Highlights from the MelaTAMP project: Methods and results

Kilu von Prince, Manfred Krifka and Ana Krajinović

Cologne, 28 March 2019
Overview

• In this talk, we will introduce our methods and some of our results in doing
Overview

- In this talk, we will introduce our methods and some of our results in doing
  - comparative research,
Overview

- In this talk, we will introduce our methods and some of our results in doing
  - comparative research,
  - based on corpus data,
Overview

• In this talk, we will introduce our methods and some of our results in doing
  • comparative research,
  • based on corpus data,
  • in combination with storyboard-based elicitations
Overview

In this talk, we will introduce our methods and some of our results in doing
- comparative research,
- based on corpus data,
- in combination with storyboard-based elicitations
- on modality, aspect and tense in Oceanic,
Overview

- In this talk, we will introduce our methods and some of our results in doing
  - comparative research,
  - based on corpus data,
  - in combination with storyboard-based elicitations
  - on modality, aspect and tense in Oceanic,
  - with a focus on habituality.
Overview

• In this talk, we will introduce our methods and some of our results in doing
  • comparative research,
  • based on corpus data,
  • in combination with storyboard-based elicitations
  • on modality, aspect and tense in Oceanic,
  • with a focus on habituality.

• We want to facilitate a discussion on which types of data are particularly useful in descriptive, comparative and theoretical research on low-resource languages.
MelaTAMP languages

Corpora were imported from their original ELAN, FLEX and Toolbox formats to ANNIS (Krause & Zeldes, 2016) through Druskat (2018).
Habituals in Oceanic

• Question: How do Oceanic languages express habitual aspect (as in *I teach at college, lions eat meat*)?
Habituals in Oceanic

- Question: How do Oceanic languages express habitual aspect (as in *I teach at college, lions eat meat*)?
- (Published) answer (*von Prince et al.*, 2019):
Habituals in Oceanic

• Question: How do Oceanic languages express habitual aspect (as in *I teach at college, lions eat meat*)?

• (Published) answer (*von Prince et al.*, 2019):
  • primarily through reduplication or
Habituals in Oceanic

• Question: How do Oceanic languages express habitual aspect (as in *I teach at college, lions eat meat*)?

• (Published) answer (*von Prince et al., 2019*):
  • primarily through reduplication or
  • imperfective aspect marking or
Habituals in Oceanic

• Question: How do Oceanic languages express habitual aspect (as in *I teach at college, lions eat meat*)?
• (Published) answer (*von Prince et al., 2019*):
  • primarily through reduplication or
  • imperfective aspect marking or
  • a combination of both;
Habituals in Oceanic

- Question: How do Oceanic languages express habitual aspect (as in *I teach at college, lions eat meat*)?
- (Published) answer (*von Prince et al., 2019*):
  - primarily through reduplication or
  - imperfective aspect marking or
  - a combination of both;
  - sometimes not at all.
Habituals in Oceanic

• Question: How do Oceanic languages express habitual aspect (as in *I teach at college, lions eat meat*)?

• (Published) answer (*von Prince et al.*, 2019):
  • primarily through reduplication or
  • imperfective aspect marking or
  • a combination of both;
  • sometimes not at all.

• Additional finding: habituality can be a property of longer passages in addition to individual clauses (compare *Carlson & Spejewski*, 1997).
Habituals: examples

(1) Daakaka, future habitual: Reduplication plus imperfective auxiliary

Ka ya=p  du es-esi  ngok teenem a  nye ka  na
ASR 3PL=POT stay REDUP-see 2SG  home  and 1SG ASR 1SG
w=i   ten   dóór   kyun.
POT=COP assigned.to dark.bush just
“[People] shall see you in the village and I, I will go to the bush.” (1347)

(2) Mavea, future habitual: imperfective auxiliary

me ro  nno me ko-lo-suruvu  atano  na  me ko-lo-taua
FUT then 2SG FUT 2SG-IMPF-sleep ground but FUT 2SG-IMPF-put
otoli na  atano
egg  LOC ground
“You will sleep on the ground, and you will lay eggs on the ground.” (06016.060)
Habituals: Translation-based searches

Example search

Searching for `ft=/.*(always|usually|used to| would).*/`
Habituals: Translation-based searches

Example search

Searching for `ft=/.*(always|usually|used to| would).*/`

Example hit:

(3) Daakaka, past habitual:

`sivi te bwe bangbang myane nya na ene
lorikeet dist cont play with 3d comp def
“Lorikeet used to play with the other two.”`
Habituals: Genre-based searches/ browsing I

- Descriptions of animals are a good source for generic statements. In our corpora, we can search for them in text-level metadata.
Habituals: Genre-based searches/ browsing 1

- Descriptions of animals are a good source for generic statements. In our corpora, we can search for them in text-level metadata.

