

How dense can you get?  
Correlates of information density as factors in  
cross-linguistic comparison

Kilu von Prince

April 11, 2018  
Saarbrücken

## Cross-linguistic variation in maximum density



The simple, unselfish joy that something good has happened to someone else

## Variation in expressions of possibility

- (1) *nye na=m kuowilye [ka na=p vyan tiye*  
1SG 1SG=REAL know COMP 1SG=POT go kill  
*suw-uk kyun]*  
REF.PRON-1SG.POSS just  
“but I **can** beat him by myself”

## The upshot

- Expressions that are functionally equivalent across languages may differ significantly in terms of their **complexity**.

## The upshot

- Expressions that are functionally equivalent across languages may differ significantly in terms of their complexity.
- In typological surveys about modal expressions, multi-clausal expressions have typically been excluded.

## The upshot

- Expressions that are functionally equivalent across languages may differ significantly in terms of their complexity.
- In typological surveys about modal expressions, multi-clausal expressions have typically been excluded.
- This exclusion has not been addressed explicitly and theoretically motivated.

## The upshot

- Expressions that are functionally equivalent across languages may differ significantly in terms of their complexity.
- In typological surveys about modal expressions, multi-clausal expressions have typically been excluded.
- This exclusion has not been addressed explicitly and theoretically motivated.
- Instead, implicit assumptions about **correlates** of complexity were probably the motivation for their exclusion.

# The upshot

- Expressions that are functionally equivalent across languages may differ significantly in terms of their complexity.
- In typological surveys about modal expressions, multi-clausal expressions have typically been excluded.
- This exclusion has not been addressed explicitly and theoretically motivated.
- Instead, implicit assumptions about correlates of complexity were probably the motivation for their exclusion.
- In this talk, I will discuss these implicit assumptions and explore how the notion of **grammaticalization** relates to the idea of **information density**.



## Expressions of possibility

(2) Martha  $\left\{ \begin{array}{l} \text{can} \\ \text{may} \\ \text{might} \\ \text{could} \end{array} \right\}$  win the race.

(3) It is  $\left\{ \begin{array}{l} \text{possible} \\ \text{likely} \\ \text{probable} \\ \text{conceivable} \\ \text{imaginable} \\ \text{thinkable} \\ \vdots \end{array} \right\}$  that Martha will win the race.

# The scope of typological studies

Structures such as *it is possible that* are systematically excluded from typological studies on modal expressions.

	mood	AFF	PART	AUX	clt	ADV	per.	der.	tag	case	noun	sub.
1	+	+	+	+	+	-	+	+	-	-	-	-
2	+	+	+	+	+	-	-	-	-	-	-	-
3	+	+	+	+	+	-	+	+	-	-	-	-
4	+	+	+	+	-	+	-	-	+	+	-	-
5	+	+	+	+	-	+	+	-	-	-	+	?

**Table:** Modal expressions discussed in different studies; per.: periphrastic; der.: derivation; sub.: subordinating verbal structures; clt: clitic

1: Bybee *et al.* (1994)

2: Palmer (2001)

3: Hengeveld (2004)

4: de Haan (2006)

5: Narrog (2016)

## Excluding multi-clausal structures

(4) It is  $\left\{ \begin{array}{l} \text{possible} \\ \text{likely} \\ \text{probable} \\ \text{conceivable} \\ \text{imaginable} \\ \text{thinkable} \\ \vdots \end{array} \right\}$  that Martha will win the race.

- The exclusion of multi-clausal structures is **not explicitly addressed** in most typological accounts.

## Excluding multi-clausal structures

(4) It is  $\left\{ \begin{array}{l} \text{possible} \\ \text{likely} \\ \text{probable} \\ \text{conceivable} \\ \text{imaginable} \\ \text{thinkable} \\ \vdots \end{array} \right\}$  that Martha will win the race.

- The exclusion of multi-clausal structures is not explicitly addressed in most typological accounts.
- Presumably, the rationale behind this is that multi-clausal structures are not sufficiently **grammaticalized**.

