Areal semantics and noun classification in northern Australia

John Mansfield john.Mansfield@uzh.ch

Ruth Singer rsinger@unimelb.edu.au

Leipzig CLDF Workshop
December 2023
Areal semantics and colexification

• Some regions of the world have distinctive conceptual structures

• ‘Areal semantics’  
  (Koptjevskaja-Tamm & Liljegren 2017; Schapper & Koptjevskaja-Tamm 2022; François 2022)

• Investigated via colexification studies / CLICS  
  (François 2008; List et al 2019)
YELLOW ≈ LOCUST.BEAN
(Segerer & Vanhoove 2022)

YELLOW is associated with
LOCUST.BEAN

<table>
<thead>
<tr>
<th>Language</th>
<th>Glottocode</th>
<th>Country</th>
<th>Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shanga (Mande, NC; McCallum Jones 2017)</td>
<td>shan1282</td>
<td>Nigeria</td>
<td>kpå’a’i (kpå’a ‘locust bean’, ‘y’ ‘water’)*</td>
</tr>
<tr>
<td>Kabiye (Gur, NC; CLNK 1999)</td>
<td>kabi1261</td>
<td>Togo</td>
<td>sota lm (sota ‘locust bean’, lm ‘water’)</td>
</tr>
<tr>
<td>Zodi (Chadic, AA; Caron 2002)</td>
<td>dass1243</td>
<td>Nigeria</td>
<td>jà bëhki (jà ‘water’, bëhki ‘locust-bean tree’)</td>
</tr>
<tr>
<td>Southern Samo (Mande, NC; Sil Burkina 2003)</td>
<td>sout2844</td>
<td>Burkina Faso</td>
<td>kùsi (‘dry locust bean pods’)</td>
</tr>
<tr>
<td>Yanda Dom (Dogon, NC; Heath 2013)</td>
<td>yand1257</td>
<td>Mali</td>
<td>lôl-pùidd ‘yellow flour from fruit of néré tree bean’ (lôl ‘locust bean’, pùidd ‘flower, powder’)</td>
</tr>
<tr>
<td>Sar (Central Sudanic, NS; Palayer 1992)</td>
<td>sarr1246</td>
<td>Chad</td>
<td>ndûûp mâtta (ndûûp ‘flour’, mâtta ‘locust bean’)</td>
</tr>
</tbody>
</table>
| Konyagi (Atlantic, NC; Sachot Santos 1996) | wame1240 | Senegal | yâr-yêyì (‘locust tree foliage’; yâr is a noun class prefix, yêyì is the stem of ‘locust bean’).
FIRE ≈ FIREWOOD ≈ TREE (Schapper et al 2016)

FIRE is associated with FIREWOOD / TREE
(a) Conceptual transmission with lexical transmission: \textit{WATER}≈\textit{RAIN}

WATER=RAIN

\textit{kura}  \hspace{1cm} \textit{kuru}

Murrinhpatha  \hspace{1cm}  Ngan’gi

(b) Conceptual transmission without lexical transmission: \textit{YEAR}≈\textit{MONSOON}

YEAR=MONSOON

\textit{thangku}  \hspace{1cm}  \textit{warri}

Murrinhpatha  \hspace{1cm}  Marri Tjevin

WATER=RAIN

YEAR=MONSOON

YEAR=MONSOON

YEAR=MONSOON
Areal semantics and colexification

• E.g. SUN \approx EYE + DAY (Urban 2010)
  * mata hari (Malay/Indo)

• SUN is partly associated with EYE
  SUN is partly associated with DAY

• These ‘concepts’ are no longer atomic (cf. Tversky 1977)
  They have internal structure, reflected in structured lexemes
Compounding and classification
Compounding and classification

• Many languages have pervasive compounding in the lexicon (aka binomial+ lexemes)

• E.g. Cambodian, Vietnamese... German!