**Example search**

```
meta::gn = "explanation nature" & rtx
```
Habituals: Genre-based searches/ browsing

- Descriptions of animals are a good source for generic statements. In our corpora, we can search for them in text-level metadata.

Example search

```
meta::gn = "explanation nature" & rtx
```
Habituals: Genre-based searches/ browsing 1

- Descriptions of animals are a good source for generic statements. In our corpora, we can search for them in text-level metadata.

Example search

\texttt{meta::gn = "explanation nature" & rtx}

Example hit

(4) Daakaka, generic:

\texttt{mwe pwer-pwer yan leveno kyun te pwe en-en}
\texttt{REAL REDUP-stay on veno.tree just DISC CONT REDUP-eat}

\texttt{ar an}

\texttt{place DEF}

“[The pheasant dove] lives on the veno tree and eats there.”
Habituals: Genre-based searches/ browsing II

- Biographic and historic accounts are good sources for past habituals.
Habituals: Genre-based searches/ browsing II

- Biographic and historic accounts are good sources for past habituals.
- Most importantly, we have compiled lists of similar stories, plots and tropes, with significant overlap between corpora.
## Table of comparable texts

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Swamphen and Hawk</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>2</td>
</tr>
<tr>
<td>Child abducted by devil</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>2</td>
</tr>
<tr>
<td>Child defeats devil</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>2</td>
</tr>
<tr>
<td>Race between two animals</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>4</td>
</tr>
<tr>
<td>Rat and octopus/turtle</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>4</td>
</tr>
<tr>
<td>Two rocks arguing</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>2</td>
</tr>
<tr>
<td>Child of a snake</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>4</td>
</tr>
<tr>
<td>Human origin of coconut</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>4</td>
</tr>
<tr>
<td>Devil and Dove</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>Chicken and Megapode</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>2</td>
</tr>
<tr>
<td>Abduction of a magical girl</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>2</td>
</tr>
<tr>
<td><strong>Stories per language</strong></td>
<td><strong>7</strong></td>
<td><strong>3</strong></td>
<td><strong>7</strong></td>
<td><strong>8</strong></td>
<td><strong>5</strong></td>
<td></td>
</tr>
</tbody>
</table>
Near-minimal cross-linguistic pairs through comparable texts

(5) Daakaka, future habitual: Reduplication plus imperfective auxiliary

\[
\text{Ka } ya=p \quad du \quad es-esi \quad ngok \quad teenem \quad a \quad nye \quad ka \quad na \quad POT=\text{stay REDUP-see 2SG home and 1SG ASR 3PL=POT stay REDUP-see 2SG home and 1SG ASR}}\]

\[
w=i \quad ten \quad dóór \quad kyun.\]

POT=COP assigned.to dark.bush just

“[People] shall see you in the village and I, I will go to the bush.” (1347)

(6) Mavea, future habitual: imperfective auxiliary

\[
\text{me fut no me ko-lo-suruvu atano na me ko-lo-taua FUT then 2SG FUT 2SG-IMPF-sleep ground but FUT 2SG-IMPF-put otoli na \quad atano egg LOC ground}
\]

“You will sleep on the ground, and you will lay eggs on the ground.” (06016.060)
### Our tagset

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Tags</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clause type</td>
<td>clause</td>
<td>assertion, question, directive; embedded: proposition, conditional, e.question, temporal, adverbial, attributive</td>
</tr>
<tr>
<td>Temporal domain</td>
<td>time</td>
<td>past, future, present</td>
</tr>
<tr>
<td>Modal domain</td>
<td>mood</td>
<td>factual, counterfactual, possible</td>
</tr>
<tr>
<td>Aspectual domain</td>
<td>event</td>
<td>bounded, ongoing, repeated, stative</td>
</tr>
<tr>
<td>Polarity</td>
<td>polarity</td>
<td>positive, negative</td>
</tr>
</tbody>
</table>

**Table:** Tag set of the MelaTAMP project.
Habituals: TAMP tagging

- Prioritizing comparable subcorpora, we tagged individual texts for clause type, mood, temporal reference, aspect and polarity.

Daakaka, past habitual: ya=t3p=dist du stay tiye kill barar pig tevy-an side.of 3s poss na comp ya=t3p=dist du stay se hook tóó wild.cane "They used to kill a pig for the occasion of spear-throwing."
Habituals: TAMP tagging

- Prioritizing comparable subcorpora, we tagged individual texts for clause type, mood, temporal reference, aspect and polarity.

