

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

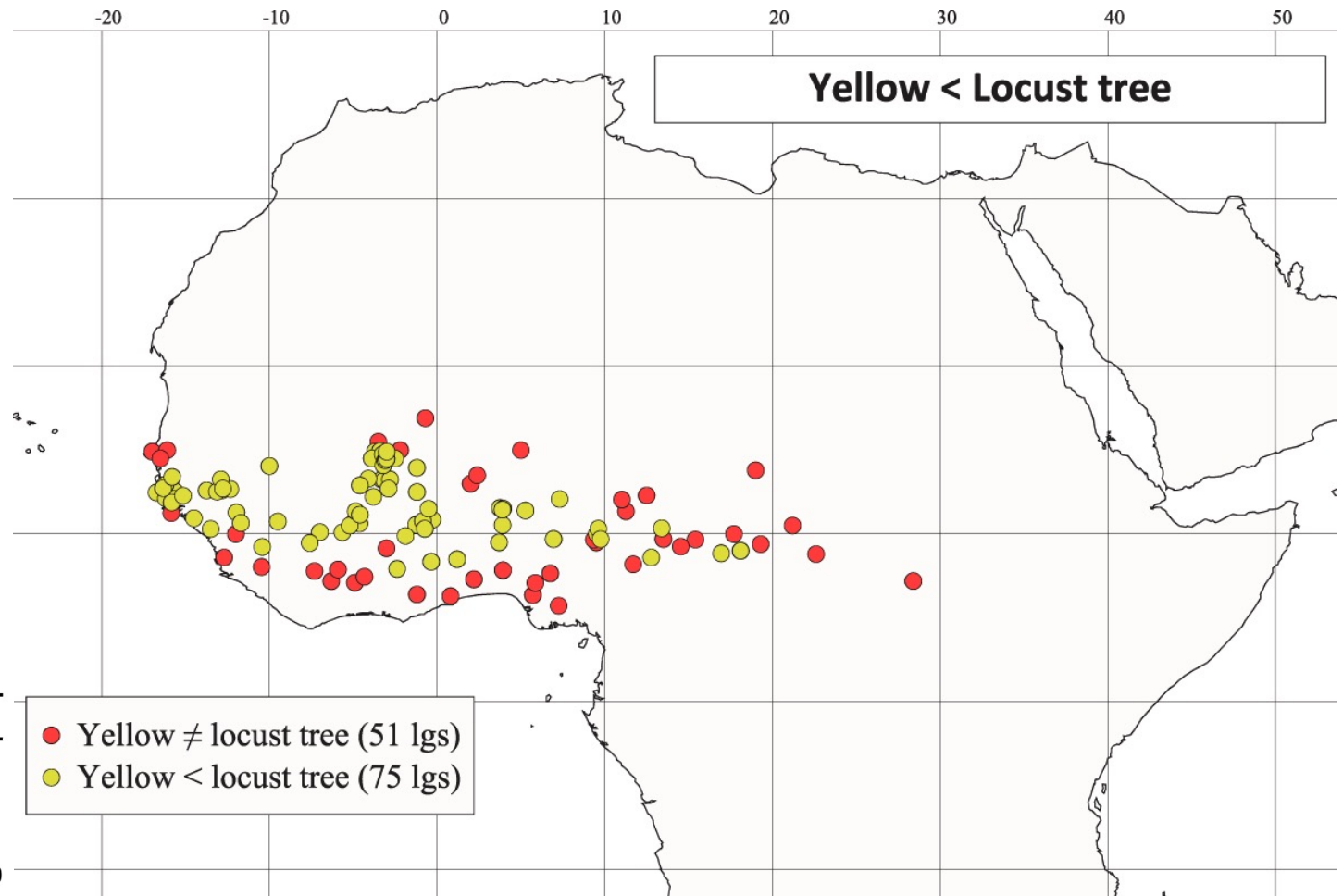


Areal semantics and colexification

- 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)

