Wordlists as a test case for lexical interoperation

Jeff Good
University at Buffalo
jcgood@buffalo.edu
Overview

- Introduce word lists as a data structure
- Show how word lists can inform the development of
  - Annotation schemes for semantic fields
  - A linked data “cloud” for linguistics
Word lists

- Superficially word lists appear to be a very simple data structure

The few works from Mbuk (see p.33) can follow for comparison. The Mbuk say they were formerly at Kiyaki, a few miles away near Su-Bum that Mbuk families once passed it. Not likely (5-10 mi).
Superficially word lists appear to be a very simple data structure. They can be represented as a pair with labels for each half of the pair.

<table>
<thead>
<tr>
<th>Bunaki</th>
<th>kosan</th>
<th>Mbuik</th>
</tr>
</thead>
<tbody>
<tr>
<td>ītál</td>
<td>beliya</td>
<td>etali</td>
</tr>
<tr>
<td>inā</td>
<td>bínā</td>
<td>inā</td>
</tr>
<tr>
<td>itiy</td>
<td>bstwèn</td>
<td>ottè</td>
</tr>
<tr>
<td>isiy</td>
<td>bifuś</td>
<td>0180</td>
</tr>
<tr>
<td>fùmadʒāŋ</td>
<td>bùnētyá</td>
<td>0nainuta</td>
</tr>
<tr>
<td>dzāŋ</td>
<td>džàŋ</td>
<td>o nyan</td>
</tr>
<tr>
<td>fùmadʒofo</td>
<td>bingatūn</td>
<td>o gbüga</td>
</tr>
<tr>
<td>dʒōfis</td>
<td>dʒōfà</td>
<td>ndzoni</td>
</tr>
<tr>
<td>dʒōfis ntsülü biy</td>
<td>mbàŋfí</td>
<td>bān fiá</td>
</tr>
<tr>
<td>mbàŋnà</td>
<td>bān busí</td>
<td></td>
</tr>
<tr>
<td>gbwi</td>
<td>bān won</td>
<td></td>
</tr>
</tbody>
</table>

The few words from Mbuik (see p. 33) can follow for comparison. The Mbuik say they were formerly at Kiyaki, a few miles away near Su Bum, that the Mu buik have been there for a long time. It is not likely (from whatever source) that they were ever there.

Chilver and Kaberry (1974:40)

<table>
<thead>
<tr>
<th>BUNA</th>
<th>KOSI</th>
<th>MBAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>búnči</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**GLOSS**

<table>
<thead>
<tr>
<th>WORD</th>
<th>GLOSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>fùmadʒāŋ</td>
<td>KOSI</td>
</tr>
<tr>
<td>dzāŋ</td>
<td>Mbuik</td>
</tr>
<tr>
<td>fùmadʒofo</td>
<td>Mbuik</td>
</tr>
<tr>
<td>dʒōfis</td>
<td>Mbuik</td>
</tr>
<tr>
<td>dʒōfis ntsülü biy</td>
<td>Mbuik</td>
</tr>
<tr>
<td>mbàŋfí</td>
<td>Mbuik</td>
</tr>
<tr>
<td>mbàŋnà</td>
<td>Mbuik</td>
</tr>
<tr>
<td>gbwi</td>
<td>Mbuik</td>
</tr>
</tbody>
</table>
Word lists

• Superficially word lists appear to be a very simple data structure
• They can be represented as a pair
• With labels for each half of the pair

<table>
<thead>
<tr>
<th>GLOSS</th>
<th>WORD</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>fûmadʒûṇṇ</td>
</tr>
</tbody>
</table>
Word list structures

- Structures more complex than a simple pair are also attested
Structures more complex than a simple pair are also attested.
Key features

- Two important questions
  - What is a gloss?
  - What is the nature of the relationship between a gloss and the target word?
Key features

- Two important questions

  - What is a gloss?
  - What is the nature of the relationship between a gloss and the target word?