- For our study on habituals, the aspectual category: repeated was especially relevant.

Daakaka, past habitual: 
\[
yā=ð du stay tiiye barar tevy-an na comp ya=ð du stay se hóó wild.cane.
\]
"They used to kill a pig for the occasion of spear-throwing." (5210)

- The same tag also covers iterative contexts (The children were bouncing up and down on.)

- A documentation of our tag set can be found at https://wikis.hu-berlin.de/melatamp/Main_page.
Habituals: TAMP tagging

- Prioritizing comparable subcorpora, we tagged individual texts for clause type, mood, temporal reference, aspect and polarity.
- For our study on habituals, the aspectual category: repeated was especially relevant.

(7) Daakaka, past habitual:

\[ ya=t \text{ du tiye barar tevy-an na ya=t du} \]

\[ 3P=\text{DIST stay kill pig side.of-3s.POSS COMP 3P=DIST stay} \]

\[ se \text{ tôó} \]

hook wild.cane

“They used to kill a pig for the occasion of spear-throwing.” (5210)
Habituals: TAMP tagging

- Prioritizing comparable subcorpora, we tagged individual texts for clause type, mood, temporal reference, aspect and polarity.
- For our study on habituals, the aspectual category: repeated was especially relevant.

(7) Daakaka, past habitual:

\[
\begin{align*}
ya &= t & du & tiye & barar & tevy-an & na & ya &= t & du \\
3P &= \text{DIST} & \text{stay} & \text{kill} & \text{pig} & \text{side.of-3S.Poss} & \text{Comp} & 3P &= \text{DIST} & \text{stay} \\
se & & tóó & \text{hook} & \text{wild.cane} & \\
\end{align*}
\]

“They used to kill a pig for the occasion of spear-throwing.” (5210)

- The same tag also covers iterative contexts (The children were bouncing up and down on.)
Habituals: TAMP tagging

- Prioritizing comparable subcorpora, we tagged individual texts for clause type, mood, temporal reference, aspect and polarity.
- For our study on habituals, the aspectual category: repeated was especially relevant.

(7) Daakaka, past habitual:

\[ ya=t \ du \ tiye \ barar \ tevy-an \ na \ ya=t \ du \]
\[ 3P=\text{DIST} \ stay \ kill \ pig \ side.of-3s.Poss \ COMP \ 3P=\text{DIST} \ stay \]
\[ se \ \text{tóó} \]

hook wild.cane

“They used to kill a pig for the occasion of spear-throwing.” (5210)

- The same tag also covers iterative contexts (*The children were bouncing up and down on.*)
- A documentation of our tag set can be found at [https://wikis.hu-berlin.de/melatamp/Main_page](https://wikis.hu-berlin.de/melatamp/Main_page).
Inter-annotator consistency

(Tjuka et al., submitted)
Inter-annotator consistency in habitual contexts

- Inter-annotator consistency was particularly low for the repeated aspectual tag.

\begin{itemize}
  \item hinage also ta dup-paisowa "we work hard too" (Saliba: Tautolowaiya_01AG_0048)
  \item kamna-da feeling-1INCL.POSS te near.SP se 3PL.SBJ yababa "and we feel tired" (Saliba: Tautolowaiya_01AG_0049)
\end{itemize}
Inter-annotator consistency in habitual contexts

• Inter-annotator consistency was particularly low for the repeated aspectual tag.
• The main reason for this is that habituality can be a property of passages, which combines with a variety of clause-level aspects:
Inter-annotator consistency in habitual contexts

- Inter-annotator consistency was particularly low for the repeated aspectual tag.
- The main reason for this is that habituality can be a property of passages, which combines with a variety of clause-level aspects:
Inter-annotator consistency in habitual contexts

- Inter-annotator consistency was particularly low for the repeated aspectual tag.
- The main reason for this is that habituality can be a property of passages, which combines with a variety of clause-level aspects:

\[(8)\]

\[\text{a. hinage ta dup-paisowa} \]

also \[1\text{INCL.SBJ DUP-work}\]

“we work hard too” (Saliba: Tautolowaiya_01AG_0048)
Inter-annotator consistency in habitual contexts

- Inter-annotator consistency was particularly low for the repeated aspectual tag.
- The main reason for this is that habituality can be a property of passages, which combines with a variety of clause-level aspects:

