

In search of episodic units:

A corpus-based study of Wan and Chuvash

Tatiana Nikitina

March 21, 2024

Workshop on Larger Discourse Units



Are larger-than-sentence units real?

Are larger-than-sentence units real?

- All units of analysis are problematic to some degree, with different kinds of challenges, cf. *word* or *sentence*.

Are larger-than-sentence units real?

- All units of analysis are problematic to some degree, with different kinds of challenges, cf. *word* or *sentence*.
- How do we make sure that our “units” are not imaginary (in text, corpus, language), and that they are cross-linguistically applicable?

Are larger-than-sentence units real?

- All units of analysis are problematic to some degree, with different kinds of challenges, cf. *word* or *sentence*.
- How do we make sure that our “units” are not imaginary (in text, corpus, language), and that they are cross-linguistically applicable?
- “The primary drawback to the episode/paragraph approach lies in the difficulty of providing explicit and structure-independent means of identifying episodes and episode boundaries” (Tomlin 1987: 475).

Are larger-than-sentence units real?

- All units of analysis are problematic to some degree, with different kinds of challenges, cf. *word* or *sentence*.
- How do we make sure that our “units” are not imaginary (in text, corpus, language), and that they are cross-linguistically applicable?
- “The primary drawback to the episode/paragraph approach lies in the difficulty of providing explicit and structure-independent means of identifying episodes and episode boundaries” (Tomlin 1987: 475).
- If such units exist, they should correspond to the use of **specific structural means**, and at the same time, be associated with **specific meaning or function**.

- Develop a **method** for identifying morphosyntactic phenomena that are sensitive to episodic units in narratives.

- Develop a **method** for identifying morphosyntactic phenomena that are sensitive to episodic units in narratives.
- Ideally, such a method should be:
 - cross-linguistically applicable,
 - based on objective criteria and form-independent definitions.

- Develop a **method** for identifying morphosyntactic phenomena that are sensitive to episodic units in narratives.
- Ideally, such a method should be:
 - cross-linguistically applicable,
 - based on objective criteria and form-independent definitions.
- The method should rely on **quantitative analysis** of corpus data; and it should be **allowed to fail** to yield results in case of insufficient or low quality data.

- Annotate traditional stories for thematic and structural units;

Proposed method

- Annotate traditional stories for thematic and structural units;
- examine the **frequency of use** of grammatical markers and discourse particles in sequence-internal clauses, compare it to their frequency in sequence-initial clauses;

Proposed method

- Annotate traditional stories for thematic and structural units;
- examine the **frequency of use** of grammatical markers and discourse particles in sequence-internal clauses, compare it to their frequency in sequence-initial clauses;
- look for **asymmetries in the distributions** that would suggest that a particular marker is sensitive to episodic structure (signals for either episode boundaries or continuity within segments);

Proposed method

- Annotate traditional stories for thematic and structural units;
- examine the **frequency of use** of grammatical markers and discourse particles in sequence-internal clauses, compare it to their frequency in sequence-initial clauses;
- look for **asymmetries in the distributions** that would suggest that a particular marker is sensitive to episodic structure (signals for either episode boundaries or continuity within segments);
- assess the method's **cross-linguistic applicability**: comparison with data from a typologically different language.

- Data from the multilingual SpeechReporting corpus (Nikitina et al. 2022): <http://discoursereporting.huma-num.fr/corpus.html>

- Data from the multilingual SpeechReporting corpus (Nikitina et al. 2022): <http://discoursereporting.huma-num.fr/corpus.html>
- The Wan subcorpus:
 - a South Mande language spoken in Côte d'Ivoire,
 - a thriving storytelling tradition,
 - surrounded by neighbors with equally thriving traditions (some of them also represented in the corpus: Mwan, Guro).
 - isolating, tonal, rigid S(Aux)OVX.

- Data from the multilingual SpeechReporting corpus (Nikitina et al. 2022): <http://discoursereporting.huma-num.fr/corpus.html>
- The Wan subcorpus:
 - a South Mande language spoken in Côte d'Ivoire,
 - a thriving storytelling tradition,
 - surrounded by neighbors with equally thriving traditions (some of them also represented in the corpus: Mwan, Guro).
 - isolating, tonal, rigid S(Aux)OVX.
- The Chuvash subcorpus:
 - a Turkic language spoken in Russia,
 - storytelling no longer widely practiced,
 - recordings from semi-professional storytellers,
 - agglutinating, vowel harmony, flexible SOV.

- clause-level annotation in ELAN,
- a combination of criteria for narrative analysis (Labov 1972) and thematic unity (Dijk 1981),
- some of the categories are culture- and genre-specific,
- with special attention to narrative event sequences.

Structural elements external to the storyline (1)

Structural elements external to the storyline (1)

- **Frame:** elements motivated by the story's traditional structure

Structural elements external to the storyline (1)

- **Frame:** elements motivated by the story's traditional structure
 - the title song;

Structural elements external to the storyline (1)

- **Frame:** elements motivated by the story's traditional structure
 - the title song;
 - opening and closing formulas:
 - introductory song + 'This [song] has a meaning';
 - 'I'm going to explain it to you';
 - 'Let me kill it'.

Structural elements external to the storyline (1)

- **Frame:** elements motivated by the story's traditional structure
 - the title song;
 - opening and closing formulas:
 - introductory song + 'This [song] has a meaning';
 - 'I'm going to explain it to you';
 - 'Let me kill it'.

 - 'That was my lie of the evening';
 - 'If you see me silent, I am finished';
 - 'I am pouring it to X's stomach'...

Structural elements external to the storyline (1)

- **Frame:** elements motivated by the story's traditional structure
 - the title song;
 - opening and closing formulas:
 - introductory song + 'This [song] has a meaning';
 - 'I'm going to explain it to you';
 - 'Let me kill it'.

 - 'That was my lie of the evening';
 - 'If you see me silent, I am finished';
 - 'I am pouring it to X's stomach'...
 - directions helping the narrator manage transition from one structural element to another (e.g., the song):
 - 'Let us do the song';
 - 'Now let it dry out'.

Structural elements external to the storyline (1)

- **Frame:** elements motivated by the story's traditional structure
 - the title song;
 - opening and closing formulas:
 - introductory song + 'This [song] has a meaning';
 - 'I'm going to explain it to you';
 - 'Let me kill it'.

