An integrated approach to online dictionary and ontology building for Austronesian Languages in Taiwan

D. Victoria Rau, Wheaton College, U.S.A.
Meng-Chien Yang, Providence University, Taiwan
Hui-Huan Ann Chang, Providence University, Taiwan
Maa-Neu Dong, National Museum of Natural Sciences, Taiwan
Outline

1. Introduction
2. Trinitarian Model
3. Online Dictionaries
4. Participatory wiki dictionary
5. Yami Fish Ontology
6. Conclusion
CIP in Taiwan

- A four-year dictionary project
- Construction of indigenous knowledge for material development
Indigenous language revitalization

• Yami corpora
  Yami language archive
  http://yamiproject.cs.pu.edu.tw/yami
  Yami e-Learning
  http://yamiproject.cs.pu.edu.tw/elearn

• ➔ A “trinitarian” model
The Trinitarian Model

- Language activists
- Linguists
- Computer scientists
Online dictionaries

• Three versions of Yami online dictionaries

1. Digital Archiving Yami Language Documentation (funded by SOAS)
   http://yamiproject.cs.pu.edu.tw/yami/database.htm

2. Yami e-Learning (funded by SOAS)
   http://yamiproject.cs.pu.edu.tw/elearn/search.php

3. Yami Online Dictionary (funded by CIP)
   (1) Lexique Pro software version
   (2) The participatory wiki dictionary
   http://yamibow.cs.pu.edu.tw
Digital Archiving Yami Language Documentation

- Keyword search from the texts gathered for digital archiving Yami language documentation
Yami e-Learning

- A concise online Yami-Chinese-English dictionary
The Lexique Pro software version

- 1786 lexical entries
- 780 roots
- 1006 derivatives
The Lexique Pro software version

• An English index
The Lexique Pro software version

- An index organized by Chinese pinyin spelling
The Lexique Pro software version

- An index organized by semantic categories
The Participatory Wiki dictionary

- The structure of Web 2.0 style version dictionary
The Participatory Wiki dictionary
The Participatory Wiki dictionary

![Image of the Participatory Wiki dictionary interface]
The Participatory Wiki dictionary

<table>
<thead>
<tr>
<th>Data type</th>
<th>Value</th>
<th>Additional Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry</td>
<td>baki</td>
<td></td>
</tr>
<tr>
<td>Upload sound of Entry</td>
<td>Current file</td>
<td>Browse</td>
</tr>
<tr>
<td>Upload graph of Entry</td>
<td>Current file</td>
<td>Browse</td>
</tr>
<tr>
<td>Root</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chinese definition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example</td>
<td>baki o tama tama mu.</td>
<td></td>
</tr>
<tr>
<td>Upload sound of Example</td>
<td>Current file</td>
<td>Browse</td>
</tr>
<tr>
<td>Chinese Definitions and Examples</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Participatory Wiki dictionary
The Participatory Wiki dictionary

<table>
<thead>
<tr>
<th>User's Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Name</strong></td>
</tr>
<tr>
<td><strong>E-mail address</strong></td>
</tr>
<tr>
<td><strong>Username</strong></td>
</tr>
<tr>
<td><strong>Password</strong></td>
</tr>
<tr>
<td><strong>Language/Dialect</strong></td>
</tr>
<tr>
<td><strong>Tribe / Location</strong></td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
</tr>
<tr>
<td><strong>Language Use</strong></td>
</tr>
</tbody>
</table>

*Items marked with an * are required unless stated otherwise*
The search page of the web 2.0 style dictionary-1
The search page of the web 2.0 style Yami (Tao) Dictionary project.
Integration of online dictionary and ontology

• The nuts and bolts of developing an ontology of Yami fish names
Yami Fish Ontology

- Three perspectives on classification of Yami fish names
  - Chinese
  - Japanese
  - Yami

- Protégé
  - [http://protege.stanford.edu/](http://protege.stanford.edu/)
Yami Fish Ontology

- 109 Yami fish with Chinese, English, and Latin name

- Toolbox ➔ Lexique Pro ➔ Protégé

- “Ontology 101 development process” by Noy and McGuinness (2001)
Yami Fish Ontology

• Goals
  – Finding the perspective and semantics of Yami fish names
  – Reinterpreting the fish classification of Yami
  – Constructing the indigenous knowledge of fish
Methodology

• 7 steps of constructing the ontology from “Ontology 101 development process” by Noy and McGuinness (2001)

• 7 steps:
  1. Determine the domain and scope of the ontology
  2. Consider reusing existing ontology
  3. Enumerate important terms in the ontology
  4. Define classes and the class hierarchy
  5. Define the properties of classes
  6. Define the facets of the slots
  13. Create instances
(1) Determine the domain and scope of the ontology

