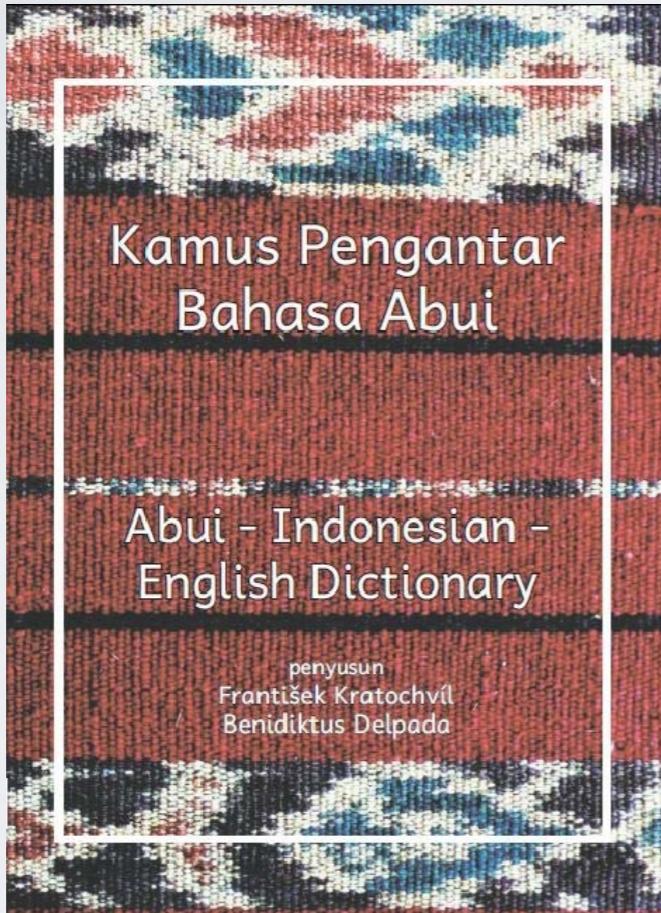


EXTENDING LEXICAL RESOURCES FOR ABUI AND SAWILA (PAPUAN, E INDONESIA)



František KRATOCHVÍL

Nanyang Technological University, Singapore

OUTLINE OF THE TALK

Lexicographic workflow for ‘small’ languages

- Wordlists
- Textual corpus
- Rapid Words
- WordNet

Purpose and users

- L1 Community
- L1 Education & Literacy development
- Linguists
 - Comparative (Austronesian Basic Vocabulary Database, ACD)
 - Typology
 - Language contact
 - Lexicography and Meaning (PanLex)

Issues

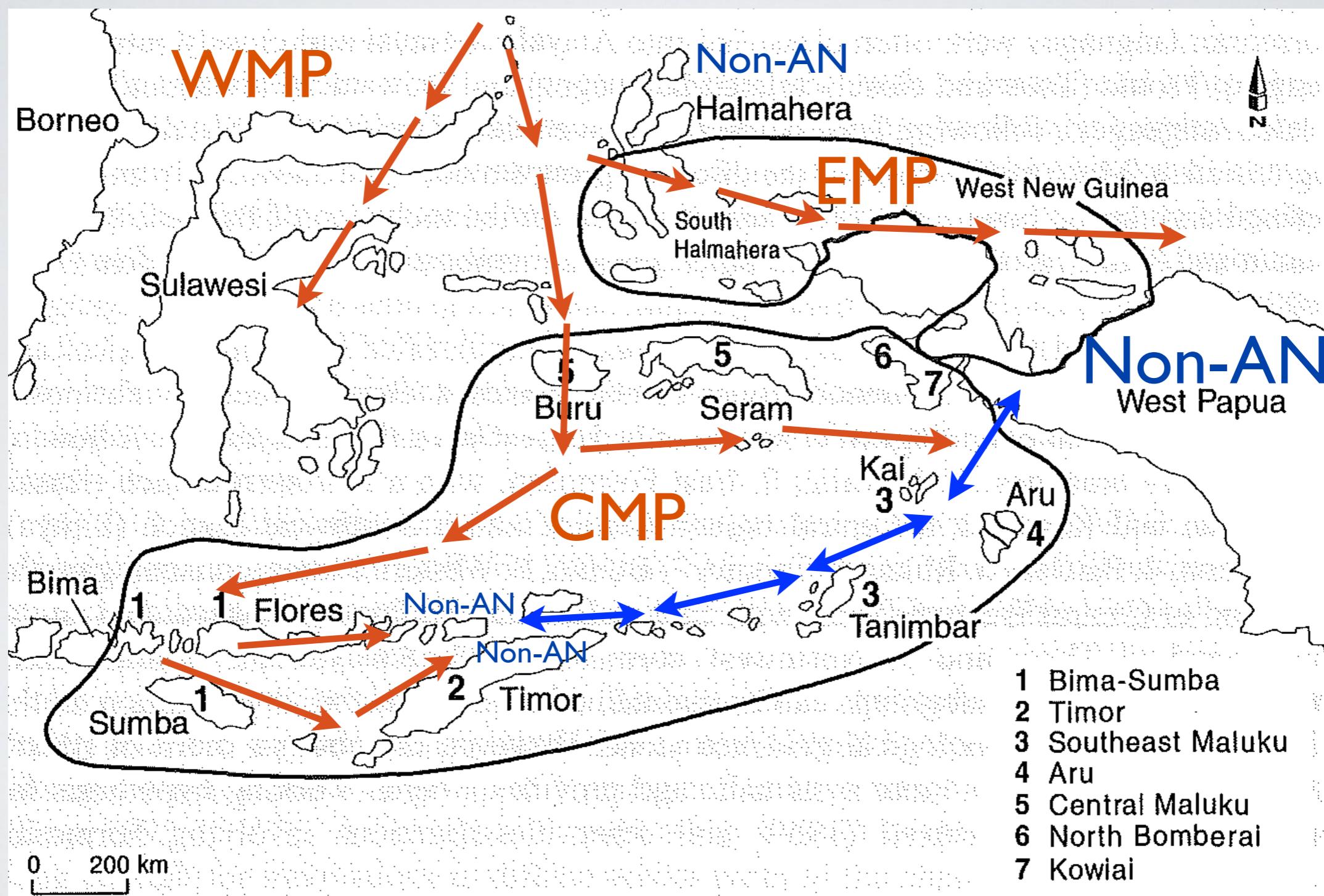
- Multiword expressions, idioms, proverbs, frequent metaphors, light verbs (many SVC)
- Organization and structure of dictionaries