 - 'That was my lie of the evening';
 - 'If you see me silent, I am finished';
 - 'I am pouring it to X's stomach'...
 - directions helping the narrator manage transition from one structural element to another (e.g., the song):
 - 'Let us do the song';
 - 'Now let it dry out'.

Structural elements external to the storyline (1)

- **Frame:** elements motivated by the story's traditional structure
 - the title song;
 - opening and closing formulas:
 - introductory song + 'This [song] has a meaning';
 - 'I'm going to explain it to you';
 - 'Let me kill it'.

 - 'That was my lie of the evening';
 - 'If you see me silent, I am finished';
 - 'I am pouring it to X's stomach'...
 - directions helping the narrator manage transition from one structural element to another (e.g., the song):
 - 'Let us do the song';
 - 'Now let it dry out'.

Structural elements external to the storyline (2)

- **Coda:** explanation of the story's meaning and implications, at the end of the story (preceding the closing formula)

Structural elements external to the storyline (2)

- **Coda:** explanation of the story's meaning and implications, at the end of the story (preceding the closing formula)
 - 'That is why the Hyena and the Hare are always at war';

Structural elements external to the storyline (2)

- **Coda:** explanation of the story's meaning and implications, at the end of the story (preceding the closing formula)
 - 'That is why the Hyena and the Hare are always at war';
 - 'If someone smaller than you is at the water source, you do not chase them away from it';

Structural elements external to the storyline (2)

- **Coda:** explanation of the story's meaning and implications, at the end of the story (preceding the closing formula)
 - 'That is why the Hyena and the Hare are always at war';
 - 'If someone smaller than you is at the water source, you do not chase them away from it';
 - 'That was the origin of all gossip';

Structural elements external to the storyline (2)

- **Coda:** explanation of the story's meaning and implications, at the end of the story (preceding the closing formula)
 - 'That is why the Hyena and the Hare are always at war';
 - 'If someone smaller than you is at the water source, you do not chase them away from it';
 - 'That was the origin of all gossip';
 - 'It's because she killed the man that the sick have trouble marrying'...

Story-internal structural elements (1)

- **Evaluation:** deviations from storyline adding to the story's comprehensibility or relevance to the audience

Story-internal structural elements (1)

- **Evaluation:** deviations from storyline adding to the story's comprehensibility or relevance to the audience
 - remarks relating the story to the here and now:
‘It means that if he had not done it, then if the bush spirit appeared in our field, he would kill us and take our food’;

Story-internal structural elements (1)

- **Evaluation:** deviations from storyline adding to the story's comprehensibility or relevance to the audience
 - remarks relating the story to the here and now:
 - 'It means that if he had not done it, then if the bush spirit appeared in our field, he would kill us and take our food';
 - 'He went until he reached a new village, say at this distance. And the girls, they are in, say, Kounahiri.'
 - rhetorical questions and comments evoking the audience's emotional response:
 - 'Imagine, you're eating, and they close the door on you!';

Story-internal structural elements (1)

- **Evaluation:** deviations from storyline adding to the story's comprehensibility or relevance to the audience
 - remarks relating the story to the here and now:
 - 'It means that if he had not done it, then if the bush spirit appeared in our field, he would kill us and take our food';
 - 'He went until he reached a new village, say at this distance. And the girls, they are in, say, Kounahiri.'
 - rhetorical questions and comments evoking the audience's emotional response:
 - 'Imagine, you're eating, and they close the door on you!';
 - 'Mister, when a woman likes a man, what happens?'
 - metanarrative or pseudo-narrative comments:
 - 'I'll be brief';

Story-internal structural elements (1)

- **Evaluation:** deviations from storyline adding to the story's comprehensibility or relevance to the audience
 - remarks relating the story to the here and now:
 - 'It means that if he had not done it, then if the bush spirit appeared in our field, he would kill us and take our food';
 - 'He went until he reached a new village, say at this distance. And the girls, they are in, say, Kounahiri.'
 - rhetorical questions and comments evoking the audience's emotional response:
 - 'Imagine, you're eating, and they close the door on you!';
 - 'Mister, when a woman likes a man, what happens?'
 - metanarrative or pseudo-narrative comments:
 - 'I'll be brief';
 - 'Oh, I forgot the song!'

Story-internal structural elements (1)

- **Evaluation:** deviations from storyline adding to the story's comprehensibility or relevance to the audience
 - remarks relating the story to the here and now:
 - 'It means that if he had not done it, then if the bush spirit appeared in our field, he would kill us and take our food';
 - 'He went until he reached a new village, say at this distance. And the girls, they are in, say, Kounahiri.'
 - rhetorical questions and comments evoking the audience's emotional response:
 - 'Imagine, you're eating, and they close the door on you!';
 - 'Mister, when a woman likes a man, what happens?'
 - metanarrative or pseudo-narrative comments:
 - 'I'll be brief';
 - 'Oh, I forgot the song!'
 - 'Oh, this cannot even be told!'

Story-internal structural elements (2)

- **Setting:** introduction of characters at the beginning of the story
 - ‘The members of the Klen cult were once taken to the secret bush’;

Story-internal structural elements (2)

- **Setting:** introduction of characters at the beginning of the story
 - ‘The members of the Klen cult were once taken to the secret bush’;
 - ‘There was a Hare whose cunning exceeded anyone else’s’;

Story-internal structural elements (2)

- **Setting:** introduction of characters at the beginning of the story
 - ‘The members of the Klen cult were once taken to the secret bush’;
 - ‘There was a Hare whose cunning exceeded anyone else’s’;
 - ‘There was a child, a little orphan’;

Story-internal structural elements (2)

- **Setting:** introduction of characters at the beginning of the story
 - ‘The members of the Klen cult were once taken to the secret bush’;
 - ‘There was a Hare whose cunning exceeded anyone else’s’;
 - ‘There was a child, a little orphan’;
 - ‘There was a village girl who said she would not marry anyone with any scar’.

Story-internal structural elements (2)

- **Setting:** introduction of characters at the beginning of the story
 - ‘The members of the Klen cult were once taken to the secret bush’;
 - ‘There was a Hare whose cunning exceeded anyone else’s’;
 - ‘There was a child, a little orphan’;
 - ‘There was a village girl who said she would not marry anyone with any scar’.
- **Speech:** speech by the characters.

