0.01 |
| 787. | dráma | 1,569 | 0.01 |
| 788. | dearmad | 1,568 | 0.01 |
| 789. | d'aois | 1,567 | 0.01 |
| 790. | éineacht | 1,563 | 0.01 |
| 791. | sular | 1,562 | 0.01 |
| 792. | óige | 1,560 | 0.01 |
| 793. | thiocfadh | 1,556 | 0.01 |
| 794. | dún | 1,553 | 0.01 |
| 795. | seoladh | 1,552 | 0.01 |
| 796. | liosta | 1,551 | 0.01 |
| 797. | déantar | 1,550 | 0.01 |
| 798. | theach | 1,544 | 0.01 |
| 799. | pointe | 1,539 | 0.01 |
| 800. | tslí | 1,539 | 0.01 |
| 801. | forbairt | 1,538 | 0.01 |
| 802. | iomaí | 1,537 | 0.01 |
| 803. | beatha | 1,536 | 0.01 |
| 804. | airgeadais | 1,535 | 0.01 |
| 805. | gael | 1,534 | 0.01 |
| 806. | ndeireadh | 1,534 | 0.01 |
| 807. | agaibh | 1,533 | 0.01 |
| 808. | dtiocfadh | 1,529 | 0.01 |
| 809. | baineann | 1,528 | 0.01 |
| 810. | abair | 1,527 | 0.01 |
| 811. | scríofa | 1,527 | 0.01 |
| 812. | thír | 1,526 | 0.01 |
| 813. | tuiscint | 1,524 | 0.01 |
| 814. | insint | 1,523 | 0.01 |
| 815. | saoirse | 1,523 | 0.01 |
| 816. | coitianta | 1,520 | 0.01 |
| 817. | mícheál | 1,518 | 0.01 |
| 818. | b'fhearr | 1,514 | 0.01 |
| 819. | tae | 1,514 | 0.01 |
| 820. | taithí | 1,514 | 0.01 |
| 821. | bhord | 1,511 | 0.01 |
| 822. | ceapadh | 1,508 | 0.01 |
| 823. | scrúdú | 1,506 | 0.01 |
| 824. | theas | 1,506 | 0.01 |
| 825. | trua | 1,506 | 0.01 |
| 826. | moladh | 1,505 | 0.01 |

| N | Word | Freq. | % |
|---|---|---|---|
| 827. | iníon | 1,504 | 0.01 |
| 828. | bhfear | 1,503 | 0.01 |
| 829. | marbh | 1,500 | 0.01 |
| 830. | tadhg | 1,498 | 0.01 |
| 831. | cos | 1,495 | 0.01 |
| 832. | intinn | 1,487 | 0.01 |
| 833. | soiléir | 1,484 | 0.01 |
| 834. | smaointe | 1,482 | 0.01 |
| 835. | déanfaidh | 1,479 | 0.01 |
| 836. | mall | 1,478 | 0.01 |
| 837. | téann | 1,477 | |
| 838. | beagnach | 1,476 | |
| 839. | línte | 1,473 | |
| 840. | ndearna | 1,472 | |
| 841. | chúl | 1,467 | |
| 842. | dearcadh | 1,466 | |
| 843. | mhaithe | 1,466 | |
| 844. | bhíos | 1,465 | |
| 845. | eorpach | 1,465 | |
| 846. | phointe | 1,465 | |
| 847. | séamas | 1,465 | |
| 848. | sheasamh | 1,463 | |
| 849. | tarraingt | 1,462 | |
| 850. | crann | 1,461 | |
| 851. | inár | 1,461 | |
| 852. | mheas | 1,461 | |
| 853. | tionchar | 1,455 | |
| 854. | siopa | 1,454 | |
| 855. | teideal | 1,453 | |
| 856. | cló | 1,452 | |
| 857. | déanann | 1,452 | |
| 858. | foilsíodh | 1,452 | |
| 859. | seirbhísí | 1,450 | |
| 860. | shaothar | 1,449 | |
| 861. | tharraing | 1,448 | |
| 862. | eoin | 1,447 | |
| 863. | chonaill | 1,444 | |
| 864. | coláiste | 1,443 | |
| 865. | measa | 1,443 | |
| 866. | bhfíor | 1,442 | |
| 867. | dochar | 1,441 | |
| 868. | fhoireann | 1,441 | |
| 869. | conaire | 1,440 | |
| 870. | théann | 1,438 | |
| 871. | láithreach | 1,437 | |
| 872. | uladh | 1,437 | |
| 873. | aigne | 1,434 | |
| 874. | leag | 1,433 | |
| 875. | te | 1,432 | |
| 876. | deireanach | 1,431 | |
| 877. | imní | 1,431 | |
| 878. | scór | 1,431 | |
| 879. | shúile | 1,428 | |
| 880. | chlann | 1,426 | |
| 881. | t-aon | 1,426 | |
| 882. | teastáil | 1,425 | |
| 883. | aonair | 1,420 | |
| 884. | beagán | 1,419 | |
| 885. | ithe | 1,419 | |