BACKGROUND LANGUAGES OF ALOR-PANTAR

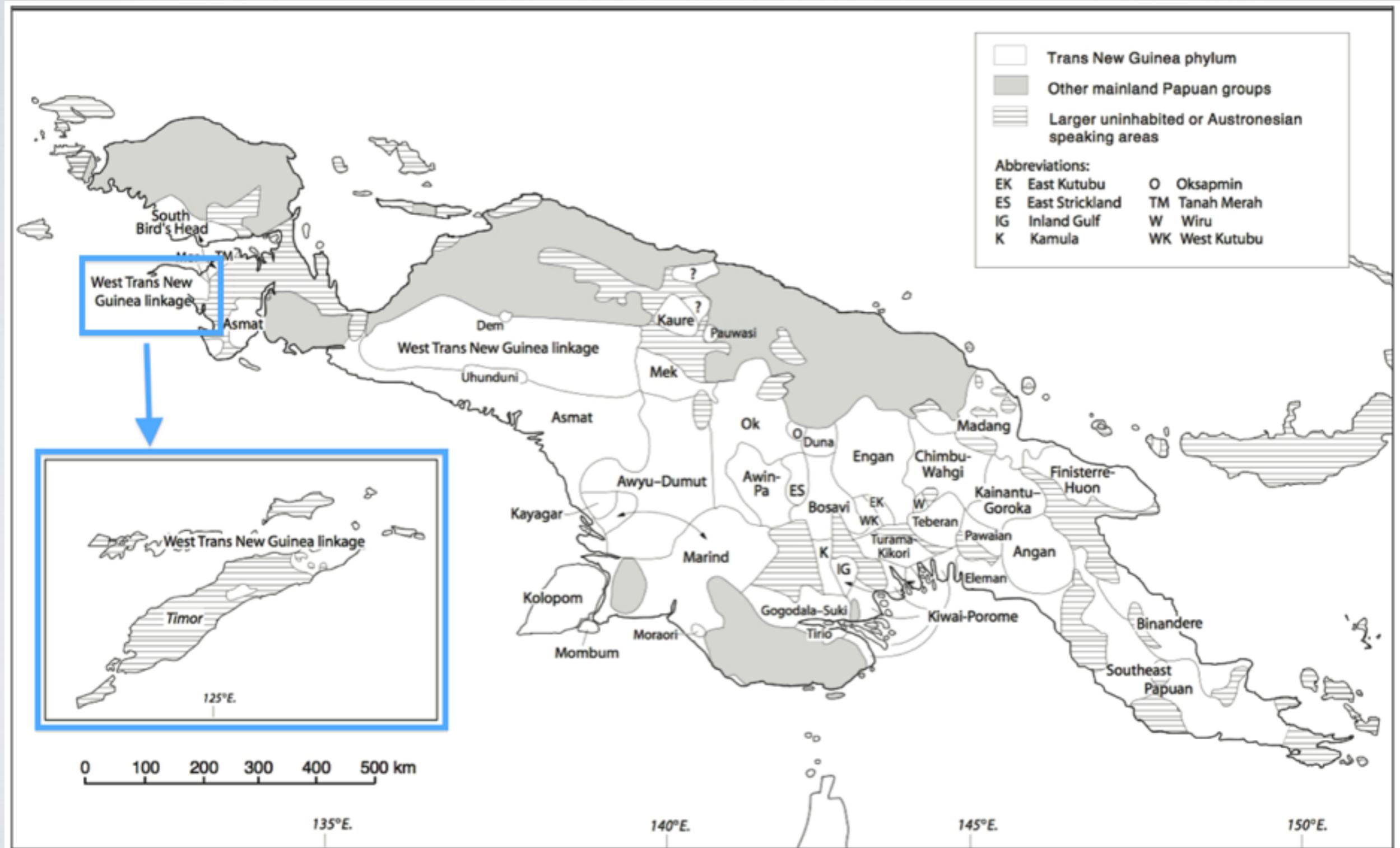
EASTERN INDONESIA - GEOGRAPHIC LOCATION OF ALOR-PANTAR



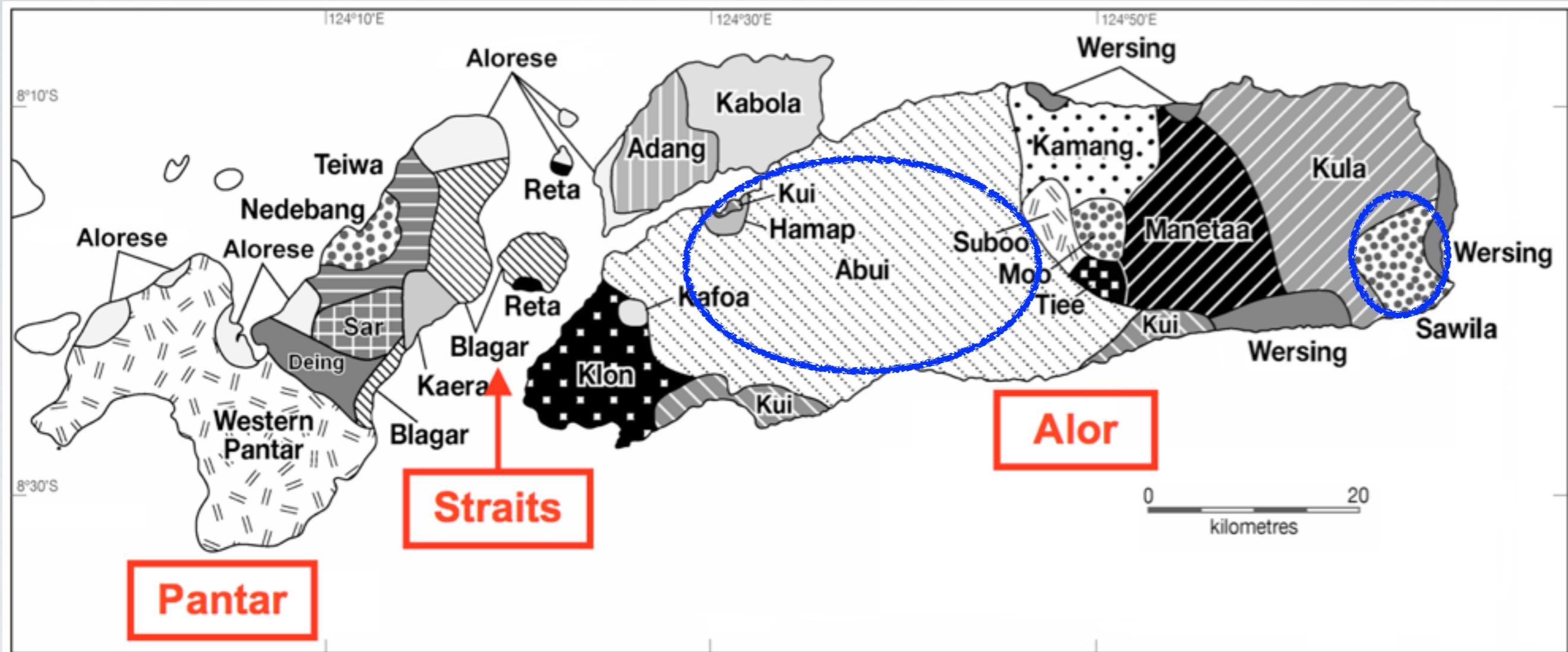
BODY-PART BASED EMOTION PREDICATES IN EAST INDONESIA (KLAMER 1999; MUSGRAVE 2006; DIX GRIMES N.D.)



LINGUISTIC SITUATION (NOT SUPPORTED IN HOLTON ET AL. 2012)



LINGUISTIC SITUATION IN ALOR-PANTAR ARCHIPELAGO



LINGUISTIC SITUATION

1. Papuan outlier (some 1000 km from the New Guinea mainland)
2. tentatively linked with Trans New Guinea (TNG) family - western Bomberai peninsula languages (Ross 2005) based on pronominal evidence *>>> not supported in Holton et al 2012*
3. small languages (max. 20,000 speakers, some < 1,000)
4. surrounded by Austronesian languages
5. long history of genetic admixture (Mona et al. 2009)
6. possibly long-lasting language contact and linguistic convergence (Holton et al. to appear)

LEXICOGRAPHIC WORKFLOW FOR ‘SMALL’ LANGUAGES

SERVING MANY MASTERS

Wordlist: for comparative linguists

- Expected: a list of diachronically stable items (Swadesh 100, 200, 400 or any other custom size, depending on the language family and tradition)
- There is a reasonable expectation that such wordlist will be included in grammar sketches, or published

Language description: anthropological view

- There is a view that language description should be based on ‘natural data’.
- Usually this means ‘monologue’
- Elicitation is seen as problematic by some (to some extent discouraging from systematic lexical work)
- Result: poor dictionaries that never get published (little or no perceived value)

Corpus: failing the ‘functional credo’

- usually below 100k words, usually leading to impressionistic statements about the grammar

Swadesh 100-word list

| | | | |
|------------|----------------|--------------|------------|
| 1. I | 30. blood | 59. know | 88. green |
| 2. thou | 31. bone | 60. sleep | 89. yellow |
| 3. we | 32. grease | 61. die | 90. white |
| 4. this | 33. egg | 62. kill | 91. black |
| 5. that | 34. horn | 63. swim | 92. night |
| 6. who? | 35. tail | 64. fly | 93. hot |
| 7. what? | 36. feather | 65. walk | 94. cold |
| 8. not | 37. hair | 66. come | 95. full |
| 9. all | 38. head | 67. lie | 96. new |
| 10. many | 39. ear | 68. sit | 97. good |
| 11. one | 40. eye | 69. stand | 98. round |
| 12. two | 41. nose | 70. give | 99. dry |
| 13. big | 42. mouth | 71. say | 100. name |
| 14. long | 43. tooth | 72. sun | |
| 15. small | 44. tongue | 73. moon | |
| 16. woman | 45. fingernail | 74. star | |
| 17. man | 46. foot | 75. water | |
| 18. person | 47. knee | 76. rain | |
| 19. fish | 48. hand | 77. stone | |
| 20. bird | 49. belly | 78. sand | |
| 21. dog | 50. neck | 79. earth | |
| 22. louse | 51. breasts | 80. cloud | |
| 23. tree | 52. heart | 81. smoke | |
| 24. seed | 53. liver | 82. fire | |
| 25. leaf | 54. drink | 83. ash | |
| 26. root | 55. eat | 84. burn | |
| 27. bark | 56. bite | 85. path | |
| 28. skin | 57. see | 86. mountain | |
| 29. flesh | 58. hear | 87. red | |

