



Makonde ujamaa

http://www.africanoracle.org/a-1_PAIDM_website/OPPTY_Catalog-2.html

Mtu ni watu
(Swahili proverb)
'A person is people'



ARC CENTRE OF EXCELLENCE FOR
THE DYNAMICS OF LANGUAGE

The said and the unsaid in social cognition: the design logic of SCOPIC, a parallax corpus

Nicholas Evans & Danielle Barth

Nicholas.Evans@anu.edu.au

Danielle.Barth@anu.edu.au

CoEDL / ANU

DGfS Workshop

Corpus-Based Typology: Spoken Language from a
Cross-Linguistic Perspective
Hamburg



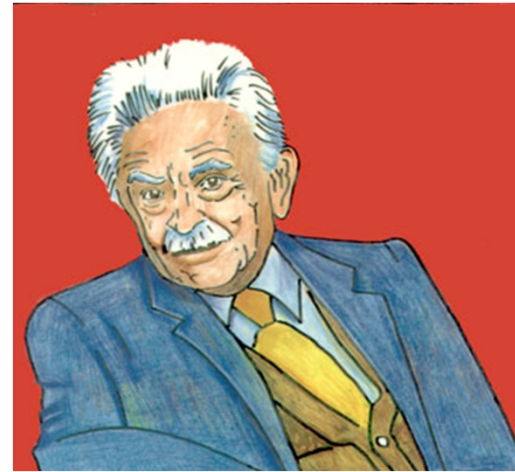
Diversity of the unsaid



Cada pueblo calla unas cosas 'para' poder decir otras. Porque 'todo' seria indecible.

'Each people stays silent about some things in order to say others. Because everything would be unsayable.'

Ortega y Gasset



Jede Sprache hat ihr eigenes Schweigen.
'Every language keeps it own silences.'

Elias Canetti

But corpus studies tend to focus on the said, not the unsaid....

Canetti, Elias. 1942-3. *Aufzeichnungen*.

Ortega y Gasset, José. 1957. *Miseria y Esplendor de la Traducción*. Madrid.



What is social cognition?



social cognition: 'the sum of those processes that allow individuals ... to interact with one another' (Frith & Frith 2007)

This must take into account both

- (a) social relationships (e.g. father of, ingroup, etc. – influencing *formulation of reference*)
- (b) psychological states (e.g. belief, desire, attention, emotional state), including those of the interlocutors with respect to the unfolding discourse – influencing *epistemic framing*



Distributing the unsaid across languages: a word in Dalabon



Wekewarrkahmolkkûndokan

‘They snuck around.’

Or

‘I’m afraid that the two of them, who are in odd-numbered generations with respect to one another, might be sneaking around (i.e. going around unbeknownst to someone who should know); by choosing this form of words, I hereby indicate that one of those I am referring to is a mother-in-law’s brother or comparable relative.’

We-ke-warrkah-molkkûn-doka-n

APPRehensive-they.twoDISHARMONIC-wrong.place-unbeknownst-go[respect.form]-NonPast

Now attention to social relations is being expressed in two places: by the choice of ‘disharmonic’ pronominal prefix *ke-* (as opposed to *barra-* for ‘harmonic’ relationships) and the use of the *drebuyno* root *doka* for ‘go’ (appropriate when talking to, or about, certain kinds of in-law). Note also three other categories relevant to social cognition – apprehensive *we-*, the ‘wrong place’ prefix *warrkah-*, and the ‘unbeknownst’ prefix *molkkûn-* – which we will return to later



Parallel corpora for comparing what gets coded across languages

Some existing methods:

Parallel elicitation, questionnaires

Disadvantage: artificial data

Translation as source of parallel texts (e.g. comparison of Bible translations, Tintin comics etc.)

Disadvantage: distortion by emphases and structures of original language)

Common stimulus set (e.g. Nijmegen-style pictures or videos; Pear Story film or Frog Story pictures)

Disadvantage: speakers are reactive in their discourse structures, following prompts in the order given by the investigator)



Parallel problem-solving

The Social Cognition Parallax Interview Corpus (SCOPIC)



Edited by Danielle Barth & Nicholas Evans



Language Documentation & Conservation Special Publication No. 12

Do speakers choose to say different things in different languages?

Examples:

- Map task
- Director-matching task
- **Family Problems Story (> SCOPIC Project)**

Goal: to generate broadly parallel discourse ('parallax corpus'), including both narrative and dialogic elements, while leaving speakers free to produce spontaneous material

Narrative problem-solving encourages naturalistic speech, and the task design lets people encode whatever social-cognition relevant categories they choose

San Roque, Lila, Alan Rumsey, Lauren Gawne, Stef Spronck, Darja Hoenigman, Alice Carroll, Julia Miller & Nicholas Evans. 2012. Getting the story straight: language fieldwork using a narrative problem-solving task. *Language Documentation and Conservation* 6:134-173.



Four parts of the task



1. Describing Pictures
2. Ordering Pictures
3. Telling story to new participant
 - a. Telling Story in third person
 - b. Telling Story in first person



SCOPIC languages



The SCOPIC Project: language selection



Danielle Barth
SCOPIC Co-leader

- | | |
|--|--|
| <ol style="list-style-type: none"> 1. Amharic (Semitic, Ethiopia, M. Amberber & H. Woldemariam) 2. Auslan (Signed Language, Australia, G. Hodge) 3. Avatime (Kwa, Ghana, S. van Putten) 4. Balinese (Austronesian, Indonesia, W. Arka) 5. Bislama (Vanuatu Creole, S. Schnell) 6. Dalabon (Gunwinyguan, Australia, N. Evans) 7. Duna (Trans-New Guinean, New Guinea Highlands, L. San Roque) 8. English (Indo-European, Australia, B. Kelly) 9. German (Indo-European, Germany, A. Schalley) 10. Hoocąk (Siouan, North America, I. Hartmann x) 11. Idi (Pahoturi River, PNG, V. Gast, O. Tykhostup) 12. Japanese (Altaic, H. Narrog) | <ol style="list-style-type: none"> 12. Khalkha-Mongol (Mongolian, Mongolia, D. Guntsetseg & E. Skribnik) 13. Kogi (Arwako-Chibchan, Columbia, H. Bergqvist x & D. Knuchel) 14. Komnzo (Yam, Southern New Guinea, C. Döhler x) 15. Kriol (Australian Creole, G. Dickson) 16. Ku Waru (Trans-New Guinea, Southern Highlands, A. Rumsey) 17. Matukar Panau (Oceanic, North New Guinea, D. Barth) 18. Murrinhpatha (Southern Daly, Australia, J. Mansfield) 19. Sanzhi Dargwa (Nakh-Daghestanian, Caucasus, D. Forker) 20. Sherpa (Tibeto-Burman, Nepal, B. Kelly) 21. Tok Pisin (Melanesian creole, PNG, D. Barth) 22. Vera'a (Oceanic, Vanuatu, S. Schnell) 23. Yurakaré (isolate, Bolivia, S. Gipper) |
|--|--|

More recent additions include Ilokano (Yukinori Kimoto), |Gui (Hitomi Ono), Indonesian (Asako Shiohara, Yanti), Sibe (Norikazu Kogura), Jinghpaw (Keita Kurabe), Korean (Seongha Rhee) through an affiliate project at TUFS (Tokyo University of Foreign Studies)



