

Tagging tense and modality in Oceanic

Kilu von Prince
Humboldt-Universität zu Berlin

2. Juni 2017
FEAST Saarbrücken

Structure of the talk

① Background

The MelaTAMP project
Motivation

② Modal meanings

Categorizing modal meanings
Modal flavors

③ Problems

TAM in Oceanic

④ Solutions

Linking tense and modality

The MelaTAMP project

We (Manfred Krifka, Ana Krajinović, Stephan Druskat, two SAs and me) investigate

- expressions of **tense**, **aspect**, **mood** and **polarity**
- in seven **Oceanic** languages of **Melanesia**
- based on **corpus data**
- supplemented by elicitations (mostly using storyboards)

Tense, aspect, mood and polarity

TAMP in Daakaka:

- (1) ***mwe** poo yan lee te **to** esi sye tuswa*
 REAL climb at tree DISC NEG.REAL see something one
 “he climbed a tree but didn’t see a thing” (1480)

The main question is: Which semantic properties determine the distribution of TAMP markers in the subject languages?

Oceanic languages of Melanesia



Describing TMA categories

Little is known about TMA categories in Oceanic languages.

Dahl (2007)

“Even in the case of comparatively well-described languages, which constitute a small minority, the information found in reference grammars and more specialized publications tends to be insufficient and often misleading. This is in particular the case for grammatical categories such as tense, mood, aspect, number, definiteness, case etc.,[...]”

Modal force and modal flavor

- (2) a. Visitors **may** leave their shoes on.
- b. The train **may** have been late.

- (3) a. Visitors **must** take their shoes off.
- b. The train **must** have been late.

Modal force and modal flavor

- (2) a. Visitors **may** leave their shoes on. ┌ possibility
b. The train **may** have been late. └
- (3) a. Visitors **must** take their shoes off. ┌ necessity
b. The train **must** have been late. └

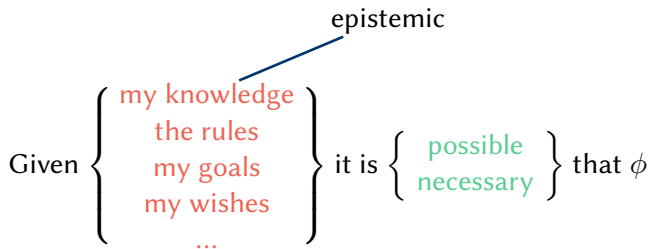
Modal force and modal flavor

- (2) a. Visitors **may** leave their shoes on.]
b. The train **may** have been late.] deontic
- (3) a. Visitors **must** take their shoes off.]
b. The train **must** have been late.] epistemic

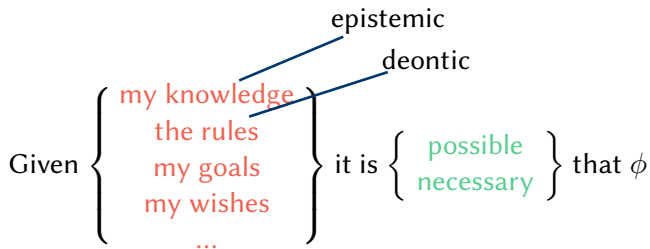
Modal **force** and modal **flavor**

Given $\left\{ \begin{array}{l} \text{my knowledge} \\ \text{the rules} \\ \text{my goals} \\ \text{my wishes} \\ \dots \end{array} \right\}$ it is $\left\{ \begin{array}{l} \text{possible} \\ \text{necessary} \end{array} \right\}$ that ϕ

Modal force and modal flavor



Modal force and modal flavor



Tagging modality

Previous work on tagging modal senses has focused on the disambiguation between modal flavors.¹

Ruppenhofer & Rehbein (2012) assume the following senses for English modals:

