Irish LFG Grammar Development

Sebastian Sulger

Universität Konstanz

August 28, 2008
1 General remarks on Irish

2 Two phenomena of Irish Morpho-Syntax
   - Lenition
   - Eclipse

3 Irish Morphology in xfst
   - Verbs
   - Nouns

4 Morphology - Syntax Interface

5 LFG Grammar
Irish - language history

- among the oldest languages of Europe
Irish - language history

- among the oldest languages of Europe
- fragments (inscriptions in stones) date from the 4th to the 6th century
Irish - language history

- among the oldest languages of Europe
- fragments (inscriptions in stones) date from the 4th to the 6th century
- related to Scottish Gaelic and Manx (which is regarded as extinct)
Irish - language history

- among the oldest languages of Europe
- fragments (inscriptions in stones) date from the 4th to the 6th century
- related to Scottish Gaelic and Manx (which is regarded as extinct)
- Indo-Germanic > Celtic > Insular Celtic > Goidelic ("Gaelic") > Irish (Scottish Gaelic, Manx)
Irish today

- first official language of Ireland
Irish today

- first official language of Ireland
- minority language; spoken predominantly in the so-called Gaeltachtaí (population: 91,862, 60% thereof: Irish in everyday life)
Irish today

- first official language of Ireland
- minority language; spoken predominantly in the so-called Gaeltachtaí (population: 91,862, 60% thereof: Irish in everyday life)
- 1,8 million people (one out of three) on the island speak Irish "to a certain degree"
Dispersion of the *Gaeltachtaí*

**Figure:** Map of Ireland including *Gaeltachtaí*
Irish today: standard language

- fear: Irish could be endangered by the discrepancies between the dialects - creation of the standard form *An Caighdeán Ofigiúil* ("The Official Standard", OS)
Irish today: standard language

- fear: Irish could be endangered by the discrepancies between the dialects -> creation of the standard form *An Caighdeán Ofigiúil* ("The Official Standard", OS)
- combination of syntactic and lexical elements of all dialects
Irish today: standard language

- fear: Irish could be endangered by the discrepancies between the dialects → creation of the standard form An Caighdeán Ofigiúil ("The Official Standard", OS)
- combination of syntactic and lexical elements of all dialects
- pronunciation is based on the Connacht dialect
Irish today: standard language

- fear: Irish could be endangered by the discrepancies between the dialects -> creation of the standard form *An Caighdeán Ofigiúil* (”The Official Standard”, OS)
- combination of syntactic and lexical elements of all dialects
- pronunciation is based on the Connacht dialect
- my implementation uses this standard form
Irish today: standard language

- fear: Irish could be endangered by the discrepancies between the dialects → creation of the standard form An Caighdeán Ofigiúil ("The Official Standard", OS)
- combination of syntactic and lexical elements of all dialects
- pronunciation is based on the Connacht dialect
- my implementation uses this standard form
- OS Irish: *Conas atá tú? ’How are you?’*
Lenition

- Initial Mutation
Lenition

- Initial Mutation
- a somewhat "strong" consonant is "weakened" (lat. *lenis* 'soft')
Lenition

- Initial Mutation
- A somewhat "strong" consonant is "weakened" (lat. lenis 'soft')

Affected consonants:

- \( b (/b/) \rightarrow bh (/v/) \)
- \( c (/k/) \rightarrow ch (/kx/) \)
- \( d (/d/) \rightarrow dh (/g/) \)
- \( f (/f/) \rightarrow fh (\emptyset) \)
- \( g (/g/) \rightarrow gh (/\gamma/) \)
- \( m (/m/) \rightarrow mh (/v/) \)
- \( p (/p/) \rightarrow ph (/f/) \)
- \( s (/s/) \rightarrow sh (/h/) \)
- \( t (/t/) \rightarrow th (/h/) \)

- Orthography: \( h \)
Lenition

- consonants that are not affected: $h, l, n, r$
Lenition

- consonants that are not affected: $h, l, n, r$
- lenition is realized in various lexical and syntactic contexts
Lenition

