

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 I

- 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 I

- 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 I

- 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
```

Example hit

- (4) Daakaka, generic:

mwe pwer-pwer yan leveno kyun te pwe en-en

REAL REDUP-stay on veno.tree just DISC CONT REDUP-eat

ar an

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

Story	Naf.	Sal.	Daa.	Dal.	Mav.	#
Swamphen and Hawk	+	-	-	+	-	2
Child abducted by devil	+	-	-	+	-	2
Child defeats devil	+	-	-	+	-	2
Race between two animals	+	+	+	+	-	4
Rat and octopus/turtle	+	-	+	+	+	4
Two rocks arguing	+	-	-	+	-	2
Child of a snake	-	+	+	+	+	4
Human origin of coconut	+	-	+	+	+	4
Devil and Dove	-	-	+	-	+	2
Chicken and Megapode	-	-	+	-	+	2
Abduction of a magical girl	-	+	+	-	-	2
Stories per language	7	3	7	8	5	

Near-minimal cross-linguistic pairs through comparable texts

- (5) 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)

- (6) 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)

Our tagset

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

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.

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.

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=DIST stay kill pig side.of-3S.POSS COMP 3P=DIST stay
se tóó
 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:

ya=*t* *du* tiye barar tevy-an na ya=*t* *du*
 3P=DIST stay kill pig side.of-3S.POSS COMP 3P=DIST stay
 se 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.*)

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.

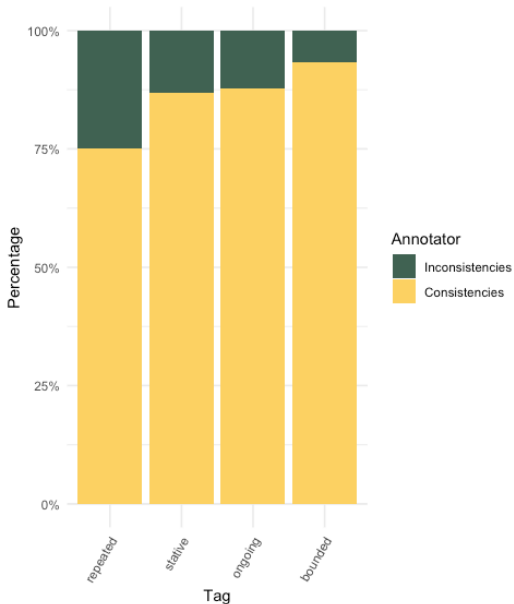
(7) Daakaka, past habitual:

ya=t du tiye barar tevy-an na ya=t du
 3P=DIST stay kill pig side.of-3S.POSS COMP 3P=DIST stay
se 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.

Inter-annotator consistency



(Tjuka et al., submitted)

Inter-annotator consistency in habitual contexts

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

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) a. *hinage ta* *dup-paisowa*
also 1INCL.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 **mooring** 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



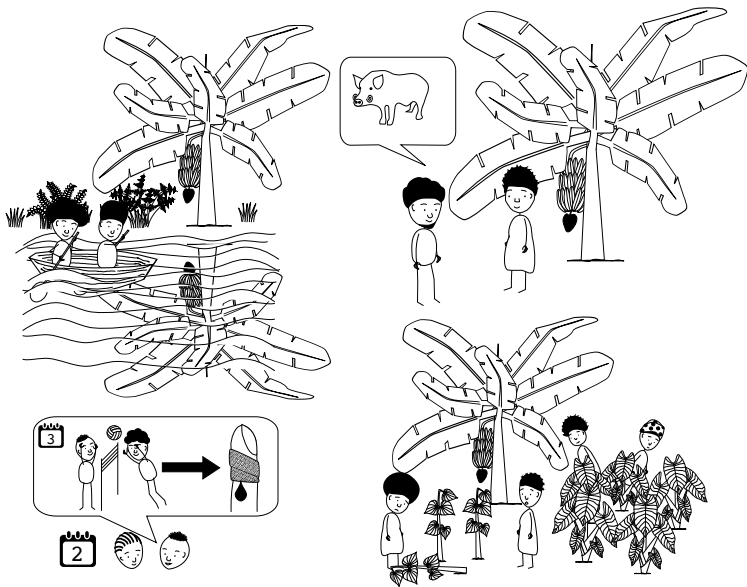
Storyboards: Bill vs. the Weather III



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
nyoo.
 3PL
 “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 I

- Burton, Strang & Lisa Matthewson. 2015. Targeted construction storyboards in semantic fieldwork. In Ryan Bochnak & Lisa Matthewson (eds.), *Methodologies in semantic fieldwork*, 135–156. Oxford University Press.
- Carlson, Greg N. & Beverly Spejewski. 1997. Generic passages. *Natural Language Semantics* 5(2). 101.
- Druskat, Stephan. 2018. ToolboxTextModules (Version 1.1.0). doi:10.5281/zenodo.1162207.
<https://doi.org/10.5281/zenodo.1162207>.
- Druskat, Stephan & Kilu von Prince. accepted. A software-driven workflow for the reuse of language documentation data in typological studies. In *Proceedings of the workshop on computational methods for endangered languages*, .

References II

- Krajinović, Ana. 2018. Comparative study of conditional clauses in Nafsan. In Brenda H. Boerger & Paul Unger (eds.), *SIL Language and Culture Documentation and Description 41 (Proceedings of COOL 10)*, <https://www.sil.org/resources/publications/entry/76827>.
- Krajinović, Ana. accepted. The semantics of perfect in Nafsan and implications for typology. *Proceedings of the semantics of African, Asian, and Austronesian languages (Triple A) 5, 27–29 June 2018, Konstanz* .
- Krause, Thomas & Amir Zeldes. 2016. ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1). 118–139.
- von Prince, Kilu, Ana Krajinović, Manfred Krifka, Valérie Guérin & Michael Franjeh. 2018. Mapping irreality: Storyboards for eliciting tam contexts. In Anja Gattnar, Robin Hörnig & Melanie Störzer (eds.), *Proceedings of linguistic evidence 2018*, .

References III

- von Prince, Kilu, Ana Krajinović, Anna Margetts, Valérie Guérin & Michael Franjeh. 2019. Habituals in Oceanic languages of Melanesia. *STUF - Language Typology and Universals* 72(1). 21–66.
- von Prince, Kilu & Anna Margetts. submitted. Expressing possibility in Saliba-Logea and Daakaka .
- Tjuka, Annika, Lena Weißmann & Kilu von Prince. submitted. Investigating habitual aspect in corpora from language documentation. submitted to LTC'19.
- Vander Klok, Jozina. 2013. Bill vs. the weather. *Totem Field Storyboards* Retrieved from <http://www.totemfieldstoryboards.org> on May 31, 2018.
- Vander Klok, Jozina. to appear. Exploring the interaction of modality and temporality through the storyboard *Bill vs. the weather*. *Semantic Field Methods* .