Story-internal structural elements (2)

- **Setting:** introduction of characters at the beginning of the story
 - 'The members of the Klen cult were once taken to the secret bush';
 - 'There was a Hare whose cunning exceeded anyone else's';
 - 'There was a child, a little orphan';
 - 'There was a village girl who said she would not marry anyone with any scar'.
- **Speech:** speech by the characters.
- **Song:** songs representing speech or actions.

Story-internal structural elements (2)

- **Setting:** introduction of characters at the beginning of the story
 - ‘The members of the Klen cult were once taken to the secret bush’;
 - ‘There was a Hare whose cunning exceeded anyone else’s’;
 - ‘There was a child, a little orphan’;
 - ‘There was a village girl who said she would not marry anyone with any scar’.
- **Speech:** speech by the characters.
- **Song:** songs representing speech or actions.
- **Narrative event:** description of the story’s events, typically in the order in which they occurred

Story-internal structural elements (2)

- **Setting:** introduction of characters at the beginning of the story
 - ‘The members of the Klen cult were once taken to the secret bush’;
 - ‘There was a Hare whose cunning exceeded anyone else’s’;
 - ‘There was a child, a little orphan’;
 - ‘There was a village girl who said she would not marry anyone with any scar’.
- **Speech:** speech by the characters.
- **Song:** songs representing speech or actions.
- **Narrative event:** description of the story’s events, typically in the order in which they occurred
 - ‘And the old man sat down under the rice straw’;
 - ‘And he took a gun’;
 - ‘and put the front of it out a little’...

- **Initial event:** the storyline runs into a temporal discontinuity (the event sequence is interrupted by an indeterminate period of time)
 - ‘God said: Tomorrow, come tomorrow. || And the morning came.’

- **Initial event:** the storyline runs into a temporal discontinuity (the event sequence is interrupted by an indeterminate period of time)
 - ‘God said: Tomorrow, come tomorrow. || And the morning came.’
- **Non-initial event:** temporal and often causal continuity with the preceding event.
 - ‘When the morning came, he went to God. God told him...’

- **Initial event:** the storyline runs into a temporal discontinuity (the event sequence is interrupted by an indeterminate period of time)
 - ‘God said: Tomorrow, come tomorrow. || And the morning came.’
- **Non-initial event:** temporal and often causal continuity with the preceding event.
 - ‘When the morning came, he went to God. God told him...’
- **Tail-Head elements:** repeated reference to the same event.
 - ‘And the morning came. When the morning came, he went to God.’

Comparing distributions

- ELAN files are easily searchable: possible to compare frequencies across different types of units.

Comparing distributions

- ELAN files are easily searchable: possible to compare frequencies across different types of units.
- Since this study focuses on episodic structure, i.e. thematic units: only narrative event sequences taken into account.

Comparing distributions

- ELAN files are easily searchable: possible to compare frequencies across different types of units.
- Since this study focuses on episodic structure, i.e. thematic units: only narrative event sequences taken into account.
- Within narrative event sequences, differences in distributions are sought between **sequence-initial** (first clause in the episode) vs. **non-initial** clauses.

Comparing distributions

- ELAN files are easily searchable: possible to compare frequencies across different types of units.
- Since this study focuses on episodic structure, i.e. thematic units: only narrative event sequences taken into account.
- Within narrative event sequences, differences in distributions are sought between **sequence-initial** (first clause in the episode) vs. **non-initial** clauses.
- Common suspects:
 - the choice of referring expressions (pronouns vs. nouns, Fox 1987),
 - conjunctions and discourse particles (Bakker 1993),
 - backgrounding devices such as nominalization (Mithun 2002)
 - the choice of tense-aspect (Schiffrin 1981),
 - prosodic means (pauses) and fillers (not addressed here)...

Results of the comparison

(only 26 of the 82 annotated stories analyzed so far)

Parts of speech

	Initial	Non-initial	(% initial)
pronoun	308		
noun	288		
verb	288		
conjunction	177		
determiner	145		
postposition	99		
adverb	61		
sentence-final particle	60		
auxiliary	45		
aspectual particle	31		
ideophone	28		
adjective	25		
interjection	11		

Parts of speech

	Initial	Non-initial	(% initial)
pronoun	308	1069	22%
noun	288	689	29%
verb	288	840	26%
conjunction	177	514	26%
determiner	145	359	29%
postposition	99	306	24%
adverb	61	125	33%
sentence-final particle	60	85	41%
auxiliary	45	165	21%
aspectual particle	31	75	29%
ideophone	28	158	15%
adjective	25	78	24%
interjection	11	69	14%

Parts of speech

	Initial	Non-initial	(% initial)
pronoun	308	1069	22%
noun	288	689	29%
verb	288	840	26%
conjunction	177	514	26%
determiner	145	359	29%
postposition	99	306	24%
adverb	61	125	33%
sentence-final particle	60	85	41%
auxiliary	45	165	21%
aspectual particle	31	75	29%
ideophone	28	158	15%
adjective	25	78	24%
interjection	11	69	14%

Major asymmetries

	Initial	Non-initial	(% initial)
pronoun	308	1069	22%
adverb	61	125	33%
sentence-final particle	60	85	41%
auxiliary	45	165	21%
ideophone	28	158	15%
interjection	11	69	14%

Referring expressions

- The **ratio** of referring expressions (nouns + pronouns) to verbs is the same in initial and non-initial event clauses: 2 to one.

	Initial	Non-initial
noun + pronoun	596	1758
verb	288	840
(% verbs)	(33%)	(32%)

Referring expressions

- The **ratio** of referring expressions (nouns + pronouns) to verbs is the same in initial and non-initial event clauses: 2 to one.

