

EXPLORING AND DISSEMINATING FIELD DATA USING HYPERLEX

Steven Bird

University of Pennsylvania

<http://www.ldc.upenn.edu/sb>

SOME BACKGROUND

1. starved, or drowned?
2. online linguistic datasets
3. creation, management, dissemination, citation
4. native speakers, field linguists, university researchers

HYPERLEX

- `www.ldc.upenn.edu/hyperlex`
- a Perl/CGI script – portable multimedia GUI
- lexicons and paradigms: text, audio, images
- visualizations of data
- basis for published linguistic research
- remote access (e.g. Connell's aspiration study)
- Yoruba, Manding, Tamil, Nahuatl, English

APPLICATION TO TAMIL

x trans	remove எடுத்து கொண்டு போ to move from a position occupied
1 intr	roll உருள் <i>uru</i> revolve and move over a flat surface, as a ball, a wheel, etc. பந்து ரோட்டில் உறுண்டு போனைதப் பார்த்தேன் <i>pandu roTTie uruNTu poonade paaTeen</i> I saw the ball roll down the street
3 intr	raise உயர்த்து <i>oyattu</i> , <i>oyarttu</i> move or cause to move upward, or to a higher position பலகையை கொஞ்சம் உயர்த்தி பிடித்துக் கொடேன் <i>palahaye konjam oyattippiTiccu kuTeen</i> Could you raise this plank for me? நிமிர்த்து <i>nimirrttu</i>
3 tr	rub தடவு <i>taTavu</i> press and move back and forth, as the hands on s.o.'s limbs முதுகைத் தடவிக்கொடுத்தால் குழந்தை தூங்கி விடும் <i>mudhat taTavikkuTutta kozhande tuunguTum</i> If you rub his back, the child will fall asleep rush அவசரப்படுத்து <i>avasarpapaTuttu</i> cause to move or act with vigorous haste or violence சினிமாவுக்கு வேற்றாகுது என்று என்னை தயாராக அவசரப்படுத்தினான் <i>sinimaaku leeTtaaraduNNu enne tayaaraha avasarpapaTuttinaan</i> He rushed me to get ready as it was getting late for the movie

LINGUISTIC RESOURCES: CREATED BY AND *For* MOTHER TONGUE SPEAKERS



An editing session with the Yemba 'academie'



Presentation to a chief

APPLICATIONS TO ORTHOGRAPHY

Description	Example Phrase
shallow tone marking	ḡmò' mènzwí pò èwàḡkhà pí métét é lé ḡjé'é ḡḡōḡ tē'
zero tone marking	ḡmò' menzwí pò ewàḡkhà pí metet e le ḡjé'e ḡḡōḡ tē'
omit most frequent tone	ḡmò' menzwí pò ewàḡkhà pí métét é lé ḡjé'é ḡḡōḡ tē'
mark most stable tone	ḡmò' mènzwí pò èwàḡkhà pí metet e le ḡjé'e ḡḡōḡ tē'
mark tone changes	ḡmò' menzwí pò èwàḡkhà pí metet e le ḡjé'e ḡḡōḡ tē'
mark first tone only	ḡmò' mènzwí pò èwàḡkhà pí métét é lé ḡjé'e ḡḡōḡ tē'
conflate mid and high	ḡmò' menzwí pò ewàḡkhà pí métét é lé ḡjé'é ḡḡōḡ tē'
mark minimal pairs	ḡmò' mènzwí pò ewàḡkhà pí metet é le ḡjé'e ḡḡōḡ tē'
mark isolation tone	ḡmò' menzwí pò ewàḡkhà pí métét e lé ḡjé'é ḡḡōḡ tē'
mark underlying tone	ḡmò' menzwí pò ewàḡkhà pí metét e le ḡjé'e ḡḡōḡ tē'

EXPLORATION USING HYPERLEX

Zeroth order

Retrieve a set of entries according to a search expression

First order

Select entries by one criteria and display them according to another (cf data transformation using SQL)

Second order

Select entries depending on the existence of similar entries elsewhere in the database.

WEAKNESSES OF HYPERLEX

- complexity of query form
- knowledge of data structure is required
- fixed schema, limited browsing, no updates
- gifs for special characters
- character strings vs words
- multiple matches are not possible
- logic of the query form
- data is hidden behind a CGI interface

FROM FIELD TO WEB TO PUBLICATION, AND BACK AGAIN

1. Verb paradigms by hand (Fig 1)
2. Online form - ASCII version (Fig 2)
3. Online form - browser version (Fig 3)
4. Published form - a complex tabulation (Fig 4)
5. Web version - parallels published form (Fig 5)

CONCLUSIONS

- coupled software development and fieldwork
- obtain plenty of data without drowning in it
- dictionary; improved orthography
- expressive interface, but needs to be improved
- integrates well with pencil- & paper-based research
- queries can be embedded in documents – preserving the relationship between a tabulation and the underlying database
- research is accountable and reproducible