Issues in the creation of a trilingual dictionary for Sasak, eastern Indonesia

Peter K. Austin

Endangered Languages Academic Programme
SOAS, University of London

ICLDC
Lexicography Workshop
February 2011
A new Sasak dictionary

Overall goals:

- to create a trilingual Sasak-Indonesian-English dictionary that will be of practical use to speakers on Lombok

- in doing so, to develop ways to deal with the complex individual, geographical and social variation in the language in a politically acceptable manner
This talk

- Sasak language and people
- Language variation
- Previous dictionaries
- Our current approach
Location
History

- 14th century — Majapahit Javanese kingdom
- 16th century — Islamicisation of eastern Lombok
- 17th century — Karangasem Balinese kingdom
- 19th century — Dutch colonial war with Balinese
- 20th century — 1942-45 Japanese war

1948 Republic of Indonesia
Population and Languages

- Sasak — 90% of Lombok population (2.5m.)
- Sasak spoken by majority of population across the island in family and village domains but has no formal status, no literacy functions for most speakers
- Bahasa Indonesia is language of education, media, government, business, literacy, status (diglossia)
- Rural-urban split in language shift
Languages and Dialects

- Sasak
- Bahasa Indonesia

- Kawi (early Mod. Javanese) exists as literary language used in texts, poetry, music and drama

- Sasak shows wide regional variation which has been documented since 19th century, including through dialect surveys of Teeuw 1948 and Mahsun 2000.

- Sasak has social variation, distinguishing base kasar from base alus (latter heavily influenced by Javanese and Balinese) marked by lexical choice
Six challenges

- Challenge 1 – dialects
- Challenge 2 – phonological variation
- Challenge 3 – semantic differences
- Challenge 4 – speech levels
- Challenge 5 – contact
- Challenge 6 – sociology and politics
Challenge 1: Dialects

- Sasak speakers recognise five ethno-linguistically named varieties (based on shibboleth words for ‘like that-like this’) and give rough geographical locations for them:
  - *Ngenó-Ngené* — central west coast and central east to north east coast;
  - *Menó-Mené* — around the towns of Puyung and Praya, central Lombok;
  - *Nggetó-Nggeté* — around Suralaga and Sembalun in the north-east;
Challenge 1: Dialects

- **Kutó-Kuté** — around Bayan in the north (also called *Ngenó-Mené*);

- **Meriaq-Meriku** — south central area around Bonjeruk and Sengkol (Pujut), and near Selong in east

This does not exhaust the actual forms found on the ground:

- **Menu-Meni** — south-east near Ganti)
Challenge 1: Dialects

Speakers characterise the dialect differences by pointing to lexical divergence and differences in grammatical forms, eg. clitic pronouns, aspect/mood markers, prepositions:

- **gin** ~ **gen** ~ **éaq** ~ **iaq** ~ **jaq** ~ **aq** ‘will’
- **òjòk** ~ **jòk** ~ **aning** ~ **timpaq** ~ **ngaró** ‘to’
- **=mèq** ~ **=bi** ~ **=de** ~ **=m** ~ **=ò** ‘you (clitic)’

However, dialect surveys by Teeuw in 1948 and Mahsun in 2000 at the village level show that there are no clear geographical distributions to allow isoglosses to be drawn.
Peta 58 'mendung'
Challenge 2: Phonological variation

- Previous studies and my research shows that there is phonological variation within and across ‘dialects’ even where there are no gross lexical differences:

Vowels

variation between i and é, u and ó, ó and ò, eg. irung ~ érung ~ iróng ~ éróng ‘nose’; uléq ~ óléq ‘to return’; tókól ~ tòkòl ‘to sit’

variation between a ~ e ~ ã, eg. mata ~ mate ~ mete ~ mạtã ‘eye’; aïq ~ eiq ‘water’
Challenge 2: Phonological variation

Consonants

variation between final -r and final -h, and final glottal stop -q and final -h, eg. *dengar ~ dengah ‘to hear’; *kókóq ~ kókóh ‘river’

variation between medial -r- and nothing (-Ø), and medial -r- and -d-, eg. *daraq ~ daq ‘blood’, *irup ~ érup ~ idup ~ édup ‘alive’, *irung ~ érung ~ iróng ~ éróng ~ iung ‘nose’ (but not *idung)
Challenge 3: Semantic differences

A given form may show differences in referential or predicational meaning depending on location. This is most extreme for deictics:

Eg. across sources for Sasak we find:

‘south’ baret ~ bat ~ daya ~ daye ~ lauq ~ bòngkòt

‘west’ baret ~ bat ~ daye

timuq ‘east’, ‘north’, ‘south’
Challenge 3: Semantic differences

But there are also subtle semantic differences in other lexical fields that we have only just begun to explore
Challenge 4: Speech Levels