	Initial	Non-initial
noun + pronoun	596	1758
verb	288	840
(% verbs)	(33%)	(32%)

- The **proportion** of pronouns is significantly higher in non-initial clauses ($p < 0.001$):

	Initial	Non-initial	Total
pronoun	308 (22%)	1069	1377 (100%)
noun	288 (29%)	689	977 (100%)

A closer look at pronouns

- No effect for demonstrative vs. 3SG personal pronouns (difference *not* significant):

	Initial	Non-initial	Total
<i>lɛ</i> 'that one'	54 (27%)	145	199 (100%)
3SG	186 (24%)	603	789 (100%)

A closer look at pronouns

- No effect for demonstrative vs. 3SG personal pronouns (difference *not* significant):

	Initial	Non-initial	Total
<i>bě</i> 'that one'	54 (27%)	145	199 (100%)
3SG	186 (24%)	603	789 (100%)

- A strong effect of pronominal form: free-standing pronoun vs. lengthening ($p < 0.005$).

	Initial	Non-initial	Total
\acute{e} = 3SG.SUBJ	99 (21%)	373	472 (100%)
\grave{a} = 3SG	36 (29%)	89	125 (100%)
: = +3SG	16 (12%)	114	130 (100%)

A closer look at conjunctions

- The conjunction *bé* 'then' is the most frequent word in narrative sequences, yet it shows no asymmetry in its distribution:

	Initial	Non-initial	Total
<i>bé</i> 'then'	147 (25%)	451	598 (100%)
all words	1615 (25%)	4735	6350 (100%)

A closer look at conjunctions

- The conjunction *bé* ‘then’ is the most frequent word in narrative sequences, yet it shows no asymmetry in its distribution:

	Initial	Non-initial	Total
<i>bé</i> ‘then’	147 (25%)	451	598 (100%)
all words	1615 (25%)	4735	6350 (100%)

- The conjunction *ké* ‘if/when’ shows a significant bias toward initial event clauses ($p < 0.01$):

	Initial	Non-initial	Total
<i>bé</i> ‘then’	147 (25%)	451	598 (100%)
<i>ké</i> ‘if/when’	20 (47%)	23	43 (100%)

A closer look at sentence-final particles

- Particles overrepresented in sequence-initial clauses:

	Initial	Non-initial
$m\bar{o}$	16	11
$g\bar{e}$	14	9
$g\bar{e} + m\bar{o}$	6	1
$w\grave{a}$	8	2

A closer look at sentence-final particles

- Particles overrepresented in sequence-initial clauses:

	Initial	Non-initial
$m\bar{o}$	16	11
$g\bar{e}$	14	9
$g\bar{e} + m\bar{o}$	6	1
$w\grave{a}$	8	2

- Event nominalization (also a backgrounding device):

	Initial	Non-initial
\acute{e} NMLZ	12	14

A closer look at sentence-final particles

- Particles overrepresented in sequence-initial clauses:

	Initial	Non-initial
$m\bar{o}$	16	11
$g\bar{e}$	14	9
$g\bar{e} + m\bar{o}$	6	1
$w\grave{a}$	8	2

- Event nominalization (also a backgrounding device):

	Initial	Non-initial
\acute{e} NMLZ	12	14

- Particles overrepresented in non-initial clauses:

	Initial	Non-initial
$w\acute{a}$ NEG	0	32
$d\acute{e}$ 'as soon as'	0	9

Linkage between clauses in narrative sequences

- Independent assessment: coding all Narrative Events for the way in which they are related to the preceding clause.

- Independent assessment: coding all Narrative Events for the way in which they are related to the preceding clause.
- Categories used:
 - general-purpose linker ('then', clausal 'and');
 - other sentence-level linkers (conjunctions);
 - nominalization;
 - backgrounding devices such as converbs;
 - no overt linkage.

Linkage between clauses in narrative sequences

- Independent assessment: coding all Narrative Events for the way in which they are related to the preceding clause.
- Categories used:
 - general-purpose linker ('then', clausal 'and');
 - other sentence-level linkers (conjunctions);
 - nominalization;
 - backgrounding devices such as converbs;
 - no overt linkage.
- The check against the independent annotation confirms the same general trends.

Linkage between clauses in narrative sequences

- Episode breaks are correlated with the use of backgrounding particles and nominalization.

	Initial	Non-initial	Total
particles	43 (61%)	27	70 (100%)
nominalization	7 (50%)	7	14 (100%)
No linkage	104 (20%)	407	511 (100%)

Linkage between clauses in narrative sequences

- Episode breaks are correlated with the use of backgrounding particles and nominalization.

	Initial	Non-initial	Total
particles	43 (61%)	27	70 (100%)
nominalization	7 (50%)	7	14 (100%)
No linkage	104 (20%)	407	511 (100%)

- They are not correlated with the use of the general-purpose conjunction *bé* 'then':

	Initial	Non-initial	Total
<i>bé</i> 'then'	81 (20%)	332	413 (100%)
No linkage	104 (20%)	407	511 (100%)

A closer look at adverbs

- Adverbs overrepresented in initial events: weakly grammaticalized

	Initial	Non-initial
<i>tāííí</i> ‘until’	19	11
<i>yá</i> ‘there, then’	5	6
<i>wānè</i> ‘there’	4	5
<i>sé</i> ‘just when’	5	0

A closer look at adverbs

- Adverbs overrepresented in initial events: weakly grammaticalized

	Initial	Non-initial
<i>tāííí</i> ‘until’	19	11
<i>yá</i> ‘there, then’	5	6
<i>wānè</i> ‘there’	4	5
<i>sé</i> ‘just when’	5	0

- Adverbs overrepresented in non-initial clauses: heavy lexical content + ‘(not) yet’

	Initial	Non-initial
<i>ségé-ségé</i> ‘completely’	4	25
<i>sá</i> ‘over there’	0	8
<i>pí</i> ‘(not) yet, still’	0	6

Another check: ten most frequent words

	Initial	Non-initial
<i>h��</i> ‘then’	147	
<i>��</i> 3SG.SUBJ	99	
<i>��</i> DEF	97	
<i>��</i> 3PL	61	
<i>g��</i> go:PAST	41	
<i>��</i> 3SG	36	
<i>h��</i> ‘that one’	36	
<i>g��</i> ‘say’	33	
<i>��</i> PERF	33	
<i>g��</i> ‘in’	23	