Using SCOPIC for inter-language comparability



- Organized along functional domains
 - Definitions of cross-linguistic categories are primarily based on semantic/functional grounds, but then tied to morphosyntactically-specified realisations in particular languages (e.g. benefactive as applicative in one language, case choice in another)
 - Distinguishing between categories intralinguistically may be along language-specific lines (i.e. Indirect Speech v. Direct Speech v. Mixed-Speech)
- We have domain specific categories (Tags) that are discussed and debated in in-person meetings to capture the real uses of these phenomena in each language in the sample
- For particular analyses, we may re-group and scale up categories, i.e. possessed v. non-possessed human referents



Language Specificity

- Tags are created to try to capture specific language categories where reasonable and where there is expectation that another language could (have) also use(d) that category
- Keeping track of and adding in language specific information:
 - In each annotation, we include a citation form-like language specific term
 - For each annotation, a researcher can add a note where there is a usage that requires more comment
 - For each domain in each language, we make a list of the unique tag-term combination and give information about those categories



Cross-linguistic tags for coding human referents (sample)

How to formulate reference?
Many choices are available...

KN_: Kin or other relational noun for close human relationship
e.g. KN_dad

PKN_: Possessed kin (or other close relations) noun
e.g. PKN_3s.wife

PGF_: a possessed family group term
PGF_friend.3s for 'his friends'

GN_: Generic noun
GN_people GN_man

RN_: Social Role (restrictive category), should indicate some kind of job or role in society
RN_policeman

DES_: one-word descriptor used to characterize people based on stable characteristics
DES_gringo DES_fat.one

GD_: a dyad group, based on internal relations within group like father-son, mother-children
GD_Liebepaar GD_father.son.pair

GF_: of kin / comparable social group like a family, a team, group of friends,
GF_family.of.three

GA_: a group characterized by a salient member and then an associated plural or dual
GA_father.with GA_Henrik.and.co GA_hiroko.tachi

GE_: group with members expressed exhaustively like boys-girls
GE_boys.girls GE_men.women



Cross-linguistic tags for coding human referents (sample)

The screenshot displays the ELAN software interface for the file 'SocCog-mjk08-sd_kir_ru_3b_20170330.eaf'. The interface includes a menu bar, a toolbar with playback controls, and a main workspace with several tracks. The top track shows the audio waveform. Below it, the transcription tracks are as follows:

Track Name	Text
English [8]	han ilo girek ti tanong
Rebecca [28]	matan wai o
English 2 [19]	han tamat brun sa panindo tamago o?
Card [34]	we don't know what he is thinking
Reported Speech [70]	about money or...
RS Notes [29]	her husband is angry with her (giving her anger) and she is crying or?
Referents [14]	CE
HumRef Notes [7]	CPF_nong
PP [41]	EPD_buru.sa.pan
PP Notes [19]	QPM_tam
Ben [14]	NEG, PD
Ben Notes [8]	PKN_tamat.3
Stance [21]	TYD_nong
St Notes [8]	FNI_buru.sa
	BNI_tam
	PD, 1p.incl
	SOMK_tamat.3s.buru.sa.pan
	VCUB_ti.tanong
	NEG, ta- 1.pl.in



For each domain in each language, we have metadata about each unique TAG-term combination

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	TAG_term_combination	English_Translation	Term_Notes														
2	DES_dabok	big	when used as referring expression, means 'big one(s)'														
3	DES_dino	old	when used as referring expression, means 'old one(s)'														
4	DES_tibud	white person	can also mean white or Western people generally														
5	ELAB		elaborated noun phrase, some examples in text are relative clauses to describe doctors and nurses, as in 'the w														
6	GE_aim.aipainim	boys and girls	compound of the noun aim 'boy' and aipain 'girl' modified with the special plural marker -im, as a compound n														
7	GE_aipainim.aim	girls and boys	compound of the noun aipain 'girl' with the special plural marker -im and aim 'boy', as a compound means 'chi														
8	GE_painim.aipainim	women and children	compound of the nouns pain 'woman' inflected with the special plural marker -im and aipain also inflected with														
9	GF_femili	family	Tok Pisin word for 'family'														
10	GF_lain	group	Tok Pisin word meaning group, used for a group of friends, or relatives, or clan members, etc.														
11	GF_turanud	friends	a term for a group of friends, grammatically plural														
12	GN_aim	boy(s)	the generic term for 'boy', when possessed indirectly it means 'son'														
13	GN_aim.natun	little boy(s)	a generic two-word term meaning 'little boy' or 'young boy'														
14	GN_aipain	girl(s)	the generic term for 'girl', when possessed indirectly it means 'daughter'														
15	GN_aipainim	children	morphemes are aipain 'girl' and -im a special plural marker that only occurs on pain and aipain, means 'children'														
16	GN_natun.pain	girl(s)	a two-word term that literally means 'small woman' but more accurately means 'girl', the adjective natun here is														
17	GN_pain	woman/women	the generic term for 'woman' or 'women', when possessed indirectly it means 'wife'														
18	GN_pain.aipainim	woman and children	compound of the noun pain 'woman' and aipainim 'children', often means 'family'														
19	GN_painim	women	the generic term for 'woman' with the (optional) special plural marker -im, if possessed it would mean 'wife'														



Dimensions of (sub)corpus comparison

4 basic dimensions of comparison:

Across languages

Across speakers

Across scenes/configurations

Across task phases

(Others are possible, e.g. between mixed-sex vs same-sex dyads, peer vs age-asymmetric dyads, equal-status vs unequal-status dyads etc. So far we have not gathered enough material to do this for any language



Referring to persons: the influence of language



Referring to persons: choices in formulation

Kogi

hěki hate-dweba hěki a-skwá

DEM=SW **grandfather-old** DEM=SW **3SG.POSS-son**

ezhi a-hwäsgwi hálde=ki ahí munzhi

or **3SG.POSS-father.in.law** DEM=SW **3SG.POSS** **woman/wife**

‘This is the grandfather. This is his [the old man’s] son. Or his [the young man’s] father in law. This one is his [the young man’s] wife.’ SocCog_kog01-CNC_130619_1 - 00:00:11-00:00:18

