The Association Experiment
Morphological Typology (from WALS)
Semantic Compositionality

When signs are put together …
What is the resulting meaning?
Semantic Compositionality

'bicycle lane'

'to bicycle shop'
The Association Operator

Monadic Association Operator

A ( X )
'entity associated with X'
in most languages,
observable in genitive construction
The Association Operator

Polyadic Association Operator

\( A( X, Y ) \)

'entity associated with X and Y'

in most languages, a default rule for compositional semantics
The Association Operator

A ( BICYCLE, THATAWAY )

'entity associated with bicycle and thataway'
Where the Association Operator is Found

- phylogenetically
in the language of captive great apes

Where the Association Operator is Found

- phylogenetically
  in the language of captive great apes

<table>
<thead>
<tr>
<th>Orangutan (Chantek) using ASL</th>
<th>Bonobo (Kanzi) using lexigrams</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOU PULL</td>
<td>LIZ HIDE</td>
</tr>
<tr>
<td>COME CHANTEK</td>
<td>HIDE AUSTIN</td>
</tr>
<tr>
<td>BEARD PULL</td>
<td>WATER HIDE</td>
</tr>
<tr>
<td>PULL BEARD</td>
<td>HIDE PEANUT</td>
</tr>
</tbody>
</table>

Where the Association Operator is Found

- phylogenetically
- **ontogenetically**
  in the language of young infants

Where the Association Operator is Found

- phylogenetically
- **ontogenetically**
  in the language of young infants

Allison 1;8

[playing with toy pig inside toy truck; pig is hurt by sharp corner of truck]

Hurt knee
Hurt truck

(Bloom 1973)
Where the Association Operator is Found

- phylogenetically
- ontogenetically
- grammatically

as a universal design feature of language

```
tilfanti leavraham

A ( TELEPHONE, ABRAHAM )

'entity associated with telepone and Avraham'

* 'Beavers build dams'

'I telephoned Abraham'
```
The Association Operator

- phylogenetically
- ontogenetically
- grammatically
- typologically

Languages vary with respect to the extent to which associational semantics is supplemented with additional rules of compositional semantics.

Highly associational languages have fewer such rules, while highly articulated languages have more such rules.
Riau Indonesian

Ayam  makan
CHICKEN  EAT

'The chicken is eating'
'Someone is eating the chicken'
'Someone is eating with the chicken'
'Someone is eating because of the chicken'

'The chicken that is eating'
'Where the chicken is eating'
'Why the chicken is eating'
Ayam
CHICKEN
makan
EAT

A ( CHICKEN, EAT )

'entity associated with chicken and eat'
But to what extent are other languages like Riau Indonesian?
The Association Experiment

Goal: measuring the availability of apparently associational interpretations interpretations that are not obtained by the application of construction-specific rules, and which therefore may plausibly be characterized as resulting from the application of the association operator

Languages studied: isolating apparent SVO word order

Semantic domain studied: thematic roles
The Association Experiment

Constructions sought:

Apparently associational interpretations involving thematic roles:

**Bare Peripheral (BPer)**
zero-marked obliques and other more extraneous thematic roles

**Bare Patient Preceding (BPatP)**
apparent OV word order
The clown is drinking the book

4 possible responses

Test picture

alternative picture
Testing for BPer (following, extraneous)  

**Badut minum buku**

?  

test picture  

alternative picture
Testing for BPer (following, extraneous)
Testing for BPer (following, extraneous) stimulus 15

Anh hê uồng sách

test picture alternative picture
Testing for BPer (following, extraneous) stimulus 15

小丑飲書

? test picture alternative picture
Testing for BPer (following, extraneous) stimulus 15

Kitenga tchi ǂxanù

test picture alternative picture
Testing for BPer (preceding, extraneous)

The money is happy

alternative picture

stimulus 8

test picture
Testing for BPer (preceding, extraneous)

stimulus 8

Duit gembira

alternative picture

test picture
Testing for BPer (following, oblique) stimulus 14

The soldier is cutting the axe

alternative picture  

?  

test picture
Testing for BPer (following, oblique) stimulus 14

Tentara potong kampak

alternative picture  test picture
Testing for BPer (preceding, oblique)  

stimulus 6

The chairs are jumping

alternative picture

test picture
Testing for BPer (preceding, oblique)  stimulus 6

Kursi loncat

alternative picture  test picture
Testing for **BPatP**

The dog is drawing

? text picture alternative picture

stimulus 3
Testing for BPatP

stimulus 3

Anjing lukis

? text picture alternative picture
Testing for **BPatP** (plus Ag) 

**The car is pushing the woman**

- text picture
- alternative picture

stimulus 11
Testing for BPatP (plus Ag)

Mobil dorong cewek

stimulus 11

Text picture

Alternative picture
### Numbers of Subjects (as of 4.4.2008)