WORDLISTS



Morris Swadesh
1909-1967

An American linguist working on the indigenous languages of the Americas
Banned from US academia during the McCarthy era

Basic vocabulary - thought to be diachronically stable - i.e. changing slower than other lexical items and unlikely to be borrowed. Morris Swadesh produced word lists of various sizes that he deemed to contain the basic vocabulary. Examples:

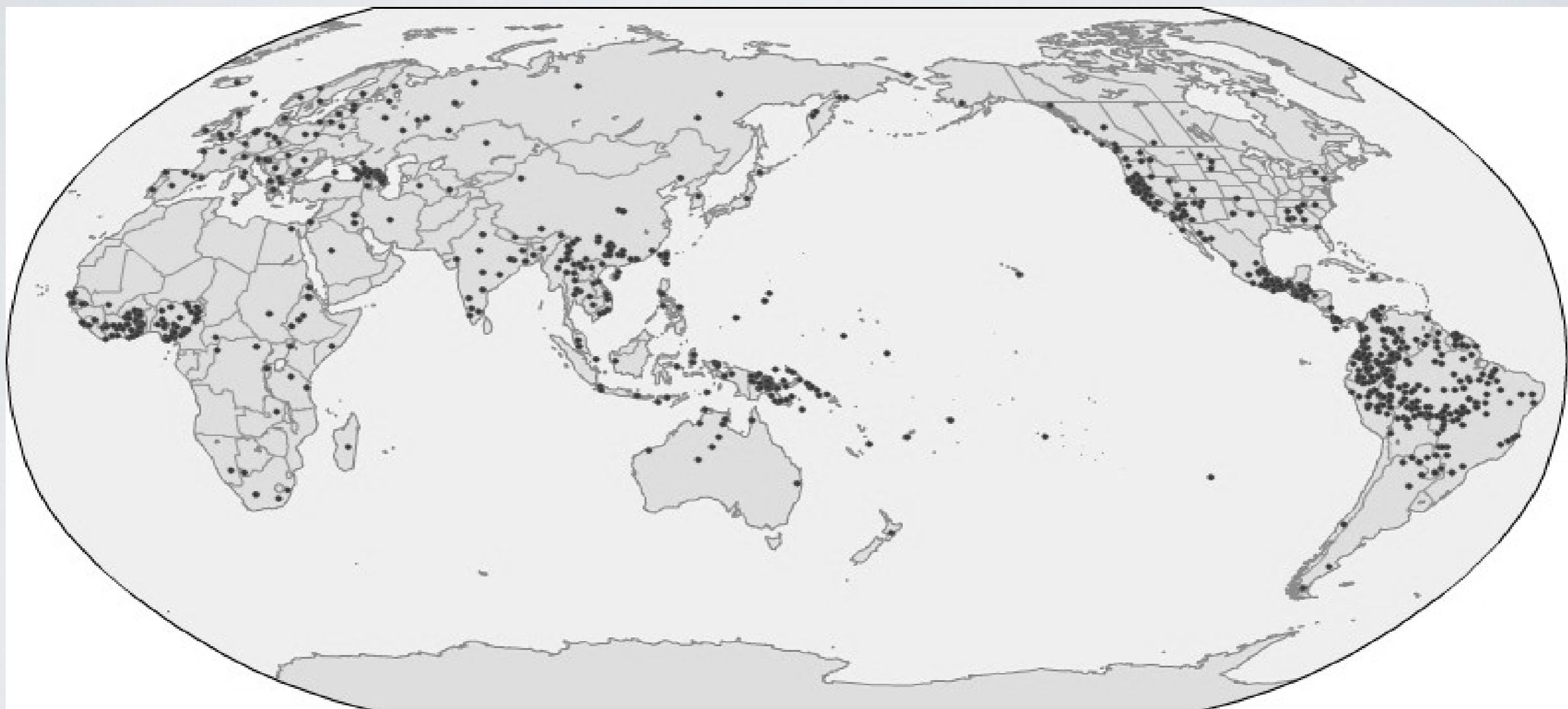
Austronesian Basic Vocabulary Database

Genealogical Language List

World Loanword Database

DIACHRONIC STABILITY (HOLMAN ET AL. 2008)

The dots on the map represent the 876 languages sampled whose wordlists were evaluated for their relative similarity and checked for how often these words were borrowed.



Item stability refers to the degree to which words for an item are retained over time and not replaced by another lexical item from the language itself or a borrowed element: words resistant to replacement are more stable than otherwise. (Holman et al. 2008:334)

TOP DIACHRONICALLY STABLE WORDS (HOLMAN ET AL. 2008)

| <i>Rank</i> | <i># in list</i> | <i>Meaning</i> | <i>Stability</i> | <i>Rank</i> | <i># in list</i> | <i>Meaning</i> | <i>Stability</i> |
|-------------|------------------|----------------|------------------|-------------|------------------|----------------|------------------|
| 1 | 22 | *louse | 42.8 | 30 | 11 | *one | 27.4 |
| 2 | 12 | *two | 39.8 | 31 | 41 | *nose | 27.3 |
| 3 | 75 | *water | 37.4 | 32 | 95 | *full | 26.9 |
| 4 | 39 | *ear | 37.2 | 33 | 66 | *come | 26.8 |
| 5 | 61 | *die | 36.3 | 34 | 74 | *star | 26.6 |
| 6 | 1 | *I | 35.9 | 35 | 86 | *mountain | 26.2 |
| 7 | 53 | *liver | 35.7 | 36 | 82 | *fire | 25.7 |
| 8 | 40 | *eye | 35.4 | 37 | 3 | *we | 25.4 |
| 9 | 48 | *hand | 34.9 | 38 | 54 | *drink | 25.0 |
| 10 | 58 | *hear | 33.8 | 39 | 57 | *see | 24.7 |
| 11 | 23 | *tree | 33.6 | 40 | 27 | bark | 24.5 |
| 12 | 19 | *fish | 33.4 | 41 | 96 | *new | 24.3 |
| 13 | 100 | *name | 32.4 | 42 | 21 | *dog | 24.2 |
| 14 | 77 | *stone | 32.1 | 43 | 72 | *sun | 24.2 |
| 15 | 43 | *tooth | 30.7 | 44 | 64 | fly | 24.1 |
| 16 | 51 | *breasts | 30.7 | 45 | 32 | grease | 23.4 |
| 17 | 2 | *you | 30.6 | 46 | 73 | moon | 23.4 |
| 18 | 85 | *path | 30.2 | 47 | 70 | give | 23.3 |
| 19 | 31 | *bone | 30.1 | 48 | 52 | heart | 23.2 |
| 20 | 44 | *tongue | 30.1 | 49 | 36 | feather | 23.1 |
| 21 | 28 | *skin | 29.6 | 50 | 90 | white | 22.7 |
| 22 | 92 | *night | 29.6 | 51 | 89 | yellow | 22.5 |
| 23 | 25 | *leaf | 29.4 | 52 | 20 | bird | 21.8 |
| 24 | 76 | rain | 29.3 | 53 | 38 | head | 21.7 |
| 25 | 62 | kill | 29.2 | 54 | 79 | earth | 21.7 |
| 26 | 30 | *blood | 29.0 | 55 | 46 | foot | 21.6 |
| 27 | 34 | *horn | 28.8 | 56 | 91 | black | 21.6 |
| 28 | 18 | *person | 28.7 | 57 | 42 | mouth | 21.5 |
| 29 | 47 | *knee | 28.0 | 58 | 88 | green | 21.1 |

TIME/PEOPLE/RESULT EVALUATION - ABUI

- František Kratochvíl: 2003-2008 (5 years, 3 years intensive)
- Benidiktus Delpada (1/2 year intensive)
- Published and distributed to the community in 2008.
- Financial support of Leiden University and Kebupaten Alor

[Side-product of the corpus \(Multi-Dictionary Formatter in SIL Toolbox\)](#)

Version 1: 2,700 lemmas, definitions in Indonesian and English, example sentences, synonyms, inflectional paradigms; short grammar sketch in Indonesian

Version 2: 3,000 lemmas, as above + Semantic Domains

Version 3: in 2016 or later (15,000 lemmas, remaining as above, WordNet linked)

ABUI LEXICON DEVELOPMENT STAGES

lemma + pronunciation
wordclass (ind)
definition (ind/eng)

afu [ʔa'fu] kb. ikan. fish. **Na afu tahai.** Saya mencari ikan, saya berusaha sebagai nelayan. I am fishing. lit.: I search fish.

