

Saying without 'say': A corpus-based approach to the typology of non-canonical speech reports

*Izabela Jordanoska¹, Ekaterina Aplonova², Guillaume Guitang³, Olga Kuznetsova⁴
and Songfola Lacina Silue⁵*

¹CNRS-LACITO, ²CNRS-LLACAN, ³Université Libre de Bruxelles, ⁴ILS RAS,

⁵INALCO



Introduction: saying without ‘say’

- In many languages reported speech can be introduced **without** any speech verbs or quotative markers:
 - 1) John entered the room. “Has this dog been fed already?”
(English, D’Arcy 2015)
- Such constructions have been described for **individual** languages:
 - English (Indo-European; Mathis and Yule 1994, Vandelanotte 2009, a.o.)
 - Umpithamu (Pama-Nyungan; Verstraete 2011)
 - Ungarinyin (Worroran; Spronck 2017)
 - Hungarian (Finno-Ugric; Teptiuk 2019)

Towards a typology

- Güldemann (2008) observed that this occurs in **multiple** African languages.
- 31 out of 39 of the African languages in his sample had some variant of this non-canonical construction.
- So how universal is this phenomenon?

Towards a typology

- Güldemann (2008) observed that this occurs in **multiple** African languages.
- 31 out of 39 of the African languages in his sample had some variant of this non-canonical construction.
- So how universal is this phenomenon?

Let's begin to make a typology of non-canonical speech reports!

In this talk we look at a convenience sample of 10 languages.

Reported speech: some definitions

- “Reported speech is speech within speech, utterance within utterance, and at the same time also speech about speech, utterance about utterance.” (Vološinov, 1973: 115)
- Special syntactic domain with a characteristic set of properties (Spronck & Nikitina 2019)
- Structure:

Mary said - “Let me eat cake.”

M(atrix clause) - **R**(eported clause)

- M ~ QI “a segmentally discrete linguistic expression which is used by the reporter for the orientation of the audience to signal in his/her discourse the occurrence of an adjacent representation of reported discourse” (Güldemann, 2008: 11).

“The literature on D[irect] R[eported] D[iscourse] and its linguistic expression rarely treats the phenomenon of Q[uotative] I[ndex]-less constructions in any detail.”

Güldemann (2008:38)

Previous literature: non-canonical reported speech

While non-canonical speech reports seem to be cross-linguistically common, they need something in the context that makes the appearance of reported speech **predictable**. This can be:

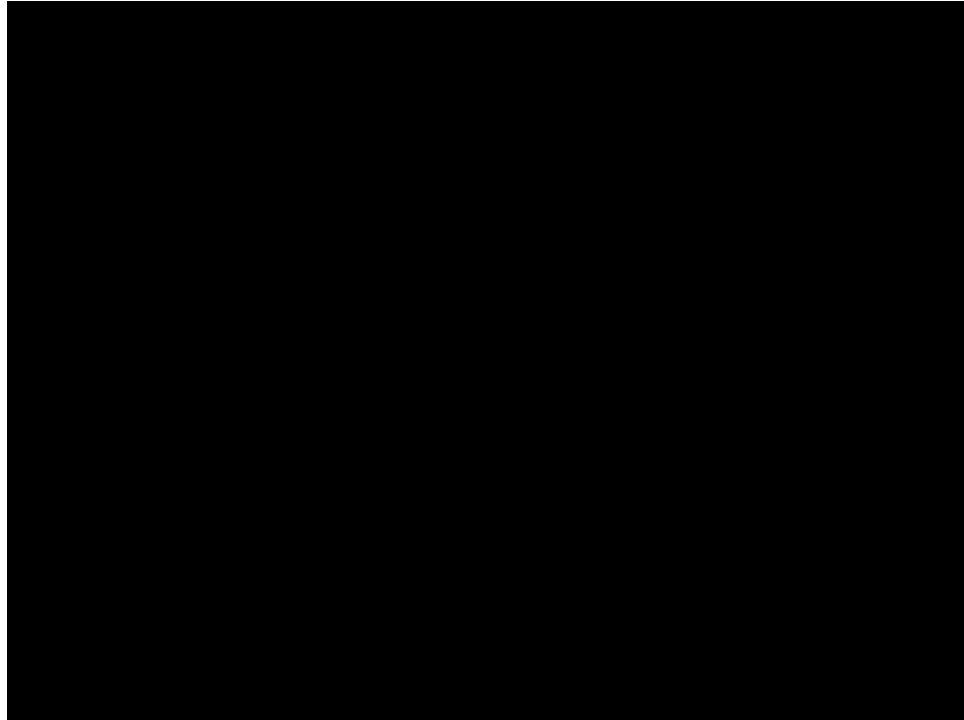
- A 'perspectivizing clause' (coined by Verstraete 2011): a **non-speech verb** which can be: motion, perception, inchoative or psych verb.
- Backgrounded **adverbial** clauses and **event sequence** markers.
- Within **dialogues**.

