Examples: how to get them, how they are used by others

Eline Visser
How do I build a linguistic corpus?
Which corpus data am I actually using?
Which grammar data will others actually use?
Overview

• Grammar writer’s perspective
  – stats from 4 grammars
  – year
  – type  
  – clarity

• Grammar user’s perspective
  – typologists questionnaire
  – five principles for good examples
Grammar writer’s perspective
Do year, type and clarity of recording influence the amount of examples used in my grammar?
<table>
<thead>
<tr>
<th>Title</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>A grammar of Palula</td>
<td>Henrik Liljegren</td>
</tr>
<tr>
<td>A grammar of Yauyos Quechua</td>
<td>Aviva Shimeiman</td>
</tr>
<tr>
<td>A grammar of Rapa Nui</td>
<td>Paulus Kieviet</td>
</tr>
<tr>
<td>Attributive constructions in North-Eastern Neo-Aramaic</td>
<td>Ariel Gutman</td>
</tr>
<tr>
<td>Advances in the study of Siouan languages and linguistics</td>
<td>Catherine Rudin (ed), Bryan James Gordon (ed)</td>
</tr>
<tr>
<td>Tone in Yongning Na: Lexical tones and morphotonology</td>
<td>Alexis Michaud</td>
</tr>
</tbody>
</table>
- Variables

- Results
  - all independent variables have an effect on dependent variable
  - year, type and clarity all influence the amount of examples that are used
Kalamang
• 107 texts
• 70,000 words (avg 670 per text)
• 1050 examples

examples per 1000 words
• range: 0 – 50 (one outlier)
• average 16,4
amount of used examples
year of collection

- 2015: 0
- 2017: 45
- 2018: 15
- 2019: 35

Frequency
clarity

- **bad**
- **normal**
- **clear**
type of text

- Conversation
- Narrative
- Descriptive
amount of used examples

• 34 texts
• 46,000 words (avg 1350 per text)
• 1650 examples

examples per 1000 words
• range: 0 – 76 (one outlier)
• average 40
Pite Saami

- Joshua Wilbur
amount of used examples

• 23 texts
• 17,000 words (avg 730 per text)
• 236 examples

examples per 1000 words
• range: 0 – 43 (one outlier)
• average 19
Rapa Nui

• Paulus Kieviet
amount of used examples

• 91 texts
• 990,000 words (avg 11,000 per text)
• 1750 examples

examples per 1000 words
• range: 0 – 24
• average 4
<table>
<thead>
<tr>
<th>Language</th>
<th>no. of texts</th>
<th>words</th>
<th>words per text</th>
<th>no. of examples</th>
<th>examples per 1000 words</th>
<th>average examples per 1000 words</th>
<th>recording years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kalamang</td>
<td>107</td>
<td>70000</td>
<td>670</td>
<td>1050</td>
<td>0-50</td>
<td>16</td>
<td>2010s</td>
</tr>
<tr>
<td>Pichi</td>
<td>34</td>
<td>46000</td>
<td>1350</td>
<td>1650</td>
<td>0-76</td>
<td>40</td>
<td>2000s</td>
</tr>
<tr>
<td>Pite Saami</td>
<td>23</td>
<td>17000</td>
<td>730</td>
<td>240</td>
<td>0-43</td>
<td>19</td>
<td>2000s, 2010s</td>
</tr>
<tr>
<td>Rapa Nui</td>
<td>91</td>
<td>990000</td>
<td>11000</td>
<td>1750</td>
<td>0-24</td>
<td>4</td>
<td>1940-2000</td>
</tr>
<tr>
<td>total</td>
<td>255</td>
<td>1123000</td>
<td></td>
<td>4690</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
clarity (Kalamang only)

all differences significant

\( \rho = .338 \) (moderate)

\( P = .000 \)

\( H(2) = 11.8 \)

\( P = .003 \)
year (all languages)

all differences significant

H(2) = 52.0
P = .000
year (all languages except Rapa Nui)

oldest > middle = newest

$H(2) = 12.5$

$P = .002$
type of text (all languages)

elicited < descriptive
narrative < descriptive

$H(3) = 9.335$
$P = .025$
Conclusion
• no recording is useless
• annotate well!
• focus on good speakers and audio quality
• descriptive recordings are most useful
How do others use my grammar?
How do others use my grammar?
How do typologists use grammars?
Not considered

• language/areal specialists
• language users
• language revitalisers/planners
What is a good example?

questionnaire to LingTyp
20 responses

• short or long?
• genre?
• deciding factor when two equally good examples?
• find yourself in corpus?
Five principles