- Dictionary: Form → Meaning

\[
\text{arbor}
\]
Key features

- Two important questions
  - What is a gloss?
  - What is the nature of the relationship between a gloss and the target word?

- Dictionary: Form $\rightarrow$ Meaning
- Word list: $\rightarrow$ Form
Key features

- Two important questions
  - *What is a gloss?*
  - *What is the nature of the relationship between a gloss and the target word?*

- Dictionary:  **Form** → **Meaning**
- Word list:  **Meaning?** → **Form**
• Two important questions

• What is a gloss?

• What is the nature of the relationship between a gloss and the target word?

• Dictionary: Form → Meaning

• Word list: → Form
Two important questions

What is a gloss?

What is the nature of the relationship between a gloss and the target word?

- Dictionary: Form → Meaning
- Word list: Concept → Form
Two important questions

- What is a gloss?
- What is the nature of the relationship between a gloss and the target word?

Dictionary: Form $\rightarrow$ Meaning

Word list: Concept $\rightarrow$ Form

Three pieces: Concept, Mapping, Form
Two important questions

What is a gloss?

What is the nature of the relationship between a gloss and the target word?

Dictionary: Form $\rightarrow$ Meaning

Word list: Concept $\rightarrow$ Form

Three pieces: Concept, Mapping, Form
Concepticons

- Concept lists are curated objects in their own right, which we term concepticons
- The Swadesh list
- Intercontinental Dictionary Series
- Comparable to an interlingua
- Possibly associated with a simple taxonomy
- Each concept minimally described via a label drawn from a major language
Meaning 1.21: the land

Description: 'the hard surface of the earth, when compared to the area covered by sea'

Typical context: The captain sighted land in the distance.

Semantic field: The physical world

Semantic category: Noun

Borrowed score: 0.27
Age score: 0.90
Simplicity score: 0.94

Counterpart words in the World Loanword Database

<table>
<thead>
<tr>
<th>Voc. ID</th>
<th>Vocabulary</th>
<th>Word Form</th>
<th>Original Script</th>
<th>Borrowed status</th>
<th>Borrowed score</th>
<th>Age score</th>
<th>Simplicity score</th>
<th>Source Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Swahili</td>
<td>nchi kavu</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.50</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Iraqw</td>
<td>hhapee</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Gawwada</td>
<td>piye</td>
<td></td>
<td>4. very little evidence for borrowing</td>
<td>0.25</td>
<td>1.00</td>
<td>1.00</td>
<td>biyya (Oromo)</td>
</tr>
<tr>
<td>4</td>
<td>Hausa</td>
<td>kásáa</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
Meaning 1.21: the land

Description: 'the hard surface of the earth, when compared to the area covered by sea'
Typical context: The captain sighted land in the distance.
Semantic field: The physical world
Semantic category: Noun
Borrowed score: 0.27
Age score: 0.90
Simplicity score: 0.94

Counterpart words in the World Loanword Database

<table>
<thead>
<tr>
<th>Voc. ID</th>
<th>Vocabulary</th>
<th>Word Form</th>
<th>Original Script</th>
<th>Borrowed status</th>
<th>Borrowed score</th>
<th>Age score</th>
<th>Simplicity score</th>
<th>Source Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Swahili</td>
<td>nchi kavu</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.50</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Iraqw</td>
<td>hhapee</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Gawwada</td>
<td>piye</td>
<td></td>
<td>4. very little evidence for borrowing</td>
<td>0.25</td>
<td>1.00</td>
<td>1.00</td>
<td>biyy (Oromo)</td>
</tr>
<tr>
<td>4</td>
<td>Hausa</td>
<td>kasa</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
Meaning 1.21: the land

Description: 'the hard surface of the earth, when compared to the area covered by sea'

Typical context: The captain sighted land in the distance.