Example: dialogue (from Macedonian)

Ama "Nemoj da go otvaraš," reče.
But "Don't open it!" he said.

I toj rekol: "Daj, sinko, ja da go
otvoru, tuka lošo pišue za tebe."
*And he (God) said: "Let me open it,
son, bad things about you are written
in it."*

"Kako be, dedo?"
"How is that, grandfather?"



Questions for a corpus

- Stuff we can count:
 - Do we find all of these constructions in all of our languages?
 - Which ones are the most frequent?
 - Are there more ways of introducing reported speech without an Matrix that haven't been observed yet?
- Theoretical puzzle: Can the perspectivizing clauses be considered a type of Matrix? Or are they a different construction altogether?

Methodology

SpeechReporting Corpus (Nikitina et al. 2022)

- No systematic way for searching for reported speech other than searching for specific verbs or quotative markers in most corpora...
- Unless... you have a corpus annotated for reported speech.
- SpeechReporting Corpus: Collection of corpora of 10 low-resourced languages from West Africa and Eurasia.
- Texts consist of fairy tales that are transcribed, glossed, translated and annotated for instances of reported speech in ELAN-CorpA (Chanard 2015).

Languages that we will analyse

Language	Family/Group	Where spoken
Bashkir	Turkic	Russia
Chuvash	Turkic	Russia
Udihe	Tungusic	Russia
Macedonian	Slavic	Macedonia
Gizey	Chadic	Cameroon
Guro	Mande	Côte d'Ivoire
Wan	Mande	Côte d'Ivoire
Kafire	Gur	Côte d'Ivoire
Joola	Atlantic	Senegal
Ut-ma'in	Kainji	Nigeria

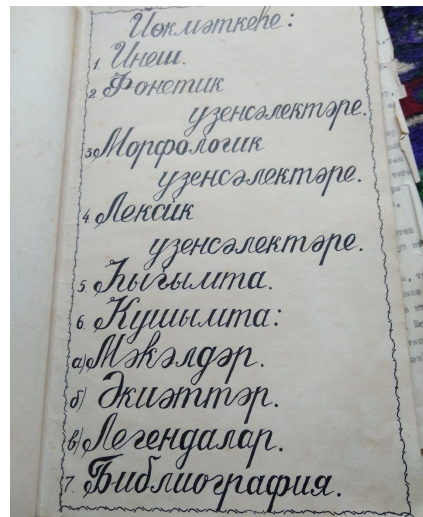
SpeechReporting database: data

- Data is very diverse: multimedia files, archived transcriptions, data from the field, written/unwritten languages, recordings of professional storytellers as well as of regular people, one or multiple participants, corpora sizes vary between ~1000 and ~50k words...



SpeechReporting database: data

- Data is very diverse: multimedia files, archived transcriptions, data from the field, written/unwritten languages, recordings of professional storytellers as well as of regular people, one or multiple participants, corpora sizes vary between ~1000 and ~50k words...



