

Workshop

**Larger discourse units in (spoken) language**

Paris, March 21-22, 2024

**ABSTRACTS**

## **Episodes are determined backwards. Addressing the problem of text-segmentation**

Jakob Egetenmeyer (University of Cologne)

Narrative and reportative discourse calls for packaging information in processable units. A large unit is the episode. It is sometimes referred to as a macro-structure (van Dijk 1980) or a macro-proposition (Miall 2004 with reference to Eco's 1984 term discourse topic). Its relevance in structural and cognitive terms has been studied for some time and with respect to a wide array of languages (see, for instance, Longacre 1979 for an interest in structural means; and Haberlandt, Berian & Sandson 1980; Zwaan, Magliano & Graesser 1995 for issues of processing). A relevant part of the research literature is interested in the linguistic material correlating with episode boundaries (Seig 2004; Riesberg et al. 2022). A key problem is determining the string of discourse which makes up an episode. Critically, in order to refrain from circularity, the episode boundaries have to be determined without a focus on potential indicators or at least with a focus on other indicators than the ones which are then correlated with those boundaries (for instance, intonation units as opposed to the realization of individual reference in the case of Seig 2004). Apparently, this is a challenging task.

An important suggestion is contributed in van Dijk (1982) who defines the episode in terms of content coherence, that is, with respect to "participants, time, location or global event or action". Miall (2004) adds the concept of "narrative twist", which, however, has a different status. The classification based on van Dijk's four basic conceptual and referential classes is highly insightful and, in principle, it may be expected to be viable in research, as it also supports the mentioned focus on the boundaries of episodes. However, it is not easily operationalized. In addition, it is difficult to use its classification to localize a certain clause within an episodic structure.

Our suggestion, first mentioned but not yet tested in Becker & Egetenmeyer (2018), is to determine episodes backwards. This principle is applied on the level of propositions on the grounds of the semantic-pragmatic relationships assumed in segmented discourse representation theory (Asher & Lascarides 2003). A crucial component is the right frontier constraint by Polanyi (1988) as it allows for a distinction of hierarchical relationships among propositions (Asher & Vieu 2005; Jasinskaja & Karagjosova 2020). According to the account, from a given point in a narrative (or reportative) text, we are able to determine a (previous) point of structural initiation, that is the point where corresponding relevant information is introduced, and which is however restricted by a domain of accessibility. The account is rooted in temporal discourse structure (Becker & Egetenmeyer 2018), but complemented by a concept of textual development beyond temporality (Egetenmeyer accepted). This allows to consider also non-eventive parts of an episode, which are sometimes excluded from the discussion (see the distinctions in Smith 2003). We test the account by analyzing structures involving competing realizations of concept classes of coherence in the sense of van Dijk (1982). We study Romance data.

### **References**

- Asher, Nicholas & Lascarides, Alex. 2003. *Logics of conversation*. Cambridge: Cambridge University Press.
- Asher, Nicholas & Vieu, Laure. 2005. Subordinating and coordinating discourse relations. *Lingua* 115 (4), 591–610.
- Becker, Martin & Egetenmeyer, Jakob. 2018. A prominence-based account of temporal discourse structure. *Lingua* 214, 28–58.

- Eco, Umberto. 1984. *The role of the reader: Explorations in the semiotics of texts*. Bloomington: Indiana University Press.
- Egetenmeyer, Jakob. Accepted a. Textual development beyond temporality and interactions between the adversative connector *pero* and the *indefinido*. In Egetenmeyer, Jakob, Sarah Dessì Schmid & Martin Becker (eds.), *Tense, aspect and discourse structure*. Berlin / Boston: De Gruyter.
- Haberlandt, Karl, Claire Berian, and Jennifer Sandson. 1980. The episode schema in story processing. *Journal of Verbal Learning and Verbal Behavior* 19, 635–650.
- Jasinskaja, Katja & Elena Karagjosova. 2020. Rhetorical Relations. In Gutzmann, Daniel et al. (eds.), *The Wiley Blackwell Companion to Semantics*, 1–29. Oxford: Wiley.
- Longacre, Robert E. 1979. The paragraph as a grammatical unit. In Givón, Talmy (ed.), *Discourse and syntax*, 115–134. New York: Academic Press.
- Miall, David S. 2004. Episode structures in literary narratives. *Journal of Literary Semantics* 33, 111-129.
- Polanyi, Livia. 1988. A formal model of the structure of discourse. *Journal of Pragmatics* 12 (5-6), 601–638.
- Riesberg, Sonja, Maria Bardají i Farré, Kurt Malcher & Nikolaus P. Himmelmann. 2022. Predicting voice choice in symmetrical voice languages. All the things that do not work for Totoli. *Studies in Language* 46 (2). 453–516.
- Seig, Mary. 2004. Episodic boundaries in Japanese and English narratives. In Moder, Carol Lynn & Aida Martinovic-Zic (eds.), *Discourse across languages and cultures*, 227–250. Amsterdam / Philadelphia: John Benjamins.
- Smith, Carlota. 2003. *Modes of discourse. The local structure of texts*. Cambridge: Cambridge University Press.
- van Dijk, Teun A. 1980. *Macrostructures. An interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, N.J.: Erlbaum.
- van Dijk, Teun A. 1982. Episodes as units of Discourse Analysis. In Deborah Tannen (ed.), *Analyzing discourse: Text and talk*, 177–195. Georgetown: Georgetown University Press.
- Zwaan, Rolf A., Joseph P. Magliano & Arthur C. Graesser. 1995. Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21 (2), 386–397.