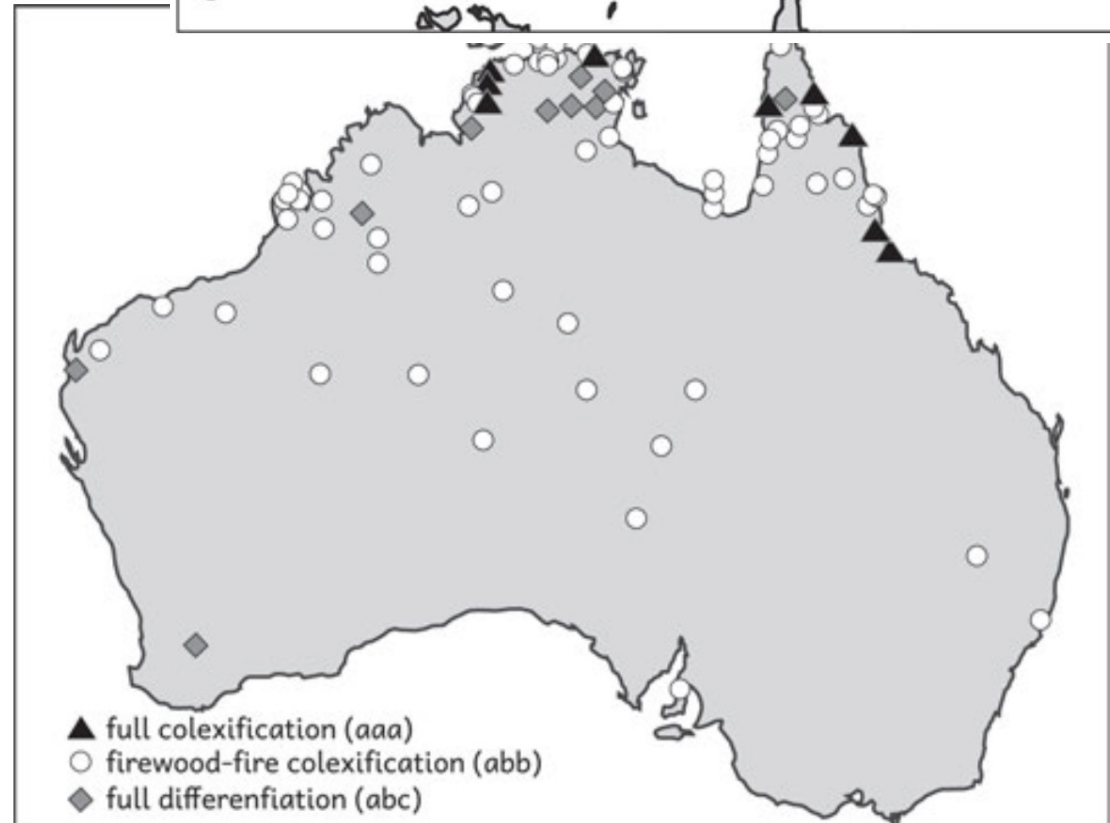
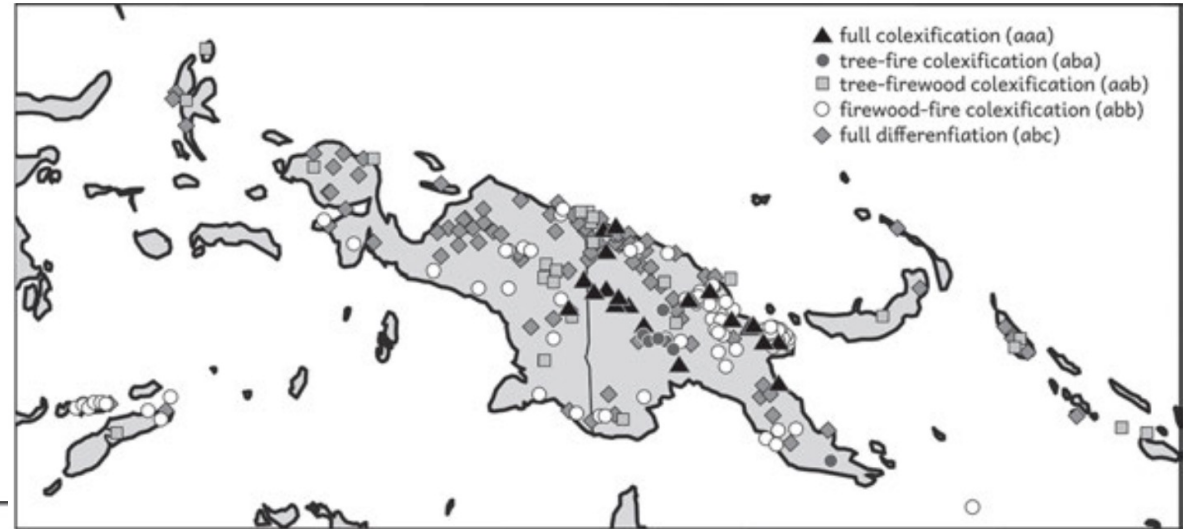
YELLOW is associated with LOCUST.BEAN



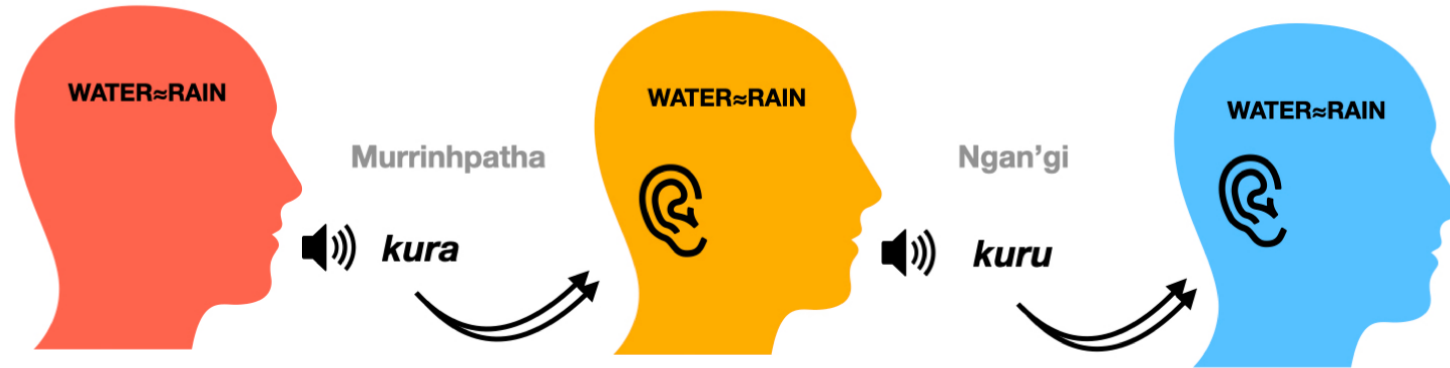
| Language | Glottocode | Country | Form |
|--|------------|--------------|---|
| Shanga (Mande, NC; McCallum Jones 2017) | shan1282 | Nigeria | <i>kpāa'i</i> (<i>kpāa</i> 'locust bean', 'i 'water') ^a |
| Kabiye (Gur, NC; CLNK 1999) | kabi1261 | Togo | <i>sɔtɔ lim</i> (<i>sɔtɔ</i> 'locust bean', <i>lim</i> 'water') |
| Zodi (Chadic, AA; Caron 2002) | dass1243 | Nigeria | <i>jā bētki</i> (<i>jā</i> 'water', <i>bētki</i> 'locust-bean tree') |
| Southern Samo (Mande, NC; SIL Burkina 2003) | sout2844 | Burkina Faso | <i>kùsi</i> ('dry locust bean pods') |
| Yanda Dom (Dogon, NC; Heath 2013) | yand1257 | Mali | <i>lòl-pūrā</i> 'yellow flour from fruit of néré tree bean' (<i>lòl</i> 'locust bean', <i>pūrā</i> 'flower, powder') |
| Sar (Central Sudanic, NS; Palayer 1992) | sarr1246 | Chad | <i>ndùjā mātā</i> (<i>ndùjā</i> 'flour', <i>mātā</i> 'locust bean') |
| Konyagi (Atlantic, NC; Sachot (Santos) 1996) | wame1240 | Senegal | <i>yæ-yéy</i> ('locust tree foliage'; <i>yæ</i> is a noun class prefix, <i>yéy</i> is the stem of 'locust bean'). |

FIRE \approx FIREWOOD \approx TREE
(Schapper et al 2016)