Semantic field: The physical world

Semantic category: Noun

Borrowed score: 0.27

Age score: 0.90

Simplicity score: 0.94

Counterpart words in the World Loanword Database

<table>
<thead>
<tr>
<th>Voc. ID</th>
<th>Vocabulary</th>
<th>Word Form</th>
<th>Original Script</th>
<th>Borrowed status</th>
<th>Borrowed score</th>
<th>Age score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Swahili</td>
<td>nchi kavu</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>Iraqw</td>
<td>hhapee</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.0</td>
</tr>
</tbody>
</table>
The relationship between concept and form is not a definition or a translation.

Rather, it is something like *counterpart*.
The relationship between concept and form is not a definition or a translation. Rather, it is something like *counterpart*.

- **CONCEPT** counterpartRelation word
- **CAT** hasCounterpart cat
Mappings

- The relationship between concept and form is not a definition or a translation.
- Rather, it is something like *counterpart*.

<table>
<thead>
<tr>
<th>CONCEPT</th>
<th>counterpartRelation</th>
<th>word</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>hasCounterpart</td>
<td>cat</td>
</tr>
<tr>
<td>PERSON</td>
<td>hasCounterpart</td>
<td>man, person</td>
</tr>
</tbody>
</table>
Mappings

- The relationship between concept and form is not a definition or a translation.
- Rather, it is something like *counterpart*.

<table>
<thead>
<tr>
<th>CONCEPT</th>
<th>counterpartRelation</th>
<th>word</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAT</td>
<td>hasCounterpart</td>
<td>cat</td>
</tr>
<tr>
<td>PERSON</td>
<td>hasCounterpart</td>
<td>man, person</td>
</tr>
<tr>
<td>PARENT’S SIBLING</td>
<td>hasSubCounterparts</td>
<td>aunt, uncle</td>
</tr>
</tbody>
</table>
• The “words” in a wordlist can be modeled as defective signs
The “words” in a wordlist can be modeled as defective signs.
Forms

• The “words” in a wordlist can be modeled as defective signs

<table>
<thead>
<tr>
<th>Form</th>
<th>Grammar</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>chien</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Schematic Sign

perro
noun
dog

Lexicon entry

6. DOG

Wordlist Entry
Lexicons and wordlists

- Interoperate with lexicons via the sign
- Infer from concept for specific applications

- If a given resource contains information beyond the form, it can be directly included

6. DOG

"dog"

chien
Lexicons and wordlists

- Interoperate with lexicons via the sign
- Infer from concept for specific applications

If a given resource contains information beyond the form, it can be directly included.
Lexicons and wordlists

- Interoperate with lexicons via the sign
- Infer from concept for specific applications

If a given resource contains information beyond the form, it can be directly included.
Lexicons and wordlists

- Interoperate with lexicons via the sign
- Infer from concept for specific applications

If a given resource contains information beyond the form, it can be directly included.
Chilver and Kaberry (1974:34)

Bird: finyê, mnyêm
Blood: alemá, bilem
Body: yur
Bone: akufê, bêkufê
Bow: kanò, bunò
Brother, etc.: mwâmo - my br. (pl. bà:)
           mwâmo - yr. br.
           el: mwatse, asse in 'mu o
           no mwatse mfa' - This man is
           the chief's brother.

Bushcow: nyam ye, tsành, nyam ye ntjûn
Calabash: kîhêr, bîhêr
Canoe: fanse, mse
Cap: kifâna, bëfànà
Cat - dom: kibò, bîbò
Wild: ngù, bângù
Cherven: kifà, bëfà [? fon's cap]
Chêr: fômûsà, múmûsà
Child: mwêli, bëmûlêli
Chêr: gûvù, ñgûvù
Chief: sì, sì
Chief: mfàn, bômfa

Described Variety of a Language

1. PERSON in English  → "person"
2. MAN in English  → "man"
3. WOMAN in English → "woman"
4. HORSE in English → "horse"
5. EWE in English   → "ewe"

Developed from Haspelmath and Tadmor (2009)
Developed from Haspelmath and Tadmor (2009)

Described Variety of a Language

Concepticon

1. PERSON in English "person"
2. MAN in English "man"
3. WOMAN in English "woman"
4. HORSE in English "horse"
5. EWE in English "ewe"