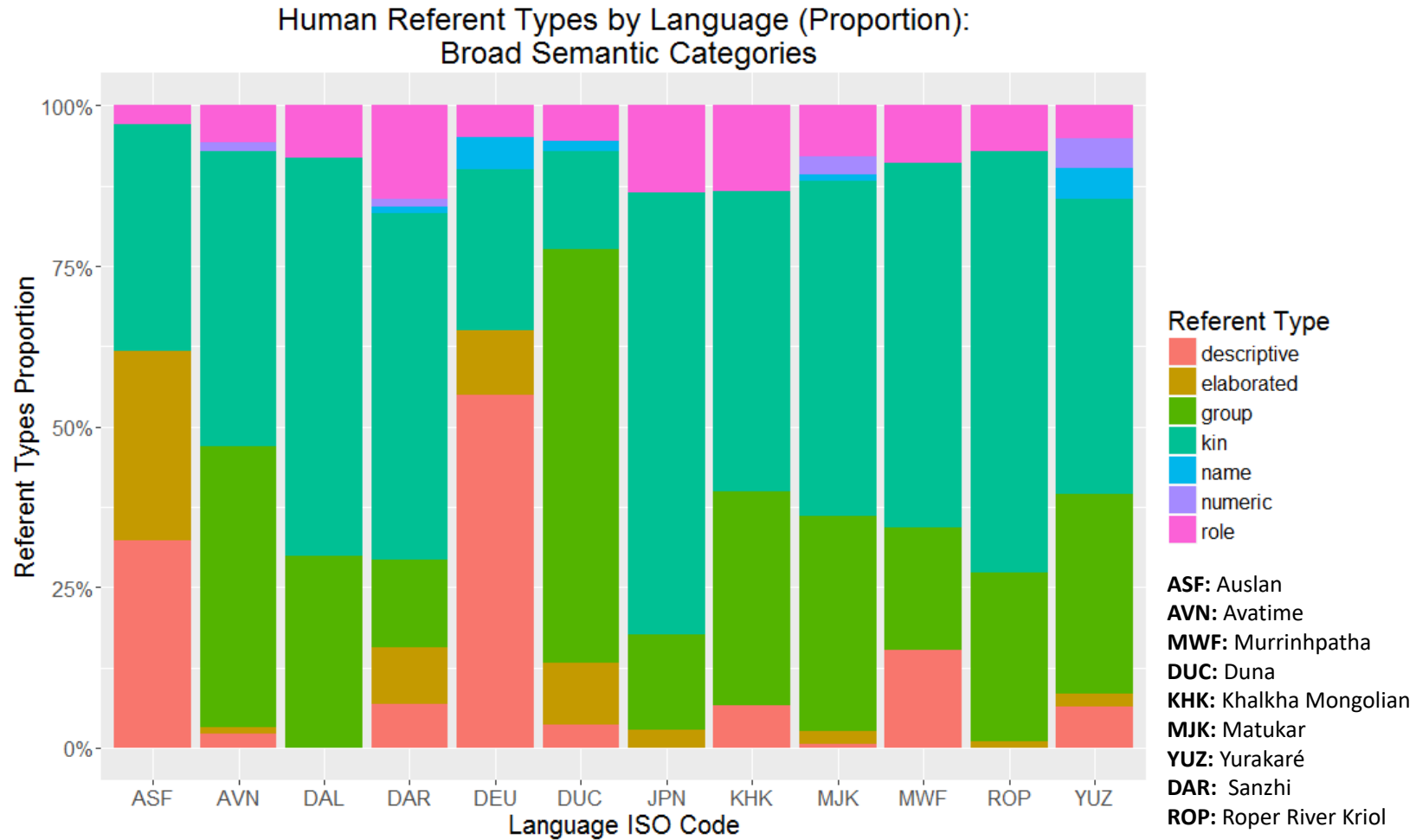
German

Da drauf sind ein älterer Mann, eine Frau, ein Kind und dieser Besuch, ein junger Mann.

‘On it (the picture) are an elderly man, a woman, a child, and this visitor, a young man.’ SocCog-deu01-hs_ks_HR_RS_PV2019-10-30 - 00:00:42-00:00:47



Human Referent Types: Distribution by Language



Human Referent Types: Distribution by Language

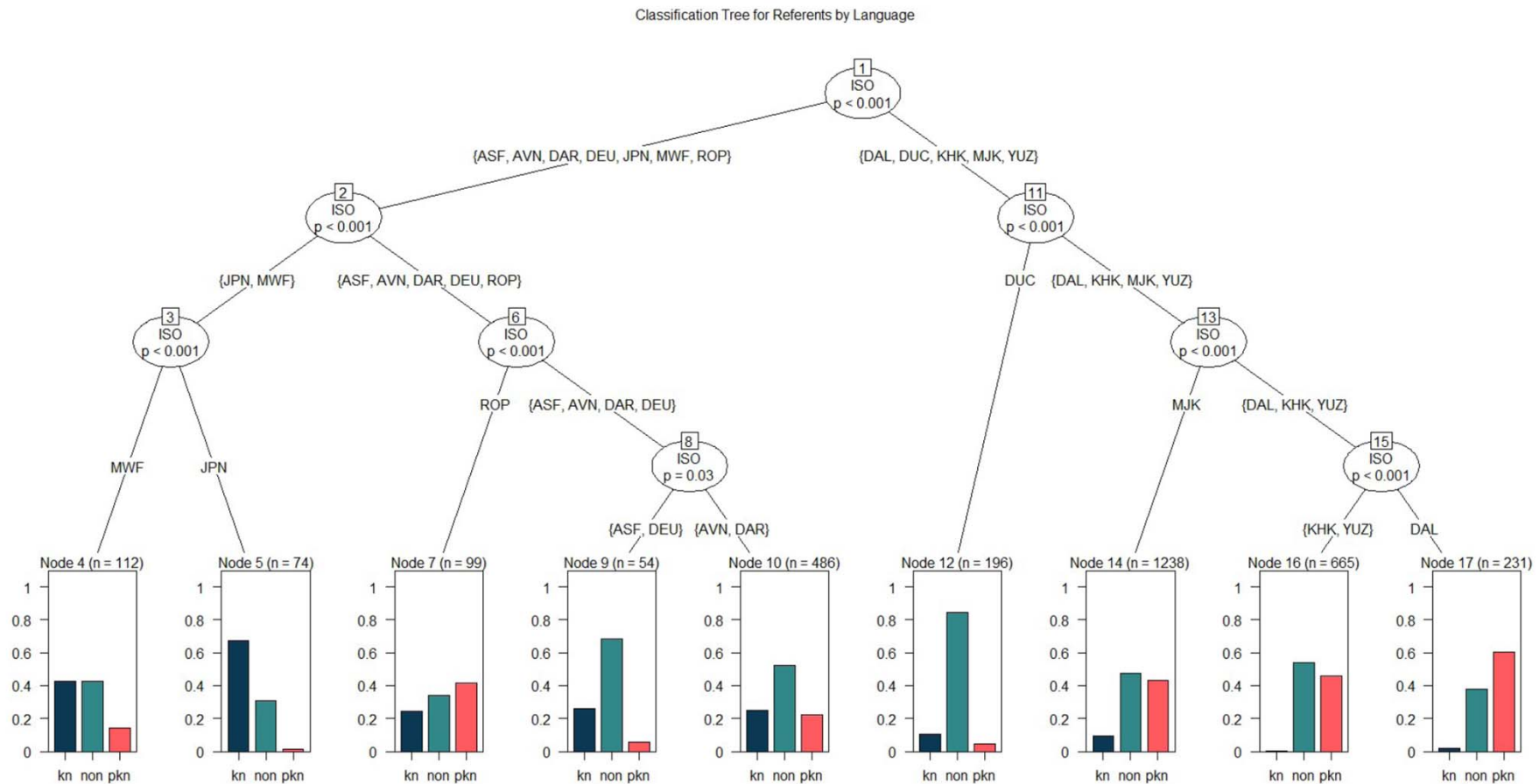
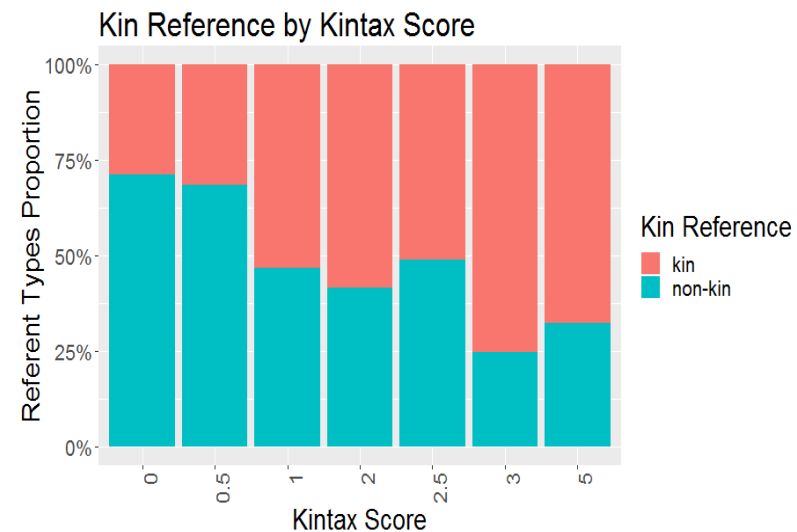
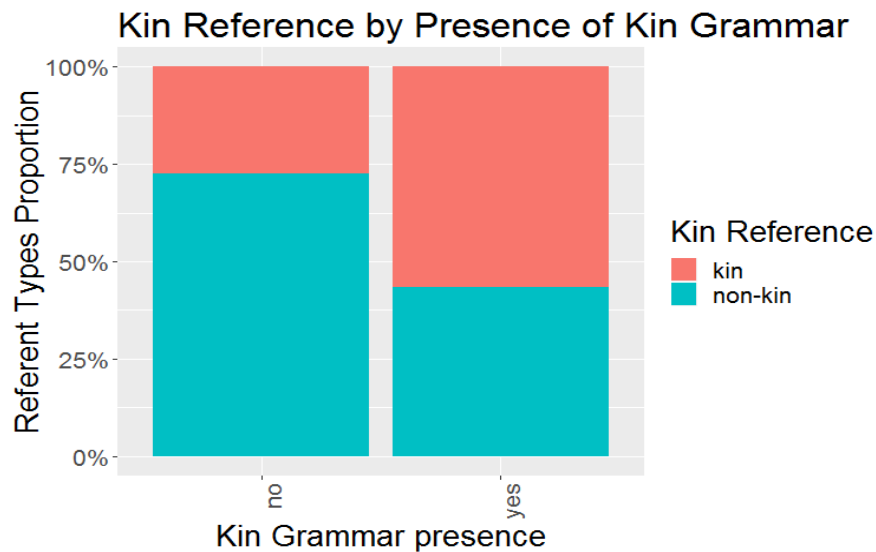