FIRE is associated with
FIREWOOD / TREE



(a) Conceptual transmission *with* lexical transmission: **WATER≈RAIN**



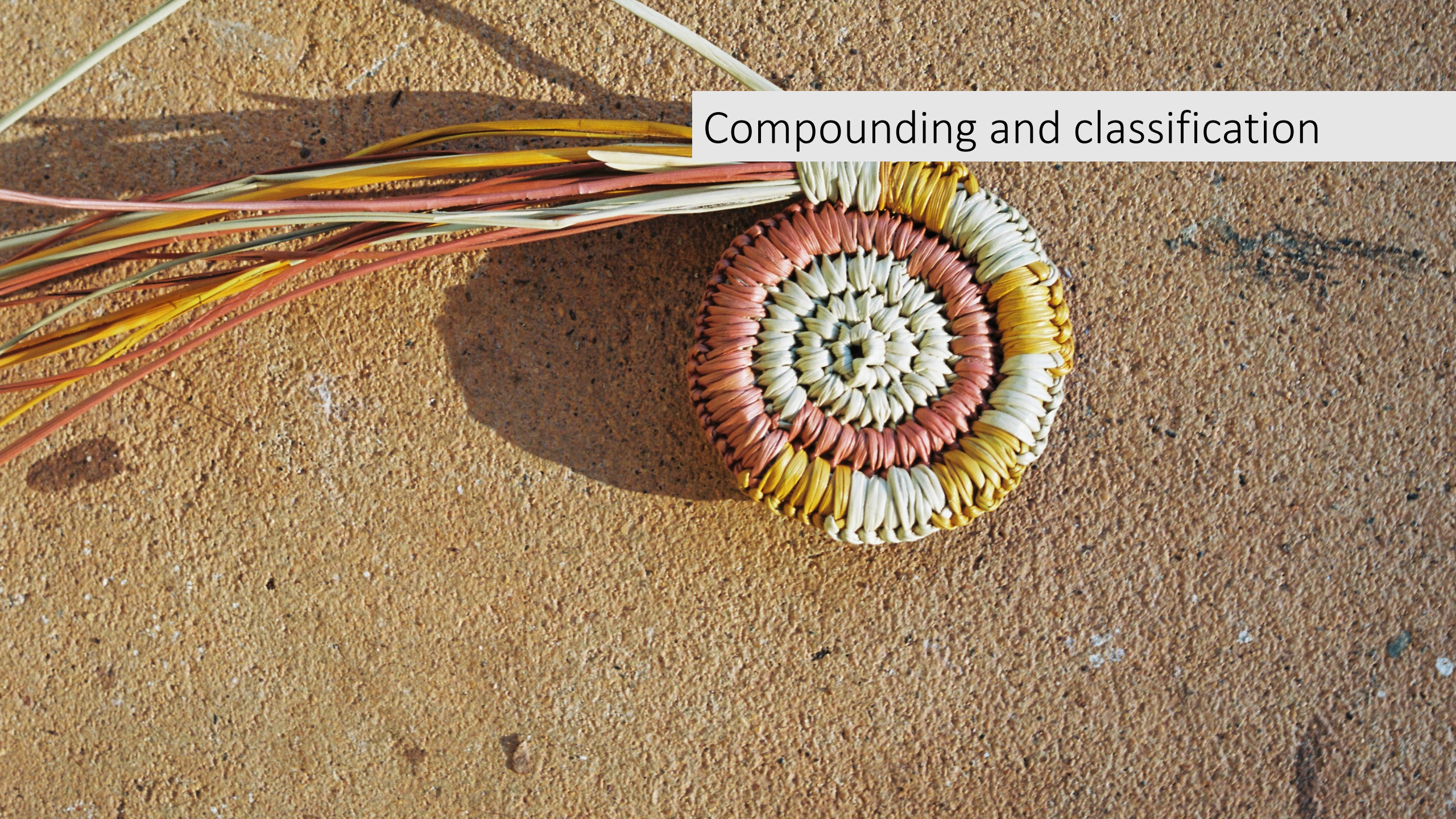
(a) Conceptual transmission *without* lexical transmission: **YEAR≈MONSOON**



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

Murrinpatha (Southern Daly)

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

| | | | |
|--------|--------------|---------|------------|
| mi | ‘veg food’ | kura | ‘water’ |
| ku | ‘animal’ | murrinh | ‘language’ |
| kardu | ‘person’ | tju | ‘weapon’ |
| nanthi | ‘thing’ | thungku | ‘fire’ |
| da | ‘place/time’ | thamul | ‘spear’ |

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:

| | | | | |
|-------------------------|-------------|------------------|-------------|-------------------|
| Annga-ma-nyi | mata | ma-lijap | mata | warlk |
| 2sg/3 VE -get-12 | VE | VE -small | VE | stick |
| “Get a small stick.” | | | | (Singer 2016: 34) |

Bleaching in Arnhem Lang

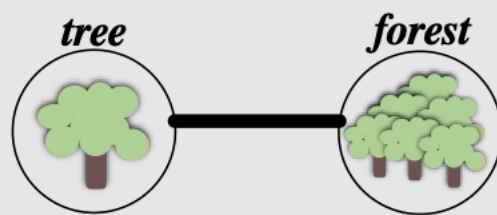
- Mawng classes: Masc, Fem, Veg, 'Land', 'Edible'
- MASC
ilijap 'boy', *ingurlaj* 'male name', *manyardi* 'song', *wurakak* 'crow'
- 'EDIBLE'
karIngurr 'pandanus nut', *mawngku* 'shelter'

Partial colexification

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

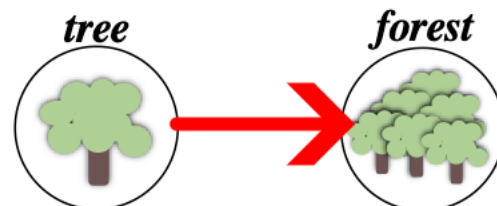
Yaqui "tree": [dʒ u j a]

Yaqui "forest": [dʒ u j a]



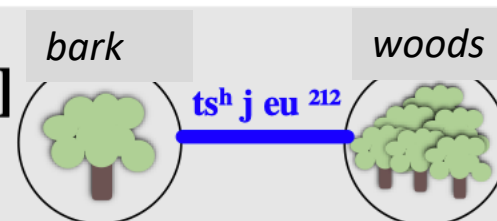
Guìlín "tree": [ɛ y²¹]

Guìlín "forest": [ɛ y²¹ l i ŋ²²]



Fúzhōu "bark": [ts^h j eu²¹² p^h w oi⁵³]