Tsakorpus: web search interface for multimedia corpora

[Multilingual corpus of the Discourse Reporting project](#)

EN | RU | ?

↓ Back to search



Search result: 115 occurrences, 110 sentences



min	šulaj	tip	ujlajəm	tigän
min	šulaj	ti	ujla	ti
min	šulaj	ti-p	ujla-a-m	ti-yan
1SG	so	sayB.CV	thinkIPFV1SG	sayPC.PST
pron	adv	v, vsuf	v, vsuf	v, vsuf
<input checked="" type="checkbox"/> RS	so	say	ø-ø-RS	say
1SG			think	

I think so – **he** said.

190719_YNK_UmnyjParen_11

qt: ['Quotative_Discourse_Report']
rp: ['Discourse_Report', 'Quotative']
speaker: SP1
typ: ['Statement']

1ely 20 documents

Some examples from our corpus

Perspectivizing clause:

2) Belie **diga-si-mi** **meteu-e-ni**

beauty eat-IM-GER finish-PST-3SG

Mam'a, **edede,** **jeu-xi** **ηeneh-e-ti?**

grandmother oh! what-DIR go-PST-3PL

‘The girl finished eating: "Grandmother, ooh, where did they go?"’ (Udihe, Tungusic)

Dialogue:

3) **Kade** **e** **Role?** **Na** **ora-nje** **jet.**

where be.PRS.3SG R. on plow-NLMZ be.PRS.3SG

“Where is Role?” “He is out plowing.”

(Macedonian, Slavic)

	0:00:50.000	00:00:51.000	00:00:52.000	00:00:53.000	00:00:54.000
ref@SP1 [61]	Baskakova II.01.026		Baskakova II.01.027		
lit@SP1 [0]					
tx@SP1 [61]	ene	Belle digasimi meteueni		Mam'a, edede, jeuxi njeneheti?	
mot@SP1 [320]	ni	Belle digasimi meteueni		Mam'a, edede, jeuxi njeneheti?	
mb@SP1 [659]	1, -n	belie digasimi meteueni		mama edede j'eu xi njeneheti	
ge@SP1 [659]	2, -3	beauty eat -IM -GER finish -PST -3SG		grandmother oh! what -DIR go -PST -3PL	
par@S [34]					
ps@SP1 [659]	T, -p, n	v -akz -conv v -TAM1 -pers		n intj pron -case v -TA -pers	
not@SP1 [0]					
ft@SP1 [61]	The girl finished eating.		"Grandmother, ooh, where did they go?"		
comments@S [7]					
rp@SP1 [47]	Non-speech Event		Discourse Report	Discourse Report	Discourse Report
typ@SP1 [33]			Greeting or Address	Expression of Attitude	Question
ad@SP1 [0]					
qt@SP1 [281]	DR Event + Discourse Report				

Methodology

- Quantitative:
 - How many reports are annotated as either 'Non-speech Events' (i.e., perspectivizing clauses) or 'Bare Quotes' (i.e., quotes that are seemingly not introduced by anything).
 - Language specific vs general frequencies?
- Qualitative:
 - Is there a structural correspondence between perspectivizing clauses and regular M clauses in the languages?

Results

Number of non-canonical constructions per language

Language	Non-canonical reported speech constructions	Total number of reported speech constructions	% of non-canonical speech constructions
Udihe	244	852	29%
Gizey	55	298	18%
Macedonian	58	417	14%
Kafire	103	803	13%
Wan	107	1789	6%
Joola	55	975	6%
Ut-ma'in	4	75	5%
Guro	10	269	4%
Chuvash	118	2893	4%
Bashkir	29	865	3%
Total	783	9236	-

General vs language-specific findings

- All languages in our corpus employ the non-canonical speech reporting strategy to some degree.
- Some more than others...
- Udihe has the highest amount of non-canonical speech reports, 244 (29%)
- 180 (74%) of these are not even introduced by a perspectivizing clause, but occur in dialogues.
- Difference in rhetorical genres?

