Guide to the annotation of reported discourse

Introduction

This document explains the principles of annotation for reported discourse in the SpeechReporting corpus, as well as the best ways to do annotation-based searches in the Tsakorpus interface.

The annotation template can be downloaded here:
http://discoursereporting.huma-num.fr/template_with_Continuation.etf
It should be cited as:
Nikitina, Tatiana, Hantgan-Sonko Abbie & Chanard Christian. 2019. *Reported speech annotation template for ELAN* (The SpeechReporting Corpus). Villejuif-Paris: LLACAN.

Annotation of reported discourse consists of four levels: the construction's syntactic type; the function of the construction's elements; the semantic type of the discourse report; and the encoding of participants inside the discourse report. They correspond to four additional ELAN tiers (in our template, qt, rp, typ and par, respectively).

This is an example of our annotation of a Bashkir sentence in an ELAN file:

In our corpus interface, annotations are reflected by background colors, frames, and pop-up windows. Explained below are the basic principles of annotation that are relevant to searching in the corpus interface.

The function of the construction's elements

Reported speech construction consists of different elements; for example, in *John said: Hello!* the reported utterance (*Hello!*) is introduced by a clause describing the speech event (*John said*).

The functions of the different elements are indicated by different background colors:
- Reported discourse is annotated as "Discourse Report" and appears in green.
- Elements introducing reported discourse are annotated as "DR Event" and appear in pink.
- Some of our languages use special grammatical markers to introduce discourse reports which are known as quotative particles or quotative verbs; such markers are annotated as "Quotative" and appear in blue.
- Reported discourse may also be introduced by a "Demonstrative", as in *He said **like this**: Hello!*; such elements appear in orange.

- Sometimes reported discourse is introduced by phrases that normally do not refer to speaking or thinking, as in *She enters the room: Hello!* Such phrases are annotated as describing a "Non-speech event" but are not marked by any particular color.
- Gray is used for elements that are not covered by any of the categories described above. The examples of such elements are rather rare and vary widely across languages. They are annotated as "Other".

In the example below (from Bashkir), *nisek juq* ('Why not?') is a Discourse report, the grammaticalized quotative verb *tip* is a Quotative, and *ajəw qəsqərəp ebärgän* ('the bear cried') is a DR Event:

nisek	juq	tip	ajəw	qəsqərəp	ebärgän
nisek	**juq**	**ti**	**ajəw**	**qəsqər**	**ebär**
nisek	juq	**ti–p**	ajəw	qəsqər–p	ebär–ɣan
how	no	**say–B.CV**	bear	cry–B.CV	leave–PC.PST
how	no	**say**	bear	cry	leave

"Why not?" – the Bear cried

In this example from Mwan, *ɓέɛ̀yí* is a Demonstrative, *káá* is a Quotative, and the portion in green is a Discourse report.

ò	pē	ɓέɛ̀yí	káá	ò	yéē	Gbàgbōò	ā	màfá	glɔ̀	mū	έ	kpέ	síέ.
ò	pē	ɓέyí	káá	ò	yéē	Gbàgbōò	ā	màfá	glɔ̀ɔ̀	mū	έ	kpέ	síέ
ò	pē	ɓέyí	káá	ò	yéē	Gbàgbōò	ā	màfá	glɔ̀ɔ̀	mū	έ	kpέ	síέ
3Pl	say	so	that	3Pl	3SG.EMPH	Gbagbo	POSS	gun	hard	PL	ART	all	spoil
pron	v	adv	conj	pron	pron	n	conj	n	adj	det	det	det	v
3Pl	say	so	that	3Pl	3SG.EMPH	Gbagbo	POSS	gun	hard	PL	ART	all	spoil

They said that all Gbagbo's heavy weapons would be destroyed.