Wordlist model
Described Variety of a Language

1. PERSON
2. MAN
3. WOMAN
4. HORSE
5. EWE

in English
"person"
"man"
"woman"
"horse"
"ewe"

Developed from Haspelmath and Tadmor (2009)
Mapping mediated by sign groups, usually of one sign

Described Variety of a Language

Concepticon

1. PERSON
2. MAN
3. WOMAN
4. HORSE
5. EWE

in English

"person"
"man"
"woman"
"horse"
"ewe"

Developed from Haspelmath and Tadmor (2009)
<entry id="_1934">
  <trait name="concept-id" value="1934" />
  <lexical-unit>
    <form lang="hig">
      <text>ɣàtʊ́ʔ</text>
    </form>
  </lexical-unit>
</entry>

<entry id="_1938">
  <trait name="concept-id" value="1938" />
  <lexical-unit>
    <form lang="hig">
      <text>ɣàtʊ́ʔ</text>
    </form>
    <lexical-unit>
      <form lang="hig">
        <text>ɣàtʊ́ʔ</text>
      </form>
    </lexical-unit>
</entry>

<skos:concept>
  <lego:conceptId>1938</lego:conceptId>
  <dc:description xsi:type="lego:default"
    lego:source="LWT"
    lego:sourceID="17.310"
    lego:label="to finish" />
  <dc:description
    lego:source="IDS"
    lego:sourceID="14.27"
    lego:label="finish" />
  <dc:description
    lego:source="UW"
    lego:id="962"
    lego:label="finish, end, vb" />
</skos:concept>
<entry id="_1934">
  <trait name="concept-id" value="1934" />
  <lexical-unit>
    <form lang="hig">
      <text>ɣàtóʔ</text>
    </form>
  </lexical-unit>
</entry>

<entry id="_1938">
  <trait name="concept-id" value="1938" />
  <lexical-unit>
    <form lang="hig">
      <text>wgyi</text>
    </form>
  </lexical-unit>
</entry>

<skos:concept>
  <lego:conceptId>1938</lego:conceptId>
  <dc:description xsi:type="lego:default" lego:source="LWT" lego:sourceID="17.310" lego:label="to finish" />
  <dc:description lego:source="IDS" lego:sourceID="14.27" lego:label="finish" />
  <dc:description lego:source="UW" lego:id="962" lego:label="finish, end, vb" />
</skos:concept>
<entry id="_1934">
  <trait name="concept-id" value="1934"/>
  <lexical-unit>
    <form lang="hig">
      <text>ɣàtóʔ</text>
    </form>
  </lexical-unit>
</entry>

<entry id="_1938">
  <trait name="concept-id" value="1938"/>
  <lexical-unit>
    <form lang="hig">
      <text>wgłyóʔ</text>
    </form>
  </lexical-unit>
</entry>

<skos:concept>
  <lego:conceptId>1938</lego:conceptId>
  <dc:description xsi:type="lego:default"
    lego:source="LWT"
    lego:sourceID="17.310"
    lego:label="to finish"/>
  <dc:description
    lego:source="IDS"
    lego:sourceID="14.27"
    lego:label="finish"/>
  <dc:description
    lego:source="UW"
    lego:id="962"
    lego:label="finish, end, vb"/>
</skos:concept>

Concepticon entry

Word list entries

XML Implementation
<skos:concept>
  <lego:conceptId>1938</lego:conceptId>
  <dc:description xsi:type="lego:default"
    lego:source="LWT"
    lego:sourceID="17.310"
    lego:label="to finish" />
  <lexical-unit>
    <form lang="hig">
      <text>ɣàtʊ́ʔ</text>
    </form>
  </lexical-unit>
  <lexical-unit>
    <form lang="hig">
      <text>wgyì</text>
    </form>
  </lexical-unit>
</skos:concept>