Abui Dictionary (2008)

afu

[ʔa'fu] *n.* fish. **Na afu tahai.** I am fishing. lit.: I search fish. *Idiom:* **afu ia** ‘may’ ‘*masa petani memanen jagung/bulan mei*’ ‘musim panen jagung/bulan mei’; *Spec:* **afu kasing** ‘split dry fish also a kind off fishes such as rayfish but has no htaili’ ‘*ikan kering terbelah juga sj. ikan spt pari tetapi tidak mempunyai ekor*’ ‘ikan bela’; **afu tulok** ‘dried small fish on a piece of bamboo’ ‘*ikan kecil-kecil yg kering telah ditusuk pd tali*’ ‘ikan tali’; **tongkol** ‘tuna fish’ ‘*ikan tongkol*’ ‘ikan tongkol’; **tafeel** ‘shark’ ‘*hiu*’ ‘ikan hiu’; **fahaai** ‘crocodile’ ‘*buaya*’ ‘buaya’; **ramaak** ‘garfish’ ‘*ikan todak*’ ‘ikan todak’; **metiing isi, fariki, song afu, tefang** ‘sardine’ ‘*ikan sarden*’ ‘ikan tembang’; **pasimara, penkar, mora, fokal ara diesing, wesing, win, yeikoi** ‘turtle’ ‘*penyu/kura-kura*’ ‘penyu’; **belo-belo, matanggolong, kaai sapaat, yeri kika, hiyeng fokal, kafiyei mona, waysa afu, silik afu, wal hawa homi kika, tan, kela afu, spidol, lamoor, pal** ‘rayfish’ ‘*ikan pari*’ ‘ikan pari’; **leipang, leber, suwai** ‘whale’ ‘*ikan paus*’ ‘ikan paus’; **ku beeka, pokdai tala, fat tafuyang, malel, leipang, timiing, maleki, liwang fariki, kila fariki, afu tool, kombong, hiyeng rama, kora-kora, kopi taka, doomo.**

Category: Fish.

← idiom

← hyponyms

semantic domain

Abui Dictionary (2014)

SAWILA LEXICON DEVELOPMENT

lemma + pronunciation →
wordclass (eng)
definition (eng/ind/alon.malay)

2,800 lemmas, example sentences,
synonyms, SD, inflection [100 pages
A4]

- František Kratochvíl: 2007-2014 (7 years, irregular, 2 years intensive)
- Anderias Malaikosa (1 day/week for 6 months)
- Aprianus Malaikosa (1 day/week for 3 months)
- Isak Bantara (checking, informant)

etymology: *hab(i)

api [?a'pi] *n.* fish; *ikan*. Malay *ikan*. **tayaa api se**. We go down fishing. *Kita mencari ikan*. Kita turun cari *ikan*. **Ganaku api li yaala galuri gako du**. The two of them were fishing with the net. *Mereka berdua sedang menjala ikan*. Dorang dua ada jala *ikan*.
Part: **gayiri** ‘mouth’ **‘mulut’** ‘mulut’; **gapiini** ‘lips’ **‘bibir’** ‘bibir’; **gawa makiiling** ‘barbel’ **‘jenggot’** ‘jenggot’; **galalaamang** ‘gill’ **‘insang’** ‘insang’; **gawa yaka gawaaramana** ‘operculum (gill cover)’ **‘kulit insang’** ‘kulit insang’; **giliili** ‘dorsal fin’ **‘sirip belakang’** ‘sirip belakang’; **giwuroka gakili lili** ‘pectoral fin’ **‘sirip samping’** ‘sirip samping’; **giwoora** ‘caudal fin (tail)’ **‘ekor’** ‘ekor’; **gitiini** ‘lateral line’ **‘garis punggung’** ‘garis punggung’.
Etym: *a[-voiceLAB][highV]. Category: **1.6.1.5.**
Fish.

examples ←

parts ←

semantic domain

EVALUATION

- Language documentation approach is slow and relies on one or two authors (usually overcommitted)
- We know of at least 20 Alor-Pantar languages but have small size dictionaries (2000+ words) for only 5 of them (Abui, Kamang, Sawila, Teiwa, and Western Pantar) created by 4 different linguists. For the remaining languages we only have wordlists. Old wordlists (from the Dutch era) become irrelevant because the villages were moved in 1960s
- None of the dictionaries contains more than 4000 words although each of them took several years to create
- Random gaps - method of lexicon collection, preferences of compilers
- Only 4 have been printed and distributed in the communities.
- No sources in Dutch, but where available difficult to use (most people studying languages of Indonesia do not speak Dutch)

RAPID WORDS
DICTIONARY DEVELOPMENT CROWDSOURCING

SEMANTIC DOMAINS - ONTOLOGY FOR DICTIONARY DEVELOPMENT

Building on the comparative dictionaries of synonyms and on thesauri tradition in lexicography. The ontology has to be coarse-grained to allow for semantic shift (intuitive and associative relationships included).

| No. | Domain | No. | Domain | No. | Domain |
|-----|-----------------------------|-----|-------------------------------|-----|---|
| 1 | Universe, creation | 4 | Social behavior (cont) | 7 | Physical actions |
| 1.1 | Sky | 4.5 | Authority | 7.1 | Posture |
| 1.2 | World | 4.6 | Government | 7.2 | Move |
| 1.3 | Water | 4.7 | Law | 7.3 | Move something |
| 1.4 | Living things | 4.8 | Conflict | 7.4 | Have, be with |
| 1.5 | Plant | 4.9 | Religion | 7.5 | Arrange |
| 1.6 | Animal | 5 | Daily life | 7.6 | Hide |
| 1.7 | Nature, environment | 5.1 | Household equipment | 7.7 | Physical impact |
| 2 | Person | 5.2 | Food | 7.8 | Divide into pieces |
| 2.1 | Body | 5.3 | Clothing | 7.9 | Break, wear out |
| 2.2 | Body functions | 5.4 | Adornment | 8 | States |
| 2.3 | Sense, perceive | 5.5 | Fire | 8.1 | Quantity |
| 2.4 | Body condition | 5.6 | Cleaning | 8.2 | Big |
| 2.5 | Healthy | 5.7 | Sleep | 8.3 | Quality |
| 2.6 | Life | 5.8 | Manage a house | 8.4 | Time |
| 3 | Language and thought | 5.9 | Live, stay | 8.5 | Location |
| 3.1 | Soul, spirit | 6 | Work and occupation | 8.6 | Parts of things |
| 3.2 | Think | 6.1 | Work | 9 | Grammar |
| 3.3 | Want | 6.2 | Agriculture | 9.1 | General words |
| 3.4 | Emotion | 6.3 | Animal husbandry | 9.2 | Part of speech |
| 3.5 | Communication | 6.4 | Hunt and fish | 9.3 | Very |
| 3.6 | Teach | 6.5 | Working with buildings | 9.4 | Semantic constituents related to verbs |
| 4 | Social behavior | 6.6 | Occupation | 9.5 | Case |
| 4.1 | Relationships | 6.7 | Tool | 9.6 | Connected with, related |
| 4.2 | Social activity | 6.8 | Finance | 9.7 | Name |
| 4.3 | Behavior | 6.9 | Business organization | | |
| 4.4 | Prosperity, trouble | | | | |

Figure 2: Top two levels of the Semantic Domains

SEMANTIC DOMAINS - 5 LEVELS OF NESTING

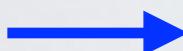
1.6 Animal

- 1.6.1 Types of animals
- 1.6.1.1 Mammal
 - 1.6.1.1.1 Primate
 - 1.6.1.1.2 Carnivore
 - 1.6.1.1.3 Hoofed animals
 - 1.6.1.1.4 Rodent
 - 1.6.1.1.5 Marsupial
 - 1.6.1.1.6 Anteater, aardvark
 - 1.6.1.1.7 Sea mammal
 - 1.6.1.1.8 Bat

1.6.1.2 Bird

- 1.6.1.3 Reptile
- 1.6.1.3.1 Snake
- 1.6.1.3.2 Lizard
- 1.6.1.3.3 Turtle
- 1.6.1.3.4 Crocodile
- 1.6.1.4 Amphibian
- 1.6.1.5 Fish
- 1.6.1.6 Shark, ray
- 1.6.1.7 Insect

- 5 levels of nesting
- very associative grouping
- easy to work with
- based on thesauri and comparative dictionaries
- multilingual tool (in 12 languages - at least partly developed)
- licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License



SEMANTIC DOMAINS

1.6.1.5 Fish

Use this domain for fish (phylum Chordata, class Osteichthyes).