You can search for elements of speech reporting constructions using the "Rs elements" box in the menu (Rs stands for Reported speech):

The construction's syntactic type

Different syntactic types of reported discourse constructions are visually represented by different frames. The types are defined by the elements the construction consists of. For example, this Bashkir construction consists of a DR Event, a Quotative element, and a Discourse Report (as explained above). This combination is annotated as "DR Event + Discourse Report + Quotative" (independent of the actual order in which the elements appear). It is represented visually as surrounded by a **red dotted line** :

nisek	juq	tip	ajəw	qəsqərəp	ebärgän
nisek	juq	ti	ajəw	qəsqər	ebär
nisek	juq	ti-p	ajəw	qəsqər-p	ebär-ɣan
how	no	say-B.CV	bear	cry-B.CV	leave-PC.PST
how	no	say	bear	cry	leave

"Why not?" – the Bear cried

In this Chuvash example, the construction consists of a Discourse Report and a Quotative marker. This type, annotated as "Discourse Report + Quotative" is represented visually by a **solid red frame** :

a	vörlas[a]	kajsan	?	tet
a CONJ	vörla VERB	kaj VERB		te VERB
a	vörla-sa	kaj-san		te-atʲ
CNJ\RUS	steal-CV_COORD	go-CV_ANT		say-PRS.3SG
CNJ\RUS	steal	go		say

А вӑ҃рлас[a] кайсан? тет.

And what if they get stolen? – he says.

This Jóoola example consists of a DR Event and a Discourse Report. This type is annotated as "DR Event + Discourse Report" and represented visually by a **solid blue frame** :

naage	yo
ek v	yo interj
n-a-ek-e	yo
LOC-NC1-say-PFV	ok
say	ok

She says yes.

3

In this Udihe example, the Discourse Report is not introduced by anything, because it is part of a larger report, and only the first sentence in that report is introduced explicitly by a special element (in this case, a DR Event which is not shown in the example). Instances of continuing reports that do not have their own marking are annotated as "Continuation" and represented visually by a solid green frame :

– 📖

sinewe	eimi	isegi
si PRON	ei VERB	isegi VERB
si–wA	ei–mi	isegi
2SG–ACC	VNEG–1SG	look_up
pron, –case1	neg, –pers	v
☑ RL–∅	∅–RS	look_up
2SG	VNEG	

I don't see you

In the next Udihe example we see again a Discourse Report that is not introduced by any other element. This time, however, it does not continue previously introduced reported speech; rather, it is understood as reported discourse in context. Such instances are annotated as "Bare Quote" and represented by a green dotted frame :

– 📖

Jeuxi	ŋeneheni?
j'eu PRON	ŋene VERB
j'eu–xi	ŋene–hA–ni
what–DIR	go–PST–3SG
pron, –case	v, –TAM1, –pers
☑ what	go

"Where did he go?"

Less frequent construction types that are not associated with any special visual representation are: "DR Event inside Discourse Report", "Quotative inside Discourse Report", and "Discourse report inside DR Event". They are only identified in a pop-up window, as illustrated below with a Bashkir example of the type "Quotative inside Discourse Report":

↓ Back to search

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

Search result: 289 occurrences, 30 sentences found in approximately 16 documents.

– 📖

ä	tölkölär	yümer	bujə	šul	qujandarðə	eðärlekläjðär	ti	haman
ä CONJ	tölkö NOUN	yümer NOUN	buj NOUN	šul PRON	qujan NOUN	eðärleklä VERB	ti VERB	haman ADV
ä	tölkö–lar	yümer	buj–hə	šul	qujan–lar–nə	eðärleklä–a–lar	ti	haman
and	fox–PL	life	length–P.3	that	hare–PL–ACC	chase–IPFV–PL	say:A.CV	still
conj	suf	life	nsuf	pron	suf, nsuf	v, vsuf, suf	v	adv
and	fox		length	that	hare	chase	say:A.CV	still

da
=la PART
=la
and
part
☑ and

haman
haman ADV
haman
still
gr: adv
trans_en: still

And foxes chase hares all the time,

Finally, languages sometimes use other constructions, which do not fit into any of the types discussed above. Such constructions are annotated as "Other" and represented visually by a [blue dotted frame] .