FIGURE 5. Classification Tree for Human Referents Possessed Kinship, Non Possessed Kinship, and Other



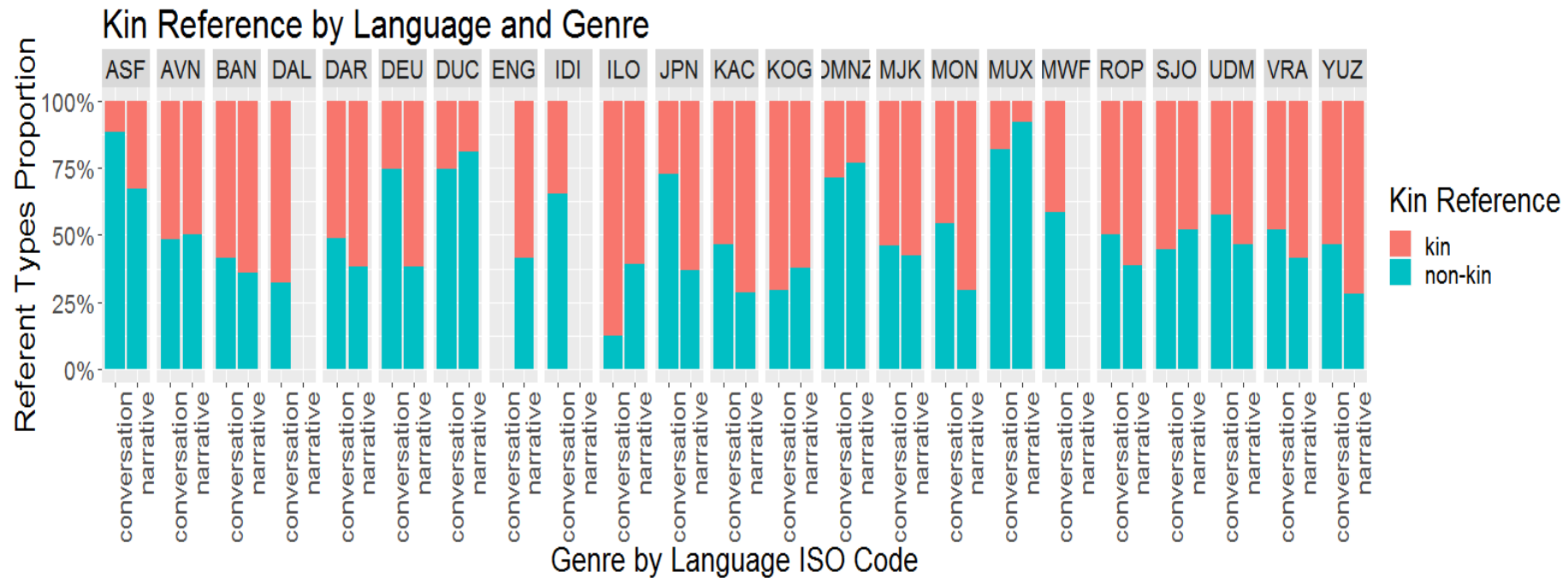
Referring to persons: the influence of grammar



The presence of grammatical constructions sensitive to kinship as a category interacts clearly with the proportion of formulations as kin ('kin reference'), whether 'kin grammar' is measured as a binary (presence or absence of 'kin grammar') or as a valued score ('kintax score', according to the number of relevant constructions available)