- consonants that are not affected: $h, l, n, r$
- lenition is realized in various lexical and syntactic contexts
- lenition needs an external trigger in front of the lenited consonant (exception: verbs in the preterite, imperfect and conditional)
consonants that are not affected: \( h, l, n, r \)

lenition is realized in various lexical and syntactic contexts

lenition needs an external trigger in front of the lenited consonant (exception: verbs in the preterite, imperfect and conditional)

example: prefix \( an- \), used for intensification of adjectives

\[ \text{beag} /b/ 'small'; \quad \text{an-bheag} /v/ 'very small' \]

\[ \text{te} /t/ 'hot'; \quad \text{an-the} /h/ 'very hot' \]
Eclipse

- Initial Mutation
Initial Mutation

- voiceless plosives and /f/ become voiced; voiced plosives become nasalized
Eclipse

- **Initial Mutation**
- voiceless plosives and /f/ become voiced; voiced plosives become nasalized
- eclipse is also possible with vowels \(\rightarrow n-\) is added in front of the vowel
Eclipse

- Initial Mutation

- voiceless plosives and /f/ become voiced; voiced plosives become nasalized

- eclipse is also possible with vowels - $n$- is added in front of the vowel

- affected consonants:
  - $b$ (/b/) $\rightarrow$ $mb$ (/m/)
  - $c$ (/k/) $\rightarrow$ $gc$ (/g/)
  - $d$ (/d/) $\rightarrow$ $nd$ (/n/)
  - $f$ (/f/) $\rightarrow$ $bhf$ (/v/)
  - $g$ (/g/) $\rightarrow$ $ng$ (/n/)
  - $p$ (/p/) $\rightarrow$ $bp$ (/b/)
  - $t$ (/t/) $\rightarrow$ $dt$ (/d/)
Eclipse

- consonants that are not affected: $h, l, m, n, r, s$
Eclipse

- consonants that are not affected: $h, l, m, n, r, s$
- eclipse is realized in various lexical and syntactic contexts
Eclipse

- consonants that are not affected: $h, l, m, n, r, s$
- eclipse is realized in various lexical and syntactic contexts
- eclipse needs an external trigger in front of the eclipsed consonant (no exceptions spotted so far)
consonants that are not affected: $h, l, m, n, r, s$

eclipse is realized in various lexical and syntactic contexts

eclipse needs an external trigger in front of the eclipsed consonant (no exceptions spotted so far)

example: prepositional phrases containing the preposition $i$

teach ($/t/$) 'house'; $i$ dteach ($/d/$) 'in a house'
Corcaigh ($/k/$) 'Cork'; $i$ gCorcaigh ($/g/$) 'in Cork'
14 adjectives, 15 verbs, 71 nouns, 15 (inflecting) prepositions
xfst System

- 14 adjectives, 15 verbs, 71 nouns, 15 (inflecting) prepositions
- 16 phonological rules written so far to account for:
xfst System

- 14 adjectives, 15 verbs, 71 nouns, 15 (inflecting) prepositions
- 16 phonological rules written so far to account for:
  - intensification of adjectives (prefixation of *an* + lenition)
xfst System

- 14 adjectives, 15 verbs, 71 nouns, 15 (inflecting) prepositions
- 16 phonological rules written so far to account for:
  - intensification of adjectives (prefixation of *an* + lenition)
  - irregular plural forms of nouns
xfst System

- 14 adjectives, 15 verbs, 71 nouns, 15 (inflecting) prepositions
- 16 phonological rules written so far to account for:
  - intensification of adjectives (prefixation of an + lenition)
  - irregular plural forms of nouns
- multiword transducer to handle phonological processes across word boundaries (Initial Mutations)
15 verbs included (the auxiliary *tá* being one of them)
Verbs

- 15 verbs included (the auxiliary tá being one of them)
- verbs may be formed synthetically or analytically
Verbs

- 15 verbs included (the auxiliary tá being one of them)
- verbs may be formed synthetically or analytically
- synthetical verb forms: person/number expressed in the ending
  
  táim 'I am'; *táim mé
  molaim 'I praise'; *molaim mé
Irish Morphology in xfst

Verbs

- 15 verbs included (the auxiliary tá being one of them)
- verbs may be formed synthetically or analytically
- synthetical verb forms: person/number expressed in the ending
  táim 'I am'; *táim mé
  molaim 'I praise'; *molaim mé
- analytical verb forms: person/number have to be expressed by a separate (pro)noun
  tá siad 'they are'
  molann sibh 'you (pl.) praise'
Verbs

