Irish Grammar Report

ParGram Fall Meeting 2010, PARC
Overview

Current state of the grammar:

- 48 rules, 353 states, 634 disjuncts
- basic VP, NP, copula and cleft constructions implemented
- lexicon includes 44 verbs with subcat frames
  → still fairly small, but growing
Contents

1. Morphology
   - A new morphology?

2. Syntax
   - Irish Verbs

3. Outlook
   - Non-adjacent mutation
Morphologies available
Morphologies available

Currently available morphologies:

- my own (small) morphology:
  + mutation tags in place
  + detailed verbal morphology (but not for all tenses)
  + various noun declensions implemented
    - small amount of actual stems

- Uí Dhonnchadha (2003):
  + very large amount of stems
  + all important tags are in place
    - improvements necessary for syntax:
      verbal morphology, mutation tags
The choice of morphology

The advantages of the large DCU/TCD morphology are clear, but:

- changing the tags can be cumbersome
- extensive ruleset is in place for adding mutation tags/changing and adding verbal morphology tags

→ it’s a morphology which was not written for LFG grammar development

→ current goal: implement own (toy) morphology with better design for LFG grammar purpose
Irish Verbs
Irish Verbs

... may occur in two different forms:

- analytic:
  - person/number features not expressed by verb
  - independent subject pronoun needed

shalaigh mé mo lámha.
dirty.Pret I my hand.Pl
‘I messed up my hands.’

*shalaigh mo lámha.
dirty.Pret my hand.Pl
Irish Verbs

... may occur in two different forms:

- analytic:
  - person/number features not expressed by verb
  - independent subject pronoun needed

  shalaigh mé mo lámha. *shalaigh mo lámha.
  ‘I messed up my hands.’

- synthetic:
  - person/number features expressed by suffixes on verb
  - independent subject pronoun not needed — even disallowed

  shalaíomar ár lámha. *shalaíomar muid ár lámha.
  ‘We messed up our hands.’
Irish Verbs

To complicate matters:

- distribution of analytic vs. synthetic forms varies across paradigms:

  verb *tuig* ‘understand’

<table>
<thead>
<tr>
<th>Features</th>
<th>Present</th>
<th>Imperfect</th>
<th>Future</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P.Sg</td>
<td><em>tuigim</em></td>
<td><em>thuiginn</em></td>
<td><em>tuigfidh</em></td>
</tr>
<tr>
<td>2P.Sg</td>
<td><em>tuigeann</em></td>
<td><em>thuigteá</em></td>
<td><em>tuigfidh</em></td>
</tr>
<tr>
<td>3P.Sg</td>
<td><em>tuigeann</em></td>
<td><em>thuigeadh</em></td>
<td><em>tuigfidh</em></td>
</tr>
<tr>
<td>1P.Pl</td>
<td><em>tuigimid</em></td>
<td><em>thuigmis</em></td>
<td><em>tuigfimid</em></td>
</tr>
<tr>
<td>2P.Pl</td>
<td><em>tuigeann</em></td>
<td><em>thuigeadh</em></td>
<td><em>tuigfidh</em></td>
</tr>
<tr>
<td>3P.Pl</td>
<td><em>tuigeann</em></td>
<td><em>thuigidis</em></td>
<td><em>tuigfidh</em></td>
</tr>
</tbody>
</table>
Irish Verbs

To complicate matters:

- ‘analytic form + pronoun’ is blocked if synthetic form exists which realizes the same features ("Morphological Blocking"):

  \[
  \text{Tuigim an fhadhb.} \quad \text{Tuigeann mé an fhadhb.}
  \]
  \[
  \text{understand.Pres.1P.Sg the problem} \quad \text{understand.Pres 1 the problem}
  \]
  \[
  \text{‘I understand the problem.’}
  \]

→ Question: How can this interaction between verb forms be captured?
Implementation challenges

Faced with two tasks:

1. block redundant use of pronoun together with synthetic form
2. block ‘analytic form + pronoun’ if synthetic form is available

Both tasks are easily accomplished using agreement templates... if certain assumptions are made.
An LFG implementation

Key assumptions:

- For both forms – synthetic and analytic – I assume that the pronoun is *pro*, realizing person/number features.
- *pro* agrees with the verb in person/number.
- Provided a detailed morphological analysis, simple agreement equations can be used to rule out ungrammatical strings.
An LFG implementation

The current implementation is based on two components:

1. morph. analyzer dealing with verbal morphology in a detailed way
   - lists both analytic and synthetic verb forms
   - analyses include morphosyntactic features such as person, number

2. grammar ruling out ungrammatical sentences using two short agreement templates
   - independent pronouns blocked when occurring with synthetic forms
   - analytic forms are blocked where inappropriate
An LFG implementation