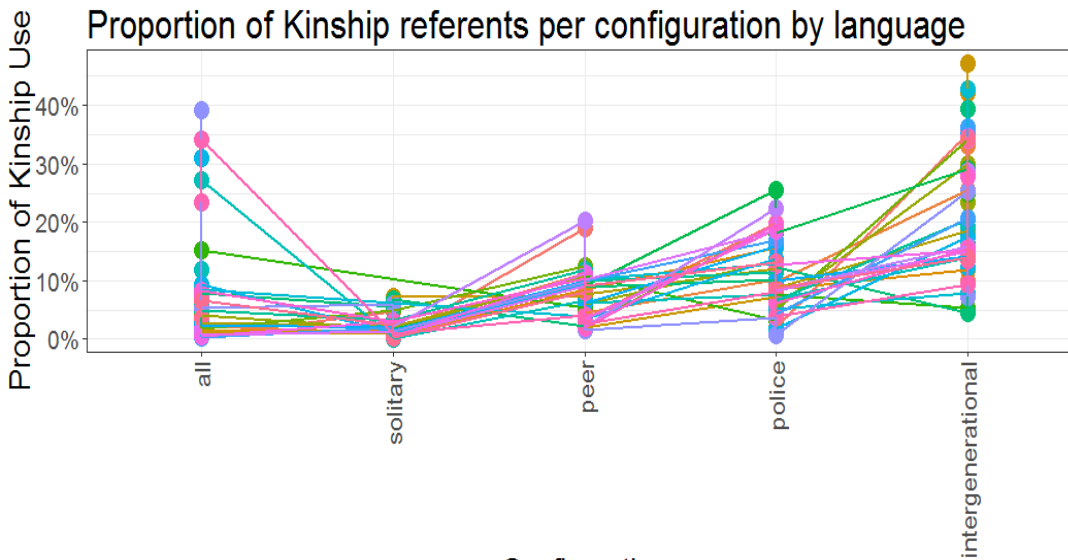
Referring to persons: the influence of task phase



Kinship formulations are significantly likely to lock in, across all languages pooled, once the narrative phases are reached (here: conversation = first 2 phases)



Character configuration and formulation



- ISO
- ASF
 - AVN
 - BAN
 - DAL
 - DAR
 - DEU
 - DUC
 - ENG
 - IDI
 - ILO
 - JPN
 - KAC
 - KOG
 - KOMNZO
 - MJK
 - MON
 - MUX
 - MWF
 - ROP
 - SJO
 - UDM
 - VRA
 - YUZ



Police

Solitary



Peer



Configuration



Intergenerational

*Except policeman



Deploying epistemic resources



Do languages differ in their deployment of resources for epistemic management?
Are these affected by task phase?
Do they vary with individuals?



Epistemic resources in Yurakaré

Evidentiality

- =*ya* 'reportative' (verbal enclitic)
- =*tiba* 'inferential' (verbal enclitic)
- =*jtë* 'assumptive' (verbal enclitic)
- shi* 'uncertain visual' (suffix)

Epistemic stance (e.g. Heritage 2012)

- =*ya* 'intersubjective epistemic judgment' (verbal enclitic)
- =*laba* 'subjective epistemic judgment' (verbal enclitic)
- =*la* 'speaker commitment' (clausal enclitic)
- =*ti* 'intersubjective commitment' (clausal enclitic)
- =*se* 'presupposition' (clausal enclitic)

Epistemic modality

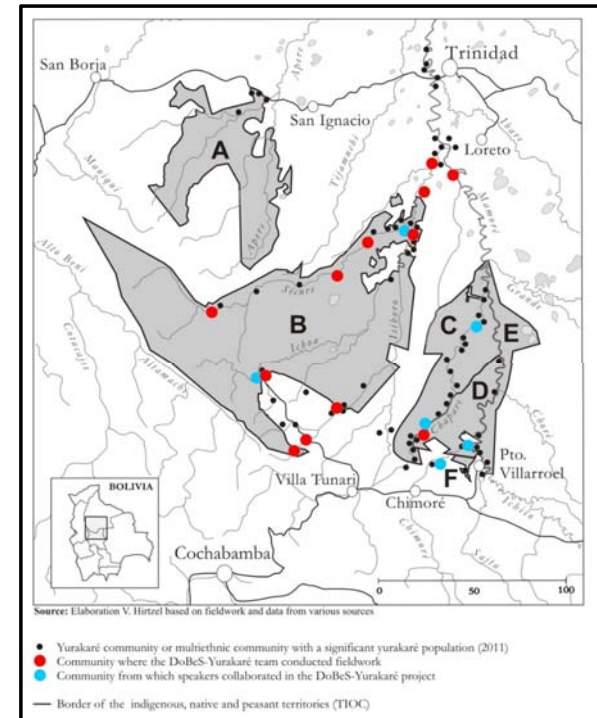
- kusu/kusuti* 'maybe' (adverb)
- nentaya* 'maybe' (adverb)

Other attitudinal markers

- =*ri/=yu* 'resignation' (clausal enclitic)
- =*ra/=ye* 'adaptive' (clausal enclitic)

Perception verb tags

- bějma* 'look imperative'
- ujampu* 'see presentative'
- no ves* 'don't you see (Spanish)'
- kalinde/kali/kay* 'watch imperative'



Sonja Gipper



Engagement in Kogi (Colombia)

Prefix	Meaning	Example	Translation
na-	Known to the speaker but inaccessible or unknown to the addressee	Kwisate nanukú. [dancing na-am-I]	'Yes (answering a question, you didn't know this), I'm really dancing.'
ni-	Known to the speaker but also accessible or known to the addressee	Kwisate ninukú. [dancing ni-am-I]	'(As you are aware) I am dancing.'
sha-	Inaccessible to the speaker, addressee has epistemic authority	Nas hanchibé shakwisatetuku? [I good sha-dancing.am.I?]	'Am I dancing well (in your opinion)?'
shi-	Accessible or surmised by the speaker, though addressee has 'epistemic authority'	Ma kwisaté shibalaw? [you dancing shi-are-you]	'Are you dancing (you look like you are)?'; 'You're dancing, eh?'
ska-	Inaccessible to the speaker, neither speaker nor hearer claim epistemic authority – used for 'I wonder' type questions where there is no expectation that the addressee will necessarily know the answer	Saki skaguatox? [what ska-doing]	'What's he doing?', '(I don't know) what he's doing (nor do I expect you to know, so let's ask him).'



Henrik Bergqvist

Dominique Knuchel



Three languages using different subcategories of epistemic marking



Card 1: Homecoming

Yurakare

ma-pëpë-∅=ya *ana*
 3PL.POSS-grandfather-3SG.SBJ=INTSUBJ DEM
 'Maybe this is their grandfather.' YUZ109-2, homecoming

Matukar Panau

main tei hanat-ama-n-da y-en-ago
 PROX bilum what-POSS-3sg-COM 3sg-Lay-R:I:IMPF

ti-ta-nong-go milo tai?
 NEG:1pl.inc-know-R:I:IMPF something DUB
 'This is a woven bag with things lying in it. We don't know what things.'