Fúzhōu "woods": [ts^h j eu²¹² l i ŋ⁵³]



Partial colexification

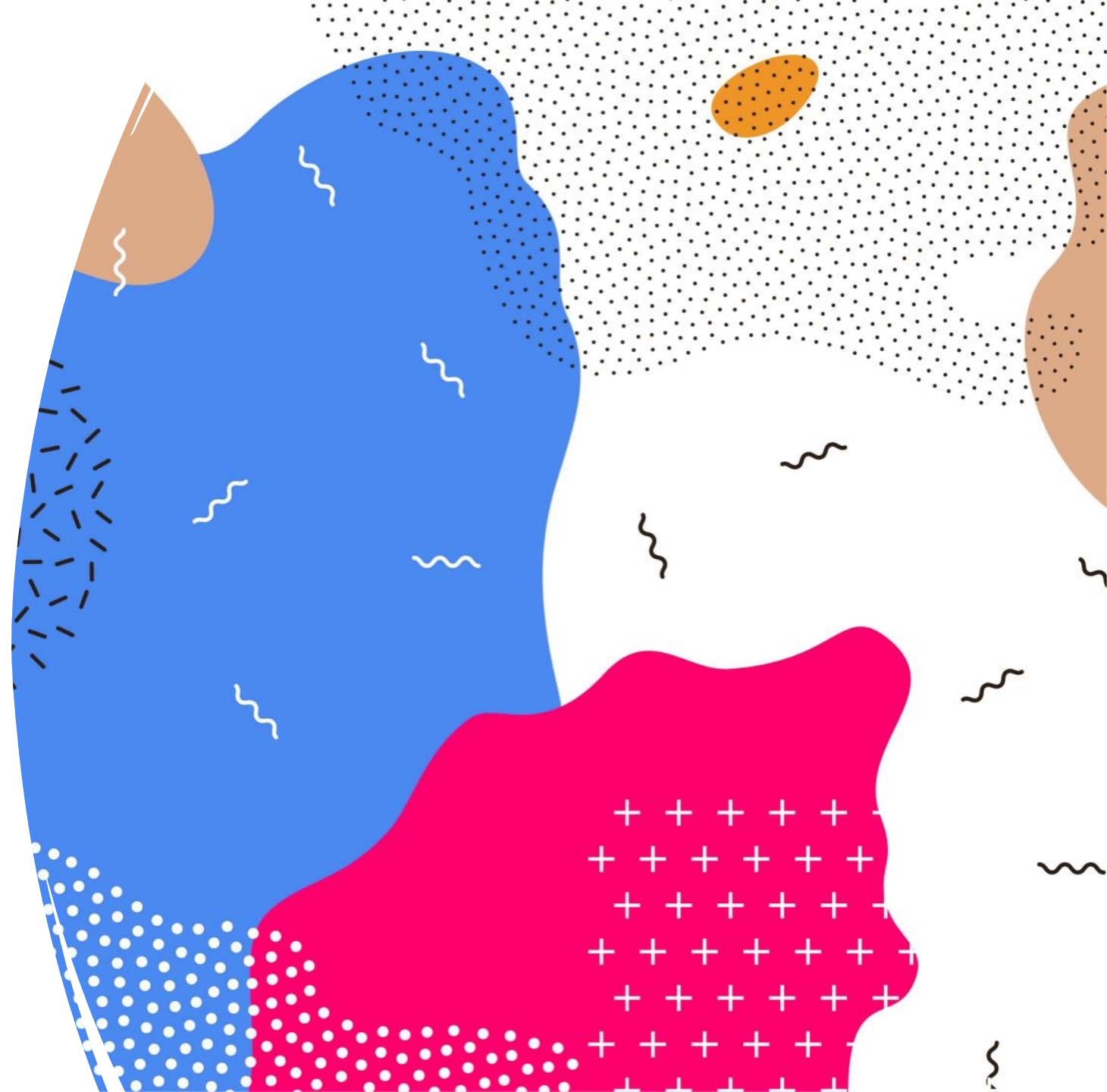
- E.g. *thangku*
YEAR \approx RAIN.SEASON
- But *da thangku* \sim *kura thangka*
TIME + RAIN.SEASON \neq WATER + RAIN.SEASON