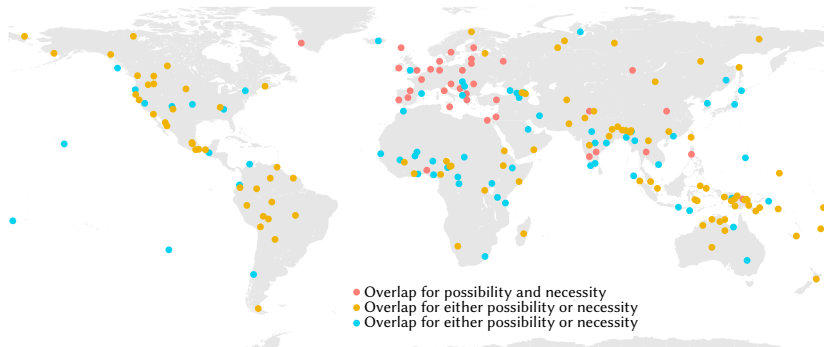
	<i>can</i>	<i>may</i>	<i>must</i>	<i>ought</i>	<i>shall</i>
epistemic	+	+	+	+	+
deontic	+	+	+	+	+
dynamic	+	-	-	-	-
optative	-	+	-	-	-
concessive	-	+	-	-	-
conditional	-	-	-	-	+

¹See also Rubinstein *et al.* (2013), Zhou *et al.* (2015) Baker *et al.* (2012) and Hacquard & Wellwood (2012)

Towards an ontology of modal flavors: probably not gonna happen



Differentiation of modal flavors cross-linguistically



According to van der Auwera & Ammann (2013), most languages outside of Europe do not conflate epistemic and deontic modality (see also Viebahn & Vetter 2016). However, that is probably a function of how well the languages are studied.

Interim summary

- Modal meanings are analyzed in terms of modal force and modal flavor.
- The basic tags used in tagging modal meanings represent flavor distinctions.
- The theoretical and empirical bases for flavor distinctions are shaky (if you ask me).

Problems with tagging modal flavors in Oceanic

- So far, efforts to tag modal flavors have been haunted by relatively low inter-annotator consistency. In examples like the following, annotators find it hard to decide between epistemic and non-epistemic readings:

Problems with tagging modal flavors in Oceanic

- So far, efforts to tag modal flavors have been haunted by relatively low inter-annotator consistency. In examples like the following, annotators find it hard to decide between epistemic and non-epistemic readings:
 - (4) It might rain tomorrow. (Given what I know? Given the weather conditions?)

Problems with tagging modal flavors in Oceanic

- So far, efforts to tag modal flavors have been haunted by relatively low inter-annotator consistency. In examples like the following, annotators find it hard to decide between epistemic and non-epistemic readings:
 - (4) It might rain tomorrow. (Given what I know? Given the weather conditions?)
- The paradigmatic expressions for which the tagsets were designed are English modal auxiliaries. But many languages do not have modal auxiliaries or any corresponding expressions.

Problems with tagging modal flavors in Oceanic

- So far, efforts to tag modal flavors have been haunted by relatively low inter-annotator consistency. In examples like the following, annotators find it hard to decide between epistemic and non-epistemic readings:
 - (4) It might rain tomorrow. (Given what I know? Given the weather conditions?)
- The paradigmatic expressions for which the tagsets were designed are English modal auxiliaries. But many languages do not have modal auxiliaries or any corresponding expressions.
- Like in English, TAM expressions in our subject languages do not appear to specify modal flavors.

Tense, aspect, mood and polarity

TAMP in Daakaka:

- (5) ***mwe** poo yan lee te **to** esi sye tuswa*
 REAL climb at tree DISC NEG.REAL see something one
 “he climbed a tree but didn’t see a thing” (1480)

The Daakaka TAM paradigm

	enclitic	proclitic	monosyllabic
Pos. Realis	<i>=m</i>	<i>mw=</i>	<i>mwe/mV</i>
Neg. Realis			<i>to</i>
Pos. Potential	<i>=p</i>	<i>w=</i>	<i>wV</i>
Neg. Potential	<i>=n</i>		<i>nV</i>
Distal	<i>=t</i>	<i>t=</i>	<i>tV</i>
Open Polarity			<i>doo</i>
Change of State			<i>bwet</i>

Table: List of TMA markers in Daakaka

The Daakaka potential

Possibility can simply be expressed by the potential marker, which is also used for assertions about the future.