• Nominal classification focuses binomial lexemes on a small set of roots
Classifiers in Daly languages

Murrinhpatha (Southern Daly)

• 10 classifiers, all independent nouns aka “generic nouns”

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>mi</td>
<td>‘veg food’</td>
</tr>
<tr>
<td>ku</td>
<td>‘animal’</td>
</tr>
<tr>
<td>kardu</td>
<td>‘person’</td>
</tr>
<tr>
<td>nanthi</td>
<td>‘thing’</td>
</tr>
<tr>
<td>da</td>
<td>‘place/time’</td>
</tr>
<tr>
<td>kura</td>
<td>‘water’</td>
</tr>
<tr>
<td>murrinh</td>
<td>‘language’</td>
</tr>
<tr>
<td>tju</td>
<td>‘weapon’</td>
</tr>
<tr>
<td>thungku</td>
<td>‘fire’</td>
</tr>
<tr>
<td>thamul</td>
<td>‘spear’</td>
</tr>
</tbody>
</table>

*Kardu parnamkutkut mi kilen* ‘people are collecting green plums’
Classifiers in Arnhem Languages

• 4 or 5 noun classes or genders

• Gender agreement is found in many areas of the language, e.g. Mawng:

\[\text{Annga} - \text{ma-nyi} \quad \text{mata} \quad \text{ma-lijap} \quad \text{mata} \quad \text{warlk}\]

\[2\text{sg/3}\text{VE}-\text{get-}12 \quad \text{VE} \quad \text{VE}-\text{small} \quad \text{VE} \quad \text{stick}\]

“Get a small stick.”

(Singer 2016: 34)
Bleaching in Arnhem Lang

• Mawng classes: Masc, Fem, Veg, ‘Land’, ‘Edible’

• MASC
  *ilijap* ‘boy’, *ingurlaj* ‘male name’, manyardi ‘song’, wurakak ‘crow’

• ‘EDIBLE’
  *karlungurr* ‘pandanus nut’, *mawngku* ‘shelter’
Partial colexification

- Part/whole relations or
- ‘Overlap colexification’ (List 2023)

Yaqui "tree": [dʒ u j a]
Yaqui "forest": [dʒ u j a]

Guìlín "tree": [ɛ y ²¹]
Guìlín "forest": [ɛ y ²¹ l i ŋ ²²]

Fúzhōu "bark": [⁴tsʰ j eu ²¹² pʰ w o i ⁵³]
Fúzhōu "woods": [⁴tsʰ j eu ²¹² l i ŋ ⁵³]
Partial colexification

• E.g. *thangku*
  YEAR ≈ RAIN.SEASON

• But *da thangku* ~ *kura thangka*
  TIME + RAIN.SEASON ≠ WATER + RAIN.SEASON
Partial colexification

- Many other nouns appear only with one classifier, e.g.:
  - *ku lawarnka* ‘WALLABY’
  - *nanthi kamarl* ‘FACE ≈ EYE’

- The classifier nouns themselves, being highly generic, are sites of massive colexification, e.g.
  - *ku* ‘MEAT≈ANIMAL’
  - *kura* ‘RAIN≈WATER’
Partial colexification

• E.g. ku WALLABY, BIRD, SNAKE, OLIVE.PYTHON...

• At the limit, this could even include more bleached class systems

• E.g. Mawng
ta karlingurr ‘pandanus nut’
ta mawngku ‘shelter’
Comparing noun class systems
Indo-European noun classes

McCarthy et al 2020

(a) German, $K = 3$
(b) Spanish, $K = 2$
Partition, entropy and mutual information

\[ H(p) \overset{\text{def}}{=} - \sum_{a \in A} p(a) \log p(a) \]

\[ I(A;B) \overset{\text{def}}{=} \sum_{A \in A} \sum_{B \in B} \frac{|A \cap B|}{N} \log \frac{N |A \cap B|}{|A||B|} \]

\[ AMI(A,B) \overset{\text{def}}{=} \frac{I(A,B) - \mathbb{E}[I(A';B')]}{\max I(A',B') - \mathbb{E}[I(A';B')] } \]
A study of northern Australia

From Bowern (ed) 2023 Oxford Guide to Australian Languages
## Concepts and classes