<skos:concept>
  <lego:conceptId>_1938</lego:conceptId>
  <dc:description
    lego:source="IDS"
    lego:sourceID="14.27"
    lego:label="finish" />
  <lexical-unit>
    <form lang="hig">
      <text>wgyívò</text>
    </form>
  </lexical-unit>
  <lexical-unit>
    <form lang="hig">
      <text>wgyì</text>
    </form>
  </lexical-unit>
</skos:concept>

<skos:concept>
  <lego:conceptId>_1938</lego:conceptId>
  <dc:description
    lego:source="UW"
    lego:id="962"
    lego:label="finish, end, vb" />
  <lexical-unit>
    <form lang="hig">
      <text>ɣàtʊ́ʔ</text>
    </form>
  </lexical-unit>
  <lexical-unit>
    <form lang="hig">
      <text>wgyívò</text>
    </form>
  </lexical-unit>
  <lexical-unit>
    <form lang="hig">
      <text>wgyì</text>
    </form>
  </lexical-unit>
</skos:concept>
<entry id="_1934">
  <trait name="concept-id" value="1934" />
  <lexical-unit>
    <form lang="hig">
      <text>ɣàtʊ́ʔ</text>
    </form>
  </lexical-unit>
</entry>

<entry id="_1938">
  <trait name="concept-id" value="1938" />
  <lexical-unit>
    <form lang="hig">
      <text>wgyì</text>
    </form>
  </lexical-unit>
  <lexical-unit>
    <form lang="hig">
      <text>wgyìvò</text>
    </form>
  </lexical-unit>
</entry>

<skos:concept>
  <lego:conceptId>1938</lego:conceptId>
  <dc:description
    xsi:type="lego:default"
    lego:source="LWT"
    lego:sourceID="17.310"
    lego:label="to finish" />
  <dc:description
    lego:source="IDS"
    lego:sourceID="14.27"
    lego:label="finish" />
  <dc:description
    lego:source="UW"
    lego:id="962"
    lego:label="finish, end, vb" />
</skos:concept>

Word list entries

XML Implementation
Building on concepticons

- Concepticons were a kind of standard before linguists thought about “standards”
- They are the closest semantic equivalent to the International Phonetic Alphabet
- Standardizing how to interpret specific concepts should be relatively easy
- These can serve as the basis for more controversial semantic categorizations
Meaning 1.21: the land

Description: 'the hard surface of the earth, when compared to the area covered by sea'

Typical context: The captain sighted land in the distance.

Semantic field: The physical world

Semantic category: Noun

Borrowed score: 0.27

Age score: 0.90

Simplicity score: 0.94

Counterpart words in the World Loanword Database

<table>
<thead>
<tr>
<th>Voc. ID</th>
<th>Vocabulary</th>
<th>Word Form</th>
<th>Original Script</th>
<th>Borrowed status</th>
<th>Borrowed score</th>
<th>Age score</th>
<th>Simplicity score</th>
<th>Source Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Swahili</td>
<td>nchi kavu</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.50</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Iraqw</td>
<td>hhapee</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Gawwada</td>
<td>pие</td>
<td></td>
<td>4. very little evidence for borrowing</td>
<td>0.25</td>
<td>1.00</td>
<td>1.00</td>
<td>biyya (Oromo)</td>
</tr>
<tr>
<td>4</td>
<td>Hausa</td>
<td>kasa</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
### Meaning 1.21: the land

**Description:** 'the hard surface of the earth, when compared to the area covered by sea'

**Typical context:** The captain sighted land in the distance.

**Semantic field:** The physical world

**Semantic category:** Noun

**Borrowed score:** 0.27

**Age score:** 0.90

**Simplicity score:** 0.94

### Counterpart words in the World Loanword Database

<table>
<thead>
<tr>
<th>Voc. ID</th>
<th>Vocabulary</th>
<th>Word Form</th>
<th>Original Script</th>
<th>Borrowed status</th>
<th>Borrowed score</th>
<th>Age score</th>
<th>Simplicity score</th>
<th>Source Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Swahili</td>
<td>nchi kavu</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.50</td>
<td>0.50</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Iraqw</td>
<td>hhaapee</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td></td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Gawwada</td>
<td>piye</td>
<td></td>
<td>4. very little evidence for borrowing</td>
<td>0.25</td>
<td>1.00</td>
<td>1.00</td>
<td>biyy (Oromo)</td>
</tr>
<tr>
<td>4</td>
<td>Hausa</td>
<td>kasa</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>
Meaning 1.21: the land

Description: ‘the hard surface of the earth, when compared to the area covered by sea’

Typical context: The captain sighted land in the distance.