Udihe complex unmarked dialogues with 2 participants

- *Mam'a, ki-ki, bi bimie, ki-ki.*

‘Grandmother, it is me.’

- *Utemile mam'atigii emegize!*

‘Then return to your grandmother!’

- *Mam'a, ki-ki-ki, ono emegizemi, ki-ki-ki, bua xueni gajani osimi.*

‘Grandmother, how will I come back, I have become a wild bird.’

- *Aja, emegie ñugasilegei.*

‘Well, come I will kiss you.’

Udihe complex unmarked dialogues with 3 participants

- *Kekui, xuñaziwehi geenehemi.*
'Cuckoo, I have come for your sister.'
- *Keeku, keeku.*
'Cuckoo, cuckoo.'
- *Keku jewe diganaini?*
'What does the cuckoo say?'
- *Xuñazixi bimi, gaitide tugezi emeweje!*
'If you have a sister, send her here at once!'

Participants: young man, cuckoo, cuckoo's sister

Corroboration findings

The languages in our sample corroborate what has been observed before:

The non-speech verb types:

- Perspectivizing clauses can be motion verbs, perception verbs, mental state verbs and inchoative verbs.

The other contexts:

- When there is no perspectivizing clause, non-canonical speech reports can occur in dialogues and with event sequence markers.

Perspectivizing clauses: examples from previous lit

- 4) policeman **yongki-n** **ina** “Wanthawa **iluwa**”
policeman come-PST 3PL.NOM IGNOR.LOC 3SG.NOM
‘The policemen came. “Where is he?”’ (Umpithamu, Verstraete 2011:499)

- 5) azzal **kezdtem,** “Hogy jó csapat a TF.”
DEM.D.COM start.PST.1SG COMP good team DEF PN
‘I started with that (that) TF is a good team’ (Hungarian, Teptiuk 2019:62)

Novel findings

- We also find other verb types occurring in the ‘perspectivizing clause’ slot that haven’t been described yet: **action** type verbs such as ‘give’, ‘take’, ‘bring’, ‘throw’ and **position** verbs: ‘sit’, ‘lie’, ‘stand’.
- We also find other context types that have not been described yet: Descriptions of the scene (that usually triggers a reaction from the speaker) and interruptions by the narrator to describe a certain situation or give more details.