Politeness is encoded lexically in basic vocabulary, both open class and closed class items, through suppletion in sets of *biase*, *madie*, *alus* items plus a system of honorifics:

- **biase** — low  
  eg. *mate* ‘eye’

- **madie** — middle  
  penenteng

- **alus** — high  
  penyerminan

Honorifics: honour and humble  
eg. *atur* ‘to give’

*ican* 'to receive'
<table>
<thead>
<tr>
<th></th>
<th>Low</th>
<th>Mid</th>
<th>High</th>
<th>Honorific</th>
<th>Humble</th>
</tr>
</thead>
<tbody>
<tr>
<td>eat</td>
<td>mangan</td>
<td>kelór</td>
<td>madaran</td>
<td>majengan</td>
<td></td>
</tr>
<tr>
<td>see</td>
<td>gitaq</td>
<td>seréóq</td>
<td>sermin</td>
<td>cingakin</td>
<td></td>
</tr>
<tr>
<td>say</td>
<td>uni</td>
<td>base</td>
<td>manik</td>
<td></td>
<td>matur</td>
</tr>
<tr>
<td>go</td>
<td>laló</td>
<td>lumbar</td>
<td>margi</td>
<td>lampaq</td>
<td></td>
</tr>
<tr>
<td>give</td>
<td>bèng</td>
<td></td>
<td>ican</td>
<td>atur</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Mid</td>
<td>High</td>
<td>Honorific</td>
<td>Humble</td>
</tr>
<tr>
<td>--------</td>
<td>-------</td>
<td>-------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>body</td>
<td>awak</td>
<td>perane</td>
<td>dèwèk</td>
<td>batang</td>
<td>rage</td>
</tr>
<tr>
<td>head</td>
<td>òtak</td>
<td>sirah</td>
<td>tendes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eye</td>
<td>mate</td>
<td>penenteng</td>
<td>penyerminan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hand</td>
<td>ime</td>
<td></td>
<td></td>
<td>gading</td>
<td></td>
</tr>
<tr>
<td>what</td>
<td>ape</td>
<td></td>
<td></td>
<td>napi</td>
<td></td>
</tr>
</tbody>
</table>
Challenge 5: Contact – Kawi

Kawi is the literary language but lexical items are still used for hyper-politeness, eg.

*nurge* (alus) > *ampure* (kawi) ‘apology’
*maté* (biase) > *ninggal* (alus) > *mangket* (kawi) ‘to die’
*awak* (biase) > *rage* (alus) > *pragayan* (kawi) ‘body’

Any account of modern Sasak has to include Kawi lexicon
Challenge 5: Contact – Arabic

Sasak is full of vocabulary adopted from Arabic since the 13th century, much of it in the religious domain but including common lexicon:

- people’s names
- days of the week
- In Ngeno-ngene the second person masculine pronoun *ante* is a loan from Arabic

All this material must be included in the dictionary.
Challenge 6: Sociology and politics

- Traditional Sasek social castes on Lombok:
  - *mènak* 1st caste (the highest), cf. *radèn*
  - *prewangse* 2nd caste
  - *jajarkarang* 3rd caste - commoners
  - *sepangan* lowest caste, servants of *mènak*

- Dominant religious role for Islam

- Contemporary society built on modernism and ethos of education and attainment, but *adat* remains a strong influence and *mènak* have important role
Previous Sasak dictionaries

Goris 1938 – Ngeno-ngene with dialect and other references (see sample)

Nazir Thoir – only Nggeto-nggete data

Nell Staff – the midden approach (see sample)

SIL dictionary – only Ampenan data, no speech levels data
Our approach

• there is no ‘standard’ Sasak and selecting one variety would lead to political problems, cf. Goris
• users want to include as much variation as possible with lots of cross-referencing of several types (social, geographical, stylistic, loans)
• provide metadata about where lexical items come from – in database identity of speaker and recorder is included for every lexical entry and sentence (can be aggregated to village level as desired)
• in Toolbox files metadata for speakers and recorders is also included with every sentence
Example from database: ‘nose’

Example index of text sentences by speaker
Future developments

- mobile devices introduce new platforms for potential dictionary development, e.g. Wunderkammer
- GPS-aware smart phones and tablet computers exist or are on the horizon
- since the lexicon is comprehensively tagged for sources, by adding GPS co-ordinates to the database we could create apps that generate individual dictionaries for device locations on the fly
- could have applications in education and teaching of Sasak in schools, cf. Azhar 2000
References


Thank you