Semantic field: The physical world

Semantic category: Noun

Borrowed score: 0.27
Age score: 0.90
Simplicity score: 0.94

Counterpart words in the World Loanword Database

<table>
<thead>
<tr>
<th>Voc. ID</th>
<th>Vocabulary</th>
<th>Word Form</th>
<th>Original Script</th>
<th>Borrowed status</th>
<th>Borrowed score</th>
<th>Age score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Swahili</td>
<td>nchi kavu</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>Iraqw</td>
<td>hhapee</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
Meaning 1.21: the land

Description: 'the hard surface of the earth, when compared to the area covered by sea'

Typical context: The captain sighted land in the distance.

Semantic field: The physical world

Semantic category: Noun

Borrowed score: 0.27
Age score: 0.90
Simplicity score: 0.94

Counterpart words in the World Loanword Database:

<table>
<thead>
<tr>
<th>Voc. ID</th>
<th>Vocabulary</th>
<th>Word Form</th>
<th>Original Script</th>
<th>Borrowed status</th>
<th>Borrowed score</th>
<th>Age score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Swahili</td>
<td>nchi kavu</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td>0.5</td>
</tr>
<tr>
<td>2</td>
<td>Iraqw</td>
<td>hhapee</td>
<td></td>
<td>5. no evidence for borrowing</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
Concepticon alternatives?

- There are alternatives, e.g., LMF SenseAxes
- The value of wordlist concepticons
  - Intended to be universally applicable
  - Already applied to a substantial portion of the world’s languages
  - Have a natural place in the descriptive linguist’s workflow
- At the same time, they are not designed for translation
Wordlists as linked data

- Wordlists used for language comparison were an early kind of **linked data**
- Such data has become valuable in the context of the Semantic Web
- The simple structure of wordlists makes visible what is needed to link lexical data
- Wordlist concepts provide
  - An **annotation system** for semantic linking
  - A **common target** for the links