Partial colexification

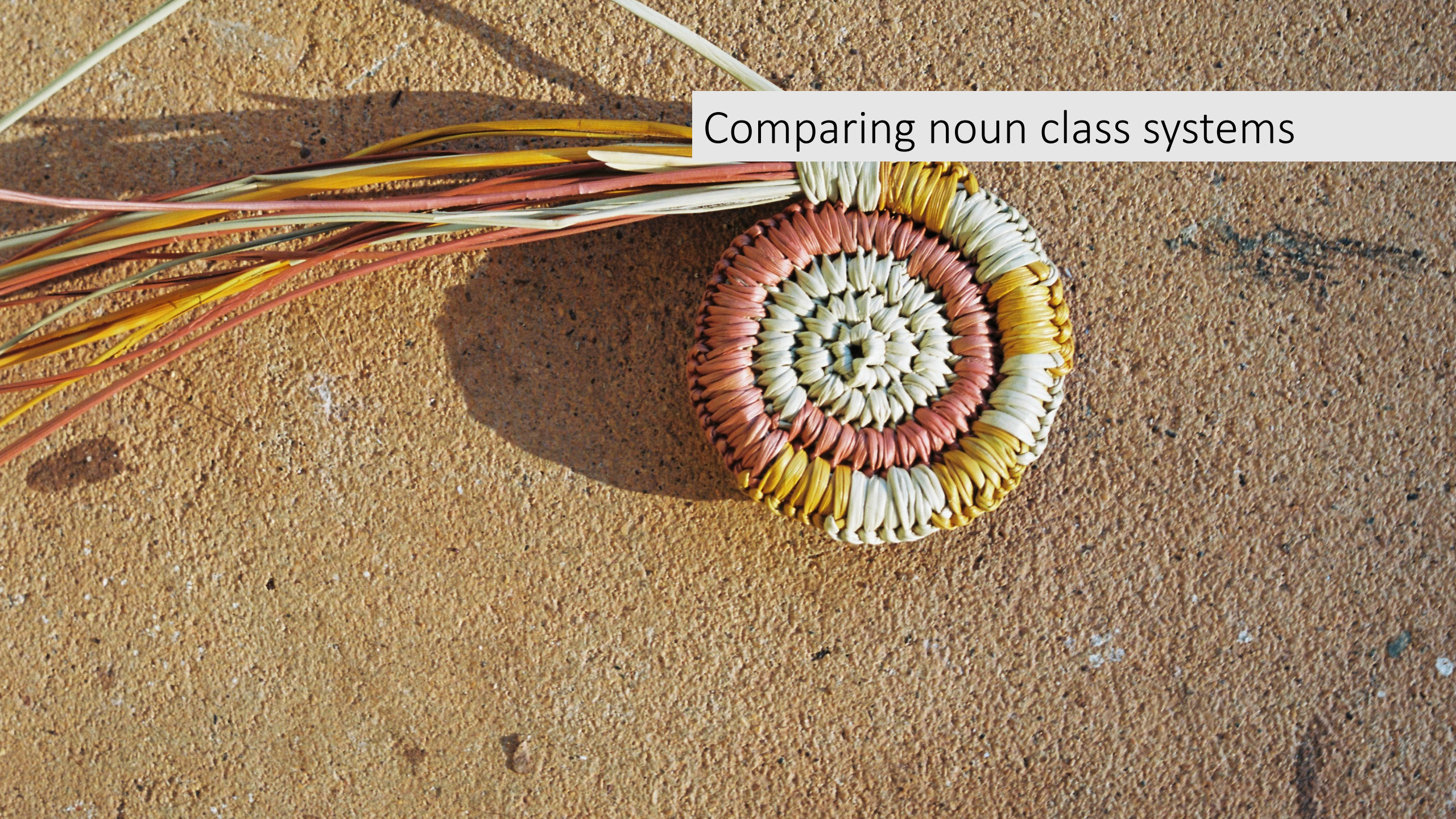
- Many other nouns appear only with one classifier, e.g.:
ku lawarnka 'WALLABY'
nanthi kamarl 'FACE ≈ EYE'
- The classifier nouns themselves, being highly generic, are sites of massive colexification, e.g.
ku 'MEAT≈ANIMAL'
kura 'RAIN≈WATER'

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 karIngurr 'pandanus nut'
ta mawngku 'shelter'