<table>
<thead>
<tr>
<th>Language</th>
<th>total</th>
<th>passed</th>
<th>distr/s</th>
<th>total</th>
<th>passed</th>
<th>distr/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>63</td>
<td>63</td>
<td></td>
<td>63</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Hebrew</td>
<td>66</td>
<td>66</td>
<td></td>
<td>66</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Papiamentu</td>
<td>46</td>
<td>45</td>
<td></td>
<td>46</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Sranan</td>
<td>63</td>
<td>60</td>
<td></td>
<td>63</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Bislama</td>
<td>50</td>
<td>47</td>
<td></td>
<td>50</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Twi</td>
<td>60</td>
<td>55</td>
<td></td>
<td>60</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Fongbe</td>
<td>48</td>
<td>45</td>
<td></td>
<td>48</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Yoruba</td>
<td>66</td>
<td>62</td>
<td></td>
<td>66</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Meyah</td>
<td>67</td>
<td>63</td>
<td></td>
<td>67</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Vietnamese</td>
<td>88</td>
<td>80</td>
<td></td>
<td>88</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>Cantonese</td>
<td>61</td>
<td>60</td>
<td></td>
<td>61</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Sundanese</td>
<td>44</td>
<td>39</td>
<td></td>
<td>44</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>Minangkabau</td>
<td>77</td>
<td>71</td>
<td></td>
<td>77</td>
<td>71</td>
<td></td>
</tr>
<tr>
<td>Mentawai</td>
<td>38</td>
<td>37</td>
<td></td>
<td>38</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Standard Indonesian</td>
<td>133</td>
<td>123</td>
<td></td>
<td>133</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>Kuala Lumpur Malay</td>
<td>78</td>
<td>76</td>
<td></td>
<td>78</td>
<td>76</td>
<td></td>
</tr>
<tr>
<td>Kuching Malay</td>
<td>32</td>
<td>31</td>
<td></td>
<td>32</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Bengkulu</td>
<td>35</td>
<td>32</td>
<td></td>
<td>35</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>Papua Malay</td>
<td>71</td>
<td>65</td>
<td></td>
<td>71</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Riau Indonesian</td>
<td>169</td>
<td>164</td>
<td></td>
<td>169</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>Siak Malay</td>
<td>83</td>
<td>75</td>
<td></td>
<td>83</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>Jakarta Indonesian</td>
<td>104</td>
<td>99</td>
<td></td>
<td>104</td>
<td>99</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>1575</strong></td>
<td><strong>1488</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experimental Results
Experimental Results

BPer
BPatP

Stand I
Conclusions I

- languages range from highly articulated to highly associational
- within the typology of isolating languages Riau Indonesian is mid-range
- highly associational languages present us with a clear view of an evolutionary fossil

Question

- why do languages vary from highly articulated to highly associational?
The Articulation Index

A numerical measure of how much obligatory grammatical marking a language has

Calculated on a representative text by counting:

- the number of morphemes
- the number of concepts belonging to major ontological types (thing, property, activity)

and then calculating the ratio:

morphemes / concepts
## The Articulation Index

Sample calculation of articulation index (based on stimulus 3)

<table>
<thead>
<tr>
<th>Language</th>
<th>Sentence</th>
<th>Concepts</th>
<th>Morphemes</th>
<th>Articulation Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riau Indonesian</td>
<td>Anjing lukis</td>
<td>2</td>
<td>2</td>
<td>1.0</td>
</tr>
<tr>
<td>Bislama</td>
<td>Dog i dro</td>
<td>2</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>Papiamentu</td>
<td>E kachó ta pinta</td>
<td>2</td>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td>English</td>
<td>The dog is drawing</td>
<td>2</td>
<td>6</td>
<td>3.0</td>
</tr>
</tbody>
</table>
The Articulation Index

Sample calculation of articulation index
(based on stimulus 3)

<table>
<thead>
<tr>
<th>concepts</th>
<th>morphemes</th>
<th>articulation index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riau Indonesian</td>
<td>Anjing lukis</td>
<td>2</td>
</tr>
</tbody>
</table>

"What strikes the learner of Malay is the complete lack of those typically Indo-European properties — gender, inflection, conjugation. It is like diving into a bath of pure logic. Everything is pared to a minimum."

Anthony Burgess

Language Made Plain
The Articulation Index

Sample calculation of articulation index (based on stimulus 3)

<table>
<thead>
<tr>
<th>Language</th>
<th>Sentence</th>
<th>Concepts</th>
<th>Morphemes</th>
<th>Articulation Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Riau Indonesian</td>
<td>Anjing lukis</td>
<td>2</td>
<td>2</td>
<td>1.0</td>
</tr>
<tr>
<td>Bislama</td>
<td>Dog i dro</td>
<td>2</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>Papiamentu</td>
<td>E kachó ta pinta</td>
<td>2</td>
<td>4</td>
<td>2.0</td>
</tr>
<tr>
<td>English</td>
<td>The dog is drawing</td>
<td>2</td>
<td>6</td>
<td>3.0</td>
</tr>
</tbody>
</table>
Articulation Index & BPer Interpretations

The graph illustrates the relationship between articulation index and BPer interpretations. The data points are plotted on a grid with the x-axis representing articulation index values from 0.75 to 3, and the y-axis representing BPer interpretations from 0 to 100. A downward trend line is drawn through the data points, indicating a negative correlation between the two variables.
Articulation Index & BPatPrec Interpretations
Articulation Index, Population Size & BPer Interpretations

