Habitual aspect as a property of text spans

Annika Tjuka, Lena Weißmann, and Kilu von Prince

May 19th, 2019
The 6th LTC Workshop on Less-Resourced Languages

Outline

The MelaTAMP project

Data

Method

Results of Inter-Annotator Agreement

Habituality

Conclusion

The MelaTAMP project

The MelaTAMP project

Figure 1: Subject languages of the MelaTAMP project.

The MelaTAMP project

- comparative research
- based on corpus data
- investigation of modality, aspect, tense, and polarity (TAMP) in Oceanic languages

The focus of this talk is on our study of habitual aspect.

Habituals in Oceanic

- (1)   a.   I teach at college.
  - b.   lions eat meat

- Question: How do Oceanic languages express habitual aspect?

Habituals in Oceanic

(1) a. I teach at college.
  b. lions eat meat

- Question: How do Oceanic languages express habitual aspect?

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

Examples

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

(3) *me   ro   nno me ko-lo-suruvu   atano na me*
     FUT then 2SG FUT 2SG-IMPF-sleep ground but FUT
     ko-lo-taua   otoli na atano
     2SG-IMPF-put egg LOC ground
     "You will sleep on the ground, and you will lay eggs on the
     ground." (Mavea 06016.060)

Data

Corpora

- Corpora of the following languages were considered in this study: Daakaka, Dalkalaen, Mavea, Nafsan, Saliba-Logea
- Texts were recorded during fieldwork sessions with speakers of the respective language.
- Annotation includes morpheme-by-morpheme glosses, part-of-speech-tags, translations into English, metadata on speakers, text genre, and the circumstances of the recording

Identifying comparable Texts

1. Translation-based searches for keywords: *used to/ would/ always/ usually/ often*
2. Genre-based searches with text-level data: "explanation nature", "story"
3. Biographic and historic accounts as a sources for past habituals

Identifying comparable Texts

1. Translation-based searches for keywords: *used to/ would/ always/ usually/ often*
2. Genre-based searches with text-level data: "explanation nature", "story"
3. Biographic and historic accounts as a sources for past habituals

Result: A list of 26 similar stories, plots and tropes with significant overlap between corpora.

Overview

	Total		Tagged	
Language	#Texts	#Tok.	#Texts	#Clauses
Daakaka	119	68k	5	141
Dalkalaen	114	34k	6	658
Mavea	61	45k	3	634
Nafsan	110	65k	6	363
Saliba-Logea	214	150k*	6	157
Total	618	362k	**26**	**1953**

Table 1: Corpora included in this study; Tok: tokens; tag.: tagged; *of the 150k tokens in this corpus, about 70k are fully annotated.

Method

Segmentation

- Prioritizing of the comparable subcorporus (26 texts)
- Segmentation of the texts into annotation units, which often correspond to sentences
- Further subdivision of these units into clauses for TAMP annotation (1953 clauses in total)

Tagging

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 2: Tag set of the MelaTAMP project, see
https://wikis.hu-berlin.de/melatamp/Main_page.

Tags in the Aspectual Domain

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

  (4)   *ya=t*   *du*   *tiye* *barar* *tevy-an*   *na*
         3P=DIST stay kill   pig   side.of-3S.POSS COMP
         ya=t   *du*   *se*   *tóó*
         3P=DIST stay hook wild.cane
         "They used to kill a pig for the occasion of
         spear-throwing." (Daakaka: 5210)

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

Tagging Example

(5)   *ya=t   du tiye barar tevy-an   na*
        3P=DIST stay kill pig   side.of-3S.POSS COMP
        ya=t   du se tóó
        3P=DIST stay hook wild.cane
        "They used to kill a pig for the occasion of spear-throwing."
        (Daakaka: 5210)

- clause: `assertion`
- time: `past`
- mood: `factual`
- event: `repeated`
- polarity: `positive`

Results of Inter-Annotator Agreement
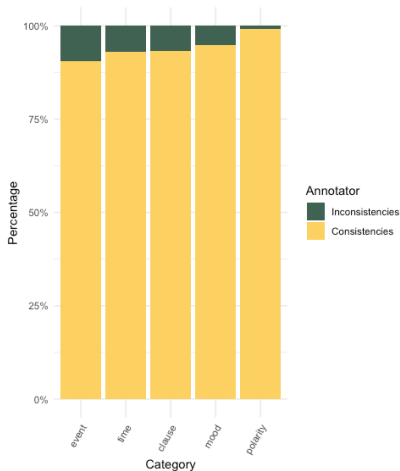
Results in each Category

Figure 2: Percentages of total inter-annotator consistencies (yellow) and inconsistencies (green) in each TAM category of the tag set.