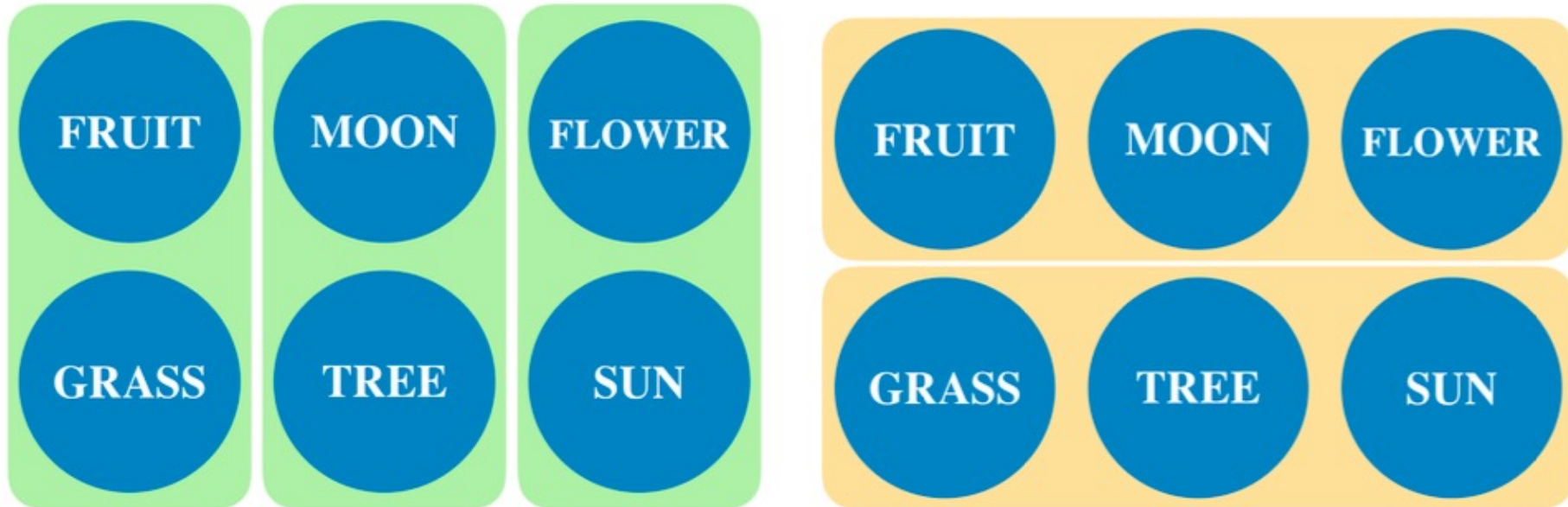


Comparing noun class systems



Indo-European noun classes

McCarthy et al 2020



(a) German, $K = 3$

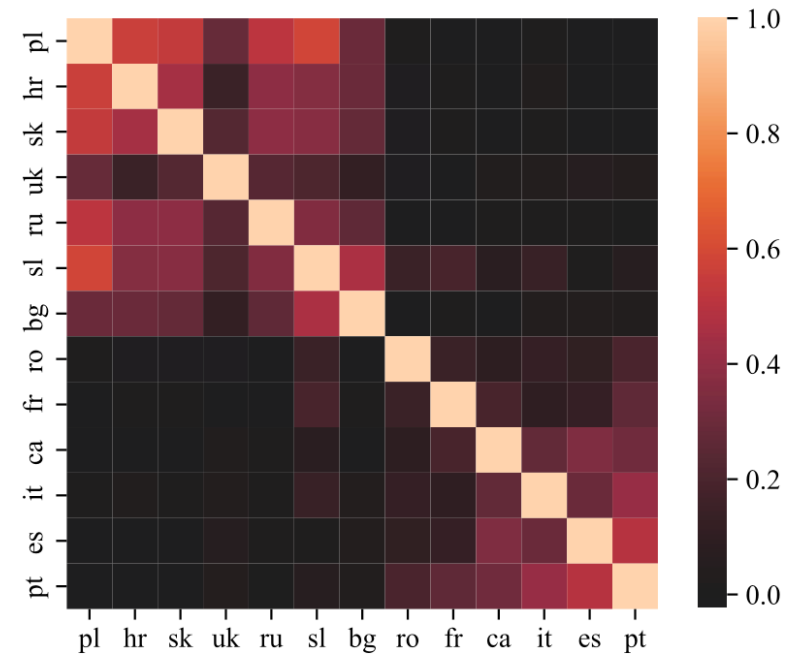
(b) Spanish, $K = 2$

Partition, entropy and mutual information

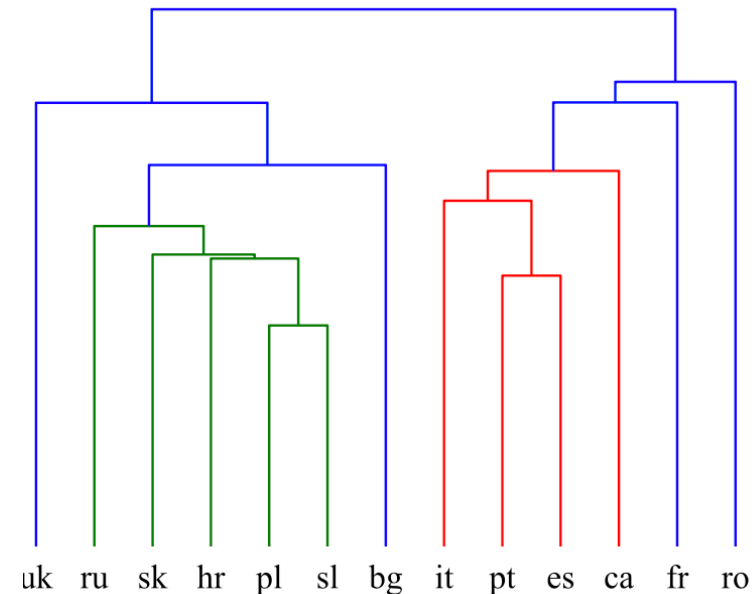
$$H(p) \stackrel{\text{def}}{=} - \sum_{a \in \mathcal{A}} p(a) \log p(a)$$

$$I(A;B) \stackrel{\text{def}}{=} \sum_{A \in \mathcal{A}} \sum_{B \in \mathcal{B}} \frac{|A \cap B|}{N} \log \frac{N |A \cap B|}{|A| |B|}$$

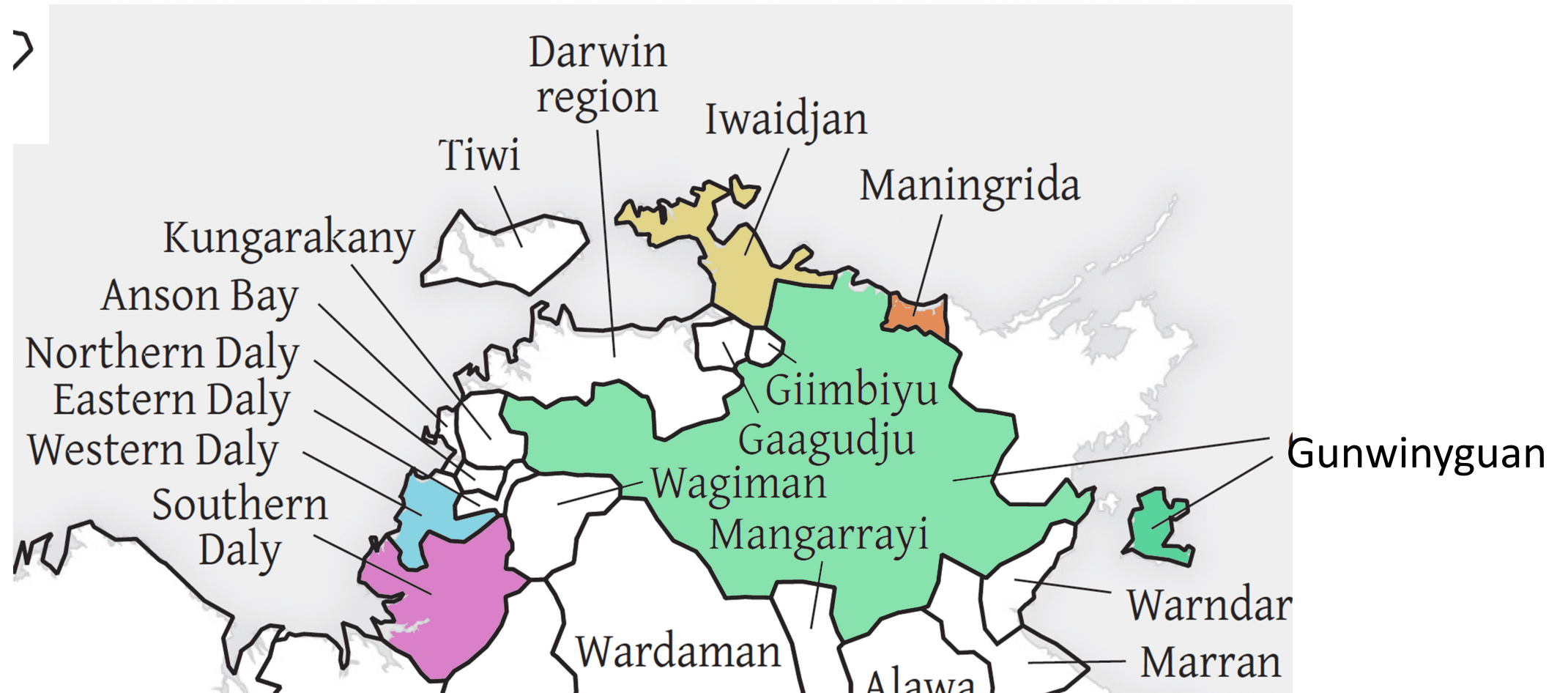
$$\text{AMI}(A,B) \stackrel{\text{def}}{=} \frac{I(A;B) - \mathbb{E}[I(A';B')]}{\max I(A',B') - \mathbb{E}[I(A';B')]}$$