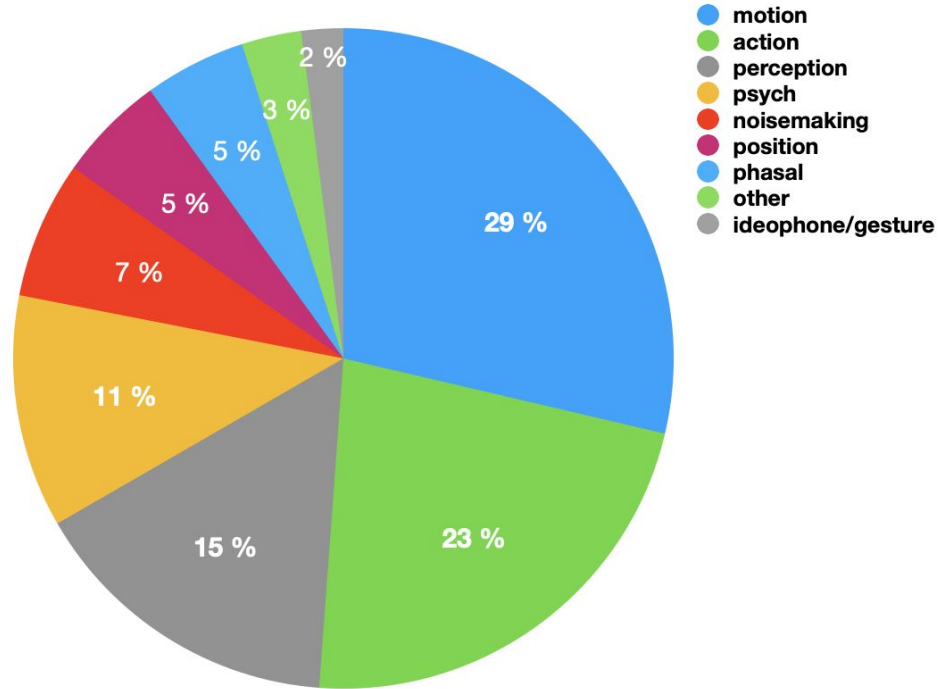
Perspectivizing clauses

- We counted all clauses of which contain a **verb** of which the agent/subject corresponds to the **speaker** of the immediately **adjacent** speech report.
- Motion verbs are the most frequent verbs in perspectivizing clauses.
- Action verbs, like 'give', which haven't been described introducing speech before, are also frequent in this construction.
- Thus this phenomenon is actually **more broad** than previously known.

Perspectivizing clauses with other verbs

- 6) O may **n-a-ŋar** **ji-iba-ol.**
REL as_well LOC-NC1-take NC11a-knife-3SG.DO
“U-gab-or-al.”
1PL.INCL-distribute-ASSOC-INCL
‘The other took his knife. “Let’s share.”’ (action, Joola)
- 7) Qujan qujan **tor-yan** da “Juqindetölkö-bikä **min hine min**
hare hare stand-PC.PST and no yet fox-lady 1SG 2SG 1SG
bögön joqlamanəm,”
today sleep-NEG-PST-1SG
‘But the Hare stood and: “No, Fox, I didn’t sleep today,”’ (position, Bashkir)

Perspectivizing clauses: verb types



Background clauses

Güldemann (2008): Speech reportes with no quotative index can be immediately preceded by a **background clause** or **marker of event sequence**:

8) dáà “bòi, Dí Bar nè uà?”

then “father, what creaks here?”

(Ngambay, Vandame 1963:162, as cited in Güldemann 2008:44,
no gloss provided by author)

In reported dialogues: no verb, no sequence marker

9) “Nyingan=karra rimij wunja2-ma-ni-ra.”

2sg=MAYBE steal 3nwsq.O:2sg.S-TAKE-PST-1sg.IO

“Anja-ku rimij ngin-w2a2-y2i-ø-nu-ngarri nginmaji

what-DAT steal 1sg.S-IRR-BE-PRS-2sg.IO-SUB 1sg MUST

buluba wu1-ra jadarn.”

