Habitual aspect as a property of text spans

Annika Tjuka*, Lena Weißmann†, Kilu von Prince*

*Humboldt-Universität zu Berlin
Unter den Linden 6, 10099 Berlin
{tjukanni, kilu.von.prince}@hu-berlin.de

†Freie Universität Berlin
Kaiserswerther Str. 16-18, 14195 Berlin
lena.weissmann@fu-berlin.de

Abstract

The Oceanic languages of Melanesia are generally small, low-resource languages, of which very little primary data is available. For our study on tense, aspect, and modality (TAM), we have access to richly annotated corpora from seven endangered Oceanic languages. In this paper, we describe some of the methods we used to investigate the category of habitual aspect in these languages. We show that information can be recovered from the English translations and from metadata on genres. For a more in-depth study, we relied on clause-based tags labeling clause type, tense, aspect, mood, and polarity. The process of tagging aspect, in particular, revealed the theoretically and practically important fact that habituality is sometimes a property of larger spans of texts (Carlson and Spejewski, 1997), rather than just a property of clauses, and can combine with more specific clause-level aspect.

1. Introduction

Our understanding of the Oceanic languages of Melanesia has so far been based mostly on descriptive accounts rather than primary data, since no corpora and recordings existed until recently. For some of these languages, high-quality corpora have now become available, but their exploration is still in its infancy. Our comparative study on tense, aspect, and modality (TAM) in Oceanic Languages is based on such corpora of five Oceanic languages of Vanuatu and Papua New Guinea: Daakaka, Dalkalaen, Mavea, Nafsan, Saliba-Logea. Their speaker populations ranges from about 30 (Mavea) to around 5000 (Nafsan). TAM-related meanings are often expressed obligatorily within the verbal complex, sometimes in more than one place. Thus, Mavea has three pre-verbal slots for expressing TAM values; in addition, some subject-agreement markers also express the difference between realis and irrealis modalities and reduplication can be used to express pluractionality (see Table 1). By contrast, Saliba-Logea only uses optional particles to express TAM-related meanings.

In this paper, we discuss one of several studies of the MelaTAMP project (von Prince et al., 2018; von Prince and Margetts, to appear), with a focus on habitual aspect.

The combination of three data types enabled us to identify habitual contexts and compare them across our subject languages: the pre-existing metadata on content and genre, which we unified for better comparability; the also pre-existing English translations, and clause-based tags labeling TAM categories. On the basis of the metadata on genre, content, and English translations, we selected a subset of comparable texts. To these we then assigned clause-based tags labeling tense, aspect, and modality categories. This assignation was largely based on the English translations. In the following, we will report on this process.

In particular, we will focus on the process of tagging aspect: in checking inter-annotator consistency, we found that habitual contexts are particularly prone to mismatches. As a closer analysis revealed, the reason for this is apparently that habituality can be a property of longer stretches of text, which can combine with a variety of clause-level aspects. This phenomenon seems to be independent of the language under investigation because it can also be observed in English (Carlson and Spejewski, 1997).

In fact, a tagging system optimized for the detection of habituality may require an additional structural unit between the text level and the clause level. The same may hold for hypothetically, reported speech, and narrative past. Our findings also have important implications for our understanding of habituality as a frame.

2. Data and methods

The data of our project consists of corpora of seven languages in total, five of which were taken into account for this study.[1] Each corpus contains a variety of texts which were recorded during fieldwork sessions with speakers of the respective language. They are richly annotated, with morpheme-by-morpheme glosses, pos tags, translations into English, as well as metadata on speakers, text genre, and the circumstances of the recording. In addition, we enriched parts of the corpora with our own tag set for TAM values. For optimal facilities for searching and analysis, we imported all corpora to the ANNIS platform (Krause and Zeldes, 2016). We used Pepper modules (Druskat, 2018) to import the texts from their native SIL Toolbox format.

---

[1]The corpora of the MelaTAMP project are held and versioned in a git repository (MelaTAMP, 2017). The repository itself is private and currently only accessible by members of the project team. Published versions of each corpus are available (von Prince, 2013a; von Prince, 2013b; Krifka, 2013; Guérin, 2006; Thieberger, 2006; Franjieh, 2013; Margetts et al., 2017).

SBJ.AGR	COND	NEG	IT/INCPT	NUM	IMPF	REDUP-	**Verb**	ADV	TR	OBJ
i-, ...	*mo-*	*sopo-*	*me-/pete-*	*r-/tol-*	*l(o)-*				*=i*	*=a*/NP

Table 1: The verbal complex in Mavea (Guérin, 2011).

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

2.1. Searching genres and translations

In an initial stage of exploration, we looked for keywords in the English translations that would be indicative of habitual contexts such as "used to/ would/ always/ usually/ often". We then checked whether the verbal morphology we observed in the matches also occurred in other aspectual contexts.