(c) Adjusted Mutual Information



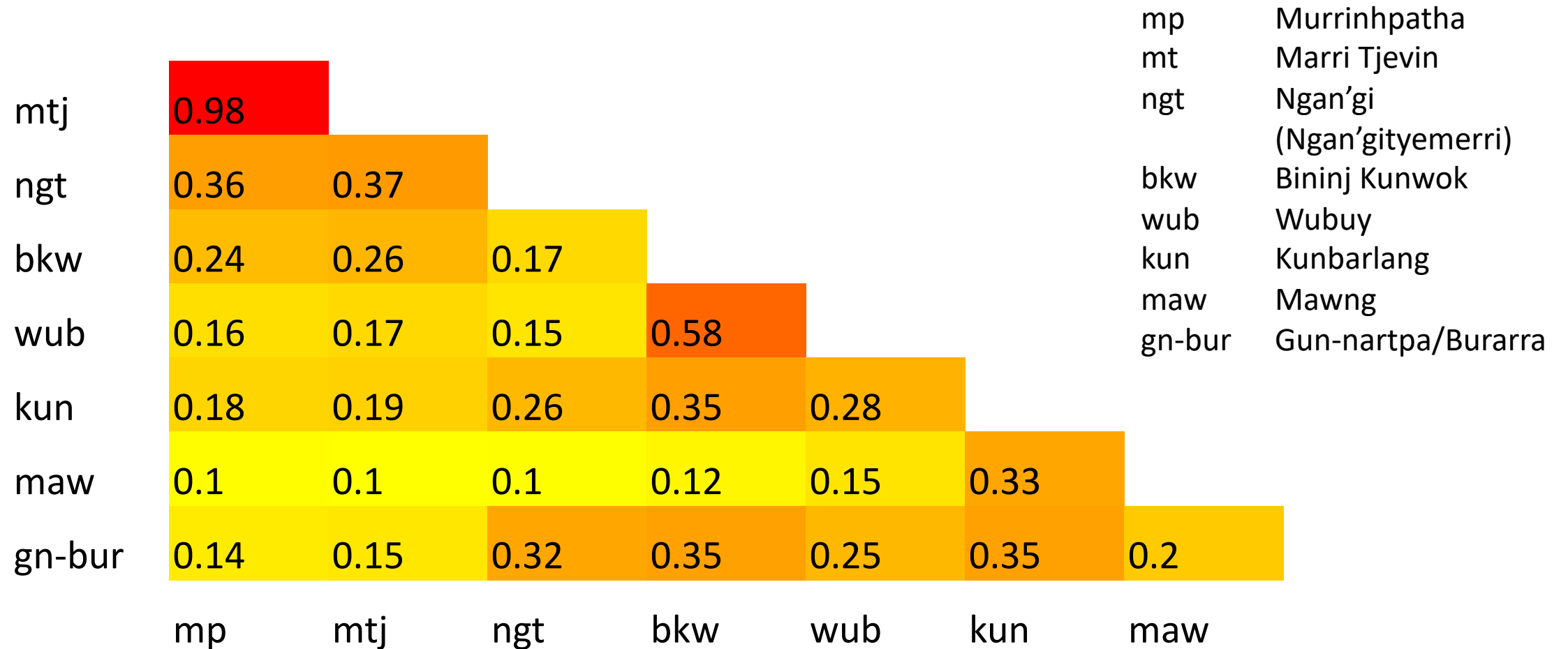
A study of northern Australia



Concepts and classes

| | Daly | | Arnhem | |
|-----------|-----------------------|-----------------------|-----------------------|---------------------|
| | Murrinhpatha | Marri Tjevin | Wubuy | Mawng |
| BIRD | <i>ku</i> (ANIM) | <i>awu</i> (ANIM) | <i>ngarra-</i> (NEUT) | <i>na-</i> (MASC) |
| FISH | <i>ku</i> (ANIM) | <i>awu</i> (ANIM) | <i>ngarra-</i> (NEUT) | <i>na-</i> (MASC) |
| WASP | <i>ku</i> (ANIM) | <i>awu</i> (ANIM) | <i>ana-</i> (FEM) | <i>niny-</i> (FEM) |
| FIREWOOD | <i>thungku</i> (FIRE) | <i>tjendji</i> (FIRE) | <i>ngarra-</i> (NEUT) | <i>ma-</i> (VEG) |
| FIRESTICK | <i>thungku</i> (FIRE) | <i>tjendji</i> (FIRE) | <i>ngarra-</i> (NEUT) | <i>niny-</i> (FEM) |
| ALCOHOL | <i>kura</i> (LIQ) | <i>wudi</i> (LIQ) | <i>ngarra-</i> (NEUT) | <i>nung-</i> (LAND) |

Adjusted Mutual Information (AMI)