(1) What general words refer to animals that live in the water?

- *fish, water animal, sea creature*

synonyms

(2) What species of freshwater fish are there?

- *trout, minnow, catfish, mudfish, perch*

(3) What species of saltwater fish are there?

- *anchovy, barracuda, bass, blackfish, bonefish, bream, carp, cod, codfish, crappie, darter, flatfish, flounder, gar, garfish, garpike, goldfish, grouper, guppy, halibut, herring, lungfish, mackerel, marlin, pickerel, pike, piranha, pollack, pompano, sailfish, salmon, sardine, sea horse, shad, sole, spearfish, sturgeon, sucker, sunfish, swordfish, tarpon, tuna, turbot, whitefish, whiting, wrasse*

hyponyms

(4) What words refer to the very small animals in the ocean?

- *plankton*

associated food source

(5) What words describe where a fish lives?

- *freshwater, saltwater, pelagic, marine, aquatic*

location

(6) What words describe fish?

- *bony, finned, jawed, jawless, lobe-finned, poisonous, snouted*

feature, property

SEMANTIC DOMAINS

1.1 Sky

Use this domain for words related to the sky.

(1) What words are used to refer to the sky?

- *sky, firmament, canopy, vault*

synonyms

(2) What words refer to the air around the earth?

- *air, atmosphere, airspace, stratosphere, ozone layer*

part-whole

(3) What words are used to refer to the place or area beyond the sky?

- *heaven, space, outer space, ether, void, solar system*

spatially adjacent

(4) What words describe something in the sky or something that happens in the sky?

- *aerial, heavenly, atmospheric, stratospheric, celestial,*

derived modifiers

(5) What words describe the appearance of the sky?

- *blue, clear, black (night), cloudy, cloud filled, brilliant, dark, night sky, foreboding, stormy, dark, star studded*

common states

(6) What words refer to the edge of the sky where it meets the ground?

- *horizon, skyline*

associated location

(7) What words refer to something in the sky?

- *heavenly body, celestial body, luminary*

part-whole

(8) What words refer to the lights that appear in the northern (or southern) sky?

- *aurora borealis, northern lights*

part-whole

(9) What words refer to something being in the sky?

- *aloft, up in the sky, up in the air,*

location

SEMI-AUTOMATIC LINKING OF SEMANTIC DOMAINS AND WORDNET (MUHAMMAD ET AL. 2014)

| LVL | Example | # IDs | ID linked to WN | | ≥ 1 relation | | ≥ 2 relation | | monosemous | |
|-----|---------------|-------|-----------------|-----|-------------------|-----|-------------------|-----|------------|-----|
| | | | eng | ind | eng | ind | eng | ind | eng | ind |
| 1 | 1: universe | 9 | 3 | 4 | 3 | 4 | 1 | 2 | 1 | 1 |
| 2 | 1.1: sky | 68 | 54 | 46 | 27 | 27 | 6 | 13 | 7 | 7 |
| 3 | 1.1.1: sun | 419 | 252 | 237 | 73 | 74 | 16 | 32 | 33 | 29 |
| 4 | 1.1.1.1: moon | 985 | 702 | 605 | 90 | 69 | 8 | 8 | 86 | 65 |

Available under Creative Commons licence from Rapid Words: <http://rapidwords.net/>

SEMI-AUTOMATIC LINKING OF SEMANTIC DOMAINS AND WORDNET (MUHAMMAD ET AL. 2014)

| English | | | Indonesian | | |
|------------|--|----------------|------------|------------|------------------|
| 8.4.1 | 15113229-n | period of time | 8.4.1 | 15115926-n | jangka waktu |
| Hyponym | | | Hyponym | | |
| 8.4.1.2 | 14484516-n | day | 8.4.1.2 | 14484516-n | hari |
| 8.4.1.3 | 15135996-n | week | 8.4.1.3 | 15135996-n | minggu |
| 8.4.1.4 | 15206296-n | month | 8.4.1.4 | 09358226-n | bulan |
| 8.4.1.5 | 00294884-v | season | 8.4.1.5 | 15239292-n | musim |
| 8.4.1.6 | 15201505-n | year | 8.4.1.6 | 15201505-n | tahun |
| 8.4.1.7 | 15248564-n | era | 8.4.1.7 | 15248564-n | zaman |
| not mapped | | | not mapped | | |
| 8.4.1.1 | 08266849-n; 06487395-n; 15173479-n | Calendar | 8.4.1.1 | 15173479-n | Kalender |
| 8.4.1.8 | NIL | Special days | 8.4.1.8 | NIL | Hari-hari khusus |

Table 4: Example of a good 2nd level mapping

SEMI-AUTOMATIC LINKING OF SEMANTIC DOMAINS AND WORDNET (MUHAMMAD ET AL. 2014)

| | English | | Indonesian |
|------------|--|------------|--|
| 8.3.3.3.4: | colors of the spectrum | 8.3.3.3.4: | rentetan warna yang diuraikan oleh cahaya |
| 3: | language and thought | 3: | bahasa dan pikiran |
| 9.4: | semantic constituents related to verbs | 9.4: | konstituen atau unsur semantik yang berkaitan dengan |
| 1.3.5: | solutions of water | 1.3.5: | larutan air |
| 2.3.1.9: | something used to see | 2.3.1.9: | sesuatu yang digunakan untuk melihat |

Table 2: **SD** main words not mapped to **WN**

- Benefits of the mapping: we could import definitions from PWN and Bahasa WordNet (if these can be used in such way).
- The definitions would serve as a model for native definitions and the false definitions can be manually rejected by the native speakers.
- We can also automatically suggest synonyms and these can be again approved or rejected.

OLD DICTIONARY PRINTED ACCORDING TO SD SYSTEM ALONG WITH THE QUESTIONNAIRES



OLD DICTIONARY PRINTED ACCORDING TO SD SYSTEM ALONG WITH THE QUESTIONNAIRES



TEAMS WORK THROUGH ASSIGNED SECTIONS OF THE DICTIONARY ALONG WITH THE SD QUESTIONNAIRE WRITING DOWN NEW WORDS AND CORRECTIONS



WE ARE USING A LOCAL SCHOOL - CHILDREN GOT 3 DAYS OFF, EVERYONE IS HAVING FUN



HOW TO MAKE THE RECORDINGS



EACH WORD COLLECTED IS RECORDED WITH AN EXAMPLE SENTENCE



Nomor Bidang Makna: _____

Nama Bidang Makna: _____

Diisi oleh _____

Penutur Asli (huruf inisial) _____

Awal Tanggal: _____ Jam: _____

Kata atau istilah

- 2.6 A.2.1. 1. Hat hawel
 2. Hlamuna haring
 3. Hea hemama.
 4. Hwang homania, Tiang dei hotihua.
 5. Huda amangra.
 6. Hiang ria.
 7. Hapating.
 8. Hion hyaning bila hahanra.
 9. Hoku matik / mukha matik
 10. Hei halua - halua
 11. He iilia
 12. He hafiring
 13. He haikha.
 14. Hok mihi
 15. Hoban hawaii
 16. Seng abet.
 17. Seng abet mayal manik.
 18. Pasing tamme.
 19. Sebalik hanyal / tangal hanyal.
 20. Dara tung neida naha.
 21.
 2.6.4.3. 22. Ting haleeta.
 * 23. Homi hapa.
 24. Hong feng, ting heng.
 25. He ama dana hapa.
 26. He ama daayadi
 * 27. Homi hapa.
 28. Hongfeng
 29. Hongfeng rofi
 30. Hion hyaning ting haleeta hei.