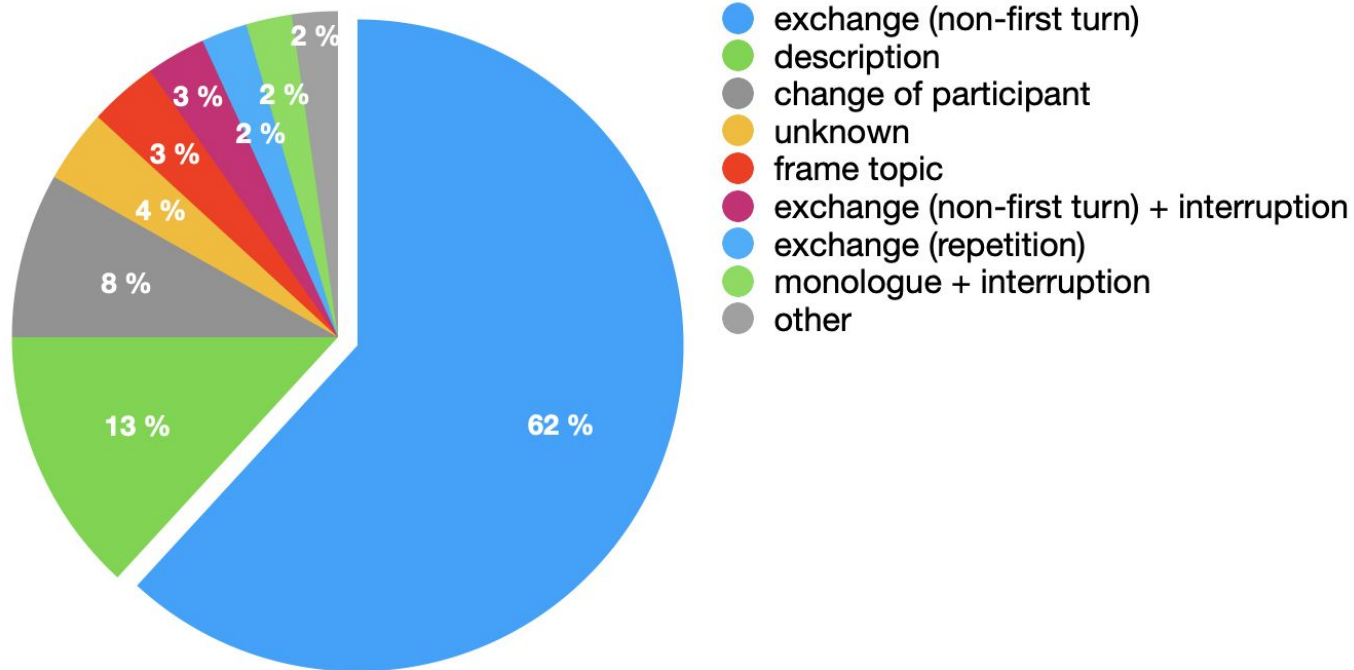
look.around-ITRV 3nwsq.O-go.to properly

“Maybe you’re the one stealing my things.”

“Why would I steal from you? Why would I rob you? You should look around for it properly.”

(Ungarinyin, Spronck 2017:125)

And when there is no perspectivizing clause?



Context types

We distinguished the following context types in all cases where there was no perspectivizing clause:

- Exchange (i.e., dialogues)
- Frame topic (i.e., backgrounded adverbial and prepositional clauses and event sequence markers)
- Change of participant
- Description
- Interruption

Context types

We distinguished the following context types in all cases where there was no perspectivizing clause:

- Exchange (i.e., dialogues)
- Frame topic (i.e., backgrounded adverbial and prepositional clauses and event sequence markers)
- **Change of participant**
- **Description**
- **Interruption**

Exchange

The speaker is reacting to what another speaker has said.

Context: “Fatou, sit.”

10) “Hani.”

“No.”

(Joola)

Context: “Where is Role?”

11) “Na ora-nje jet.”

on plow-NLMZ be.PRS.3SG

“He is out plowing.”

(Macedonian)

Frame topic

Scene-setting topics, adverbs, prepositional phrases, event sequence markers

Context: And then:

12) “Ezeṇefi w'ali.”
VNEG-FUT-1PL.INCL-1PL.INCL fight
“We will not fight.”

(Udihe)

Context: In the winter:

13) “Vörmönö kaj!”
forest-ACC/DAT go
“Go to the forest!”

(Chuvash)

Description

A clausal description of the setting that does not involve some action by the speaker. Usually the speaker reacts to the scene that is being described.

Context: [There is] a cart, there is a cart there or something... /coughing/.

14) “Ajta.”

“Let’s go.”

(Chuvash)

Context: It is like sorcery that will make the whole room smell bad!

15) “Nier a-pa-lol ë-íne a-hu a-u fu-ton
thus NC1-friend-3SG.POSS NC1-man NC1-DEF NC1-DEF NC7-room
fa-fu fa-pul-e me.”

NC7b-DEF NC7b-rot-PFV COMP

“So my friend, you are the one is making the bad smell.”

(Joola)

Interruption

There is a dialogue structure, but the narrator interrupts it to make a comment.