- 15 verbs included (the auxiliary *tá* being one of them)
- verbs may be formed synthetically or analytically
  - synthetical verb forms: person/number expressed in the ending
    - *táim* 'I am'; *táim mé*
    - *molaim* 'I praise'; *molaim mé*
  - analytical verb forms: person/number have to be expressed by a separate (pro)noun
    - *tá siad* 'they are'
    - *molann sibh* 'you (pl.) praise'
- problems for syntax: verb and subject in a single word; sometimes synthetical, sometimes analytical, even within one paradigm
Irish Morphology in xfst

Verbs

15 verbs included (the auxiliary tá being one of them)
verbs may be formed synthetically or analytically

synthetical verb forms: person/number expressed in the ending
- táim 'I am'; *táim mé
- molaim 'I praise'; *molaim mé

analytical verb forms: person/number have to be expressed by a separate (pro)noun
- tá siad 'they are'
- molann sibh 'you (pl.) praise'

problems for syntax: verb and subject in a single word; sometimes synthetical, sometimes analytical, even within one paradigm

Nouns

- 71 nouns included
Nouns

- 71 nouns included
- five declinations, each with several sub-groups
Nouns

- 71 nouns included
- five declinations, each with several sub-groups
- declination classes and sub-groups differ in forming the genitive and plural forms (among other differences)
Nouns

- 71 nouns included
- five declinations, each with several sub-groups
- declination classes and sub-groups differ in forming the genitive and plural forms (among other differences)
- approach: multiple lexicons for declinations and sub-groups
71 nouns included
five declinations, each with several sub-groups
decision classes and sub-groups differ in forming the genitive and plural forms (among other differences)
approach: multiple lexicons for declinations and sub-groups
nouns are arranged in these lexicons and labeled accordingly; labels trigger phonological rules to form the correct plural/genitive
Nouns

- 71 nouns included
- five declinations, each with several sub-groups
- declination classes and sub-groups differ in forming the genitive and plural forms (among other differences)
- approach: multiple lexicons for declinations and sub-groups
- nouns are arranged in these lexicons and labeled accordingly; labels trigger phonological rules to form the correct plural/genitive
- works: implemented for the nominative plural
problem: (lexical) triggering of Initial Mutations
Morphology - Syntax Interface

- problem: (lexical) triggering of Initial Mutations
- Initial Mutations operate across word boundaries
problem: (lexical) triggering of Initial Mutations
Initial Mutations operate across word boundaries
approach (cf. Mittendorf & Sadler (2006)):
problem: (lexical) triggering of Initial Mutations

Initial Mutations operate across word boundaries

approach (cf. Mittendorf & Sadler (2006)):
  - list words in lexicon with all mutation variants (radical, lenition, eclipse), tagged accordingly
problem: (lexical) triggering of Initial Mutations

Initial Mutations operate across word boundaries

approach (cf. Mittendorf & Sadler (2006)):

- list words in lexicon with all mutation variants (radical, lenition, eclipse), tagged accordingly
- list triggers in lexicon with governed mutation (e.g. preposition $i$: $+\text{Ecl}+$ for Eclipse)
problem: (lexical) triggering of Initial Mutations
Initial Mutations operate across word boundaries
approach (cf. Mittendorf & Sadler (2006)):
- list words in lexicon with all mutation variants (radical, lenition, eclipse), tagged accordingly
- list triggers in lexicon with governed mutation (e.g. preposition $i$: $+\text{Ecl}+$ for Eclipse)
- set up multiword transducer that ensures that lexical mutation conditions are satisfied by the input string
Example: *i 'in*, *teach 'house*

\[
xfst[1]: \text{up } i \\
+\text{Rad}+i+\text{Prep}+\text{PrepNoP}+\text{Ecl}+ \\
\]

\[
xfst[1]: \text{up } teach \\
+\text{Rad}+\text{teach}+\text{Noun}+3P+Sg+Masc \]

\[
\text{CS 1: } P:23 \\
\text{MUT_SFX_BASE:16 } P_{\text{BASE}}:10 \text{ } P_{\text{SFX_BASE}}:6 \text{ } \text{PINFL}_{\text{SFX_BASE}}:4 \text{ } \text{MUT}_{\text{SFX_BASE}}:2 \\
+\text{Rad}:17 \text{ } i:9 \text{ } +\text{Prep}:7 \text{ } +\text{PrepNoP}:5 \text{ } +\text{Ecl}:3 \\
\]