The following example from Bashkir contains two DR Events followed by evidential quotatives (*ti*), which does not correspond to any of the recognized construction types and is annotated as "Other".

bəna	tölkö	qəsqərəp	jörögän	ti	–ana	majðə	ajəw	ašaɣan	,	tip	äjtä	ti
bəna	tölkö	qəsqər	jörö	ti	ana	maj	ajəw	aša		ti	äjt	ti
bəna	tölkö	qəsqər–p	jörö–ɣan	ti	ana	maj–nə	ajəw	aša–ɣan		ti–p	äjt–a	ti
there	fox	cryB.CV	goPC.PST	say:A.CV	there	butterACC	bear	eatPC.PST		sayB.CV	tellIPFV	say:A.CV
pron	n	v, vsuf	v, vsuf	v	part	n, nsuf	n	v, vsuf		v, vsuf	v, vsuf	v
there	fox	cry	go	say:A.CV	there	butter	bear	eat		say	tell	say:A.CV

täne	majlə
tän	maj
tän–hə	maj–lə
bodyP.3	butterADJ
n, nsuf	n, suf
body	butter

and the Fox went and screamed: "The Bear ate all the butter, he has butter on his body!"
evidential quotatives, two DR Events

You can search for construction type using the "Syntactic type" field in the menu:

Multilingual corpus of the Discourse Reporting project   EN |

Word #1

Word:
Lemma:
POS:
Gloss:

Translation (en):
Participant (par):

Speaker:
Rs elements:
Syntactic type (qt) ✓
Semantic type (typ)   DR_Event_Discourse_Report
Analyses   DR_Event_Discourse_Report_Quotative
Position in sentence   Quotative_Discourse_Report
Language/tier   Quotative
                          Bare_Quote
                          Continuation
                          DR_Event_inside_Discourse_Report
Full-text search: Quotative_inside_Discourse_Report   tch
                          Discourse_Report_inside_DR_Event
Search sentences   Other   S

5

Semantic type of the Discourse Report

Discourse Reports are annotated for their semantic type, indicated in pop-ups in the top right corner. The semantic types include the following options:
- Statement,
- Question,
- Command,
- Greeting or Address,
- Expression of Attitude,
- Feeling or Thought,
- Story,
- Song,
- Partial,
- Embedded.

The following examples illustrate some of the semantic types:

- **Statement** (the most common type across the corpus) - example from Ut-ma'in:

- **Question** (Gizey):

- **Command** (Gizey):

↓ Back to search

qt: ['DR_Event_Discourse_Report_Quotative']
rp: ['Quotative', 'Discourse_Report']
speaker:
typ: ['Command']

Search result: 404 occurrences, 43 sentences found in approximately 7 documents.

ʔāl	ʔāzī	kālīj	gū	vàjt	vɔ̄ʔ	vàmū	.
ʔāl CONJ	ʔāzī PRON	kāl VERB	gū PART	vàjt ADV	vɔ̄ʔ ADV	['=àm', 'v'] ADP	
ʔāl	ʔāzī	kāl-īj	gū	vàjt	vɔ̄ʔ	v=àm-ū	
quot1	3pl	go-dest	rev	today	at_home	of=3sm-fv	
CNJ	PRN	V, PTCL	PTCL	ADV	ADV	PREP, PRN	
quot1	RL	go	rev	today	at_home	∅-RS-∅	
	3pl					=3sm, of	

vàmū
['=àm', 'v'] ADP
v=àm-ū

That they should come to his house today.

- **Greeting or Address** (Chuvash):

↓ Back to search

qt: ['DR_Event_Discourse_Report_Quotative']
rp: ['Discourse_Report', 'Quotative']
speaker: SP1
typ: ['Greeting_or_Address']

Search result: 271 occurrences, 38 sentences found in approximately 8 documents.

– repʲattə	! – terë	.
repʲattə NOUN	te VERB	
repʲattə	te-r-ë	
guys\RUS	say-PST-3	
n	v, suff	
guys\RUS	say	

terë
te VERB
te-r-ë
say-PST-3
gr: v, suff
trans_en: say

– Репьаттӑ! – терĕ.