Dalabon

Nûnda kardû kah-dunkûn-daddinj, kardû nadjomorrwu
 DEM maybe 3sgREAL-in.gaol-be.inPIPF maybe policeman

bukahng-munkuyunj, "djah-dudjmû kirdikird-ngu-kah, duwe-ngu-kah"
 3sg>3sg.REAL.SEQ-send:PPF 2sg-returnIMP wife-2sgPOSS-LOC f.in.law-2sg-LOC
 'This one, maybe he was in gaol, maybe the policeman had released him,
 "you go back to your wife, to your father-in-law"' (MTDL20120612_01)



Epistemic marking in one language but not in another



Card 3: Sitting Drinking

Yurakare

kusu nij ma-bashti-∅=laba
 maybe NEG 3PL.POSS-wife-3SG.SBJ=SUBJECTIVE.EPISTEMIC.JUDG
 '(I reckon) maybe they are not their wives.'
 (SocCog-YUZ105-1, sitting drinking)

Dalabon

nidjarra bulahlng-kolhngu-ninj, kirdikird burrkunh,
 this.one 3pl>3sg.REAL.SEQ-drink-PIPF woman two

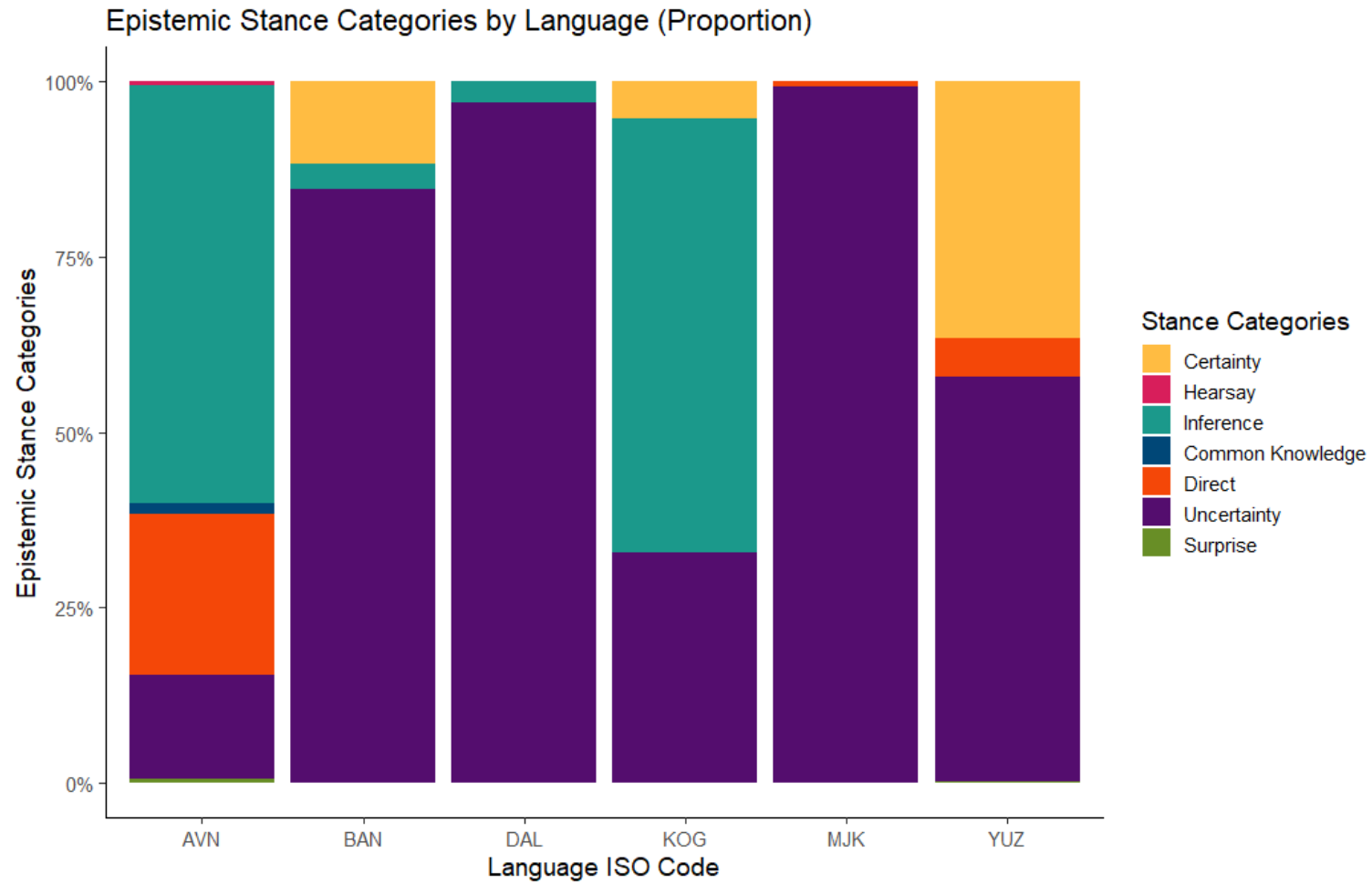
nunh_kanh waluHwalum-be kah-di nidjarra mah bulhdjarn
 DEM south.from 3sgREAL-be.NPST this.one also middle

nûnda_kanda biyi_kirdikird-no kahnun
 DEM man wife-3sgPOSS DEM

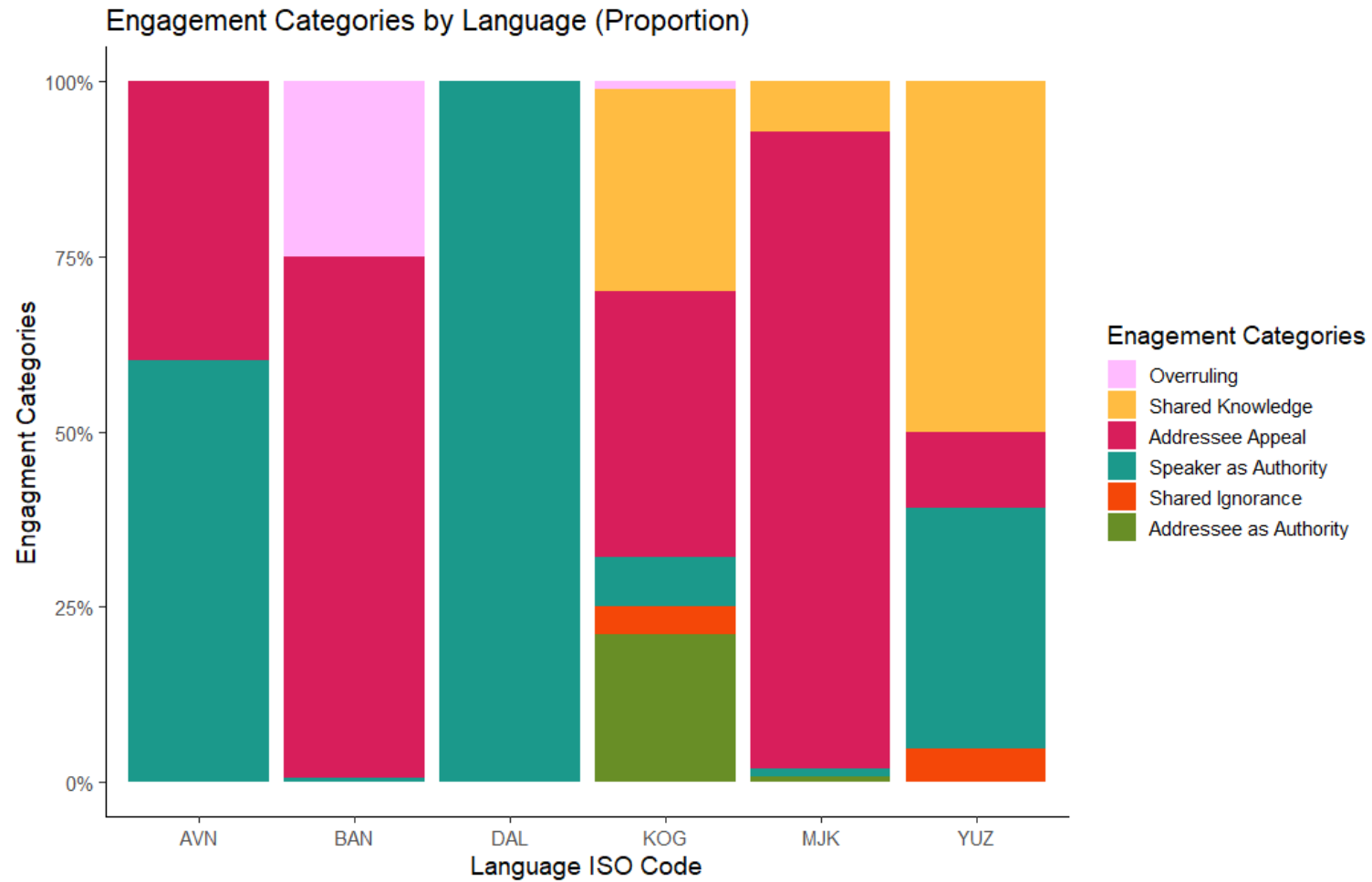
'Here they are all drinking, two women, and this one looking from the south, and also this one in the middle, this is a husband and wife.'" (MTDL20120612_01)