<skos:Concept rdf:about="http://www.purl.org/linguistics/lego/concept/539">
  <skos:prefLabel>bird</skos:prefLabel>
  <lego:hasCounterpart>
    <gold:LinguisticSign rdf:about="http://www.purl.org/linguistics/Ngong_ChilverPhyllis1974/12">
      <lego:fromLanguage>
      </lego:fromLanguage>
      <gold:hasForm>
        <gold:formUnit><gold:stringRep>fini</gold:stringRep></gold:formUnit>
      </gold:hasForm>
    </gold:LinguisticSign>
  </lego:hasCounterpart>
  <lego:hasCounterpart>
      <lego:fromLanguage>
      </lego:fromLanguage>
      <gold:hasForm>
        <gold:formUnit><gold:stringRep>mùní</gold:stringRep></gold:formUnit>
      </gold:hasForm>
    </gold:LinguisticSign>
  </lego:hasCounterpart>
</skos:Concept>
</rdf:RDF>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:lego="http://purl.org/linguistics/lego/
  xmlns:gold="http://purl.org/linguistics/gold/
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://www.purl.org/linguistics/lego/concept/539">
    <skos:prefLabel>bird</skos:prefLabel>
    <lego:hasCounterpart>
      <gold:LinguisticSign rdf:about="http://www.purl.org/linguistics/Ngong_ChilverPhyllis1974/12">
        <lego:fromLanguage>
        </lego:fromLanguage>
        <gold:hasForm>
          <gold:formUnit><gold:stringRep>fini</gold:stringRep></gold:formUnit>
        </gold:hasForm>
      </gold:LinguisticSign>
    </lego:hasCounterpart>
    <lego:hasCounterpart>
        <lego:fromLanguage>
        </lego:fromLanguage>
        <gold:hasForm>
          <gold:formUnit><gold:stringRep>mùní</gold:stringRep></gold:formUnit>
        </gold:hasForm>
      </gold:LinguisticSign>
    </lego:hasCounterpart>
  </skos:Concept>
</rdf:RDF>
<skos:Concept rdf:about="http://www.purl.org/linguistics/lego/concept/539">
  <skos:prefLabel>bird</skos:prefLabel>
  <lego:hasCounterpart>
    <gold:LinguisticSign rdf:about="http://www.purl.org/linguistics/Ngong_ChilverPhyllis1974/12">
      <lego:fromLanguage>
      </lego:fromLanguage>
      <gold:hasForm>
        <gold:formUnit><gold:stringRep>fini</gold:stringRep></gold:formUnit>
      </gold:hasForm>
    </gold:LinguisticSign>
  </lego:hasCounterpart>
  <lego:hasCounterpart>
      <lego:fromLanguage>
      </lego:fromLanguage>
      <gold:hasForm>
        <gold:formUnit><gold:stringRep>mùní</gold:stringRep></gold:formUnit>
      </gold:hasForm>
    </gold:LinguisticSign>
  </lego:hasCounterpart>
</skos:Concept>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:lego="http://purl.org/linguistics/lego/
  xmlns:gold="http://purl.org/linguistics/gold/
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://www.purl.org/linguistics/lego/concept/539">
    <skos:prefLabel>bird</skos:prefLabel>
    <lego:hasCounterpart>
      <gold:LinguisticSign rdf:about="http://www.purl.org/linguistics/Ngong_ChilverPhyllis1974/12">
        <lego:fromLanguage>
        </lego:fromLanguage>
        <gold:hasForm>
          <gold:formUnit><gold:stringRep>fini</gold:stringRep></gold:formUnit>
        </gold:hasForm>
      </gold:LinguisticSign>
    </lego:hasCounterpart>
    <lego:hasCounterpart>
        <lego:fromLanguage>
        </lego:fromLanguage>
        <gold:hasForm>
          <gold:formUnit><gold:stringRep>mùní</gold:stringRep></gold:formUnit>
        </gold:hasForm>
      </gold:LinguisticSign>
    </lego:hasCounterpart>
  </skos:Concept>
</rdf:RDF>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:lego="http://purl.org/linguistics/lego/
  xmlns:gold="http://purl.org/linguistics/gold/
  xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <skos:Concept rdf:about="http://www.purl.org/linguistics/lego/concept/539">
    <skos:prefLabel>bird</skos:prefLabel>
    <lego:hasCounterpart>
      <gold:LinguisticSign rdf:about="http://www.purl.org/linguistics/Ngong_ChilverPhyllis1974/12">
        <lego:fromLanguage>
        </lego:fromLanguage>
        <gold:hasForm>
          <gold:formUnit><gold:stringRep>fini</gold:stringRep></gold:formUnit>
        </gold:hasForm>
      </gold:LinguisticSign>
    </lego:hasCounterpart>
    <lego:hasCounterpart>
        <lego:fromLanguage>
        </lego:fromLanguage>
        <gold:hasForm>
          <gold:formUnit><gold:stringRep>mùní</gold:stringRep></gold:formUnit>
        </gold:hasForm>
      </gold:LinguisticSign>
    </lego:hasCounterpart>
  </skos:Concept>
</rdf:RDF>
RDF Implementation
RDF Implementation
The idea of a unified concepticon has garnered the most interest

Some possibilities

- Refining associated taxonomies
- Linking to resources like WordNet

A nice example of a traditional resource that is already almost “web-ready”
Acknowledgments

• Shakthi Poornima—Graduate Assistant
• National Science Foundation Award BCS-0753321