– Guys!

"Ребята!"

- **Expression of Attitude** is most commonly represented by interjections, as in the following example from Chuvash:

↓ Back to search

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

Search result: 1175 occurrences, 169 sentences found in approximately 15 documents.

oj	, tet	, pitë	n[ö]maj	polkkə	pötertëm	, tet	.
oj INTJ	te VERB	pitë ADV	nömaj ADV	polkkə NOUN	pöter VERB	te VERB	
oj	te-atʲ	pitë	nömaj	polkkə	pöter-r-əm	te-atʲ	
INTJ\RUS	say-PRS.3SG	very	much	regiment	be_finished+CAUS-PST-1SG	say-PRS.3SG	
intj	v, suff	adv	adv	n	v, suff	v, suff	
INTJ\RUS	say	very	much	regiment	∅-∅-RS	say	
					be_finished+CAUS		

tet
te VERB
te-atʲ
say-PRS.3SG
gr: v, suff

Ой, тет, питĕ н[ӑ˚]май полккӑ пĕ˚тертĕм, тет.

– Oh, – he says, – I have destroyed very many regiments.

7

- **Feeling or Thought** (Chuvash):

qt: ['DR_Event_Discourse_Report']
rp: ['Discourse_Report', 'DR_Event']
speaker: SP1
typ: ['Feeling_or_Thought']

↓ Back to search

Search result: 133 occurrences, 19 sentences found in approximately 9 documents.

mën	tövas	pət^j	ëntë	?	şotlat^j		şotlat^j		. ..	për	molkatçi
mën PRON	tu VERB	pət^j VERB	ëntë PART		şot NOUN		şot NOUN			për NUM	molkatç NOUN
mën	tu–as	pət^j	ëntë		şot–la–at^j		şot–la–at^j			për	molkatç–ë
what	do–PC_FUT	is_needed	already		thought–VBLZ–PRS.3SG		thought–VBLZ–PRS.3SG			one	hare–ANAPH
prn	v, suff	v	prt		n, suff		n, suff			num	n, suff
what	do	is_needed	already		thought		thought			one	hare

kalat^j	. ..	për	çin	tövat^j	. ..	për	mën	ëntë	. ..	pəxkalas[a]
kala VERB		për NUM	çin NOUN	tu VERB		për NUM	mën PRON	ëntë PART		pəx VERB
kala–at^j		për	çin	tu–at^j		për	mën	ëntë		pəx–kala–sa
speak–PRS.3SG		one	person	do–PRS.3SG		one	what	already		look–ITER–CV_COORD
v, suff		num	n	v, suff		num	prn	prt		v, suff
speak		one	person	do		one	what	already		look

çyrerë	çyrerë	, çyrerë	çyrerë	te
çyre VERB	çyre VERB	çyre VERB	çyre VERB	ta PART
çyre–r–ë	çyre–r–ë	çyre–r–ë	çyre–r–ë	ta
walk–PST–3	walk–PST–3	walk–PST–3	walk–PST–3	and
v, suff	v, suff	v, suff	v, suff	prt
walk	walk	walk	walk	and

te
ta PART
ta
and
gr: prt
trans_en: and

Мĕн тă°вас–пăть ĕнтĕ? Шотлать–шотлать... Пĕр мо[...] пĕр çын тă°вать... пĕр мĕн ĕнтĕ... пăхкалас[а] çÿрерĕ–çÿрерĕ, çÿрерĕ–çÿрерĕ те:

What is to be done now? He is thinking and thinking... One hare says... one person does... one that one... he walked and walked [around] inspecting, and:

- Annotation "**Story**" is for evidential uses ("They say that once upon a time…") as well as impersonal naming uses ("This is called X"):