Terjemahan dalam Bahasa Indonesia

- Membesarkan anak lewat makan minum
Yaitu sayang
Menjadi Ibu, bapa.
Jelap memperhatikan / memedulikan
Mengajar dan mendidik ^{dan mendidik} untuk dimengerti
Relasi seorang untuk diketahui.
Menasihati untuk diingat.
Membetulkan / memperbaiki kesalahan.
Bantai menganduh anak.
Anak yang selalu di manjakan / memanjanya.
Bengakaihan anak / tidak memperhatikan.
Memperlakukan anak dengan kasar.
Bengadopsi / mengangkat anak orang untuk ^{dipilih}
Benjaya / mengancam/mengintimidasi anak.
Seseorang yg dihukuskan untuk menjaga
Remuda
Remuda - pemudi / remaja
Super dan akil balik.
Wira beranak / melahirkan.
Belum culup umur, belum matang, masih mudah
Belum berpengalaman.
Orang dewasa peria dan wanita.
Obasan dewasa / matang / kebaya.
Sebutan orang yang sudah dewasa / adupun
Masih mempunyai hairid.
Hairid sudah mati / tidak ada sind day
Dewasa.
~~Dewasa~~.
Dewasa seutuhnya.
Besar-Besarnya mencapai kedewasaan / seumat

PROGRESS REPORT

- 7 days (20+ participants): 15,000 words

To be done:

- add and/or verify the Indonesian gloss
- add English gloss
- provide pronunciation and example
- record the phonetic form
- link to synonyms and antonyms
- provide additional information about inflectional paradigm, etc.
- record dialectal forms
- print and check the dictionary update
- timeline: next 2-3 years

MULTIWORD EXPRESSIONS EMOTION AND COGNITION PREDICATES

ONOMASIOLOGY AND COMPOSITIONALITY

| English | Italian | Rumanian | Czech | Indonesian | Abui |
|---------|---------|----------|-------|------------|---|
| hand | mano | mina | ruka | tangan | ha-tang |
| arm | braccio | brat | | | |
| foot | piede | picioř | noha | kaki | he-toku |
| leg | gamba | | | | |
| finger | dito | deget | prst | jari | ha-tang paka (lit. 3poss-arm longish.fruit) |
| toe | | | | | he-toku paka (lit. 3poss-leg longish.fruit) |



paka ‘longish fruit, pod, (deriving fingers), dreadlocks, etc

ABUI EMOTION AND COGNITION PREDICATES DATABASE (490 ITEMS)

- Common Template: [Possessor-Body.Part]Experiencer Predicate

| | | |
|---------------|-------------------|----|
| <i>-mi</i> | inside | 96 |
| <i>-ieng</i> | eye | 44 |
| <i>-pong</i> | face | 30 |
| <i>-wa</i> | mouth | 24 |
| <i>-tang</i> | arm, hand | 23 |
| <i>-wei</i> | ear | 9 |
| <i>-weti</i> | tooth | 8 |
| <i>-aaka</i> | physical strength | 7 |
| <i>-kin</i> | gut | 7 |
| <i>-iyek</i> | butt | 6 |
| <i>-aking</i> | breath | 6 |
| <i>-wea</i> | blood | 5 |
| <i>-took</i> | stomach | 4 |

| | | |
|--------------------|------------|-----|
| <i>-nooting</i> | soul | 4 |
| <i>-paai</i> | stamina | 4 |
| <i>-ui</i> | back | 4 |
| <i>-toku</i> | leg | 4 |
| <i>-mool</i> | appearance | 3 |
| <i>-rai -iyeng</i> | chest | 3 |
| <i>-min</i> | nose | 3 |
| <i>-moi</i> | voice | 3 |
| <i>-mata</i> | waist | 2 |
| <i>-pikai</i> | head | 2 |
| <i>-faai</i> | side | 1 |
| no body part | | 173 |

ABUI - GUTS (*ha-kin*) - BODILY CONDITION ~ BEHAVIOR ~ CHARACTER

satiated = big guts

ha-kin foka
3inal-gut big

satiated = basket guts

ha-kin buot
3inal-gut basket

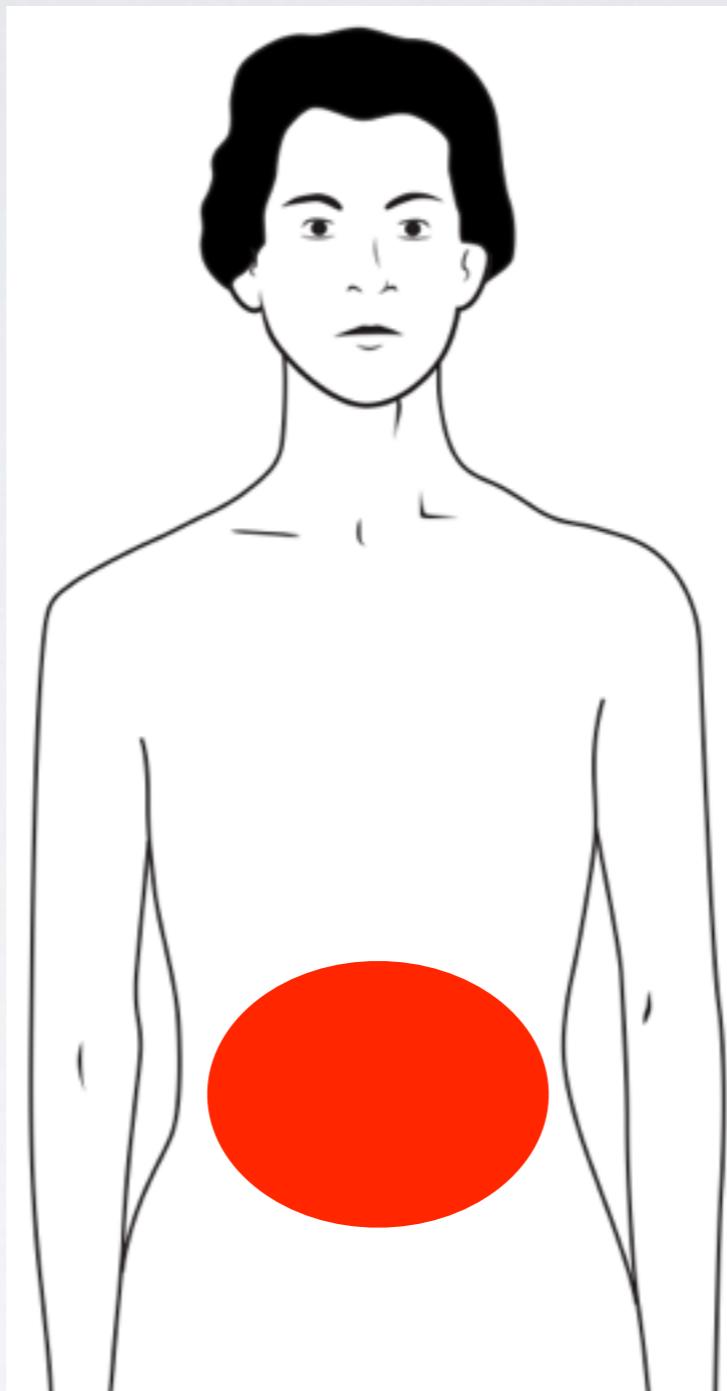
satiated = basket guts

ha-kin pakai
3inal-gut basket

↑
no control
[bodily condition]



lack of self-control
[character]



overeat = guts hurt

ha-kin ha-rik
3inal-gut 3pat-hurt

↑
no control
[bodily process]



[behaviour]

image from Van Staden and Majid 2006:160

ABUI - CHEST (*ha-rai h-iyeng*) - EMOTION ~ BEHAVIOR

nervous = chest moves in shocks

ha-rai h-iyeng ading-adin-ra
3inal-chest 3inal-eye RED[move.in.shock]-REACH.CNT

no control
[emotion]

