Using rule-based computational linguistics for Australian languages: Electronic resources for Murrinh-Patha

Melanie Seiss & Rachel Nordlinger

University of Konstanz & Melbourne University

7th European Australianists Workshop 2012

03./04.04.2012
How computational linguistics and Australian languages can profit from each other

**Australian languages**

- another way of describing and conserving a language
- applications useful for, e.g.:
  - promoting literacy among language speakers (e.g. Arrernte Footy, Lareau et al. 2011)
  - promoting language skills for language learners

**Computational linguistics**

- test existing tools and methods on different language types
- test tools and methods on languages without other resources (corpora etc.)
- develop new tools and methods
A Brief Introduction to Murrinh-Patha

Murrinh-Patha

- polysynthetic
- non-Pama-Nyungan
- Southern Daly subgroup, together with Ngan’gityemerri (Green 2003)
- spoken in and around Wadeye, NT by approx. 2500 speakers
- lingua franca of region (still acquired by children, still spoken in every day life)
What makes Murrinh-Patha so difficult?

- bipartite verbal structure
- complicated verbal template
- complex number system
- morphophonemics
Murrinh-Patha bipartite verbs

- MP verbs (mostly) consist of a *classifier stem* and a *lexical stem*

  (1) *manganta*
  
  mangan - rta
  
  3sgS.SNATCH(9).nFut - hug
  
  ‘He/she hugged him/her.’

- Classifier stem: inflected for subject person, number, tense; ‘rather general’ meaning

- Lexical stem: uninflecting, ‘more specific’ meaning
Inflection on classifier stems

Encoded in portmanteau forms:

- subject person/number marked on classifier stem
- 4-way number contrast: singular, 1. inclusive, dual, plural
- 3-way person contrast
- 5-way tense/aspect contrast: non-Future (nFut), Past Imperfective (Plmpf), Future (Fut), Future Irrealis (FutIrr), Past Irrealis (PstIrr)

⇒ more than 50 forms per paradigm
The Murrinh-Patha verbal template & dependencies

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class.</strong></td>
<td>RR</td>
<td><strong>SubjN/Obj</strong></td>
<td>RR</td>
<td>IBP</td>
<td><strong>Lex</strong></td>
<td>TNS</td>
<td>Adv/Prt</td>
<td><strong>SubjN/ObjN</strong></td>
<td>Adv/Prt</td>
</tr>
</tbody>
</table>

Class: classifier stem, marked for tense, aspect & subject number  
SubjN: subject number markers for dual & paucal subject  
Obj: object agreement marker  
ObjN: object number marker for dual & paucal  
RR: reflexive / reciprocal marker  
IBP: incorporated body part  
Lex: lexical stem  
TNS: tense marker  
Adv: Adverbial  
Prt: Particle

(adapted from Blythe 2009)
The Murrinh-Patha verbal template & dependencies

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class.</td>
<td>RR</td>
<td>SubjN/Obj</td>
<td>RR</td>
<td>IBP</td>
<td>Lex</td>
<td>TNS</td>
<td>Adv/Prt</td>
<td>SubjN/ObjN</td>
<td>Adv/Prt</td>
</tr>
</tbody>
</table>

Class: classifier stem, marked for tense, aspect & subject number
SubjN: subject number markers for dual & paucal subject
Obj: object agreement marker
ObjN: object number marker for dual & paucal
RR: reflexive / reciprocal marker
IBP: incorporated body part
Lex: lexical stem
TNS: tense marker
Adv: Adverbial
Prt: Particle

(adapted from Blythe 2009)
Murrinh-Patha Number System

Complex number system for subject and object:

- singular (sg), dual (du), paucal (pauc), plural (plural)
- sibling vs. non-sibling (in dual and paucal only)
- gender: female (fem) vs. male (in dual and paucal only)
marked by a combination of the classifier stem and separate morphemes (Nordlinger 2010a)

Example: ‘They saw it.’

(2a) *Bam-ngintha-ngkardu* ‘They 2 fem non-sib’
(2b) *Bam-nintha-ngkardu* ‘They 2 male non-sib’
(2c) *Pubamka-ngkardu* ‘They 2 sibling’
(2d) *Pubamka-ngkardu-ngime* ‘They paucal fem non-sib’
(2e) *Pubamka-ngkardu-neme* ‘They paucal male non-sib’
(2f) *Pubamkardu (Pubam-ngkardu)* ‘They plural, they paucal sib’
Object Marking

- Direct and indirect object marking on the verb
- Same categories as for subject marking
- Discontinuous object markers for non-sibling categories

(3a) *Bam-ngo-nga-kardu* ‘He/she saw me.’
(3b) *Bam-nganku-nga-kardu-ngintha* ‘He/she saw us (2 fem non-sib).’
(3c) *Bam-nganku-nga-kardu* ‘He/she saw us (2 sibling).’
(3d) *Bam-nganku-nga-kardu-ngime* ‘He/she saw us (paucal fem non-sib).’
(3e) *Bam-pun-nga-kardu* ‘He/she saw us (plural/paucal sib).’
Competing subject and object markers:

- Subject number and object marker compete for same slots
- Object marker has priority

(5a) `bam -ngintha -ngkardu`
3sgS.SEE(13).nFut -du.f -see
‘They two (non-sibl) saw him/her.