Context: “You should make this one to put in the sauce of the guest to have breakfast first. Then we we will tie up the body.” Hare heard it, [but] Hare did not say anything.

16) “M̄ gá rā lèè pjē lē cjǣ=ʔǣ nī à wòlò
2SG COND IPFV DEM3.SG do put sauce=DEF2.SG in CIPRT 1PL
m̄=ʔǣ wòlò kè wó nǎ wó
mouth=DEF2.SG 1PL CFPRT 1PL.SBJV.AFF then 1PL.SBJV.AFF
ká rā kú=w pwūū.”
later IPFV dead=DEF1.SG tie.IPFV

“If you had made this one to put in the sauce to take the breakfast, before we begin to prepare the body.” (Kafire)

Change of participant

Like a perspectivizing clause, this context involves a clause with a verb, however, in this case it is not the agent/subject of this event that also is the speaker of the following speech report, but some other participant involved in the event.

Context: He went back and came to his sister.

17) “Nexus'e emegi-jeu?”

younger_sibling come-REV-Q

"Little brother, you have returned?"

(Udihe)

Context types: findings

- Thus, besides dialogues, **descriptions** of the scene is also a commonly used way to introduce non-canonical reported speech.
- Dialogues can also be **interrupted** by the narrator to give a comment.
- And even perspectivizing clauses do not always predict which participant will be the speaker (contra Verstraete 2011 and Spronck 2017)
 - Certain verbs “put a specific participant into perspective and thus anticipate a shift to their speech or thought” (Verstraete, 2011: 498-499).
 - Spronck (2017): co-referentiality between little-m (= perspectivizing clause) and reported speaker.

The nature of the perspectivizing clauses

Is the perspectivizing structure universal?

- If yes, then we expect the majority of these constructions to look like [CLAUSE R], independent of the typological features of the language.
- This is a **universally available construction** independent of the grammar of the language.
- In line with Spronck (2017): there are no syntactic dependencies between little-m and R in Ungarinyin.

Is the perspectivizing structure universal?

- Alternative view, in the spirit of CxG: the perspectivizing clauses get **interpreted as introducers** of reported discourse, because they appear in a structure that usually introduces discourse.
- Similar to constructions such as ‘Paul crunches into it’ in which ‘crunch’ is inserted in the canonical slot for ‘bite’.
- In this case we expect them to **only** appear in structures which are canonical of speech reporting for any given language.

How to find this out

- What is the canonical order of M and R in the language?
- Do perspectivizing clauses adhere to this order?

Evidence from our corpus

- In most languages the order of both normal M clauses and perspectivizing clauses is [CLAUSE R], so this doesn't provide any negative evidence. But!
- **Turkic and Udihe:** While in all languages, perspectivizing clauses (can) precede the speech report, even the Turkic languages (Chuvash and Bashkir) in which the quotative verbs strictly follow the speech report.
- **Macedonian:** In Macedonian, the quotative verb speech verb can occur in multiple positions in relation to the reporting clause, however perspectivizing clauses **only** precede the speech report

Chuvash perspectivizing clause vs quotative verb

17) Olaşkintæn **axlatsa** **jatëë.**
remainder-POSS_3-ABL groan-CV_COORD let-PST-3

“Poşalstë orəx an tivër.”

please\RUS other PROH touch-IMP.2PL

‘At the end she started moaning. “Please do not touch [me] anymore”.’

18) “Ep sik-se təop tö-va-təp,” **tet.**
1SG jump-CV_COORD kiss do-PRS-1SG say-PRS.3SG

“I will jump [and] kiss [her],” he says.’

Macedonian **perspectivizing clause** vs **quotative verb**

- 19) **Posla-la** jambolija-ta, uboo, i: “Ajde da...
spread.AOR-LPTCP.F shag_carpet-DEF.F pretty-N and come_on SUBJ
ajde, ajde, kum-e, bujrum, ajde.”
come_on come_one godfather-VOC welcome come_one
‘She rolled out the rug, nicely, and: “Come, come, godfather, welcome, come.”’