Story (Jóola Eegimaa)

guogol	Fatou	; Fatou jawol	nabugol	nevonol	; Fatou ndakkay	fúrësór	;
ek VERB	NOUN	NOUN	bug NOUN	nevon ADV	NOUN	rës VERB	
gu–ek–ol		er–3SG.POSS	n–a–bug–ol	nevon–ol	kkay ADV	fu–rës–or	
3PL–say–3SG.DO			LOC–NC1–descendent–3SG.DO	alone–3SG.DO	n–a–kkay	NC7–play–REFL	
say			descendent	alone	LOC–NC1–leave	play	
					leave		

Fatou
NOUN
trans_en:

qt: ['DR_Event_Discourse_Report']
rp: ['DR_Event', 'Discourse_Report']
speaker: SP1
typ: ['Story']

her name was Fatou. Fatou is the only girl of her mother. Fatou will never play.

- Annotation "**Song**" subsumes representations of sound rather than meaning. Typical examples are sung portions of the text (which may often be uninterpretable or include nonce words), but only if they represent character speech (some songs represent action, and are not annotated). Also common are onomatopoeias representing sounds produced by non-human characters (again, when introduced as instances of reported speech), as in the following example from Udihe:

Keku	digankini:	-	Keeku,	keeku,	keeku.	-	Keku	jewe	diganaini?
keku	digana		keeku	keeku	keeku				digana
keku	digana–ki–ni		keku	keku	keku				digana–ini
cuckoo	tell–PST–3SG	-	cuckoo	cuckoo	cuckoo				tell–3SG
n	v, –suf	–suf	onomat	onomat	onomat				v, –suf
cuckoo	tell		cuckoo	cuckoo	cuckoo				tell

keeku.
keku
keku
cuckoo
gr: onomat

qt: ['DR_Event_Discourse_Report']
rp: ['Discourse_Report']
speaker: SP1
typ: ['Song']

The Cuckoo said: "Cuckoo, cuckoo, cuckoo[...] cuckoo say?"

- **Partial** is used to annotate syntactically incomplete reports, which cannot stand on their own and do not have a semantic type on their own. These include interruptions where a speech report is left unfinished. The following example is from Chuvash:

- **Embedded** is used to identify Discourse reports embedded under other reports (for example, when a character in the story reports speech by another character). This type overrules any other semantic type that a discourse report may have (Statement, Question, etc.); the commentary tier can be used for additional information if needed. The following is an example of the Embedded category in Joola Eegimaa:

You can search for semantic type using the corresponding field in the search menu:

Reference to participants

In the Participant tier (par) we annotate all elements referring to participants in the current or reported speech event, such as current and reported speakers and listeners. The elements referring to such participants are typically pronouns and agreement markers on the verb. They include the following options (illustrated below with examples):

- Reported speaker (RS): *John$_i$ said **he**$_i$ was lucky* or *John$_i$ said: **I**$_i$ am lucky!*
- Reported listener (RL): *John told Mary$_i$ **she**$_i$ was lucky* or *John told Mary$_i$: **You**$_i$ are lucky!*
- Current speaker (CS): *John$_i$ told Mary **I**$_j$ was lucky*
- Current listener (CL): *John$_i$ told Mary **you**$_j$ were lucky*
- Speaker + Listener (all combinations of speakers and listeners, reported or current): *John$_i$ told Mary$_j$ **they**$_{i+j}$ were lucky* or *John$_i$ told Mary$_j$: **We**$_{i+j}$ are lucky!*

Reference to participants is reflected in the pop-ups that appear to the right of every word in the example.

In this Chuvash example, the pronoun *mana* 'me' refers to the **Reported Speaker (RS):**

mana	an	vëlerër	tolʲkkə	, – tet
man PRON	an PART	vǒler VERB	tolʲkkə ADV	te VERB
man–a	an		tolʲkkə	te–atʲ
1SG.OBL–ACC/DAT	2PL	only\RUS	say–PRS.3S
prn, suff		adv	v, suff
RS–∅		only\RUS	say
1SG.OBL			

mana
man PRON
man–a
1SG.OBL–ACC/DAT
gr: prn, suff
par: RS–∅
trans_en: 1SG.OBL

Мана ан вĕлер

just do not kill

– Только не уб... ...рит.