In addition, we utilized pre-existing metadata on genres and content to identify texts that were likely to include habitual descriptions. For example, texts that were tagged as belonging to the genre `explanation` often feature descriptions of wild-life behaviour and thus typical generic statements, such as *the banded rail lives on the ground*.

We also tagged a subset of our corpora for clause-level TAM values, and it is this part of the process that we focus on in this paper.

2.2. TAM tagging

In order to identify maximally comparable contexts across corpora, we selected the texts to be tagged according to their content: certain stories and themes are widespread throughout the region and were present in more than one corpus, such as stories about the origin of the coconut. We thus created a sub-corpus of 26 texts in five of our seven subject languages (see Table 2).

Our choice of stories may have introduced some bias to the range of TAM references we tagged. At the same time, this study is contextualized by more exploratory corpus work beyond the TAM-tagged sub-corpus, detailed studies on individual languages, and additional field work in Vanuatu, which makes us fairly confident that our findings can be generalized to a wider range of texts and genres.

Each text was segmented into annotation units, which often correspond to sentences. For TAM annotation, we further subdivided these units into clauses. In the 26 parallel texts, each clause was annotated for clause type, temporal reference, modal reference, aspect, and polarity. The following analysis of the inter-annotator consistency is based on these clause-based tags (1953 clauses in total). An example of one such clause is given in (1).

(1) *Meerin   yaapu   nyoo ya=m   du*
long.time big.man 3PL 3PL=REAL stay
"Long ago, there were great men" (Daakaka: 1388)

Each clause was tagged separately by two annotators: Annika Tjuka and Lena Weißmann. Both annotators are linguistics students training for their MA degree and fluent in English. The tag set applied consists of five categories with a total of 21 tags (see Table 3).

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 3: Tag set of the MelaTAMP project.

The tagging was mainly based on the English translation of the texts although in some cases, the glosses were considered as well. The example in (1) was tagged as follows:
- clause: `assertion`
- time: `past`
- mood: `factual`
- event: `stative`
- polarity: `positive`

After a text was tagged by both annotators, the tags of each clause were compared by one of the two annotators. Inconsistencies were noted in a table. If the decision for either one of the tags was clear, the correct tag was inserted in the final document. Doubtful cases were discussed with Kilu von Prince, who is one of two principal investigators of the project. The numbers of inconsistencies in the following discussion include both the uncontroversial and the doubtful cases.

3. Tag set and inter-annotator consistency

The analysis of the inter-annotator consistency revealed that in 817 tags, mismatches between the two annotators were present (including typos and other easily resolved mismatches). Figure 1 illustrates the inter-annotator consistency and inconsistency in each category of the tag set.

The graph shows that in the overall set of tagged texts, the inconsistencies were never higher than 11% of the total
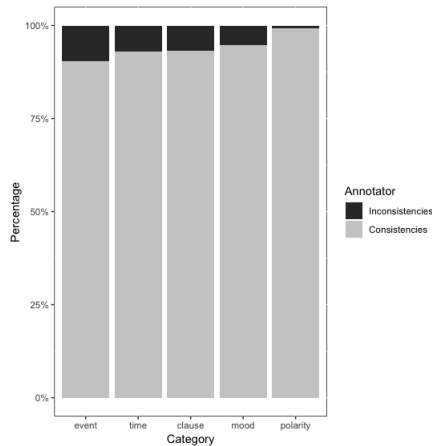
Figure 1: Percentages of total inter-annotator consistencies (light) and inconsistencies (dark) in each TAM category of the tag set.

tags in each TAM category. The category with the lowest inconsistency percentage is polarity, with 0.79% inconsistencies. This translates to an inter-annotator agreement score of $\kappa = 0.91$, which corresponds to an almost perfect reliability (Carletta, 1996). Mismatches occurred, for example, in sentences with negative verbs, as in (2).

(2) *Melee mwe   danga*
    food   REAL not.exist
    "There is no food."   (Dalkalaen: 0817)

In the categories mood, clause, and time, the inconsistency rates were considerably higher. For example, the inconsistency percentage in the mood category is 5.52% with $\kappa = 0.85$. In this category, the inter-annotator inconsistency is mainly caused by differences between the tags `factual` and `possible` in temporal sentences, such as (3). The point in question is the relative time reference which was either annotated as `future` or `past` by one of the annotators. Note that the factual domain comprises the actual present and past, whereas the `possible` tag comprises several domains, depending on the temporal reference of the clause.