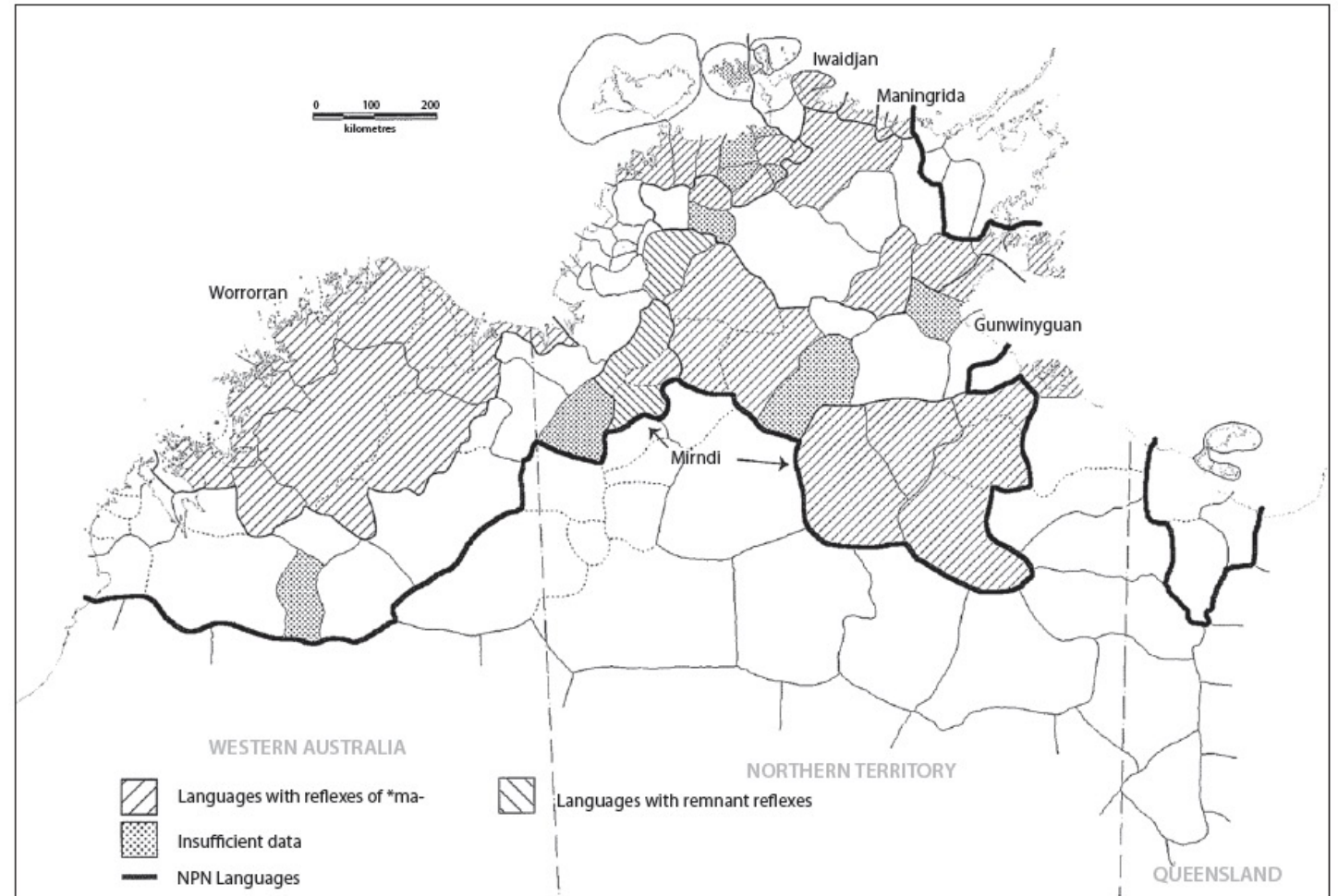


Inheritance of noun class from proto-Australian?

- 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

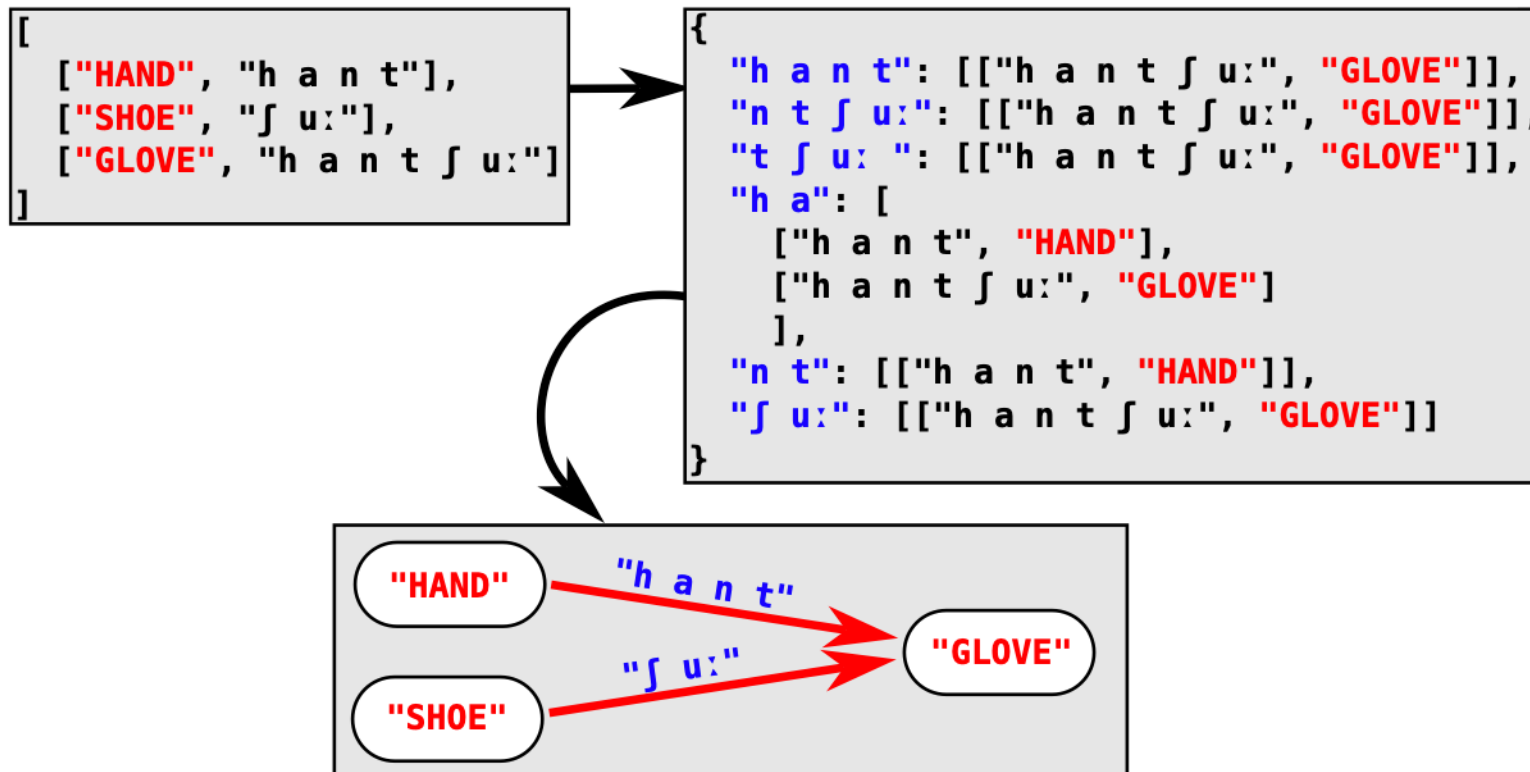
| | | |
|--------------|---------|----------|
| Pipil | 2_ashes | nex-ti |
| Pochutec | 2_ashes | nox-t |
| ProtoAztecán | 2_ashes | *nix-tli |

| | | |
|------------------------|----------|---|
| Pipil-2_ashes-1 | nex-ti | 2 |
| Pochutec-2_ashes-1 | nox-t | 2 |
| ProtoAztecán-2_ashes-1 | *nix-tli | 2 |

forms.csv
cognates.csv
(Davletshin 2012)
24

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?

THE END