Sample morphological analysis:

verb *tuig* ‘understand’ in the present paradigm

<table>
<thead>
<tr>
<th>Features</th>
<th>Verb Form</th>
<th>FSM Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1P.Sg</td>
<td><em>tuigim</em></td>
<td>tuig+Verb+Pres+1P+Sg+PronIncl</td>
</tr>
<tr>
<td>2P.Sg</td>
<td><em>tuigeann</em></td>
<td>tuig+Verb+Pres+2P+Sg</td>
</tr>
<tr>
<td>3P.Sg</td>
<td><em>tuigeann</em></td>
<td>tuig+Verb+Pres+3P+Sg</td>
</tr>
<tr>
<td>1P.Pl</td>
<td><em>tuigimid</em></td>
<td>tuig+Verb+Pres+1P+Pl+PronIncl</td>
</tr>
<tr>
<td>2P.Pl</td>
<td><em>tuigeann</em></td>
<td>tuig+Verb+Pres+2P+Pl</td>
</tr>
<tr>
<td>3P.Pl</td>
<td><em>tuigeann</em></td>
<td>tuig+Verb+Pres+3P+Pl</td>
</tr>
</tbody>
</table>

- tag +PronIncl attaches only to synthetic forms: makes sure subject is a pronominal
- detailed person/number information - also for identical forms: to enforce agreement constraints
An LFG implementation

In the grammar:

Tag +PronIncl linked to template:

PronSFX = (↑ SUBJ PRED) = ‘pro’.

- tag itself provides information that subject is a pronominal
- independent subject pronouns bring their own PRED

→ combination ‘synthetic verb form + independent subject pronoun’ is ruled out: PREDs don’t unify
An LFG implementation

In the grammar:

- **Person/number tags linked to agreement templates:**
  
  +1P V-PERS_SFX XLE @(AGR-P 1).
  +Sg V-NUM_SFX XLE @(AGR-N sg).

- **Entries call up templates, which try to assign the respective value:**

  \[\text{AGR-P(P)} = (\uparrow \text{SUBJ PERS}) = P.\]
  \[\text{AGR-N(N)} = (\uparrow \text{SUBJ NUM}) = N.\]
Independent subject pronouns provide their own person-number information (for example: *mé* - first person singular)

Analysis fails if values from analytic verb forms and subject pronouns do not match

\[
\begin{align*}
\text{tuigeann} & \quad \{+2P \parallel +3P\} & (\uparrow \text{SUBJ PERS}) = \{2 \parallel 3\} \\
\text{mé} & & (\uparrow \text{SUBJ PERS}) = 1 \\
\text{sé} & & (\uparrow \text{SUBJ PERS}) = 3
\end{align*}
\]

strings like *tuigeann mé* are ruled out...
strings like *tuigeann sé* are kept in.
An LFG implementation

Sample analysis output by the grammar:

*Tuigeann mé an fhadhb.
understand.Pres I the problem ‘I understand the problem.’

→ Analysis fails: no agreement in (^ PERS) of subject
An LFG implementation

Sample analysis output by the grammar:

*Tuigeann sé an fhadhb.*

understand.Pres he the problem

‘He understands the problem.’

→ Analysis valid: agreement in (\(^ \text{PERS}\)) and (\(^ \text{NUM}\))
Quick summary

- Using a detailed analyzer & two agreement templates, the data can be easily accounted for.
- I assume an agreement relationship between the verb form and *pro* (in whatever way *pro* is realized).
- “Morphological Blocking” is a direct effect of this relationship: no extra mechanism has to be defined.
- The approach is currently implemented for present and preterite tenses (easily extendable to other tenses).
- A 30-sentence testsuite is in place – full coverage without any overgeneration.
Future work
Possible future work includes:

- implement relative clauses; especially free relatives
- get mutation facts right – look at non-adjacent mutation
- (expanding on morphological analyzer)
- (expanding on lexicon – especially verb lexicon)
Initial Mutation in general

Two processes: *Lenition* and *Eclipse*

- Initial mutation: phonological process affecting first sound of word
- realized in various lexical and syntactic contexts
- needs an external trigger in front of the affected sound
Initial Mutation in general

Two processes: *Lenition* and *Eclipse*

- Initial mutation: phonological process affecting first sound of word
- realized in various lexical and syntactic contexts
- needs an external trigger in front of the affected sound
- example (lexical triggering): prefix *an*-: intensification of adjectives

*beag* /b/ 'small'; *an-bheag* /v/ 'very small'
*te* /t/ 'hot'; *an-the* /h/ 'very hot'

→ Lenition
Initial Mutation in general

Two processes: *Lenition* and *Eclipse*

- Initial mutation: phonological process affecting first sound of word
- realized in various lexical and syntactic contexts
- needs an external trigger in front of the affected sound
- example (lexical triggering): prefix *an-*: intensification of adjectives

  `beag /b/ 'small'; an-bheag /v/ 'very small'`
  `te /t/ 'hot'; an-the /h/ 'very hot'`  → Lenition

- example (syntactic triggering): prepositional phrases containing an NP with definite article

  `bord (/b/) 'table'; ar an mbord (/m/) 'on the table'`
  `cat (/k/ 'cat'; roimh an gcat (/g/) 'in front of the cat'`  → Eclipse
Analyzing Initial Mutation

