Variation in the occurrence of the transitive suffix in a corpus of Tok Pisin

Conor Snoek
University of Alberta
Overview

- Tok Pisin (Melanesian Pidgin)
- The Slone Wantok Corpus
- Transitive verbs in Tok Pisin
- Three possible sources of variation:
  - meaning differences
  - regional differences
  - differences over time
Melanesian Pidgin

- Tok Pisin is a language spoken in Papua New Guinea, mainly as a second language, about 121,000 speakers, although there are some first-language speakers, ca. 50,000. (SIL)

- Main lexifier language is English. Older German stratum of vocabulary. Noticeable influence from Kuanua (Tolai)

- A large number of other potential substrates.
The Slone *Wantok* Corpus

- Composed of stories sent in to a newspaper, *Wantok*, over a 30 year period
- The stories are *Ancestor stories* (*Stori Tumbuna*) and typically of traditional origins
- Information on the publication date of the story is provided
- Information on the region of origin of the story is provided (some missing data)
- Approximately 800,000 words in the corpus
Transitive verbs in Tok Pisin

• Any verb ending in \(-im\) is transitive and has a direct object:

  • *bihanim mi* ‘to follow me’

• Any transitive verb not ending in the suffix \(-im\) requires *long* before its object:

  • *lukluk long graun* ‘look on the ground’

  (Mihalic 1971:24)
Transitive verbs in Tok Pisin

There are a number of transitive verbs that appear with and without the suffix:

- **abrus** (miss/ evade), **dring** (drink), **wet** (wait), **kaikai** (eat), **bilip** (believe), **goap** (go up), **hangamap** (hang up), **holimpas** (hold), **kalap** (jump), **lain** (line up), **lukaut** (look out), **singaut** (call for), **tok** (talk), **was** (watch)

Without the suffix they “require the preposition long.”

(Mihalic 1971: 26)
Overview

- Melanesian Pidgin (TP)
- The Slone Wantok Corpus
- Transitive suffix verbs in Tok Pisin
- Three possible sources of variation:
  - meaning differences
  - regional differences
  - differences over time
Verb groups based on glossing and collocate patterns

Cases where the suffixed verbs and verbs followed by *long* differ in their semantics:

*abrus* (miss / evade), *dring* (drink), *wet* (wait), *kaikai* (eat)

Cases where the suffixed verbs and verbs followed by *long* have very similar semantics:

*bilip* (believe), *goap* (go up), *hangamap* (hang up), *holimpas* (hold), *kalap* (jump), *lain* (line up), *lukaut* (look out), *singaut* (call for), *tok* (talk), *was* (watch)
Verbs with differing semantics

*abrus* (long)  ‘to miss’

*Spia bilong tupela i no save abrus long wanpela abus.*
‘Their arrows never missed an animal.’

*abrusim*  ‘to evade’

*Sikau i abrusim gut spia.*
‘The marsupial dodged the spears well.’

The difference in these verb pairs lies with the causer of the “missing”: the source or the target.
Glossing of *singaut* and *singautim*

- **Singaut (long)**   ‘to call for someone’, ‘to call for’

*Man ya i singaut long meri long kalap i go long kanu.*
‘The man called out for the woman to jump into the canoe’

- **Singautim**   ‘to call someone’, ‘to call for’

*Em i singautim meri bilong em long stretim ol dispela kaikai.*
‘He called out for his wife to take care of all of the food’
### 10 most frequent collocates of *singaut long* and *singautim*

<table>
<thead>
<tr>
<th>Relative Frequency</th>
<th>collocate</th>
<th>Relative Frequency</th>
<th>collocate</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.076687117</td>
<td><em>ol</em></td>
<td>0.197674419</td>
<td><em>ol</em></td>
</tr>
<tr>
<td>0.036809816</td>
<td><em>em</em></td>
<td>0.127906977</td>
<td><em>olgeta</em></td>
</tr>
<tr>
<td>0.024539877</td>
<td><em>tupela</em></td>
<td>0.098006645</td>
<td><em>em</em></td>
</tr>
<tr>
<td>0.021472393</td>
<td><em>papa</em></td>
<td>0.064784053</td>
<td><em>tupela</em></td>
</tr>
<tr>
<td>0.021472393</td>
<td><em>mi</em></td>
<td>0.043189369</td>
<td><em>pikinini</em></td>
</tr>
<tr>
<td>0.018404908</td>
<td><em>wanem</em></td>
<td>0.031561462</td>
<td><em>meri</em></td>
</tr>
<tr>
<td>0.018404908</td>
<td><em>man</em></td>
<td>0.024916944</td>
<td><em>mi</em></td>
</tr>
<tr>
<td>0.018404908</td>
<td><em>dok</em></td>
<td>0.024916944</td>
<td><em>dok</em></td>
</tr>
<tr>
<td>0.015337423</td>
<td><em>pikinini</em></td>
<td>0.023255814</td>
<td><em>yu</em></td>
</tr>
<tr>
<td>0.01226994</td>
<td><em>tokples</em></td>
<td>0.021594684</td>
<td><em>man</em></td>
</tr>
</tbody>
</table>
Overview

- Melanesian Pidgin (TP)
- The Slone Wantok Corpus
- Transitive suffix verbs in Tok Pisin
- Three possible sources of variation:
  - meaning differences
  - regional differences in usage of verbs from the second group
  - differences over time
Comparing frequencies of verbs followed by long and –im suffixed verbs by Region

The chart illustrates the frequency distribution of long and –im suffixed verbs across different regions. Each bar represents a region, with the height indicating the frequency. The red bars represent verbs followed by long, while the blue bars represent verbs followed by –im. The regions are color-coded for clarity: W.Sepik, E.Sepik, Madang, Enga, S.Highlan..., Simbu, W.Highlan..., E.Highlan..., Morobe, Gulf, Western, E.NewBri..., W.NewBri..., N.Ireland, Manus, N.Solom...
Comparing frequencies of verbs followed by *long* and *–im* suffixed verbs by Region

Verbs followed by *long* in white
Verbs with *–im* suffix in pink
Overview

- Melanesian Pidgin (TP)
- The Slone Wantok Corpus
- Transitive suffix verbs in Tok Pisin
- Three possible sources of variation:
  - meaning differences
  - regional differences
  - differences over time in usage of verbs from the second group
Singautim and Singaut long over time
Comparing frequencies of verbs followed by long and –im suffixed verbs by time
Outlook

- One set of verbs show clear semantic differences
  For the other set:
  - Some indication of temporal variation
  - Some indication of regional variation
  - However, none of these factors seem to be able to explain the variation on their own
  - Substrate languages might also play a role
  - Statistical testing is necessary to gain further insights
  - It remains a work in progress
Thank you.

Conor Snoek
snoek@ualberta.ca