Another check: ten most frequent words

	Initial	Non-initial	
<i>bé</i> 'then'	147	✓	451
<i>è</i> 3SG.SUBJ	99	✓	373
<i>ě</i> DEF	97	✓	294
<i>à</i> 3PL	61	<i>gé</i> 'say'	235
<i>gā</i> go:PAST	41	<i>bě</i> 'that one'	145
<i>à</i> 3SG	36	<i>à</i> 3PL	142
<i>bě</i> 'that one'	36	<i>ň</i> PERF	124
<i>gé</i> 'say'	33	:	+3SG 114
<i>ň</i> PERF	33	<i>à</i> 3SG	87
<i>gó</i> 'in'	23	<i>ě</i> REFL	53

Another check: ten most frequent words

	Initial	Non-initial	
<i>bé</i> ‘then’	147	✓	451
<i>è</i> 3SG.SUBJ	99	✓	373
<i>ě</i> DEF	97	✓	294
<i>à</i> 3PL	61	<i>gé</i> ‘say’	235
<i>gā</i> go:PAST	41	<i>bě</i> ‘that one’	145
<i>à</i> 3SG	36	<i>à</i> 3PL	142
<i>bě</i> ‘that one’	36	<i>ň</i> PERF	124
<i>gé</i> ‘say’	33	:	+3SG 114
<i>ň</i> PERF	33	<i>à</i> 3SG	87
<i>gó</i> ‘in’	23	<i>ě</i> REFL	53

Asymmetries in the distribution of words

- Frequent words overrepresented in initial events:

		Initial	Non-initial	Total
<i>ẽ</i>	DEF	97 (33%)	294	391 (100%)
<i>gā</i>	go:PAST	41 (57%)	31	72 (100%)

Asymmetries in the distribution of words

- Frequent words overrepresented in initial events:

		Initial	Non-initial	Total
<i>ě</i>	DEF	97 (33%)	294	391 (100%)
<i>gā</i>	go:PAST	41 (57%)	31	72 (100%)

- Frequent words overrepresented in non-initial clauses:

		Initial	Non-initial	Total
<i>gé</i>	'say'	33 (12%)	235	268 (100%)
:	+3SG	16 (12%)	114	130 (100%)

A closer look at tense and aspect

A closer look at tense and aspect

- Past tense is marked by mid tone.

(1) à **zō** kōŋ gó
 3PL come:PAST village in
 ‘They came from the village.’

A closer look at tense and aspect

- Past tense is marked by mid tone.

(1) à z̄ō k̄ōŋ gó
3PL come:PAST village in

‘They came from the village.’

- Aspectual markers (and a focus marker) are placed after the VP:

(2) à zò á k̄ōŋ gó
3PL come HAB village in

‘They come from the village (habitually).’

A closer look at tense and aspect

- Past tense is marked by mid tone.

(1) à **zō** kōŋ gó
 3PL come:PAST village in
 ‘They came from the village.’

- Aspectual markers (and a focus marker) are placed after the VP:

(2) à zò **á** kōŋ gó
 3PL come HAB village in
 ‘They come from the village (habitually).’

- Auxiliaries are placed after the subject:

(3) à **ɲ** zò kōŋ gó
 3PL PERF come village in
 ‘They come from the village.’

A closer look at tense and aspect

- Past tense is marked by mid tone.

(1) à zō kōŋ gó
3PL come:PAST village in
‘They came from the village.’

- Aspectual markers (and a focus marker) are placed after the VP:

(2) à zò á kōŋ gó
3PL come HAB village in
‘They come from the village (habitually).’

- Auxiliaries are placed after the subject:

(3) à ñ zò kōŋ gó
3PL PERF come village in
‘They come from the village.’

- Some aspectual constructions involve both markers:

(4) à á zò lé kōŋ gó
3PL PROG come PROG village in
‘They are coming from the village.’

Past tense vs. aspectual constructions

- In narrative event sequences, the major distinction is between the **past tense forms** and **aspectual constructions**; the latter involve an auxiliary and/or a post-verbal aspectual marker.

	Initial	Non-initial	Total
past tense	120 (35%)	225	345 (100%)
auxiliary	45 (21%)	165	210 (100%)
post-VP particle	31 (29%)	75	106 (100%)

Past tense vs. aspectual constructions

- In narrative event sequences, the major distinction is between the **past tense forms** and **aspectual constructions**; the latter involve an auxiliary and/or a post-verbal aspectual marker.

	Initial	Non-initial	Total
past tense	120 (35%)	225	345 (100%)
auxiliary	45 (21%)	165	210 (100%)
post-VP particle	31 (29%)	75	106 (100%)

- The difference is significant for past tense forms vs. constructions with auxiliaries ($p < 0.001$), but not for past tense vs. constructions with post-verbal particles.

Past tense vs. aspectual constructions

- In narrative event sequences, the major distinction is between the **past tense forms** and **aspectual constructions**; the latter involve an auxiliary and/or a post-verbal aspectual marker.

	Initial	Non-initial	Total
past tense	120 (35%)	225	345 (100%)
auxiliary	45 (21%)	165	210 (100%)
post-VP particle	31 (29%)	75	106 (100%)

- The difference is significant for past tense forms vs. constructions with auxiliaries ($p < 0.001$), but not for past tense vs. constructions with post-verbal particles.
- The asymmetry is consistent with the opposition between the past and the perfect as the **two major narrative forms**.

Distributions of the past and the perfect

- The perfect is significantly more biased toward clauses that are not initial in narrative sequences ($p < 0.001$):

	Initial	Non-initial	Total
past tense	120 (35%)	225	345 (100%)
perfect	36 (19%)	155	191 (100%)

Distributions of the past and the perfect

- The perfect is significantly more biased toward clauses that are not initial in narrative sequences ($p < 0.001$):

	Initial	Non-initial	Total
past tense	120 (35%)	225	345 (100%)
perfect	36 (19%)	155	191 (100%)

- The choice between the past and the perfect is sensitive, in narrative event sequences, to episode boundaries.

- No overall effect of episode boundary on the use of *bé* 'then': 25% in initial positions (cf. 26% for verbs, 24% for postpositions and adjectives).

Tense/aspect and conjunctions

- No overall effect of episode boundary on the use of *bé* 'then':
25% in initial positions (cf. 26% for verbs, 24% for postpositions and adjectives).
- Correlation with the use of past tense as opposed to the perfect.