- 20) “A,” **reče**, “Itrata lisica šo mi na-pravi-la!”
ah say.PFV.3SG cunning fox what 1SG.DAT PFV-make-LPTCP.F
“Ah,” he said, “What the cunning fox has done to me!””

Perspectivizing clauses are different from M-clauses

The perspectivizing clause is a **universally available construction** independent of the grammar of the language (in line with Spronck 2017).

What licenses perspectivizing clauses?

- But we have found that the verbs that can occur in this construction are more general than previously thought.
- So what is it about these subtypes of verbs that makes them felicitous in environments normally reserved for speech verbs?
- Nonetheless, there must be some generalizations about the verbs, cf:

21a) She walks into the room. “Hello everybody!”

21b) She sleeps. “Hello everybody!”

Discussion

Conclusion

- All of the previously described types of contexts that can precede a non-canonical speech report have also been found in our sample of 10 languages.
- But these contexts are more broad than previously thought.
- Novel findings:
 - Perspectivizing clauses can contain action verbs and position verbs.
 - It is not always the agent of a motion event that is the speaker in perspectivizing clauses with a motion verb.
 - Descriptions of a scene can precede a non-canonical speech report.
 - Dialogues can be interrupted by a comment from the narrator and not marked with a Matrix clause afterwards.
- Perspectivizing clauses are syntactically different from Matrix clauses.

Open issues

- Intonation and voice quality changes have been observed in previous literature (Mathis & Yule 1994, Spronck 2017)
- Also in our corpus prosody seems to play a role, but can we systematically look at it?
- Gesture and gaze also definitely seem to play a role judging from the videos, but thus far we have no systematic way of annotating this.

Thank you!

Thank you to you, the audience, for listening,
to the organizers and reviewers for letting me talk here,
to the ERC for the money,
and most of all to our consultants and data contributors!



SpeechReporting database: data contributors



Katya Aplonova



Izabela Jordanoska



Tatiana Nikitina



Abbie Hantgan-Sonko



Elena Perekhvalskaya



Alexandre Kouadio Djamala



Lacina Songfola Silué



Rebecca Paterson



Guillaume Guitang



Olga Kuznetsova

Visit the corpus yourselves:

<http://discoursereporting.huma-num.fr/index.html>



References

- Chanard, C. (2015). ELAN-CorpA: lexicon-aided annotation in ELAN. In *Corpus-based studies of lesser-described languages*, 311–332. John Benjamins
- D'Arcy, A. (2015). Quotation and advances in understanding syntactic systems. *Annual Review of Linguistics*, 1(1), 43-61.
- Güldemann, T. (2008). *Quotative indexes in African languages*. De Gruyter Mouton.
- Mathis, T., & Yule, G. (1994). Zero quotatives. *Discourse processes*, 18(1), 63-76.
- Nikitina, T., Aplonova, E., Jordanoska, I., Hantgan-Sonko, A., Guitang, G., Kuznetsova, O., Perekhvalskaya, E. & Silué, L. (eds.) 2022. *The SpeechReporting Corpus: Discourse Reporting in Storytelling*. CNRS-LLACAN & LACITO, <http://discoursereporting.huma-num.fr/index.html>
- Spronck, S. (2017). Defenestration: Deconstructing the frame-in relation in Ungarinyin. *Journal of Pragmatics*, 114, 104-133.
- Teptiuk, D. (2019). *Quotative indexes in Finno-Ugric (Komi, Udmurt, Hungarian, Finnish and Estonian)*. Doctoral dissertation, Tartu University.
- Vandelanotte, L. (2009). *Speech and thought representation in English*. Berlin/New York: Mouton De Gruyter.
- Verstraete, J. C. (2011). The functions of represented speech and thought in Umpithamu narratives. *Australian Journal of Linguistics*, 31 (4), 491-517.