(6) *barvinye swa ka we luk teve-sye m-ada em*
 grass one ASR POT grow side.of-3s.POSS CL2-1D.IN.POSS house
 “a grass will grow next to our house” (2523)

(7) *me yan bat-en, bat-en ka wa pe-pyo*
 come on head-3s.POSS head-3s.POSS ASR POT REDUP-WHITE
 (vyen)
 probably
 “and its head, its head is white I think” (6113)

The Daakaka distal

Counterfactual conditionals

- (8) [Ka **we/ te** eli buluwu wa ge myane tomo] te **tu**
 SUB.CON POT/ DIST dig hole POT like with rat CONJ DIST
vu.
 good
 ‘If he had dug a hole like the rat, it would have been good.’

Discontinuous past:

- (9) [meu=*an na nenyu* **te** melumlum], melumlum, a
 live=NM ATT yesterday DIST quiet quiet but
*meu=*an na doma mwe yas**
 live=NM ATT today REAL hard
 ‘the life of the past was easy, it was easy, but the life of today is hard’ (4011)

Summary

Classifying modal meanings in terms of flavors has the following problems:

- Theoretically, the distinctions are not very well motivated.

Summary

Classifying modal meanings in terms of flavors has the following problems:

- Theoretically, the distinctions are not very well motivated.
- Labels proliferate, the relations between flavors are unclear.

Summary

Classifying modal meanings in terms of flavors has the following problems:

- Theoretically, the distinctions are not very well motivated.
- Labels proliferate, the relations between flavors are unclear.
- Even for languages like English, inter-coder reliability in classifying flavors is moderate at best.

Summary

Classifying modal meanings in terms of flavors has the following problems:

- Theoretically, the distinctions are not very well motivated.
- Labels proliferate, the relations between flavors are unclear.
- Even for languages like English, inter-coder reliability in classifying flavors is moderate at best.
- In our subject languages, differences in flavors do not determine differences in marking.

Branching times

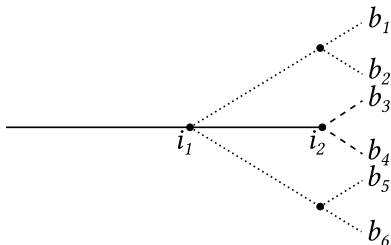


Figure: solid: the actual relative to i_2 ; dashed: the possible relative to i_2 ; dotted: the counterfactual relative to i_2

In contrast to previous work (Thomason & Gupta, 1980; Tedeschi, 1981; Placek & Müller, 2007; Ippolito, 2003, 2013), I assume that quantification over branches/ histories is not restricted to those branches that pass through the actual present i_c .

Pinning down differences between potential and distal mood

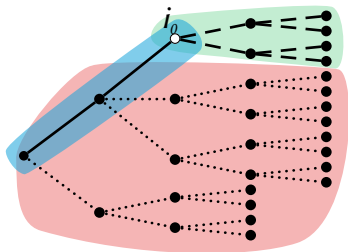


Figure: red: distal; green: potential; blue: realis;

Our tagset

Categories	Name	Values
Clause type	clause	(main): assertion (default), question, directive, other; embedded: proposition, e.question, conditional, temporal, adverbial, attributive
Temporal domain	time	past, future, present
Modal domain	mood	factual, counterfactual, possible
Event structure	event	bounded, ongoing, repeated, stative, cos (change-of-state)
Polarity	polarity	positive, negative

Applying the tags

```

\ref .0978
\tx ma wese webung wi sii, wi vyer
    ma gete, mu kukyu
\mb ma wese webung w- i sii, w- i vyer
    ma ge -te, mu ku-kyu
\ge REAL enough day POT- COP three POT- COP four REAL
    like -MEDIAL REAL REDUP-SURROUND
\ps tam v n tam- cop num tam- cop num
    tam v -*** v.pre v.tr
\ft it might have been three or four days, it surrounded
    him
\unitref 1 2 13
\clause 1 proposition
\time 1 present
\mood 1 possible
\event 1 stative
\polarity 1 positive
\keywords 1 EPISTEMIC

```


Preliminary results

- We have not yet calculated reliable measures of inter-coder agreement.