(8)  

a. *hinage ta dup-paisowa*
    also 1INCL.SBJ DUP-work
    “we work hard too” (Saliba: Tautolowaiya_01AG_0048)

b. *kamna-da te se yababa*
    feeling-1INCL.POSS near.SP 3PL.SBJ bad
    “and we feel tired” (Saliba: Tautolowaiya_01AG_0049)
Generic passages

- The observation that habituality can be a property of stretches of discourse, not just of individual clauses, is not new: Carlson & Spejewski (1997):

  (9)  
  a. My grandmother used to bake the most wonderful pies every Saturday.
  b. She went to the orchard on Shady Lane early in the morning.
  b’. The alarm clock would have gone off at 6 a.m.
  c. She then would pick a basket each of apples and peaches.
  c’. Cows would be in the orchard mooing at her.

- Something similar can be said for a number of TAM contexts, including
  - modal subordination,
  - sequence of tense,
  - reported speech,
  - present-in-the-past;
Storyboards: Introduction

• Storyboards are picture stories such that each picture can be summarized by roughly one sentence.
Storyboards: Introduction

- Storyboards are picture stories such that each picture can be summarized by roughly one sentence.
- They are useful for the generation of entire texts ...
Storyboards: Introduction

• Storyboards are picture stories such that each picture can be summarized by roughly one sentence.
• They are useful for the generation of entire texts ...
• with very specific semantic targets;
Storyboards: Introduction

- Storyboards are picture stories such that each picture can be summarized by roughly one sentence.
- They are useful for the generation of entire texts ...
- with very specific semantic targets;
- For many speakers, a storyboard elicitation is a more meaningful task than, e.g. a judgement task.
Storyboards: Introduction

- Storyboards are picture stories such that each picture can be summarized by roughly one sentence.
- They are useful for the generation of entire texts ...
- with very specific semantic targets;
- For many speakers, a storyboard elicitation is a more meaningful task than, e.g. a judgement task.
- Crucially, storyboards can be used to create near-parallel corpora across speakers and languages.
Storyboards: Bill vs. the Weather I

(Vander Klok, 2013)
Storyboards: Bill vs. the Weather II
Storyboards: Bill vs. the Weather III
Storyboards: habitual contexts in Daakaka

(10)  *taem wuoswa ka  we mas yurmiline s-an  sye mwelili nyoo ...* time some  ASR POT NEC forget  CLF-3SG.POSS thing small.PL 3PL “sometimes he would always forget his small things [that he would take with him to his work].”

(11) *A mw=i vyanten swa na  mo kuowilye ka  we yurmiline sisye* and REAL=COP person  one COMP REAL know  COMP POT forget  things 3PL nyoo. “And he was a person who would sometimes forget his things.” (lit. “...who could forget his things.”)

(12) *Mwe vyan yan mili  gyes=an, bwe yurmiline kap, ki, sanglas, REAL go on place.of work=NMLZ REAL.CONT forget cup key sunglasses at, myane sus. hat with shoes “He went to his work place, he would forget his cup, keys, sunglasses, hat and shoes.”

(13) *...a mwe yur-yurmiline [...] suku-on  ane gyes=an  nyoo.* and REAL REDUP-forget  stuff.of-3SG.POSS TR work=NMLZ 3PL “and he repeatedly forgot his things for work.”
MelaTAMP storyboards
Resources

- Totem field storyboards (26): http://totemfieldstoryboards.org
- MelaTAMP storyboards (8): https://zenodo.org, search for "melatamp storyboards"
- Publications on our corpus methods:
  - Druskat & von Prince (accepted)
  - Tjuka et al. (submitted)
- Publications on storyboards:
  - Burton & Matthewson (2015)
  - von Prince et al. (2018)
  - von Prince et al. (2019)
  - von Prince & Margetts (submitted)
  - Krajinović (2018)
  - Krajinović (accepted)
  - Vander Klok (to appear)
Resources

- Totem field storyboards (26): http://totemfieldstoryboards.org
- MelaTAMP storyboards (8): https://zenodo.org, search for "melatamp storyboards"
- Publications on our corpus methods:
  - Druskat & von Prince (accepted)
  - Tjuka et al. (submitted)
- Publications on storyboards:
  - Burton & Matthewson (2015)
  - von Prince et al. (2018)
  - von Prince et al. (2019)
  - von Prince & Margetts (submitted)
  - Krajinović (2018)
  - Krajinović (accepted)
  - Vander Klok (to appear)

Thank you!


References II


Vander Klok, Jozina. to appear. Exploring the interaction of modality and temporality through the storyboard *Bill vs. the weather. Semantic Field Methods*.