## Excluding multi-clausal structures

(4) It is  $\left\{ \begin{array}{l} \text{possible} \\ \text{likely} \\ \text{probable} \\ \text{conceivable} \\ \text{imaginable} \\ \text{thinkable} \\ \vdots \end{array} \right\}$  that Martha will win the race.

- The exclusion of multi-clausal structures is not explicitly addressed in most typological accounts.
- Presumably, the rationale behind this is that multi-clausal structures are not sufficiently grammaticalized.
- This rationale however, relies on additional assumptions about **correlates** of structural complexity.

# The correlates of structural complexity

- Structural **complexity** correlates with

# The correlates of structural complexity

- Structural complexity correlates with
  - the size/ openness of the set of paradigmatic **alternatives**,

# The correlates of structural complexity

- Structural complexity correlates with
  - the size/ openness of the set of paradigmatic alternatives,
  - semantic **specificity**



# The correlates of structural complexity

- Structural complexity correlates with
  - the size/ openness of the set of paradigmatic alternatives,
  - semantic specificity
  - and **frequency**.

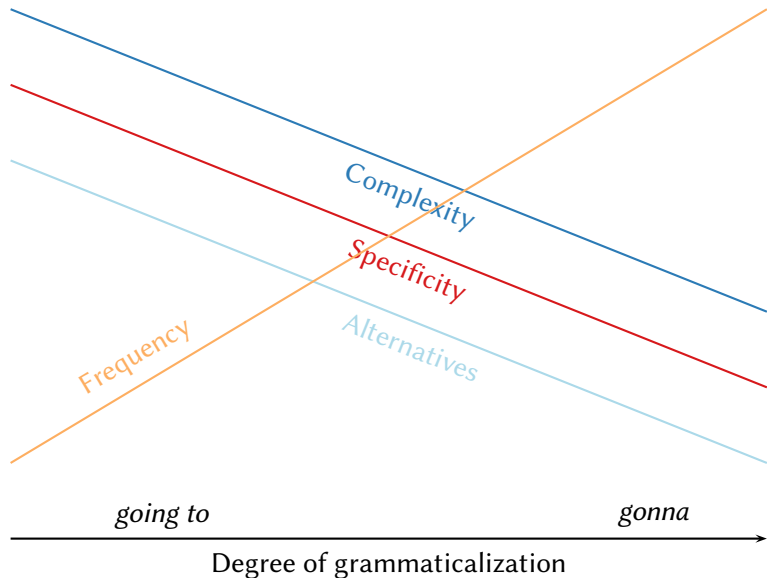
# The correlates of structural complexity

- Structural complexity correlates with
  - the size/ openness of the set of paradigmatic alternatives,
  - semantic specificity
  - and frequency.
- I would like to refer to the combination of these features as the **grammaticalization profile**.

# The correlates of structural complexity

- Structural complexity correlates with
  - the size/ openness of the set of paradigmatic alternatives,
  - semantic specificity
  - and frequency.
- I would like to refer to the combination of these features as the grammaticalization profile.
- At the same time, all these properties are also correlates of **information density**.

# Grammaticalization



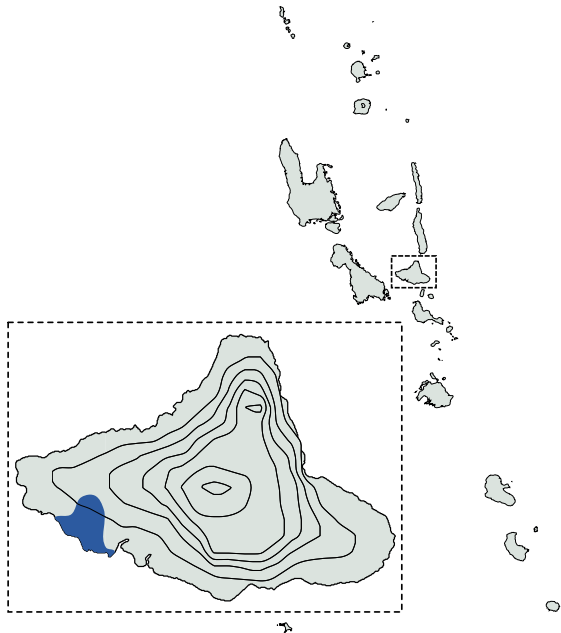
## Grammaticalization profiles

(5) Martha  $\left\{ \begin{array}{l} \text{can} \\ \text{may} \\ \text{might} \\ \text{could} \end{array} \right\}$  win the race.