In the same Chuvash example, the verb suffix (imperative plural) is analyzed as referring to the **Reported Listener (RL)**:

mana	an	vëlerër	tolʲkkə	, – tet	.
man PRON	an PART	vǒler VERB			
man–a	an	vǒler–ər			
1SG.OBL–ACC/DAT	PROH	die+CAUS–IMP.2PL			SG
prn, suff	prt	v, suff			
RS–∅	PROH	∅–RL			
1SG.OBL		die+CAUS			

vëlerër
vǒler VERB
vǒler–ər
die+CAUS–IMP.2PL
gr: v, suff
par: ∅–RL
trans_en: die+CAUS

Мана ан вĕлерĕр тольккă, – тет.

just do not kill me, – he says.

In the following two examples from Mwan, reference is made to the **Current Speaker (CS)** and the **Current Listener (CL)**, respectively:

ɓέὲ	ɓέ	pē	mí	nī	káádɔ̄ɔ̄	mí	nú	ɓέ	bā.
ɓέὲ	ɓέ	pē	mí	nī	['dɔ̄ɔ̄', 'káá']	mí	nū	ɓέ	bā
ɓέὲ	ɓέ	pē	mí	nī	káá–dɔ̄ɔ̄	mí	nū	ɓέ	bā
then	3SG.ANAPH	say	1SG.EMPH	DAT	that–say	1SG.EMPH	pstp	conj	
then	3SG.ANAPH	say	1SG.EMPH	DAT	say, that	pron			
						CS			
						1SG.EMPH			

mí
mí
mí
1SG.EMPH
gr: pron
par: CS
trans_en: 1SG.EMPH

And she tells me to come to their house.

Golu.135

Ké	yà	pèà	yòò	é	dɛ̄	zí,	yè	é	dɛ̄	là	wō
Ké	yà	pē	yòò	é	dɛ̄	zí	è	é	dɛ̄	là	wō
Ké	yà	pē–à	yòò	é	dɛ̄				dɛ̄	là	wō
and	3SG>3SG	say–PRF	3SG.COP	2SG	kil	v			ǩill	FOC.NSBJ	do
conj	pron	v, –mrph	pron	pron	v				v	pron	v
and	3SG>3SG	say	RS	CL	kil				ǩill	FOC.NSBJ	do
			3SG.COP	2SG							

é
é
é
2SG
gr: pron
par: CL
trans_en: 2SG

If he decided to kill you, he will do

Armand_Dominique.046

11

The following Chuvash example illustrates reference to **Speaker + Listener**. Here, the reported speaker (a girl's lover) refers to both himself and the girl, while addressing her:

pire	[] savəşma	tɕərmantaratʲ	, tet	, pirën	[ona]	vŏleres	pətʲ	, tet	,
pir	savəş	tɕərman	te	pir	on	vŏler	pətʲ	te	
pir-a	savəş-ma	tɕərman-tar-atʲ	te-atʲ	pir-n	on-a	vŏler-as	pətʲ	te-atʲ	
1PL.OBL-ACC/DAT	be_in_love-INF	worry-CAUS-PRS.3SG	say-PRS.3SG	1PL.OBL-GEN	3SG.OBL-ACC/DAT	die+CAUS-PC_FUT	is_needed	say-PRS.3SG	
prn, suff	v, suff	v, suff	v, suff	prn, suff	prn, suff	v, suff	v	v, suff	
Speaker_Listener-Ø	be_in_love	worry	say	Speaker_Listener-Ø	3SG.OBL	die+CAUS	is_needed	say	
1PL.OBL				1PL.OBL					

пире [] савăшма чăрмантарать, тет, пирĕн [она] вĕ°лерес-пăть, тет,

- he prevents us from making love, - he says, - we need to kill him, - he says,

You can search for annotation for participant reference using the Participant field in the search menu:

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