Tense/aspect and conjunctions

- No overall effect of episode boundary on the use of *bé* 'then': 25% in initial positions (cf. 26% for verbs, 24% for postpositions and adjectives).
- Correlation with the use of past tense as opposed to the perfect.
- Linkage in initial event clauses ($p < 0.001$):

	<i>bé</i> 'then'	No linkage
past tense	58	37
perfect	1	30

Tense/aspect and conjunctions

- No overall effect of episode boundary on the use of *h  * ‘then’: 25% in initial positions (cf. 26% for verbs, 24% for postpositions and adjectives).
- Correlation with the use of past tense as opposed to the perfect.
- Linkage in initial event clauses ($p < 0.001$):

	<i>h��</i> ‘then’	No linkage
past tense	58	37
perfect	1	30

- Linkage in non-initial event clauses ($p < 0.001$):

	<i>h��</i> ‘then’	No linkage
past tense	159	51
perfect	2	149

Conjunction complementary with the perfect

- Past tense forms, often introduced by a conjunction, tend to occur at the beginning of thematic units.

Conjunction complementary with the perfect

- Past tense forms, often introduced by a conjunction, tend to occur at the beginning of thematic units.
- The perfect is rarely introduced by a conjunction, and tends to appear unit-internally.

Conjunction complementary with the perfect

- Past tense forms, often introduced by a conjunction, tend to occur at the beginning of thematic units.
- The perfect is rarely introduced by a conjunction, and tends to appear unit-internally.
- The past and the use of the perfect are complementary strategies for signalling cohesion within units.

Conjunction complementary with the perfect

- Past tense forms, often introduced by a conjunction, tend to occur at the beginning of thematic units.
- The perfect is rarely introduced by a conjunction, and tends to appear unit-internally.
- The past and the use of the perfect are complementary strategies for signalling cohesion within units.
- Both strategies cluster together within units ($p < 0.001$):

	after past	after perfect
unit-internal past	126	6
unit-internal perfect	21	76

Conjunction complementary with the perfect

- Past tense forms, often introduced by a conjunction, tend to occur at the beginning of thematic units.
- The perfect is rarely introduced by a conjunction, and tends to appear unit-internally.
- The past and the use of the perfect are complementary strategies for signalling cohesion within units.
- Both strategies cluster together within units ($p < 0.001$):

	after past	after perfect
unit-internal past	126	6
unit-internal perfect	21	76

- This suggests that some units are held together by the perfect, while others are held together by “conjunction + past tense”.

Unexpected result

Difference from tense switching observed in English (Schiffrin 1981):

- The two major narrative tense forms are past and present.

Unexpected result

Difference from tense switching observed in English (Schiffrin 1981):

- The two major narrative tense forms are past and present.
- Tense-switching occurs at unit boundaries “maintenance of one tense groups events into one scene or episode, while switching separates events from one another” (52).

Unexpected result

Difference from tense switching observed in English (Schiffrin 1981):

- The two major narrative tense forms are past and present.
- Tense-switching occurs at unit boundaries “maintenance of one tense groups events into one scene or episode, while switching separates events from one another” (52).
- A similar function is associated with conjunctions:
“temporal conjunctions occur when there is some break between events, and suggest that they function to separate events from one another” (55)

Difference from tense switching observed in English (Schiffrin 1981):

- The two major narrative tense forms are past and present.
- Tense-switching occurs at unit boundaries “maintenance of one tense groups events into one scene or episode, while switching separates events from one another” (52).
- A similar function is associated with conjunctions: “temporal conjunctions occur when there is some break between events, and suggest that they function to separate events from one another” (55)
- All this is strikingly different from Wan, where:
 - episodes tend to start in the past and continue in the perfect,
 - the major temporal conjunction is not associated with unit boundaries.

Interim summary: biases

	Beginning of units	Unit-internally
Referring expressions:	nouns segmental 3SG	pronouns lengthening

Interim summary: biases

	Beginning of units	Unit-internally
Referring expressions:	nouns segmental 3SG	pronouns lengthening
Backgrounding:	particles nominalization	

Interim summary: biases

	Beginning of units	Unit-internally
Referring expressions:	nouns segmental 3SG	pronouns lengthening
Backgrounding:	particles nominalization	
Adverbs:	‘until’, ‘there, then’	‘completely’
Verbs:	‘go’	‘say’

Interim summary: biases

	Beginning of units	Unit-internally
Referring expressions:	nouns segmental 3SG	pronouns lengthening
Backgrounding:	particles nominalization	
Adverbs:	‘until’, ‘there, then’	‘completely’
Verbs:	‘go’	‘say’
Tense/aspect	past tense	perfect

Extending the scope: Chuvash

- Comparison is needed to test the method for cross-linguistic applicability.

- Comparison is needed to test the method for cross-linguistic applicability.
- Chuvash is a good candidate because it differs from Wan on a variety of typological parameters.

- Comparison is needed to test the method for cross-linguistic applicability.
- Chuvash is a good candidate because it differs from Wan on a variety of typological parameters.
- Data: a small portion of the corpus of traditional narratives (Nikitina 2022).

- Comparison is needed to test the method for cross-linguistic applicability.
- Chuvash is a good candidate because it differs from Wan on a variety of typological parameters.
- Data: a small portion of the corpus of traditional narratives (Nikitina 2022).
- Annotation based on the same principles, even though the strategies for linking clauses differ dramatically from those of Wan.

Selected parts of speech

	Initial	Non-initial	(% initial)
noun	233		
verb	213		
pronoun	129		
particle	109		
adverb	27		
adjective	22		
numeral	27		
interjection	8		
ideophone	0		
conjunction	5		

Selected parts of speech

	Initial	Non-initial	(% initial)
noun	233	441	35%
verb	213	468	31%
pronoun	129	191	40%
particle	109	159	41%
adverb	27	83	25%
adjective	22	64	26%
numeral	27	36	43%
interjection	8	25	
ideophone	0	23	
conjunction	5	10	

Selected parts of speech

	Initial	Non-initial	(% initial)
noun	233	441	35%
verb	213	468	31%
pronoun	129	191	40%
particle	109	159	41%
adverb	27	83	25%
adjective	22	64	26%
numeral	27	36	43%
interjection	8	25	
ideophone	0	23	
conjunction	5	10	

Referring expressions: effects of pro-drop

- The ratio of referring expressions (nouns + pronouns) to verbs is higher than in Wan, because of pro-drop.