- Small (under 100k)
- Mid-range
- Large (over 100m)
Articulation Index, Population Size & BPatPrec Interpretations

- small (under 100k)
- mid-range
- large (over 100m)
Conclusions II

- languages with lower articulation index are more highly associational
- languages with smaller populations are more highly associational

Question

- why is there so much spread Within Malay/Indonesian?
### Criteria for Sociolinguistic Distinctiveness of Malay/Indonesian Dialects

<table>
<thead>
<tr>
<th>1st half</th>
<th>Kuala Lumpur Malay</th>
<th>Kuching Malay</th>
<th>Bengkulu Malay</th>
<th>Papua Malay</th>
<th>Indonesian Riau Malay</th>
<th>Siak Malay</th>
<th>Jakarta Indonesian</th>
<th>Minangkabau Sundanese</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distinct self-appellation</strong></td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>Cannot be referred to as Bahasa Melayu/Indonesia</strong></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Referred to as bahasa daerah ('regional language')</strong></td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Never considered &quot;broken&quot; or &quot;imperfect&quot; language</strong></td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Prototypically associated with particular ethnicity</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td><strong>Most speakers are of single particular ethnicity</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td><strong>Published dictionary</strong></td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
### Criteria for Sociolinguistic Distinctiveness of Malay/Indonesian Dialects

<table>
<thead>
<tr>
<th></th>
<th>Kuala Lumpur Malay</th>
<th>Kucing Malay</th>
<th>Bengkulu Malay</th>
<th>Papua Malay</th>
<th>Indonesian Riau Malay</th>
<th>Siak Malay</th>
<th>Jakarta Indonesian</th>
<th>Minangkabau</th>
<th>Sundanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distinct conventionalized orthography</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Widespread conscious use in newspapers</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Widespread conscious use on radio</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Emblematic use in advertising</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Widespread conscious use in popular music</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Official status as vehicle of instruction in schools</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Has its own distinct acrolect</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>
Criteria for Sociolinguistic Distinctiveness of Malay/Indonesian Dialects

<table>
<thead>
<tr>
<th></th>
<th>Kuala Lumpur Malay</th>
<th>Kucing Malay</th>
<th>Bengkulu Malay</th>
<th>Papua Malay</th>
<th>Indonesian Riau Malay</th>
<th>Siak Malay</th>
<th>Jakarta Indonesian</th>
<th>Minangkabau</th>
<th>Sundanese</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sociolinguistic Distinctiveness</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>0</td>
<td>6</td>
<td>8</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Lexical Distinctiveness</td>
<td>2</td>
<td>24</td>
<td>21</td>
<td>11</td>
<td>14</td>
<td>21</td>
<td>12</td>
<td>23</td>
<td>61</td>
</tr>
<tr>
<td>Sum: 4S+L</td>
<td>18</td>
<td>36</td>
<td>37</td>
<td>25</td>
<td>14</td>
<td>45</td>
<td>44</td>
<td>79</td>
<td>117</td>
</tr>
</tbody>
</table>
Sociolinguistic Distinctiveness & BPatPrec Interpretations
Conclusions III

• more sociolinguistically distinct Malay/Indonesian dialects are more highly associational

Why?

• two reasons: one "real", one "artifactual"
The "real" reason

- the more sociolinguistically distinctive the dialect is, the more appropriate it is to characterize its population as the relevant subset of the 200m of Malay/Indonesian

- the more sociolinguistically distinctive the dialect is, the smaller its population, and therefore the higher its associationality
The "artifactual" reason

• this is not a real fact about the different dialects but rather an artifact of the experimental method

• the less sociolinguistically distinct the dialect is, the more subjects responses display interference from the standard language, which is of low associationality

• in reality, all Malay/Indonesian dialects may be of similarly high associationality
Language-Internal Sociolinguistic Variation

- up to now, all the data was from subjects conforming to

Baseline Sociolinguistic Profile

- uneducated
- low-to-middle class
- over 12 y/o
- living in community where test language is spoken
- tested in same community

... but what about other speakers?
Baseline Sociolinguistic Profile vs. University Students
Baseline Sociolinguistic Profile vs. University Students
Conclusions IV

• lesser educated speakers are more highly associational

Why?
Why Less Educated Subjects Are More Highly Associational

- **the syntactic-mode effect:**
  educated speakers ⊗ syntactic mode
  uneducated speakers ⊗ pragmatic mode

- **the freak-out effect:**
  uneducated subjects for whom task is unusual panic and simplify their task, resorting to "protolanguage"

- **the semi-literacy effect:**
  semi-literate subjects latch on to the content words but but ignore the grammatical information

- **the standard-language interference effect:**
  educated subjects are more likely to display interference from a standardized language of lower associationality
Linguists describing languages based on data from foreign students (and other expatriates and emigrés) have **systematically underestimated** the extent to which languages may be highly associational.
Badut minum buku
Over 6000 languages waiting for YOU