(3) *Panpan na   rapak   nagis*
    until   PURP 3D.RS=to point
    "Until they got to the point [. . .]" (Nafsan: 036.017)

The clause category has an inter-annotator agreement score of $\kappa = 0.854$. The 7.15% percentage of inconsistencies in the clause category are largely caused by differences in tagging adverbial sentences which were wrongly tagged as `proposition`, for example in (4). A clause is tagged as `adverbial` if it modifies the event or proposition expressed by the main clause. On the other hand, if an embedded clause functions as the argument of a subordinating verb, such as *want*, *think*, *say* or *know*, it is marked as `proposition`.

(4) *Masisipe ka   te   vyan yen or   ten*
    NAME   MOD DIST go   in   bush assigned.to

    na   tilya mees me
    COMP take food come
    "When Masisipe went to the bush **to get food**"
    (Daakaka: 5933)

Note that sentences like (4) are a special type of adverbial clauses: they indicate the purpose or goal of an action and can be called purpose clauses as well. To account for this distinction, we additionally tagged those embedded clauses with a keyword: `PURPOSE`.

The inter-annotator agreement score in the time category is $\kappa = 0.862$.

The category with the highest percentage of inter-annotator inconsistency is the event category (10.55% inconsistencies). These inconsistencies are reflected in the lowest inter-annotator agreement score: $\kappa = 0.79$. The values associated with event structure are `repeated`, `stative`, `ongoing`, and `bounded`. All descriptions of events that involve multiple instances of the same action are tagged as `repeated`. The distinctions into `bounded`, `ongoing`, and `stative` only apply to descriptions of individual events or actions. A detailed analysis of the inconsistencies in each tag of this category is given in the following section.

3.1. Mismatches in annotating habitual contexts

The comparison of the percentage of inconsistencies between each TAM category of the tag set in Figure 1 showed that the number of mismatches is greatest in the event category. A detailed analysis of this category reveals that the percentage of inter-annotator consistency in each tag (`bounded`, `ongoing`, `repeated`, `stative`) differs. Figure 2 illustrates the percentage of inter-annotator consistency and inconsistency in each tag of the event category.
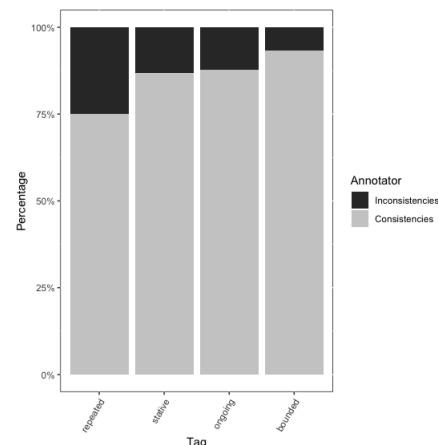

Figure 2: Percentages of total inter-annotator consistencies (light) and inconsistencies (dark) in each tag of the event category.

The `bounded` tag has the smallest percentage of inconsistency with 7.25% of a total of 1283 clauses which were marked as `bounded`. The tag `bounded` is used to refer to episodic, non-stative event descriptions of the past or future that do not refer to the internal structure of the

event. The inconsistencies result mainly from verbs which were assigned the `stative` tag by one of the two annotators. These verbs include, e.g., *sleep*, *try*, *agree*, *disappear*.

Based on a total of 158 clauses the `ongoing` tag has the second highest percentage of inconsistency (13.92% inconsistencies). Event descriptions which refer to i) the dynamics of the event's internal development, ii) episodic descriptions of the present moment, and ii) the background event for a new development in a narrative are labeled as `ongoing`. The differences in the annotation stem mostly from tagging the aspect of a sentence as `bounded` instead of `ongoing`, see for example (5).

(5) *temeli swa mwe pwer usili tes*
child one REAL stay follow sea
"the boy was following the shore" (Daakaka: 3722)

The `ongoing` tag is closely followed by the `stative` tag with 15.12% inter-annotator inconsistency in 430 clauses. If an event is marked as `stative`, its description implies little if any internal event dynamics. As in the `bounded` tag, most inconsistencies come from verbs which are tagged as `bounded` instead of `stative`, for example, *stay*, *feel*, *know*, *live*.

By far the highest inconsistency percentage compared with the other tags in the event category is found in the `repeated` tag. In our annotation, the label `repeated` extends to cases where the same action is performed by the same actor, either on the same occasion (iterative aspect) or on various occasions (habitual). Non-stative generic descriptions, as in *Stegosaurus ate plants* also fall into this domain. The mismatches in the annotation amount to 33.33% in a total of 102 clauses.

Although only a few clauses are tagged as `repeated`, the error rate is highest. Most inconsistencies in this tag result from passages which describe a habitual context, as in (6). While the context of the *passage* is habitual, individual clauses within the passage might differ with respect to their local aspectual values. Thus, in (6), the first part includes a bounded event description, while the second part is stative.