(6) It is  $\left\{ \begin{array}{l} \text{possible} \\ \text{likely} \\ \text{probable} \\ \text{conceivable} \\ \text{imaginable} \\ \text{thinkable} \\ \vdots \end{array} \right\}$  that Martha will win the race.

Expression	Complexity	Alternatives	Specificity	Frequency
<i>can</i>	low	few	low	high
<i>possible</i>	high	many	high	low

# A perspective from Ambrym, Vanuatu



## Expressing possibility in Daakaka: *kuowilye* “know”

- (7) *nye na=m kuowilye [ka na=p vyan tiye*  
1SG 1SG=REAL know COMP 1SG=POT go kill  
*suw-uk kyun]*  
REF.PRON-1SG.POSS just  
“but I **can** beat him by myself”

## Expressing possibility in Daakaka: *kuowilye* “know”

- (7) *nye na=m kuowilye [ka na=p vyan tiye*  
1SG 1SG=REAL know COMP 1SG=POT go kill  
*suw-uk kyun]*  
REF.PRON-1SG.POSS just  
“but I **can** beat him by myself”

Expresson	Complexity	Alternatives	Specificity	Frequency
<i>kuowilye</i>	high	few	low	high



## *Kuowilye*: set of alternatives

(8) *nye na=m*  $\left\{ \begin{array}{l} \textit{kuowilye} \\ \textit{wese} \end{array} \right\}$  [*ka na=p vyan tiye suw-uk kyun*]  
“but I **can** beat him by myself”

## *Kuowilye*: specificity

- (9) *Mwe meu mo kuowilye ka wa sikya dom ves?*  
REAL live REAL know MOD POT touch year how.much  
“How long can it live?” (lit. “it lives it can reach how many years”)

## *Kuowilye*: specificity

- (9) *Mwe meu mo kuowilye ka wa sikya dom ves?*  
REAL live REAL know MOD POT touch year how.much  
“How long can it live?” (lit. “it lives it can reach how many years”)

- (10)[...] *te mo kuowilye [ka we pyos-pyos ane]*  
DISC REAL know MOD POT REDUP-joke TRANS  
“then he can joke with her.”

## *Kuowilye*: frequency

Table: Frequencies in the spoken BNC

---

can_{V}	3556.63/million
possible [for,that,to]	35.64/million

---

## *Kuowilye*: frequency

Table: Frequencies in the spoken BNC

---

can_{V}	3556.63/million
possible [for,that,to]	35.64/million

---

Table: Query to the Daakaka corpus in ANNIS

---

mb="kuowilye" & mb="ka" & #1 . #2 1303.25/million

---

## Comparing grammaticalization profiles

Expresson	Complexity	Alternatives	Specificity	Frequency
<i>can</i>	low	few	low	high
<i>possible</i>	high	many	high	low

Expresson	Complexity	Alternatives	Specificity	Frequency
<i>kuowilye</i>	high	few	low	high

# Comparing grammaticalization profiles

Expresson	Complexity	Alternatives	Specificity	Frequency
<i>can</i>	low	few	low	high
<i>possible</i>	high	many	high	low

Expresson	Complexity	Alternatives	Specificity	Frequency
<i>kuowilye</i>	high	few	low	high

## Comparing grammaticalization profiles

- The Daakaka expression *kuowilye* requires a much more complex structure than English *can*.



## Comparing grammaticalization profiles

- The Daakaka expression *kuowilye* requires a much more complex structure than English *can*.
- But all other dimensions of its grammaticalization profile are much more similar to *can* than to *possible*.