- most phonological analyses assume some element in input, triggering the mutation
- triggering element: floating autosegment/morpheme, segmentally empty $\rightarrow$ not pronounced, located at right edge of triggers
Analyzing Initial Mutation

- most phonological analyses assume some element in input, triggering the mutation
- triggering element: floating autosegment/morpheme, segmentally empty → not pronounced, located at right edge of triggers
- motivation: IM usually applies in adjacent fashion
  - trigger words/trIGGERING syntactic environments *directly* before affected word
  - speaks in favor of triggering segment at right edge of trigger
Analyzing Initial Mutation

- most phonological analyses assume some element in input, triggering the mutation
- triggering element: floating autosegment/morpheme, segmentally empty → not pronounced, located at right edge of triggers
- motivation: IM usually applies in adjacent fashion
  - trigger words/triggering syntactic environments *directly* before affected word
  - speaks in favor of triggering segment at right edge of trigger

→ BUT: cases where IM is applied in a non-adjacent fashion
Non-Adjacent Mutation

Two Cases of Non-Adjacent Mutation:

1) [Possessive Pronoun + dhá ‘two’ + Noun]_{NP}
Non-Adjacent Mutation

Two Cases of Non-Adjacent Mutation:

1) \([\text{Possessive Pronoun} + \text{dhá ‘two’} + \text{Noun}]_{NP}\)

bhur   dteach
your.pl house
‘your (pl.) house’ → possessive pronoun triggers eclipsis on noun
Non-Adjacent Mutation

Two Cases of Non-Adjacent Mutation:

1) \([\text{Possessive Pronoun} + \text{dhá ‘two’} + \text{Noun}]_{NP}\)

bhur \ dteach
your.pl house
‘your (pl.) house’ → possessive pronoun triggers eclipsis on noun

dhá \ theach
two house
‘two houses’ → numeral dhá triggers lenition on noun
Non-Adjacent Mutation

Two Cases of Non-Adjacent Mutation:

1) \([\text{Possessive Pronoun} + \text{dhá ‘two’} + \text{Noun}]_{NP}\)

bhur \(d\) teach
your.pl house
‘your (pl.) house’ \(\rightarrow\) possessive pronoun triggers eclipsis on noun

dhá \(t\) teach
two house
‘two houses’ \(\rightarrow\) numeral \(\text{dhá}\) triggers lenition on noun

bhur \(\text{dhá} d\) teach
your.pl two house
‘your (pl.) two houses’ \(\rightarrow\) possessive pronoun triggers eclipsis on noun

even though the two are not adjacent
Non-Adjacent Mutation

2) \([\text{Preposition} + [\text{Noun} + \text{Conjunction} + \text{Noun}]_{NP}]_{PP}\)
Non-Adjacent Mutation

2) \([\text{Preposition} + [\text{Noun} + \text{Conjunction} + \text{Noun}]_{NP}]_{PP}\)

do bhuachaillí

to boys

‘to boys’ → preposition *do* triggers lenition on noun
Non-Adjacent Mutation

2) \([\text{Preposition} + [\text{Noun} + \text{Conjunction} + \text{Noun}]_{NP}]_{PP}\)

do bhuachaílí
to boys
‘to boys’ \(\rightarrow\) preposition \(do\) triggers lenition on noun

do bhuachaílí nó chailíní
to boys or girls
‘to boys or girls’ \(\rightarrow\) preposition \(do\) triggers lenition on both nouns
even though 2nd noun is not adjacent to preposition
Non-Adjacent Mutation

2) \([\text{Preposition} + [\text{Noun} + \text{Conjunction} + \text{Noun}]_{NP}]_{PP}\)

dó bhuachaillí
to boys
‘to boys’ \(\rightarrow\) preposition \(do\) triggers lenition on noun

dó bhuachaillí nó chailiní
to boys or girls
‘to boys or girls’ \(\rightarrow\) preposition \(do\) triggers lenition on both nouns

even though 2nd noun is not adjacent to preposition

(Lenition appears on all conjuncts within coordinated structures)
Non-Adjacent Mutation

2) \([\text{Preposition} + [\text{Noun} + \text{Conjunction} + \text{Noun}]_{\text{NP}}]_{\text{PP}}\)

do bhuachaillí
to boys
‘to boys’ → preposition do triggers lenition on noun

do bhuachaillí nó chailíní
to boys or girls
‘to boys or girls’ → preposition do triggers lenition on both nouns even though 2nd noun is not adjacent to preposition

(Lenition appears on all conjuncts within coordinated structures)

→ examples speaking against floating autosegment approach (e.g. Swingle 93, Gnanadesikan 97)

→ mutation of non-adjacent words *not explained so far*
Summary

- cases of non-adjacent mutation speak against a floating autosegment type of approach
- have to find different analysis?
  grateful for input!
- problem: data is difficult to establish
  - already conducted corpus search, using Perl
  - but: hard to define good regular expressions for this purpose
Thank you!

Are there questions?