1) clarity
2) brevity
3) context
4) accessibility
5) good translation
1) clarity

• use consistent glosses
• follow conventions

\[ ma=bon \quad kiun=bon \quad se \quad bo-t \]
\[ 3SG=COM \text{ wife.3SG.POSS=}COM \text{ I AM go-T} \]

‘He and his wife have gone.’
2) brevity

• start with short ("vanilla") example
  – no room for confusion
• don’t shun elicitation

_opa   [som  _karuok_] me  _mambara-n kon ladan kerkap_
earlier person three   _TOP stand-N_   one shirt  red

‘Earlier, there were three people standing, one [had] a red shirt.’
2) brevity: practice

IMT vault: all LangSci book examples (https://imtvault.xitio.de/)

• 40 characters
• 5 words

«This can probably be interpreted as the result of negotiating the competing motivations of naturalness, exhaustiveness and succinctness.» - Sebastian Nordhoff
3) context

• provide extralinguistic context
• facilitates interpretation

Sudah diangkat barang sama orang.
PFCT di-lift thing accompany person
[Landing at airport, arriving late at conveyor belt, passenger is worried]
‘The things may have already be taken by someone.’
4) accessibility

• add a tag
• make tag informative

(2) \textit{bal se sor=at koraru}  
dog IAM fish=OBJ bite  
‘The dog has bitten the fish.’  

(3) \textit{bal na}  
dog eat  
‘The dog eats.’  

4) accessibility

- make archive easy to navigate
  - informative titles
  - metadata
- update your email
- choose searchable archive
- think about language! (regional/international)
5) good translations

• use idiomatic sparsely
• add literal at phrase/sentence level

b.  məpəne məcədə layrik pahəlli
   mə -pa -nə mə -čə -tə layrik pa -hən -i
   3P -male -AGN 3P -small -LOC book read -CAUS -NHYP
   father to son book cause to read
   ‘The father had the book read through his son (to someone).’

c. Ramə Shyamə lihəlləmmi
   Ram -nə Shyam -tə li -hən -ləm -i
   Ram -AGN Shyam -LOC narrate -CAUS -EVD -NHYP
   Ram to Shyam caused to narrate
   ‘Ram had (a story) narrated through Shyam.’

discussion point: replicability

• be forthcoming about alternative analyses
• code for certainty?
• «thick analysis»
• gloss ≠ truth

<table>
<thead>
<tr>
<th>Language</th>
<th>Family</th>
<th>Voice marker</th>
<th>Person marker</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soninke</td>
<td>Mande</td>
<td>DETR</td>
<td>-i</td>
<td>3PL</td>
</tr>
<tr>
<td>Mandinka</td>
<td>Mande</td>
<td>ANTIP, REFL</td>
<td>i</td>
<td>3PL</td>
</tr>
<tr>
<td>Mandinka</td>
<td>Mande</td>
<td>ANTIP, REFL</td>
<td>i</td>
<td>2SG</td>
</tr>
<tr>
<td>Krongo</td>
<td>Kadugli-Krongo</td>
<td>ANTIP</td>
<td>-ti</td>
<td>3.INAN.OBL</td>
</tr>
<tr>
<td>Krongo</td>
<td>Kadugli-Krongo</td>
<td>ANTIP</td>
<td>-ti</td>
<td>1SG.NOM</td>
</tr>
<tr>
<td>Krongo</td>
<td>Kadugli-Krongo</td>
<td>ANTIP</td>
<td>-Ăktu</td>
<td>3SG.F</td>
</tr>
<tr>
<td>Koyraboro Senni</td>
<td>Songhay</td>
<td>ANTIP</td>
<td>-a</td>
<td>3SG</td>
</tr>
</tbody>
</table>

Table 3: Antipassive-person overlaps in Africa

discussion point: replicability

- be forthcoming about alternative analyses
- code for certainty?
- «thick analysis»
- gloss ≠ truth

P.QI 1 10.A.i.6–8 ἀγγέλος ἐφόρησεν μαρτυρόσχο[γα] τεν 
ματτάλα αὐλοκιδολ. τα[ρου] πετροιδαλ Ἀουαρα.

aggelos ἐγό-ιν marturos-gou-ka ten mat-t ila
angel Lord-GEN martyr-PL-ACC 3PL GEN afflict-NMLZ-DAT
aul-os-i-ol-∅ tarou petiros-idal
save-PFV-PLACT-PST1.DET-NOM 3SG Peter-COM
dou-ar-α exist-PST1-PRED

‘The angel of the Lord who saved the martyrs from their affliction 
stayed himself with Peter.’

van Gerven Oei, Vincent. presentation on Old Nubian.