Preliminary results

- We have not yet calculated reliable measures of inter-coder agreement.
- But informal, percentage-based assessments inspire our confidence that the modal and temporal tags are straightforward to assign.

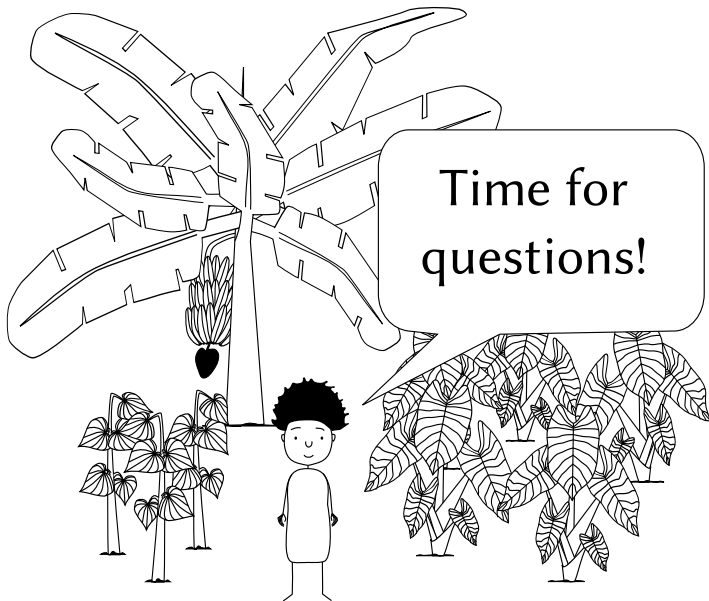
Preliminary results

- We have not yet calculated reliable measures of inter-coder agreement.
- But informal, percentage-based assessments inspire our confidence that the modal and temporal tags are straightforward to assign.
- So far, we have primarily used the tagset in order to identify contexts that are rare or unattested in the corpora.

Preliminary results

- We have not yet calculated reliable measures of inter-coder agreement.
- But informal, percentage-based assessments inspire our confidence that the modal and temporal tags are straightforward to assign.
- So far, we have primarily used the tagset in order to identify contexts that are rare or unattested in the corpora.
- We are presently developing elicitation materials to fill the existing gaps during fieldwork.

Linking tense and modality



No differentiation between modal flavors: Daakaka *mas*

- (10) *kyun te basée swa mwe ka ka wo mas i maa kyun*
 just DISC bird one REAL say ASR POT must COP dove just
 “then one bird said: ‘It must have been the dove.’” (4669)
- (11) *Ma ka: ‘Ko=∅ mas ane’*
 REAL say 2S=POT must eat
 “He said: ‘You have to eat it.’” (6202)

Flavors expressed by *mas*

	epistemic	root
<i>mas</i>	✓	✓

No differentiation between modal flavors: Daakaka *wese*, “be sufficient”

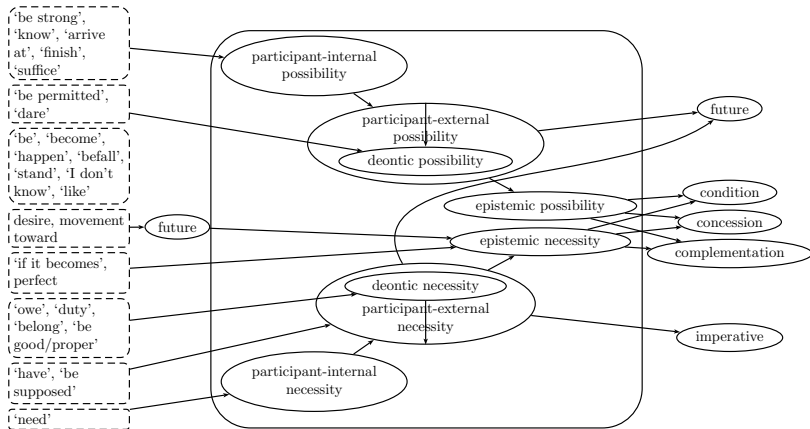
- (12) *tomo mwe vyan ongame mwe wese na ka lisepsep wa*
 rat REAL go hear REAL enough COMP MOD lisepsep POT
sikya
 touch
 “The rat felt that the lisepsep would be able to touch him.”
 (4557)