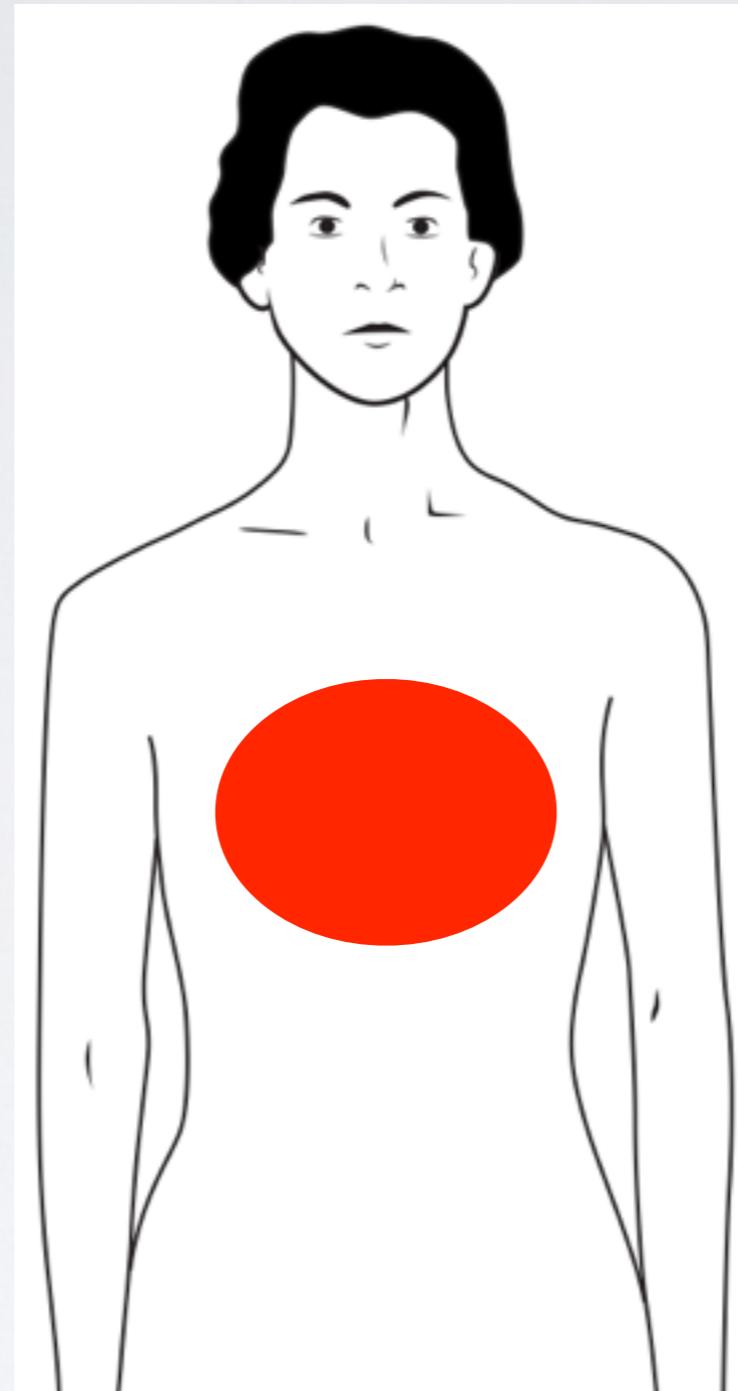


image from Van Staden and Majid 2006:160

boast = broad/exposed chest

da-rai h-iyeng teek
3i.inal-chest 3inal-eye expose

da-rai h-iyeng do-tileesing-dia
3i.inal-chest 3inal-eye 3i.rec-broad-GET.CNT

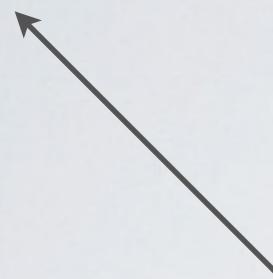
+ control
[behaviour]

ABUI - CHEST (*ha-rai h-iyeng*) - SOCIAL COGNITION - LICENSE AND TARGET

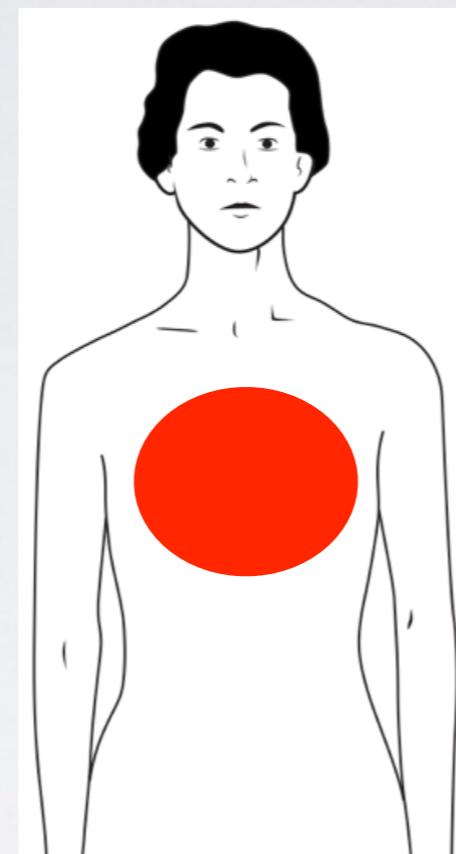
nervous = chest moves in shocks

ha-rai h-iyeng ading-adin-ra

3inal-chest 3inal-eye RED[move.in.shock]-REACH.CNT



no control
[emotion]



boast = broad/exposed chest

da-rai h-iyeng teek

3i.inal-chest 3inal-eye expose

da-rai h-iyeng do-tileesing-dia

3i.inal-chest 3inal-eye 3i.rec-broad-GET.CNT



+ control
[behaviour]

| metaphor | user M<>F | target | | | | alternatives |
|-----------------------|--------------|--------|----------|---------|-----|----------------------------------|
| | | older | same age | younger | M/F | |
| chest moves in shocks | Y | Y | Y | Y | Y | |
| broad chest | N | Y | Y | Y | Y | instead: <i>tanga faring</i> (F) |
| exposed chest | N | Y | Y | N | N | instead: <i>tanga ong</i> (F) |

Goody 1995; Evans 2011; Firth and Firth 2007

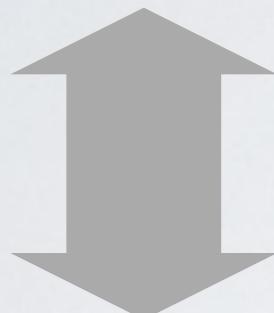
ABUI - CHEST (*ha-rai*) - INTERACTION WITH CULTURE

defenseless = empty chest

ha-rai taka

final-chest empty

↑
no control
[bodily state]

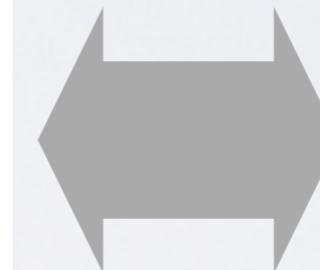
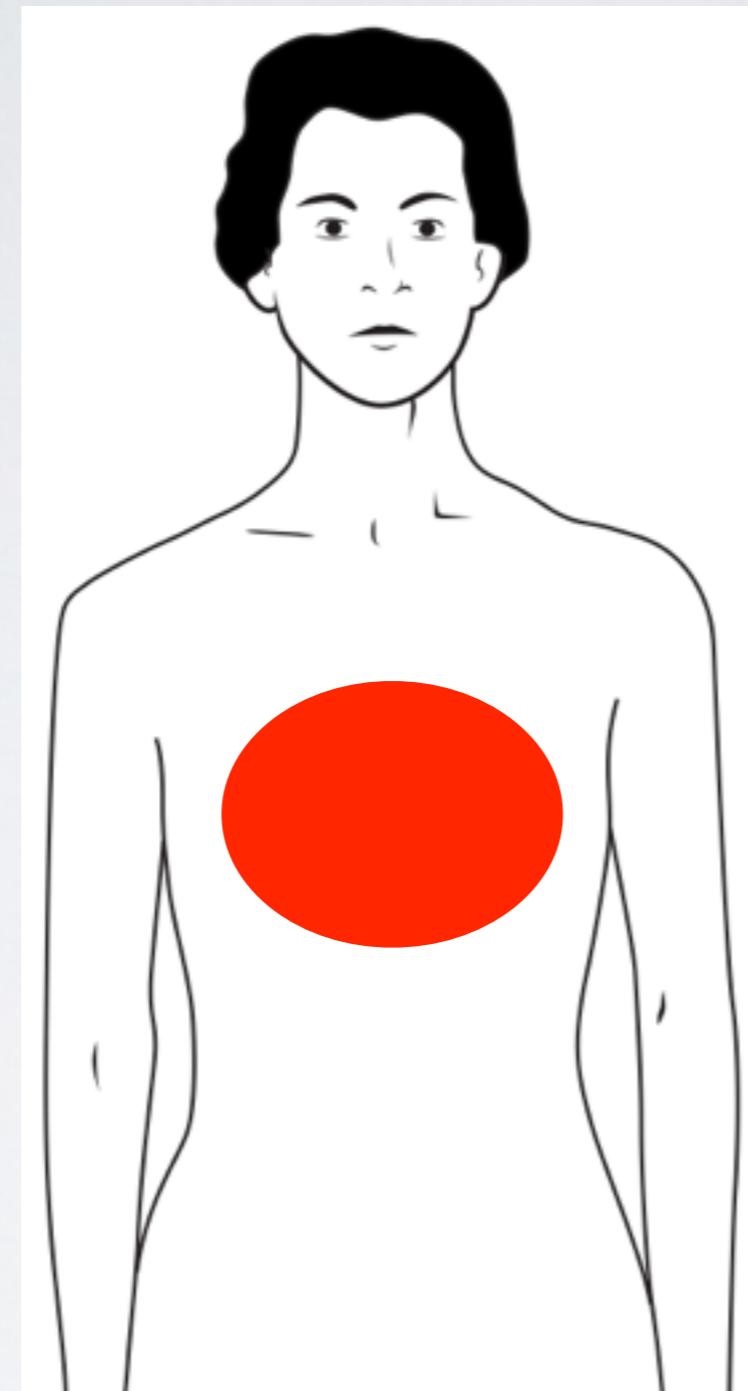