Epistemic stance subtypes by language



Engagement subcategory, by language



'Narrative problem-solving design' elicits rich epistemic marking

Ger: *achama buybu*
be_like_that story
'That's the story.'

Jul: *achu mala-∅=ye*
like_that go.SG-3SG.SBJ=ADAP.F
'That's how it goes?'

Ger: *achu mala-∅=ti*
like_that go.SG-3SG.SBJ=intersubject.commitm
'That's how it goes.'
(SocCog-YUZ104-2, discussing the whole story)



Gerónimo & Julia



Inagaki & Sonoda

ST: 多分こっちが先じゃないかな？
tabun kotchi-ga saki-janai-ka-na
possibly this.one-NOM before-COP.NEG.NPST-PRT10-PRT4
'Maybe this one is first perhaps, do you think?'

ST: だってさっき、あの、希望,絶望で終わりみたいな感じだったやん。
datte sa sakki zetsu ano kibou zetsubou-de owari mitai-na kanji-datta-yan
because short.time.ago that hope despair-with end like-COP.ATT
feeling-COP.ATT.PST-COP.NEG.NPST
'Well it's because didn't we [discuss] that hope is followed by despair.'

IK: あ、わかった。
a wakat-ta
ah understand-PST
'Understood.'

IK: でもハッピーエンドやんな？
demo happii happii-endo-yan-na
but happy happy-ending-COP.NEG.NPST-PRT4
'But isn't it actually (going to be) a happy ending?'



Different task phases, different epistemic categories

mjk02 - Part 1

CARD 1

main *tei* *hanat-ama-n-da* *y-en-ago*
PROX bilum what-POSS-3sg-COM 3sg-lay-R:I:IMPV

ti-ta-nong-go *milo* *tai?*
NEG-1pl.incl-know-R:I:IMPV something DUB

'This is a woven bag with things lying in it. We don't know what things.'



Discussion and negotiation, phase 2



mjk02 - Part 2

while arranging cards...

uyan *ha-n=lo* *das-aba,* *ilo-m* *gire-nggo?*
good CL-3sg=LOC ascend-IRR:I:FUT inside-2sg think-R:I:IMPV

Taleo Kreno - SocCog02-tk_jb_2 - 2:18-2:20.5

‘(it) will go above the good one, do you think?’



Narrative phase

mjk02 - Part 3

CARD 13

ha-di *aim* *dabok-kasman-e,* *main* *te-p?*
POSS-3pl boy big-INTSF-R:I:PFV PROX see-IRR:D

'Their child is already big, you see?'

John Bogg - SocCog-mjk02-tk_jb_bk_3 - 3:05-3:07.5



Total stance tokens per language (6 language sample)

ISO	minutes of data	tokens	tokens per minute
MJK	281.8	241	0.9
KOG	65.6	157	2.4
DAL	13.7	34	2.5
BAN	386.8	1090	2.8
AVN	60.8	267	4.4
YUZ	375.3	3574	9.5



Token ranges by speaker (6 language sample)

ISO	Minutes of data by Language	Token Ranges (by Speaker)		Speaker Ranges (tokens per minute)	
MJK	281.8	1	34	0.02	1.86
KOG	65.6	--	--	--	--
DAL	13.7	3	31	0.22	2.26
BAN	386.8	2	125	0.06	5.44
AVN	60.8	2	125	0.21	3.63
YUZ	375.3	26	459	0.39	7.53