1. Which fish are edible and inedible for Yami people?

2. Which gender can eat what kind of fish?

3. What kind of fish can be eaten by Yami elderly males?

4. What kind of fish can Yami pregnant women eat?
(2) Consider reusing existing ontologies

- The Fish Database of Taiwan (http://fishdb.sinica.edu.tw/)
(3) Enumerate important terms in the ontology

(a) classification of Yami fish:
- inedible fish: *anito*
- edible fish
  1. *raet* “bad fish”
  2. *oyod* “good fish”
  3. *kakanen no rarakeh* “fish for old men”

(b) named Yami fish:
- *ilek* “rudderfish”
- *cilat* “jackfish”

(c) Yami people:
  old men, men, and women

(d) women of three stages:
  not pregnant, pregnant, and breast feeding
(4) Define classes and the class hierarchy

<table>
<thead>
<tr>
<th>Yami fish</th>
<th>oyod</th>
<th>raet</th>
<th>kakanen</th>
<th>among no anito</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yami people</td>
<td>Good fish</td>
<td>Bad fish</td>
<td>no-rarakeh Fish for old men</td>
<td>Inedible fish</td>
</tr>
<tr>
<td>Men</td>
<td>(+) (can eat)</td>
<td>(+)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Women</td>
<td>(+)</td>
<td>(cannot eat)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Old men</td>
<td>(+)</td>
<td>(+)</td>
<td>(+)</td>
<td>–</td>
</tr>
</tbody>
</table>
(5) & (6) Define the properties of classes and slots and define the facets of the slots

The datatype properties

- CommonNameInChinese_中文俗名
- CommonNameInEnglish_英文俗名
- Family_科名
- ScientificName_英文學名
- ScientificNameInChinese_中文學名
- Synonym_同義詞
- TaoLiteralMeaning_字面意義
- TaoName_雅美名

The object properties

- inedible <-> edible
  - cannot_eat <-> cannot_be_eaten_by
  - cannot_be_eaten_by <-> cannot_eat
- edible <-> inedible
  - can_be_eaten_by <-> can_eat
  - can_eat <-> can_be_eaten_by
- has_a <-> is_a
- is_a <-> has_a
(7) Create instances (individuals)

- An Example of a class editor: *paloy* “big eye emperor”
Paloy “big eye emperor”

- OntoViz display for *paloy* fish
Yami people and fish

- An initial ontology of Yami fish
Yami fish ontology

- Hierarchy of Yami fish ontology
Good fish, bad fish, fish for old men only, and inedible fish

- OWLViz Displaying the Inferred Hierarchy
Reclassification

- OWLViz Displaying the Asserted Hierarchy
Inedible fish and fish for old men only

- OWLViz Display for anito_Class and kakanen_no_rarakeh_Class
OWLViz Display for rahet_Class and oyod_Class
Ontology browser window generated by Protégé

Indigenous Language Dictionary Project

Yami (Tao) Dictionary Project

Contents

Ontology
All Resources
All Classes
All Object Properties
All Data Properties
All Annotation Properties
Individuals

Ontology

Fish in Yami are divided into two categories: ooyd fish (for women to eat) and ræhet fish (for men to eat). The ræhet fish have a stronger fishy taste than the ooyd fish. According to a Yami legend, some women vomited after eating the ræhet fish, so they no longer eat it. As a result, this type of fish is reserved for men. Men can eat the fish meant for women, but women cannot eat men's fish. Some fish have a very strong fishy taste and this is only for the elderly to eat because they believe catching fish is hard work. It would be a waste to throw it away for its bad smell. Nowadays young people also eat this type of fish while drinking alcohol.
The OWL document generated for the Yami fish ontology

Limitations

• Total dependence on previous classifications of a semantic domain.
• No independent approach to test previous hypotheses
• Call for a text analysis
Future research

(1) selection of a text on a semantic domain

(2) reconstruction of the IK by building a network relationship of the semantic domain independently by both the linguist and the language activist to achieve high reliability

(3) transformation of the final diagram of the network relationship into the Protégé.
Text analysis of taro planting

- Analysis by Protégé of a text about taro planting

Diagram:
- **Thing**
  - **Classifying_taro_sprouts**
  - **Results_of_taro**
    - **Bad_sprouts**
      - **Reasons_to_produce_bad_sprouts**
        - soil_is_too_deep_in_the_field
        - soil_is_too_soft_in_the_field
      - **Ways_to_improve**
        - Planted_in_hard_soil
        - Planted_near_river
      - **Ways_to_treat_bad_sprouts**
        - In_the_past
        - Nowadays
    - **Good_sprouts**
Summary

• The trinitarian model of developing three Yami online dictionaries

• A wiki dictionary

• An ontology of Yami fish names, with the goal of building a collective knowledge system for the Yami language
Ongoing project

1. An online encyclopedia in Yami
2. The semantic infrastructure of the Yami language
Contributions

- Model of collaboration
- Use of technology
- Language revitalization and capacity building
Ayoy!
Thank you!

http://yamibow.cs.pu.edu.tw