\[
\text{CS 2: } N:27 \\
\text{MUT_SFX_BASE:18 } N_{\text{BASE}}:14 \text{ } N_{\text{SFX_BASE}}:8 \text{ } \text{PERS}_{\text{SFX_BASE}}:6 \text{ } \text{NUM}_{\text{SFX_BASE}}:4 \text{ } \text{GEND}_{\text{SFX_BASE}}:2 \\
+\text{Rad}:19 \text{ } \text{teach}:11 \text{ } +\text{Noun}:9 \text{ } +3P:7 \text{ } +Sg:5 \text{ } +\text{Masc}:3 \\
\]
Example: *i dteach* 'in a house’ - resulting tree

\[xfst[1]: \text{up } i\]
\[\text{+Rad+i+Prep+PrepNoP+Ecl+}\]
\[xfst[1]: \text{up dteach}\]
\[\text{+Ecl+teach+Noun+3P+Sg+Masc}\]
problem: synthetical verb forms
• problem: synthetical verb forms

• paradigm tá (directly from source file):
  tá + Verb + Pres + Pron1Sg: tám #;
  tá + Verb + Pres: tá #;
  tá + Verb + Pres: tá #;
  tá + Verb + Pres + Pron1Pl: táimíd #;
  tá + Verb + Pres: tá #;
  tá + Verb + Pres: tá #;
problem: synthetical verb forms

paradigm tá (directly from source file):
  tá+Verb+Pres+Pron1Sg:táim #;
  tá+Verb+Pres:tá #;
  tá+Verb+Pres:tá #;
  tá+Verb+Pres+Pron1Pl:táimíd #;
  tá+Verb+Pres:tá #;
  tá+Verb+Pres:tá #;

'táim' and 'táimíd' are synthetical verb forms -> labeling
approach: labels trigger templates in LFG grammar
- approach: labels trigger templates in LFG grammar
- templates assign the appropriate grammatical features to the subject
Morphology - Syntax Interface

- approach: labels trigger templates in LFG grammar
- templates assign the appropriate grammatical features to the subject
- example:
  tá+Verb+Pres+Pron1Sg:táim #;
approach: labels trigger templates in LFG grammar

- templates assign the appropriate grammatical features to the subject

example:

`tá+Verb+Pres+Pron1Sg:táim`\n
label `'+Pron1Sg'` activates template `PronSFX1Sg'` in LFG grammar;
template assigns the following to the subject:

\[
\begin{align*}
(\text{^SUBJ PRED}) &= 'pro' \\
(\text{^SUBJ CASE}) &= \text{nom} \\
(\text{^SUBJ NUM}) &= \text{sg} \\
(\text{^SUBJ PERS}) &= 1 \\
(\text{^SUBJ PRON-TYPE}) &= \text{pers}.
\end{align*}
\]
grammar is yet very small: "Grammar has 14 rules with 70 states, 77 arcs, and 77 disjuncts (77 DNF)."
grammar is yet very small: "Grammar has 14 rules with 70 states, 77 arcs, and 77 disjuncts (77 DNF)."

coverage: agreement, intransitives, transitives, passive construction (without agent)
grammar is yet very small: "Grammar has 14 rules with 70 states, 77 arcs, and 77 disjuncts (77 DNF)."

coverage: agreement, intransitives, transitives, passive construction (without agent)

grammar is constantly under construction
Example: *táim fuar*. 'I am cold.'
Example: *chaith sí úll.* 'She threw an apple.'
Some C- and F-Structures

Example: *ní chaith sí úll.* 'She does not throw an apple.'
Some C- and F-Structures

Example: *staidéareann mé an chláirseach*. 'I study the harp.'

```
CS 1: ROOT:70
  S:68        PUNCT:37
    V:49  NP:54  NP:67 .:36
      staidéareann:1  PRON:15  D:17  N:66
        mé:14  an:16  chláirseach:18

"staidéareann mé an chláirseach."
[PRED 'staidéar<[14:pro], [16:cláirseach]>'
  14 PRED 'pro'
  15 SUBJ CASE nom, NUM sg, PERS 1, PRON-TYPE pers
       54
  36 18 PRED 'cláirseach'
  37 66 OBJ CASE acc, DEF +, GEND fem, NUM sg, PERS 3
       16
       17
       67
  70 CLAUSE-TYPE decl, TENSE pres
```
Thank you!