(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)

Therefore, this tag is the source of many inconsistencies. In the case of habituality, each sentence does not necessarily stand alone but is embedded in a context which spans over two or more sentences. Sometimes the habituality of a paragraph is indicated by an initial phrase, for example, *they had [the following] habits*. When entire passages express a habitual description, these passages are often in irrealis mood, as opposed to individual habitual sentences, which are typically in realis mood.

4. Discussion

In this paper, we explored the tagging of TAM categories in corpora of five Oceanic languages. Our data is based on parallel texts of a specific genre which were enriched with additional tags for TAM categories. These tags enabled us to compare certain contexts in the languages with different TAM systems. The focus of this paper was on the process of tagging aspect.

After identifying parallel texts in the corpora, each clause of the selected text was tagged manually by two annotators. In the process of comparing the tag of each clause, mismatches were noted. The result of the inter-annotator consistency analysis showed that in a total of 1953 clauses, the annotation of the annotators differed in 817 tags. This result indicates that our tag set seems to be suitable for tagging TAM categories. In particular, the categories mood and polarity have a very low error rate which shows that they can be applied accurately through larger texts.

However, in some categories, the inconsistencies seem to occur more often than in others. Most mismatches appear in the event category with its tags `bounded`, `ongoing`, `stative`, and `repeated`. The detailed analysis of this category showed that the `repeated` tag is prone to inter-annotator inconsistencies. Due to the internal structure of passages which describe habitual contexts, the `repeated` tag might not be exhaustive.

The difficulties of tagging this event category indicate that habituality is sometimes a property of passages, or stretches of discourse, rather than a property of individual clauses, and that it can combine with a variety of clausal aspects. A system optimized for tagging aspect might therefore have to include an additional annotation layer that can span a sequence of sentences within the same text, as opposed to the binary distinction between text-level and clause-level tags we utilized in our study.

The fact that habituality can be a property of passages has been previously discussed and it has been argued that the property of genericity or habituality ranges over an entire stretch of discourse (Carlson and Spejewski, 1997). Typically, the generic passage is initiated with a summary sentence and the remainder of the discourse is related to a series of events in a temporal order (episodic sequence). The individual sentences and clauses within this sequence can have a variety of aspectual and temporal properties: they can express events prior to the sequence as well as states of affairs or progressive event descriptions. These properties are illustrated in the following example (Carlson and Spejewski, 1997, 132):

(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.

The sentences in the following episodic sequence are com-

monly marked with *used to* and *would* and the temporal relation between them is indicated with adverbs, for example, *then*, *next*, and *after that*. However, the reference time of each generic status is included in the temporal scope (Carlson and Spejewski, 1997). The authors conclude that generic passages should be analyzed as sequences of generic sentences which obtain a individual reference time.

Other accounts observe the interaction between a wider habitual context and more local aspectual values (Dickey, 2000). They demonstrate that Slavic languages differ in whether they allow for perfective verb forms in habitual contexts.

However, these observations have not yet found their way into much of the theoretical literature on aspect. Our findings indicate that they might be far more central to the nature of habituality and genericity than previously acknowledged. Moreover, it is well possible that habitual passages are just one special case of a much more general situation: there are several well-documented cases in which mood, aspect or tense are a property of entire passages or texts that combine with partially independent, local TAM values at the clause level. These include, for example, modal subordination: *A wolf might come in here. It would eat you first* (Roberts, 1989; Klecha, 2011). Indirect speech in German *Konjunktiv I* is another example:

(8) *Sie   sagte, die Ergebnisse **seien**   erfreulich.*
    3SG.F said   the results   be.KONJI pleasing
    *Es **habe**   außergewöhnlich viele*
    it have.KONJI exceptionally   many
    Einreichungen gegeben.
    submissions   give.PARTPERF
    "She said the results were pleasing. There had been [according to her] exceptionally many submissions."

Other candidates for passage-wide TAM are the present-in-the-past (*It is the year 1990. The cold war is coming to an end.*) and sequence-of-tense phenomena (*Esra was determined to win the race. She would train every day.*).

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. In practice, this means that corpus studies that assign TAM tags to text units may have to consider including an additional unit between the text level and the clause level or adjusting their tag sets to allow for combinations of passage-level and clause-level TAM values.

5. Conclusion

Tagging TAM categories in small, low-resource languages is a crucial process to get a general idea of the use of aspectual categories for specific text genres, and to compare the expression of aspect across languages. The comparison of the annotation of parallel texts in corpora in different languages highlights the difference between passage-level aspect and clause-level aspect. In some cases, habituality can span over sequences of several clauses which indicates that it can be a property of passages, not only of individual clauses.

Nevertheless, the overall tag set we use to annotate the TAM categories exhibits a high percentage of inter-annotator consistency throughout different categories. Our tag set has thus proven useful for our purposes but may have to be adjusted for more fine-grained studies, or studies with a different focus.

6. 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 Maˇvea: 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 Krajinović, 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.