(5b) `bam -ngi -ngkardu -ngintha`
3sgS.SEE(13).nFut -1sgDO -see -du.f
‘They two (non-sibl, female) saw me.

- Many more interdependencies between morphemes, see Nordlinger (2010b) for an overview.
Surface form is often different from the component parts:

(5a) mam-watha → mampatha
(5b) bam-ngkardu → bamkardu
(5c) mangan-rta → manganta
Electronic Resources for Murrinh-Patha

- **Electronic Dictionary:**
  - The electronic dictionary can process Murrinh-Patha words and phrases.
  - It decomposes the input for the user and looks up the meaning parts.

- **Translation System:**
  - The translation system takes English input and generates Murrinh-Patha output.
  - It can be used to translate simple sentences.
  - It is especially intended to learn about the Murrinh-Patha verb form and its number system.
Components of the Implementation

Murrinh-Patha:
- XLE Grammar
- XFST Morphology
- Lexicon

XFR Rewrite Rules

Graphical User Interface with Perl TK

English Pargram XLE Grammar
Resources used for the lexicon

- entries automatically extracted from Street (1989)
- Additional vocabulary from Walsh (1987), fieldnotes from Joe Blythe and Rachel Nordlinger

- Used as entries in
  - XFST Morphology
  - XLE Grammar (verbs only)
  - XFR Rewrite Rules (translation)
  - Dictionary entry (definitions & examples)
Morphology

- Finite State morphology built with XFST (Beesley & Karrtunen 2003)
- 2 level morphology:
  \[ \text{bam}+\text{class13}+3\text{P}+\text{sg}+3\text{sgDO}+\text{ngkardu}+\text{LS} : \text{bamngkardu} \]
- Inbuilt mechanisms to model the long distance dependencies between morphemes (e.g., discontinuous object markers)
- Allows for modeling of morphophonemic processes, e.g.
  \[ [ \text{n g k} \rightarrow \text{k} \parallel \text{m}_-, \text{n}_- ] \]
  \[ \rightarrow \text{bam}+\text{class13}+3\text{P}+\text{sg}+3\text{sgDO}+\text{ngkardu}+\text{LS} : \text{bamkardu} \]
XLE Grammars

- XLE Parser developed at PARC (Crouch et al. 2011, Butt et al. 1999)
- Implementation based on Lexical-Functional Grammar formalism
- Used by the ParGram-Group for large-scale grammar implementation: English, German, French, Norwegian, Japanese, Urdu, Hungarian, ...
XLE Grammar Output

"bamkardu kardu kigay"

CS 2: ROOT

S

VP

NP

V

NC

N

bamkardu kardu kigay
XFR Rewrite Rules for Translation

\[
\begin{align*}
\text{PRED} & \quad \text{‘see< (SUBJ) (OBJ)>’} \\
\text{SUBJ} & \quad \text{[PRED ‘PRO’]}
\quad \text{NUM sg} \\
\text{PERS} & \quad 3 \\
\text{OBJ} & \quad \text{[NUM ‘boy’]}
\quad \text{NUM sg} \\
\text{PERS} & \quad 3 \\
\cdots
\end{align*}
\]

\[
\begin{align*}
\text{PRED} & \quad \text{‘ngkardu< (SUBJ) (OBJ)>’} \\
\text{SUBJ} & \quad \text{[PRED ‘PRO’]}
\quad \text{NUM sg} \\
\text{PERS} & \quad 3 \\
\text{OBJ} & \quad \text{[PRED ‘kigay’]}
\quad \text{NUM sg} \\
\text{PERS} & \quad 3 \\
\text{NOMCLASS} & \quad \text{kardu} \\
\cdots
\end{align*}
\]

\[
\begin{align*}
\text{PRED} & \quad \text{(\%V, see)} \\
\Rightarrow & \\
\text{PRED} & \quad \text{(\%V, ngkardu).}
\end{align*}
\]

\[
\begin{align*}
\text{PRED} & \quad \text{(\%V, boy)} \\
\Rightarrow & \\
\text{PRED} & \quad \text{(\%V, kigay),}
\text{NOMCLASS} & \quad \text{(\%V, kardu).}
\end{align*}
\]
Example:
Input: *bamkardu kardu kigay*

Tries parsing:
1) NP: kardu kigay bamkardu
2) kardu kigay bamkardu

Only 2) gives grammatical output

script extracts information

Lookup in dictionary:

*ngkardu* + classifier 13: ‘to see’

*kigay + nounclass kardu:*

‘teenage boy’
Challenges posed by MP verbs

Implementation

Electronic Resources

Architecture of Translation system

Disambiguation Module:
- checks if the f-structure of the English input has a plural subject or object
- If no plural is present, redirects to the transfer rules
- If a plural is present, prompts the user to give more information
Disambiguation Module
Both, the Translation System and the Dictionary offer more information to the user after the initial output:

- morphological analysis
- show form with different tense, subject and object number information
- show various paradigms (keeping other information stable):
  - show form in all tenses
  - show form with all subject numbers
  - show form with all object numbers

→ Can be used to study structure, detect patterns, etc.
Future Work

- build web-based application (so far perl tk interface)
- broaden coverage
- more fine-grained feedback
- add sound files
- build applications for Murrinh-Patha speakers learning English using the same underlying Murrinh-Patha grammar
References