vulnerable = empty chest

ha-rai taka

final-chest empty

[character]



history



image from Van Staden and Majid 2006:160

PREDICATES FOR ANGER (11 multiword vs. 7 simple)

| ID | Abui (3sg) | English | Structure | SD.ID | SD.ENG | Bodypart (E) | Main predicate |
|-----|-------------------|---|-----------------------------|---------|--------|--------------|----------------|
| 127 | wea hapong hemara | amuck, go berserk, rage | wea-PAT-pong LOC-state | 3.4.2.3 | Angry | face + blood | go up |
| 201 | homu halilra | to anger, infuriate, irritate | REC-mi PAT-verb-ra | 3.4.2.3 | Angry | inside | heat up |
| 30 | hewealila | hotheaded, impulsive, hot-tempered | LOC-wea state | 3.4.2.3 | Angry | blood | hot |
| 31 | hewealaliela | hotheaded, impulsive, hot-tempered | LOC-wea temperature | 3.4.2.3 | Angry | blood | rise |
| 33 | heweahemara | be angry, rage | LOC-wea LOC-verb | 3.4.2.3 | Angry | blood | go up |
| 112 | hapongfaakda | disturb, stir up, interrupt | PAT-pong state-da | 3.4.2.3 | Angry | face | break |
| 200 | hominalok-hawaila | provoke, stir anger | REC-mi-ng PAT-verb-PAT-verb | 3.4.2.3 | Angry | inside | poke and turn |
| 202 | homihaluol-harena | resent, hold a grudge | REC-mi-PAT-verb-PAT-verb | 3.4.2.3 | Angry | inside | follow after |
| 244 | hawatafeel | throw tantrums, scream, cry a lot (child) | PAT-wa state | 3.4.2.3 | Angry | mouth | DISTR-hurt |
| 245 | dawatikai | bad-tempered, cranky | PAT-wa state | 3.4.2.3 | Angry | mouth | bite |
| 285 | dawet takai | grit one's teeth, gnash one's teeth | PAT-wet verb | 3.4.2.3 | Angry | tooth | bite |
| 357 | hebekbekda | provoke, insult, ridicule | 0 | 3.4.2.3 | Angry | 0 | damage |
| 456 | liiki beeka | irritable, choleric | 0 | 3.4.2.3 | Angry | 0 | fierce bad |
| 457 | daliikdi | offended | 0 | 3.4.2.3 | Angry | 0 | bent |
| 459 | kafering | angry, snap at | 0 | 3.4.2.3 | Angry | 0 | harsh |
| 460 | kaferingkaang | irritable, choleric | 0 | 3.4.2.3 | Angry | 0 | very harsh |
| 461 | kaferingbeeka | choleric | 0 | 3.4.2.3 | Angry | 0 | badly harsh |
| 466 | tebai | be angry at someone | 0 | 3.4.2.3 | Angry | 0 | hit each other |

ORGANISATION OF DICTIONARIES - PREFIXING LANGUAGES (MEASURING THE DEGREE OF CHANGE ALONG A SCALE)

| verb | PAT | REC | LOC | GOAL | BEN |
|-------------------------------|---|---|-------------------------------------|--------------------------------------|---------------------------------------|
| a. <i>fanga</i> ‘say’ | <i>ha-fanga</i> ‘request him’ | <i>do-fanga</i> ‘ask for him-self’ | <i>he-fanga</i> ‘say it’ | <i>noo-fanga</i> ‘scold me’ | <i>nee-fanga</i> ‘say for me’ |
| b. <i>liya</i> ‘fly’ | <i>ha-liya</i> ‘shoot it’ | <i>do-liya</i> ‘fly on his own’ | <i>he-liya</i> ‘fly on it’ | <i>noo-liya</i> ‘fly to me’ | <i>nee-liya</i> ‘fly for me’ |
| c. <i>faaling</i> ‘listen’ | * <i>ha-faaling</i> | <i>do-faaling</i> ‘listen for himself’ | <i>he-faaling</i> ‘listen to it’ | <i>noo-faaling</i> ‘listen to me’ | <i>nee-faaling</i> ‘listen for me’ |
| d. <i>wik</i> ‘carry’ | <i>ha-wik</i> ‘carry him’ (child) | <i>no-wik</i> ‘carry for myself’ | <i>he-wik</i> ‘carry it’ | <i>noo-wik</i> ‘let me carry’ | <i>nee-wik</i> ‘carry for me’ |
| e. <i>rumai</i> ‘strong’ | <i>ha-rumai</i> ‘strengthen it’ | <i>no-rumai</i> ‘I feel strong’ | <i>he-rumai</i> ‘it is strong’ | <i>noo-rumai</i> ‘rely on me’ | <i>nee-rumai</i> ‘strong for me’ |
| f. <i>fahak</i> ‘embrace’ | * <i>ha-fahak</i> | * <i>ho-fahak</i> ‘embrace it’ | <i>he-fahak</i> ‘embrace it’ | <i>noo-fahak</i> ‘hug me’ | <i>nee-fahak</i> ‘hug for me’ |

SUMMARY

- Extending resources built for Malay (Indonesian & Malaysian) for other languages spoken in Indonesia (and Malaysia)
- The number of individual languages listed for Indonesia is 719 (10% of the world's languages). Of these, 706 are living and 13 are extinct. Of the living languages, 19 are institutional, 86 are developing, 261 are vigorous, **265 are in trouble, and 75 are dying.**
- There is a wealth of Dutch produced lexical resources (from 1850s until 1950s) of great detail (large dictionaries), which are becoming obsolete, because the community and most linguists are not fluent in Dutch. Multilingual WordNet allows to give these sources a second life.
- Linking to Comparative Databases and other WordNets (tracking borrowing and semantic shift)
- Semantic typology and onomasiology (stratification of a lexical field) (Koptjevskaja-Tamm, WALS) (need for ontology and well defined relations to calculate semantic distance)
- The Domain Ontology is essential for rapid development of lexical resources. They need to be user-friendly and suitable for crowdsourcing for communities with developing literacy

REFERENCES

Kratochvíl, F. and Delpada, B., Abui - Indonesian - English Dictionary. UBB Kupang. 2008.

Muhammad Zulhelmy bin Mohd Rosman, Bond, F., and Kratochvíl, F., “Bringing together over- and under-represented languages: Linking Wordnet to the SIL Semantic Domains,” In Orav, H., Fellbaum, C., and Vossen, P. (eds.), *Proceedings of the Seventh Global WordNet Conference*, pp. 40-48. Tartu, Estonia, January 25-29, 2014. <http://gwc2014.ut.ee/>