Referring expressions: effects of pro-drop

- The ratio of referring expressions (nouns + pronouns) to verbs is higher than in Wan, because of pro-drop.
- It is higher in initial than in non-initial event clauses ($p < 0.05$):

	Initial	Non-initial
noun + pronoun	362	632
verb	213	468
(% verbs)	(37%)	(43%)

Referring expressions: effects of pro-drop

- The ratio of referring expressions (nouns + pronouns) to verbs is higher than in Wan, because of pro-drop.
- It is higher in initial than in non-initial event clauses ($p < 0.05$):

	Initial	Non-initial
noun + pronoun	362	632
verb	213	468
(% verbs)	(37%)	(43%)

- Unlike in Wan, the proportion of pronouns is higher in initial clauses (but the result is not significant at this sample size):

	Initial	Non-initial	Total
pronoun	129 (40%)	191	320 (100%)
noun	233 (35%)	441	674 (100%)

Referring expressions: effects of pro-drop

- The ratio of referring expressions (nouns + pronouns) to verbs is higher than in Wan, because of pro-drop.
- It is higher in initial than in non-initial event clauses ($p < 0.05$):

	Initial	Non-initial
noun + pronoun	362	632
verb	213	468
(% verbs)	(37%)	(43%)

- Unlike in Wan, the proportion of pronouns is higher in initial clauses (but the result is not significant at this sample size):

	Initial	Non-initial	Total
pronoun	129 (40%)	191	320 (100%)
noun	233 (35%)	441	674 (100%)

- All these effects are possibly explained by higher incidence of pro-drop in non-initial clauses.

A closer look at pronouns

- Unlike in Wan, demonstrative pronouns are more likely to appear in initial event clauses than 3SG pronouns ($p < 0.05$):

	Initial	Non-initial	Total
'this'	35 (47%)	40	75 (100%)
'that'	30 (53%)	27	57 (100%)
'that aforementioned'	19 (43%)	25	44 (100%)
'what'	15 (38%)	24	39 (100%)
3SG	15 (25%)	45	60 (100%)

A closer look at pronouns

- Unlike in Wan, demonstrative pronouns are more likely to appear in initial event clauses than 3SG pronouns ($p < 0.05$):

	Initial	Non-initial	Total
'this'	35 (47%)	40	75 (100%)
'that'	30 (53%)	27	57 (100%)
'that aforementioned'	19 (43%)	25	44 (100%)
'what'	15 (38%)	24	39 (100%)
3SG	15 (25%)	45	60 (100%)

- Perhaps demonstrative pronouns function as long-distance anaphors, and are more suitable for reference across episode boundaries?

- Unlike in Wan, conjunctions are rarely used in Chuvash (and most are borrowings from Russian).

Conjunctions and particles

- Unlike in Wan, conjunctions are rarely used in Chuvash (and most are borrowings from Russian).
- The discourse linking function is instead associated with particles:

	Initial	Non-initial	Total
<i>ta/te</i> ‘and’	44 (39%)	68	112 (100%)
<i>ëntë</i> ‘already’	35 (46%)	41	76 (100%)
<i>vara</i> ‘then’	11	6	17

A closer look at particles

- Episode boundaries are correlated with the use of the clause-final particle *ta/te* ‘and’ ($p < 0.005$).

	Initial	Non-initial	Total
<i>ta/te</i> ‘and’	30 (48%)	33	63 (100%)
No linkage	118 (27%)	317	435 (100%)

A closer look at particles

- Episode boundaries are correlated with the use of the clause-final particle *ta/te* ‘and’ ($p < 0.005$).

	Initial	Non-initial	Total
<i>ta/te</i> ‘and’	30 (48%)	33	63 (100%)
No linkage	118 (27%)	317	435 (100%)

- The particle is forward-looking, i.e. relates the unit to the one that follows.

A closer look at particles

- Episode boundaries are correlated with the use of the clause-final particle *ta/te* ‘and’ ($p < 0.005$).

	Initial	Non-initial	Total
<i>ta/te</i> ‘and’	30 (48%)	33	63 (100%)
No linkage	118 (27%)	317	435 (100%)

- The particle is forward-looking, i.e. relates the unit to the one that follows.
- The particle may have a backgrounding effect, just like the backgrounding sentence-final particles in Wan.

- The two major narrative verb forms in Chuvash are the past and the present.

Past vs. present

- The two major narrative verb forms in Chuvash are the past and the present.
- The present tense forms are more likely to appear in non-initial clauses ($p < 0.05$):

	Initial	Non-initial	Total
past	69 (36%)	125	194 (100%)
present	55 (25%)	164	219 (100%)

Past vs. present

- The two major narrative verb forms in Chuvash are the past and the present.
- The present tense forms are more likely to appear in non-initial clauses ($p < 0.05$):

	Initial	Non-initial	Total
past	69 (36%)	125	194 (100%)
present	55 (25%)	164	219 (100%)

- This distribution is similar to the one observed in Wan: the past is more likely to be used to start an episode than the alternative form.

Tense and the forward-linking particle ‘and’

- The particle *ta/te* ‘and’ is used to link a clause to the one that follows.

Tense and the forward-linking particle ‘and’

- The particle *ta/te* ‘and’ is used to link a clause to the one that follows.
- Its use is correlated with the use of the past tense ($p < 0.001$):

	Conjunction		No linkage	Total
past	40	(21%)	154	194 (100%)
present	17	(8%)	194	211 (100%)

Tense and the forward-linking particle ‘and’

- The particle *ta/te* ‘and’ is used to link a clause to the one that follows.
- Its use is correlated with the use of the past tense ($p < 0.001$):

	Conjunction		No linkage	Total
past	40	(21%)	154	194 (100%)
present	17	(8%)	194	211 (100%)

- This is similar to the situation in Wan, where the use of the past tense was correlated with the use of the conjunction (as opposed to the perfect).