| N | Word | Freq. | % |
|---|---|---|---|
| 886. | mian | 1,419 | |
| 887. | torthaí | 1,419 | |
| 888. | álainn | 1,418 | |
| 889. | údar | 1,418 | |
| 890. | bhonn | 1,416 | |
| 891. | nóta | 1,416 | |
| 892. | conradh | 1,415 | |
| 893. | dhún | 1,413 | |
| 894. | teilifíse | 1,412 | |
| 895. | bualadh | 1,410 | |
| 896. | imirt | 1,410 | |
| 897. | éire | 1,408 | |
| 898. | staidéar | 1,408 | |
| 899. | cúis | 1,406 | |
| 900. | stop | 1,406 | |
| 901. | scéim | 1,405 | |
| 902. | saghas | 1,401 | |
| 903. | gcomhair | 1,400 | |
| 904. | gairid | 1,399 | |
| 905. | éigean | 1,397 | |
| 906. | ndeachaigh | 1,395 | |
| 907. | ins | 1,394 | |
| 908. | tíortha | 1,393 | |
| 909. | uachtarán | 1,393 | |
| 910. | micil | 1,391 | |
| 911. | anonn | 1,390 | |
| 912. | gaillimhe | 1,390 | |
| 913. | fuar | 1,388 | |
| 914. | tharraingt | 1,387 | |
| 915. | ráite | 1,386 | |
| 916. | diabhal | 1,385 | |
| 917. | oscailt | 1,382 | |
| 918. | iompar | 1,381 | |
| 919. | páirtí | 1,378 | |
| 920. | éadan | 1,373 | |
| 921. | cothrom | 1,372 | |
| 922. | glas | 1,372 | |
| 923. | chinn | 1,371 | |
| 924. | mhalairt | 1,371 | |
| 925. | dr | 1,369 | |
| 926. | eatarthu | 1,369 | |
| 927. | bháis | 1,368 | |
| 928. | faide | 1,367 | |
| 929. | bheatha | 1,366 | |
| 930. | easpa | 1,366 | |
| 931. | amharc | 1,364 | |
| 932. | dteach | 1,362 | |
| 933. | háite | 1,362 | |
| 934. | rté | 1,362 | |
| 935. | áine | 1,361 | |
| 936. | ghaeltacht | 1,358 | |
| 937. | anocht | 1,356 | |
| 938. | life | 1,355 | |
| 939. | múinteoirí | 1,355 | |
| 940. | bhíodar | 1,354 | |
| 941. | díol | 1,352 | |
| 942. | chloig | 1,350 | |
| 943. | féadfaidh | 1,349 | |
| 944. | imeachtaí | 1,347 | |

| N | Word | Freq. | % |
|---|---|---|---|
| 945. | stáisiún | 1,347 | |
| 946. | sheas | 1,344 | |
| 947. | dtugtar | 1,343 | |
| 948. | ábhair | 1,340 | |
| 949. | d'fhan | 1,335 | |
| 950. | choinne | 1,334 | |
| 951. | crua | 1,334 | |
| 952. | pobail | 1,334 | |
| 953. | iontas | 1,333 | |
| 954. | fhágáil | 1,330 | |
| 955. | ros | 1,329 | |
| 956. | pictiúr | 1,327 | |
| 957. | uaithi | 1,327 | |
| 958. | láimhe | 1,326 | |
| 959. | m'athair | 1,324 | |
| 960. | sraith | 1,324 | |
| 961. | teilifís | 1,321 | |
| 962. | hiomlán | 1,320 | |
| 963. | cónaí | 1,318 | |
| 964. | cuí | 1,318 | |
| 965. | inar | 1,318 | |
| 966. | cruinn | 1,317 | |
| 967. | léim | 1,317 | |
| 968. | cheantar | 1,316 | |
| 969. | ciallaíonn | 1,315 | |
| 970. | meitheamh | 1,315 | |
| 971. | suite | 1,315 | |
| 972. | bhreatain | 1,314 | |
| 973. | aimsire | 1,311 | |
| 974. | gcionn | 1,311 | |
| 975. | chois | 1,309 | |
| 976. | áiteanna | 1,308 | |
| 977. | arm | 1,308 | |
| 978. | dúil | 1,308 | |
| 979. | bhéarla | 1,307 | |
| 980. | íoc | 1,304 | |
| 981. | mbealach | 1,302 | |
| 982. | mhair | 1,302 | |
| 983. | scríbhneoirí | 1,302 | |
| 984. | cluiche | 1,300 | |
| 985. | léaráid | 1,299 | |
| 986. | aniar | 1,298 | |
| 987. | cogadh | 1,298 | |
| 988. | fadó | 1,298 | |
| 989. | ghlacadh | 1,298 | |
| 990. | tábla | 1,297 | |
| 991. | deiridh | 1,296 | |
| 992. | dhein | 1,296 | |
| 993. | tamaill | 1,295 | |
| 994. | ceannais | 1,294 | |
| 995. | tigh | 1,294 | |
| 996. | gcúrsaí | 1,293 | |
| 997. | bád | 1,290 | |
| 998. | domhnaigh | 1,286 | |
| 999. | grúpa | 1,286 | |
| 1000. | teoranta | 1,286 | |
| 1001. | cóir | 1,285 | |
| 1002. | d'fhiafraigh | 1,284 | |
| 1003. | rang | 1,284 | |
| 1004. | eoraip | 1,283 | |
| 1005. | ciúin | 1,282 | |
| 1006. | eanáir | 1,281 | |