Inter-Annotator Agreement Score for each Category

- Polarity: $\kappa = 0.91$
- Mood: $\kappa = 0.86$
- Clause: $\kappa = 0.85$
- Time: $\kappa = 0.85$
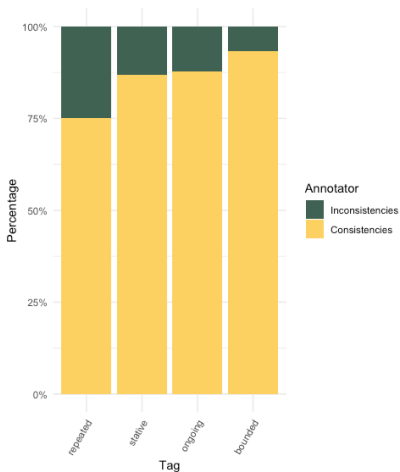- Event: $\kappa = \mathbf{0.79}$

Results in the Event Category

Figure 3: Percentages of total inter-annotator consistencies (yellow) and inconsistencies (green) in each tag of the event category.

Results in the Event Category

- 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.

Habituality

Observations

- A habitual narrative is described in consecutive sentences.
- Individual clauses within the passage differ with respect to their local aspectual values.

Observations

- A habitual narrative is described in consecutive sentences.
- Individual clauses within the passage differ with respect to their local aspectual values.

(6) 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)

Habituality as a Property of Text Passages

(7)   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.

(Carlson and Spejewski, 1997)

Theoretical Implications

- Habitual passages might just be one special case of a much more general situation:
  - modal subordination,
  - sequence of tense,
  - reported speech,
  - and present-in-the-past.

Theoretical Implications

- Habitual passages might just be one special case of a much more general situation:
    - modal subordination,
    - sequence of tense,
    - reported speech,
    - and present-in-the-past.

- Our experiences highlight that tense, aspect, and mood are properties that apply at clause-level, but also, to some extent, to larger passages, and that passage-level TAM values are partially independent from clause-level TAM values.

Implications for Tagging Aspect Categories

- Passage-wide properties need to be considered depending on the intended scope and degree of granularity.

Conclusion

Conclusion

- Tagging TAM categories gives insights into the use of aspectual categories for specific text genres.

Conclusion

- Tagging TAM categories gives insights into the use of aspectual categories for specific text genres.
- Difference between passage-level aspect and clause-level aspect can be found by comparing the annotation of comparable texts.

Conclusion

- Tagging TAM categories gives insights into the use of aspectual categories for specific text genres.
- Difference between passage-level aspect and clause-level aspect can be found by comparing the annotation of comparable texts.
- Habituality can span over sequences of several clauses.

Conclusion

- Tagging TAM categories gives insights into the use of aspectual categories for specific text genres.
- Difference between passage-level aspect and clause-level aspect can be found by comparing the annotation of comparable texts.
- Habituality can span over sequences of several clauses.

Thank you!

References

Carletta, Jean, 1996. Assessing agreement on classification tasks: the kappa statistic. Computational linguistics, 22(2):249–254.

Carlson, Greg N. and Beverly Spejewski, 1997. Generic passages. Natural Language Semantics, 5(2):101.

Dickey, Stephen M., 2000. Parameters of Slavic aspect: A cognitive approach. Center for the Study of Language and Information.

Druskat, Stephan, 2018. ToolboxTextModules (Version 1.1.0).

Franjieh, Michael, 2013. A documentation of North Ambrym, a language of Vanuatu. London: SOAS, ELAR.

Guérin, Valérie, 2006. Documentation of Mavea. London: SOAS, ELAR.

Guérin, Valérie, 2011. A grammar of Mavˇea: An Oceanic language of Vanuatu. Honolulu: University of Hawai'i Press.

Klecha, Peter, 2011. Optional and obligatory modal subordination. In Proceedings of Sinn und Bedeutung, volume 15.

Krause, Thomas and Amir Zeldes, 2016. ANNIS3: A new architecture for generic corpus query and visualization. Digital Scholarship in the Humanities, 31(1):118–139.

Krifka, Manfred, 2013. Daakie, The Language Archive. Nijmegen: MPI for Psycholinguistics.

Margetts, Anna, Andrew Margetts, and Carmen Dawuda, 2017. Saliba/Logea. The Language Archive.

MelaTAMP, 2017. Primary data repository – MelaTAMP. https://wikis.hu-berlin.de/melatamp.

von Prince, Kilu, 2013a. Daakaka, The Language Archive. Nijmegen: MPI for Psycholinguistics.

von Prince, Kilu, 2013b. Dalkalaen, The Language Archive. Nijmegen: MPI for Psycholinguistics.

von Prince, Kilu, Ana Krajinovic, Manfred Krifka, Valérie Guérin, and Michael Franjieh, 2018. Mapping Irreality: Storyboards for Eliciting TAM contexts. In Anja Gattnar, Robin Hörnig, and Melanie Störzer (eds.), Proceedings of Linguistic Evidence 2018.

von Prince, Kilu and Anna Margetts, to appear. Expressing possibility in Saliba-Logea and Daakaka. Studies in Language.

Roberts, Craige, 1989. Modal subordination and pronominal anaphora in discourse. Linguistics and Philosophy, 12:683–721.

Thieberger, Nick, 2006. Dictionary and texts in South Efate. Digital collection managed by PARADISEC.