## Comparing grammaticalization profiles

- The Daakaka expression *kuowilye* requires a much more complex structure than English *can*.
- But all other dimensions of its grammaticalization profile are much more similar to *can* than to *possible*.
- For typological studies on grammaticalization paths, semantic properties of modal expressions and many other purposes, we would therefore want to include expressions like *kuowilye*.

## Comparing grammaticalization profiles

- The Daakaka expression *kuowilye* requires a much more complex structure than English *can*.
- But all other dimensions of its grammaticalization profile are much more similar to *can* than to *possible*.
- For typological studies on grammaticalization paths, semantic properties of modal expressions and many other purposes, we would therefore want to include expressions like *kuowilye*.
- Suggestion: when defining **comparative concepts** such as *modal expression*, we might want to look for the most highly **grammaticalized** (densest?) expression available in a given language.

Thank you!

## Expressing possibility in Daakaka: potential mood

(11) *Ka w=i Ros o ka w=i Yokon.*

ASR POT=COP Ros or ASR POT=COP Yokon

“It **might** be Rose or it **might** be Yokon.”

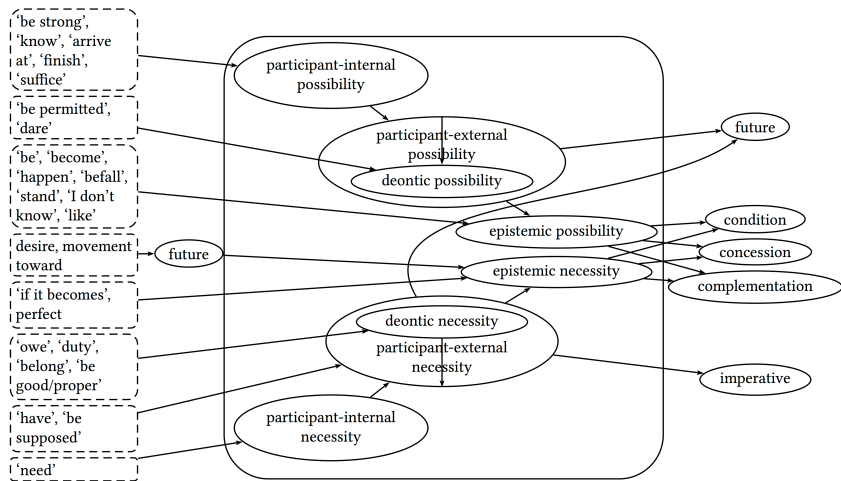
# The TAM system in Daakaka

	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>=∅</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>

SBJ.AGR	(=)TAMP	(AUX)	(REDUP-)	Verb
<i>na, ...</i>	= <i>m, ...</i>	<i>du,pwer</i>	...	...

- (12) *Ko=p tas we!*  
 2SG=POT sit first  
 “Sit down please!”

# Grammaticalization of modal expressions: Bybee *et al.* (1994)



The figure is from van der Auwera & Plungian (1998).

## References I

- van der Auwera, Johan, & Plungian, Vladimir A. 1998. Modality's semantic map. *Linguistic typology*, 2(1), 79–124.
- Bybee, J. L., Perkins, Revere, & Pagliuca, W. 1994. *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. The University of Chicago Press.
- de Haan, Ferdinand. 2006. Typological approaches to modality. In: Frawley, William (ed), *The expression of modality. The Expression of Cognitive Categories*, vol. 1. Berlin, New York: Mouton de Gruyter.
- Hengeveld, K. 2004. Illocution, mood and modality. *Pages 1190–1201 of: Booij, G., Lehmann, C., & Mugdan, J. (eds), Morphology: A handbook on inflection and word-formation*. Berlin/New York: Mouton de Gruyter.



## References II

- Narrog, Heiko. 2016. The expression of non-epistemic modal categories. *Pages 89–116 of: The Oxford Handbook of Mood and Modality*. Oxford University Press.
- Palmer, F.R. 2001. *Mood and modality*. 2 edn. Cambridge University Press.