Comparison (1)

Comparison (1)

- Non-initial clauses are more likely to refer to highly identifiable participants (pro-droppable in Chuvash, pronominal in Wan):

	Beginning of units	Unit-internally
Wan:	nouns	pronouns
Chuvash:	pronouns	nouns

Comparison (1)

- Non-initial clauses are more likely to refer to highly identifiable participants (pro-droppable in Chuvash, pronominal in Wan):

	Beginning of units	Unit-internally
Wan:	nouns	pronouns
Chuvash:	pronouns	nouns

- In Wan, such participants are coded non-segmentally:

	Beginning of units	Unit-internally
Wan:	segmental 3sg	lengthening

Comparison (1)

- Non-initial clauses are more likely to refer to highly identifiable participants (pro-droppable in Chuvash, pronominal in Wan):

	Beginning of units	Unit-internally
Wan:	nouns	pronouns
Chuvash:	pronouns	nouns

- In Wan, such participants are coded non-segmentally:

	Beginning of units	Unit-internally
Wan:	segmental 3SG	lengthening

- In Chuvash, overt pronouns are reserved for less identifiable participants, and they are overwhelmingly demonstrative:

	Beginning of units	Unit-internally
Chuvash:	demonstratives	3SG

Comparison (2)

- Both Chuvash and Wan make use of particles to relate a unit to the one that follows.

	Beginning of units	
Wan:	backgrounding particles	
Chuvash:	particles 'and', 'already'	
Wan:	nominalization	
Chuvash:	(converbs)	

Comparison (2)

- Both Chuvash and Wan make use of particles to relate a unit to the one that follows.

	Beginning of units	
Wan:	backgrounding particles	
Chuvash:	particles 'and', 'already'	
Wan:	nominalization	
Chuvash:	(converbs)	

- Both languages switch tense or aspect within units (from the past to the present or to the perfect).

	Beginning of units	Unit-internally
Wan:	past	perfect
Chuvash:	past	present

Comparison (3)

- Both languages show a correlation between tense/aspect of the clause and type of linkage.

Comparison (3)

- Both languages show a correlation between tense/aspect of the clause and type of linkage.
- Past tense forms tend to be introduced by a conjunction in Wan (the clause-initial ‘then’), by a clause-final particle in Chuvash (the forward-looking ‘and’).

Comparison (3)

- Both languages show a correlation between tense/aspect of the clause and type of linkage.
- Past tense forms tend to be introduced by a conjunction in Wan (the clause-initial ‘then’), by a clause-final particle in Chuvash (the forward-looking ‘and’).
- Non-past forms (the perfect in Wan, the present in Chuvash) are associated with absence of overt linking.

Conclusion

- The study tested a method for identifying phenomena sensitive to episode boundaries, and yielded results.

Conclusion

- The study tested a method for identifying phenomena sensitive to episode boundaries, and yielded results.
- The results can be meaningfully compared, and some of them are consistent with other differences (such as pro-drop).

Conclusion

- The study tested a method for identifying phenomena sensitive to episode boundaries, and yielded results.
- The results can be meaningfully compared, and some of them are consistent with other differences (such as pro-drop).
- The phenomena identified are the usual suspects: the choice of referring expression type, the use of particles, the choice of tense/aspect.

Conclusion

- The study tested a method for identifying phenomena sensitive to episode boundaries, and yielded results.
- The results can be meaningfully compared, and some of them are consistent with other differences (such as pro-drop).
- The phenomena identified are the usual suspects: the choice of referring expression type, the use of particles, the choice of tense/aspect.
- The method has also identified some lexical choices as sensitive to episodic structure (e.g., weakly grammaticalized vs. full-fledged contentful adverbs).

Conclusion

- The study tested a method for identifying phenomena sensitive to episode boundaries, and yielded results.
- The results can be meaningfully compared, and some of them are consistent with other differences (such as pro-drop).
- The phenomena identified are the usual suspects: the choice of referring expression type, the use of particles, the choice of tense/aspect.
- The method has also identified some lexical choices as sensitive to episodic structure (e.g., weakly grammaticalized vs. full-fledged contentful adverbs).
- There also emerged unexpected correlations between the different phenomena (the use of conjunctions and particles with past tense, as opposed to the perfect or the present).

- Bakker, Egbert J. 1993. "Boundaries, topics, and the structure of discourse: An investigation of the Ancient Greek particle *dé*." *Studies in Language* 17 (2): 275–311.
- Dijk, Teun A. van. 1981. "Episodes as units of discourse analysis." In *Analyzing discourse: Text and talk*, edited by Deborah Tannen, 177–195. Georgetown: Georgetown University Press.
- Fox, Barbara. 1987. "Anaphora in popular written English narratives." In *Coherence and grounding in discourse: Outcome of a Symposium, Eugene, Oregon, June 1984*, edited by Russell S. Tomlin, 157–174. Amsterdam: John Benjamins.
- Labov, William. 1972. "The transformation of experience in narrative syntax." In *Language in the Inner City: Studies in the Black English Vernacular*, edited by William Labov, 354–405. Philadelphia: University of Pennsylvania Press.
- Mithun, Marianne. 2002. "Rhetorical nominalization in Barbareño Chumash." *Survey Reports, Survey of California and Other Indian Languages*, no. 12: 55–63.

- Nikitina, Tatiana. 2022. "A narrative corpus of Chuvash." In *The SpeechReporting Corpus: Discourse Reporting in Storytelling*, edited by Tatiana Nikitina, Ekaterina Aplonova, Izabela Jordanoska, Ekaterina Biteeva, Abbie Hantgan-Sonko, Guillaume Guitang, Olga Kuznetsova, Rebecca Paterson, Elena Perekhvalskaya, and Lacina Silué. CNRS-LLACAN & LACITO.
- Nikitina, Tatiana, Ekaterina Aplonova, Izabela Jordanoska, Ekaterina Biteeva, Abbie Hantgan-Sonko, Guillaume Guitang, Olga Kuznetsova, Rebecca Paterson, Elena Perekhvalskaya, and Lacina Silué, eds. 2022. *The SpeechReporting Corpus: Discourse Reporting in Storytelling*. <http://discoursereporting.huma-num.fr/index.html>: CNRS-LLACAN & LACITO.
- Schiffrin, Deborah. 1981. "Tense variation in narrative." *Language* 57 (1): 45–62.
- Tomlin, Russell S. 1987. "Linguistic reflections of cognitive events." In *Coherence and grounding in discourse: Outcome of a Symposium, Eugene, Oregon, June 1984*, edited by Russell S. Tomlin, 455–479. 11. Amsterdam: John Benjamins.