- (13) *ma wese webung w=i sii, w=i vyer ma*
 REAL enough day POT=COP three POT=COP four REAL
ge=te, mu ku-kyu
 like=MED REAL REDUP-surround
 “it might have been three or four days, [the lava] surrounded
 him” (0978)

Flavors expressed by *wese*

	force	epistemic	root
<i>mas</i>	necessity	✓	✓
<i>wese</i>	possibility	✓	✓

Modality's semantic map



(van der Auwera & Plungian, 1998)

Eliciting TAM expressions

What we now is mostly based on translated examples and grammatical descriptions.

Dahl (1985: 50)

“If what we have said here is true, one may ask how anyone could ever be so stupid as to choose translations as a basis for an investigation of language use. The simple answer is that it is the only realistic method for large-scale data collection in typologically oriented linguistic research.”

This concern is also relevant to new work based on parallel Bible corpora (Asgari & Schütze, 2017).

References I

- Asgari, E., & Schütze, H. 2017. Past, Present, Future: A Computational Investigation of the Typology of Tense in 1000 Languages. *Arxiv e-prints*, Apr.
- van der Auwera, Johan, & Plungian, Vladimir A. 1998. Modality's semantic map. *Linguistic typology*, **2**(1), 79–124.
- Baker, Kathryn, Bloodgood, Michael, Dorr, Bonnie J., Callison-Burch, Chris, Filardo, Nathaniel W., Piatko, Christine, Levin, Lori, & Miller, Scott. 2012. Use of modality and negation in semantically-informed syntactic MT. *Computational linguistics*, **38**(2), 1–28.
- Dahl, Östen. 1985. *Tense and aspect systems*. Blackwell.

References II

- Dahl, Östen. 2007. From questionnaires to parallel corpora in typology. *Sprachtypologische Universalienforschung*, **60**, 172–181.
- Hacquard, Valentine, & Wellwood, Alexis. 2012. Embedding epistemic modals in English: A corpus-based study. *Semantics and pragmatics*, **5**, 1–29.
- Ippolito, M. 2003. Presuppositions and implicatures in counterfactuals. *Natural Language Semantics*, **11**(2), 145–186.
- Ippolito, Michela. 2013. *Subjunctive conditionals: a linguistic analysis*. Linguistic Inquiry Monograph, vol. 65. Massachusetts Institute of Technology.
- Placek, Tomasz, & Müller, Thomas. 2007. Counterfactuals and Historical Possibility. *Synthese*, **154**(2), 173–197.

References III

- Rubinstein, Aynat, Harner, Hillary, Krawczyk, Elizabeth, Simonson, Daniel, Katz, Graham, & Portner, Paul. 2013. Toward fine-grained annotation of modality in text. *In: Proceedings of the 10th international conference on computational semantics (iwcs), workshop on annotation of modal meanings.*
- Ruppenhofer, Josep, & Rehbein, Ines. 2012. Yes, we can!?. *In: Proceedings of the eight international conference on language resources and evaluation (lrec)*, vol. 12.
- Tedeschi, Philip. 1981. Some evidence fo a branching-futures semantic model. *In: Tedeschi, Philip, & Zaenen, Annie (eds), Tense and aspect.* Syntax and Semantics, no. 14. New York: Academic Press.

References IV

- Thomason, Richmond, & Gupta, Anil. 1980. A theory of conditionals in the context of branching time. *The philosophical review*, **89**(1), 65–90.
- van der Auwera, Johan, & Ammann, Andreas. 2013. *Overlap between situational and epistemic modal marking*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Viebahn, Emanuel, & Vetter, Barbara. 2016. How many meanings for ‘may’? the case for modal polysemy. *Philosopher’s imprint*, **16**(10), 2–26.
- Zhou, Mengfei, Frank, Anette, Friedrich, Annemarie, & Palmer, Alexis. 2015. Semantically enriched models for modal sense classification. *Pages 44–53 of: Proceedings of the emnlp 2015 workshop on linking models of lexical, sentential and discourse-level semantics*.