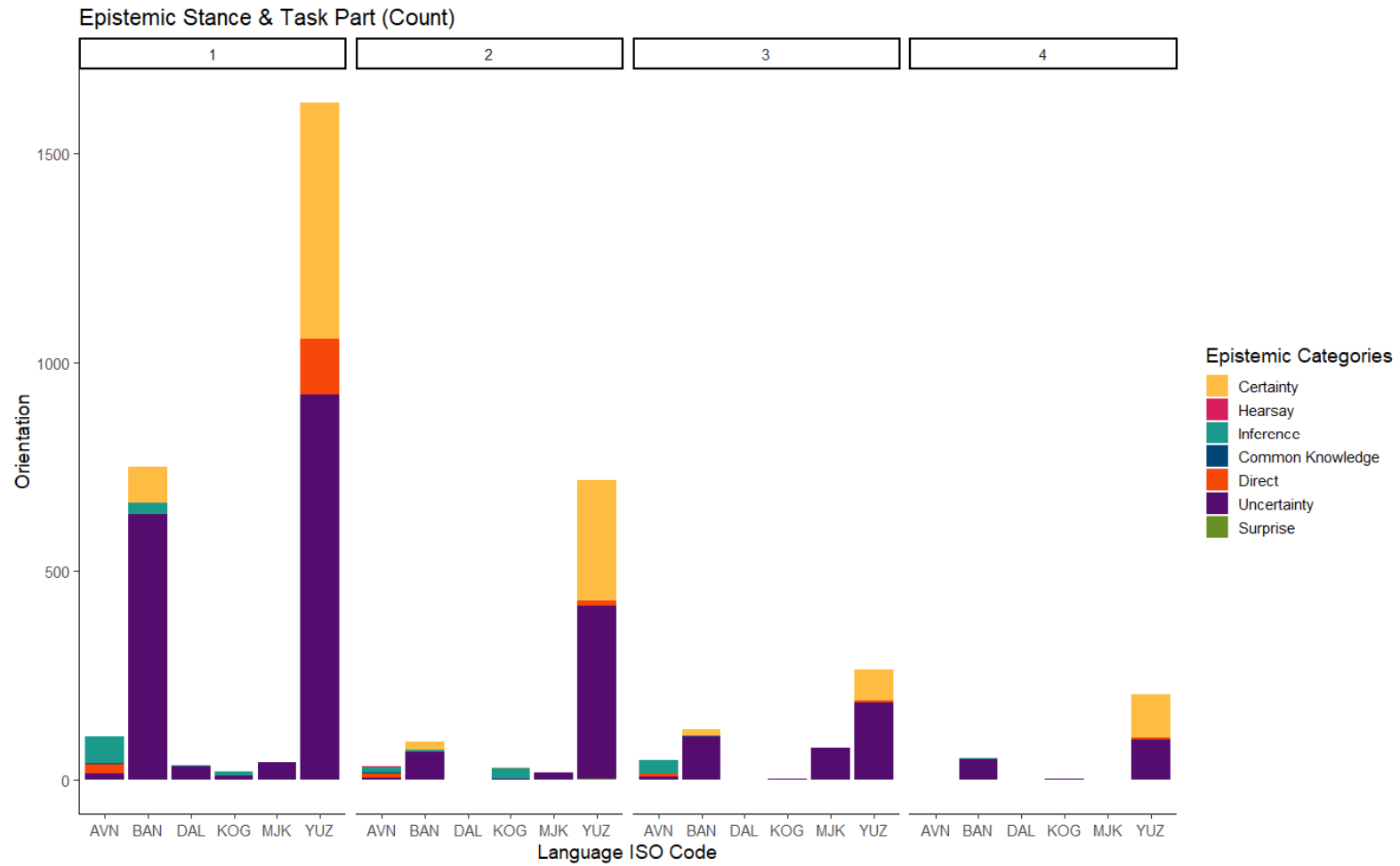


Tokens by stance subcategory

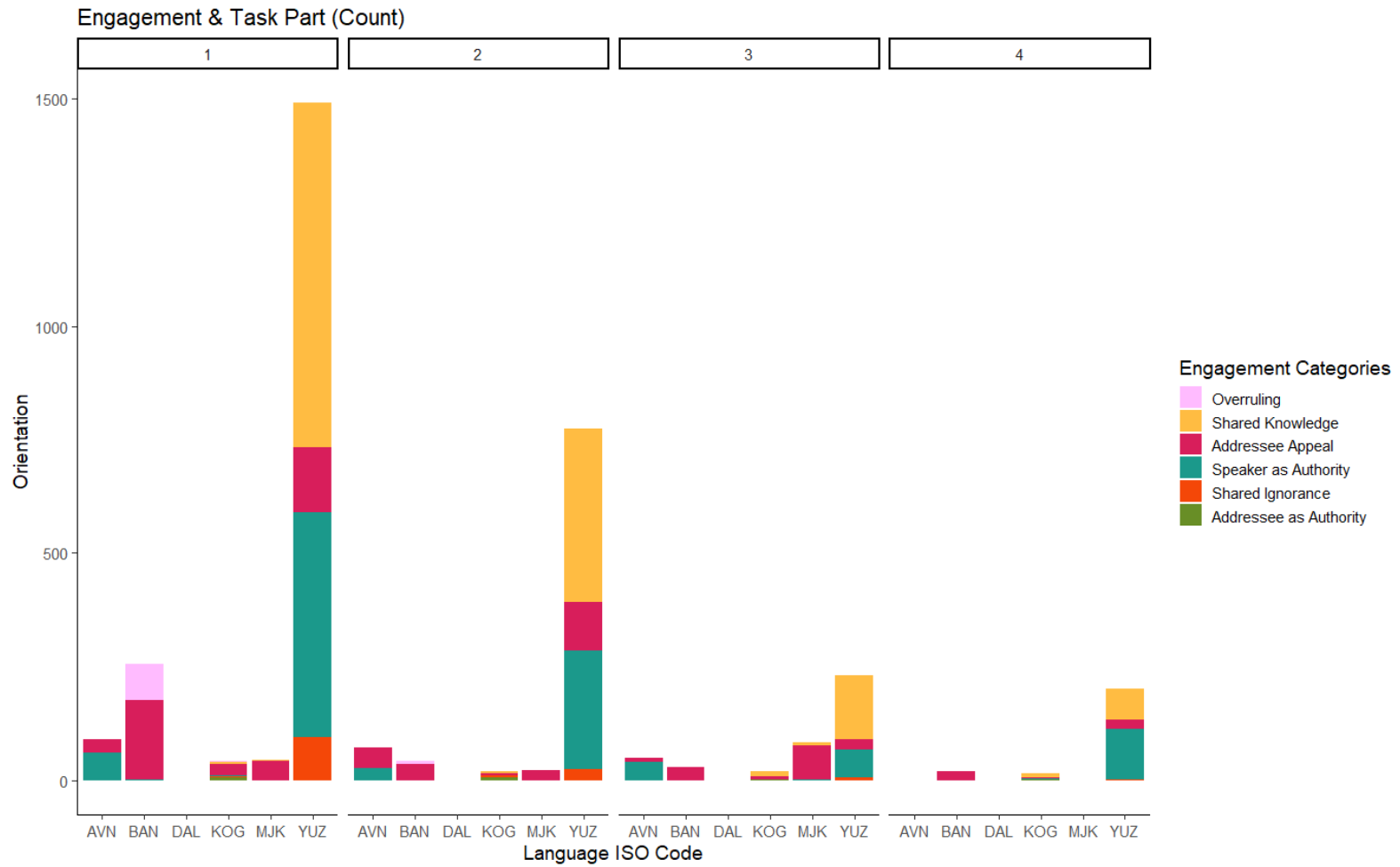
ISO	minutes of data	Tokens (total)	Tokens per minute (total)	Tokens per minute (epistemic)	Tokens per minute (engagement)
MJK	281.8	241	0.9	0.5	0.6
KOG	65.6	157	2.4	0.9	1.5
DAL	13.7	34	2.5	2.5	0.1
BAN	386.8	1090	2.8	2.6	0.9
AVN	60.8	267	4.4	3.0	3.5
YUZ	375.3	3574	9.5	7.5	7.2



Epistemic categories, by task phase



Engagement, by task phase



Langue meets parole in the moment of choice

The idea of choice links the unfolding moment of discourse with the whole apparatus of grammar, lexicon and other expressive resources that sits silently behind each moment of speech

Each such choice draws langue into parole, and each choice made to include – or not to include – some item in the unfolding parole feeds back into the vast set of summed utterance moments which feed back into the perpetual reshaping of grammar

To understand how this plays out in corpus linguistics, we need methods that include the unsaid as well as the said – choosing to characterise someone as his father or the man, or to say ‘He returned home’ vs ‘He’s returning home, isn’t he?’

Bickel’s famous ‘what is where why?’ can be recast here as ‘What is here when, why?’

By allowing us to assemble bodies of naturalistic speech, across languages, across speakers, across task phases, and across event configurations, SCOPIC helps us understand some of these questions by corpus methods

It is only one such tool, though – in this case, designed with a particular set of semantic categories in mind – and we hope the emerging field of corpus-based typology will develop many others in the quest to build more naturalistic but comparable foundations to the study of linguistic diversity



With thanks to

Australian Research Council

Alexander von Humboldt Stiftung (Anneliese Maier Forschungspreis)

Sonja Gipper

Cale Johnstone

The SCOPIC Team

Stockholm University

Universität zu Köln

Japan Society for the Promotion of Science

The Research Institute for Languages and Cultures of Asia and Africa (ILCAA)



Australian Government

Australian Research Council



**The Wellsprings of
Linguistic Diversity**



**ARC CENTRE OF EXCELLENCE FOR
THE DYNAMICS OF LANGUAGE**



THE AUSTRALIAN NATIONAL UNIVERSITY



Alexander von Humboldt
Stiftung/Foundation