<table>
<thead>
<tr>
<th></th>
<th>Daly</th>
<th>Marri Tjevin</th>
<th>Wubuy</th>
<th>Mawng</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Murrinhpatha</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIRD</td>
<td>ku (ANIM)</td>
<td>awu (ANIM)</td>
<td>ngarra- (NEUT)</td>
<td>na- (MASC)</td>
</tr>
<tr>
<td>FISH</td>
<td>ku (ANIM)</td>
<td>awu (ANIM)</td>
<td>ngarra- (NEUT)</td>
<td>na- (MASC)</td>
</tr>
<tr>
<td>WASP</td>
<td>ku (ANIM)</td>
<td>awu (ANIM)</td>
<td>ana- (FEM)</td>
<td>niny- (FEM)</td>
</tr>
<tr>
<td>FIREWOOD</td>
<td>thungku (FIRE)</td>
<td>tjendji (FIRE)</td>
<td>ngarra- (NEUT)</td>
<td>ma- (VEG)</td>
</tr>
<tr>
<td>FIRESTICK</td>
<td>thungku (FIRE)</td>
<td>tjendji (FIRE)</td>
<td>ngarra- (NEUT)</td>
<td>niny- (FEM)</td>
</tr>
<tr>
<td>ALCOHOL</td>
<td>kura (LIQ)</td>
<td>wudi (LIQ)</td>
<td>ngarra- (NEUT)</td>
<td>mung- (LAND)</td>
</tr>
</tbody>
</table>
Adjusted Mutual Information (AMI)

<table>
<thead>
<tr>
<th></th>
<th>mp</th>
<th>Murrinhpatha</th>
<th>mt</th>
<th>Marri Tjevin</th>
<th>ngt</th>
<th>Ngan’gi</th>
<th>(Ngan’gityemerri)</th>
<th>bkw</th>
<th>Bininj Kunwok</th>
<th>wub</th>
<th>Wubuy</th>
<th>kun</th>
<th>Kunbarlang</th>
<th>maw</th>
<th>Mawng</th>
<th>gn-bur</th>
<th>Gun-nartpa/Burarra</th>
</tr>
</thead>
<tbody>
<tr>
<td>mtj</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ngt</td>
<td>0.36</td>
<td>0.37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bkw</td>
<td>0.24</td>
<td>0.26</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wub</td>
<td>0.16</td>
<td>0.17</td>
<td>0.15</td>
<td>0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kun</td>
<td>0.18</td>
<td>0.19</td>
<td>0.26</td>
<td>0.35</td>
<td>0.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>maw</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.12</td>
<td>0.15</td>
<td>0.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gn-bur</td>
<td>0.14</td>
<td>0.15</td>
<td>0.32</td>
<td>0.35</td>
<td>0.25</td>
<td>0.35</td>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Inheritance of noun class from proto-Australian?

- Harvey and Mailhammer 2018
  Reconstruct a system of five noun classes for proto-Australian

I  *ji-   Human Male
II *jiny- Human Female
III *ta-   Animal
IV  *ma-   Plant
V  *ku-   Default

They treat the widespread prefixes: na- ‘Masculine’ and ngal- ‘Feminine’ as a separate derivational system

Map 3. Reflexes of *ma- “Plant”
The way ahead...

• Add class-cognacy to data

• Model for contact vs inheritance (e.g. SBayes, Ranacher et al 2021)

• Add data from another continent
CLDF data

• CLDF cognate coding

• We can use this to trace colexification as lexically transmitted

<table>
<thead>
<tr>
<th>Language</th>
<th>Form 1</th>
<th>Form 2</th>
<th>Form 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pipil</td>
<td>2_ashes</td>
<td></td>
<td>nex-ti</td>
</tr>
<tr>
<td>Pochutec</td>
<td>2_ashes</td>
<td></td>
<td>nox-t</td>
</tr>
<tr>
<td>ProtoAztecan</td>
<td>2_ashes</td>
<td></td>
<td>*nix-tli</td>
</tr>
<tr>
<td>Pipil-2_ashes-1</td>
<td></td>
<td>nex-ti</td>
<td>2</td>
</tr>
<tr>
<td>Pochutec-2_ashes-1</td>
<td></td>
<td>nox-t</td>
<td>2</td>
</tr>
<tr>
<td>ProtoAztecan-2_ashes-1</td>
<td></td>
<td>*nix-tli</td>
<td>2</td>
</tr>
</tbody>
</table>

forms.csv
cognates.csv
(Davletshin 2012)
CLDF data

- Partial colexification in CLDF (List 2023)

Can also be generated as part of data collation.

Can this be stored within the CLDF format?

What about cognacy for part of a lexeme?