## Question Sequences as Discourse Units in TED Talks

Michele Cardo and Agnès Celle (Université Paris Cité)

TED talks are an interesting discourse type which combines elements of monologue and dialogue because there is one main speaker who controls the discourse and yet there is the possibility for the speaker to invite audience participation in the form of dialogic interaction. Because a TED speaker controls the discourse and the level of dialogic conversation that may be present, it is difficult to segment TED talk transcripts into different meaningful discourse parts.

Several studies have worked toward a definition of what constitutes a discourse unit. Even the verbiage used to describe some part of a discourse varies. For example, Hymes (1967; 1972; 1974), Gumperz (2015), and Egbert and Schnur (2018) prioritize the term *speech event* to refer to some coherent part of a discourse which can be identified if a text or utterance fulfills some set of criteria (Egbert et al. (2021)). Focusing on spoken discourse, however, Egbert et al. (2021) use the term *discourse unit* to describe “functional segments of conversation” (p. 725). Egbert et al. (2021) present the following four parameters as part of their definition of a discourse unit:

1. Coherent for its overarching communicative goal.
2. Characterized by one or more communicative purposes.
3. Recognizably self-contained
4. Containing a minimum of five utterances or 100 words

TED talks count as a unique type of discourse because of the nature of TED talks to revolve around one specific theme. Thus, there is thematic coherence throughout a TED talk. Coherence being a main criterion for what can be classified as a discourse, according to Hobbs (1985), TED talks can clearly be considered a discourse in their own right. Hobbs (1985), Mann and Thompson (1987), and Grosz and Sidner (1986) point out that, “*discourse structure* is a necessary component of discourse interpretation” (Lascarides and Asher, 2008, p. 1). As a result, to successfully interpret a TED talk we must understand the underlying discourse structure. This requires identifying and segmenting discourse units in TED talks.

TED speakers often use question sequences as a discursive device throughout their presentations. Question sequences are coherent utterances with a clear communicative goal and purpose. They are easily recognizable by the fact that they contain multiple questions in a row with a distinct beginning and end. They sometimes contain five or more utterances. As such, question sequences fulfill almost all of the criteria laid out by Egbert et al. (2021) regarding what qualifies as a *discourse unit*.

This paper aims to define question sequences as a larger discourse unit within TED talks. Among other research goals, this study will consider the discourse relationship between question sequences and the rest of the TED talk and understand how speakers use question sequences to accomplish discursive and interactive goals throughout their presentations.

### References

Egbert, J., & Schnur, E. (2018). Missing the trees for the forest: The role of the text in corpus and discourse analysis. In Marchi Anna & Charlotte Taylor (eds.), *Corpus approaches to discourse: A critical review*, 159–173. New York: Routledge. <https://doi.org/10.4324/9781315179346-8>.

- Egbert, J., Stacey Wizner, Daniel Keller, Douglas Biber, Tony McEnery and Paul Baker. (2021). Identifying and describing functional units in discourse in the BNC Spoken 2014. *Text&Talk* 41(5-6), 715–737.
- Grosz, B., & Sidner, C. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, 12(3), 175–204.
- Gumperz, J.J. (2015). Interactional sociolinguistics: A personal perspective. In Deborah Tannen, Heidi Hamilton & Deborah Schiffrin (eds.), *The handbook of discourse analysis*, 309–323. West Sussex, UK: Wiley Blackwell. <https://doi.org/10.1002/9781118584194.ch14>.
- Hobbs, J.R. (1985). On the coherence and structure of discourse.
- Hymes, D. (1967). Models of the interaction of language and social setting. *Journal of Social Issues* 23(2), 8–28.
- Hymes, D. (1972). Models of the interaction of language and social life. In John Gumperz & Dell Hymes (eds.), *Directions in sociolinguistics: The ethnography of communication*, 35–71. New York: Holt, Rhinehart & Winston.
- Hymes, D. (1974). *Foundations in sociolinguistics: An ethnographic approach*. Philadelphia: University of Pennsylvania Press.
- Lascarides, A., & Asher, N. (2008). Segmented Discourse Representation Theory: Dynamic Semantics With Discourse Structure. In: Bunt, H., Muskens, R. (eds) *Computing Meaning. Studies in Linguistics and Philosophy*, vol 83. Springer, Dordrecht. [https://doi.org/10.1007/978-1-4020-5958-2\\_5](https://doi.org/10.1007/978-1-4020-5958-2_5)
- Mann, W.C., & Thompson, S. A. (1987). Rhetorical structure theory: A framework for the analysis of texts. *International Pragmatics Association Papers in Pragmatics*, 1, 79–105

### Multimodal segmentation in an oral epic song

Anna Bonifazi, Department of Linguistics, University of Cologne

This paper discusses the features cueing segmentation above and beyond the clause level in the Serbocroatian epic song P(arry) N(umber) 662, *Halil Hrnjičić and Miloš the Highwayman* (995 verses), belonging to the Milman Parry Collection of Oral Literature hosted at Harvard University.

Segmenting texts above the clause level canonically coincides with paragraphing. The visual marking of paragraphs in the page layout is usually clear (let us think of indentation), but the linguistic features enabling us to motivate paragraph boundaries may be less clear. However, already in the 70s and 80s, scholars have identified key linguistic criteria for the definition of paragraphs: unified topics, coherence and cohesion, continued pronominalization (Hinds 1977; Longacre 1979; Hofmann 1989). Crucially, paragraphs are also defined with respect to episodes: “An episode is properly a semantic unit, whereas a paragraph is the surface manifestation or the expression of such an episode” (Van Dijk 1982: 177).

All this regards written texts and texts whose segmentation is prototypically organized on the word level. How are texts paragraphed if they that are designed to be performed, i.e. they encompass more than words? The present paper offers a qualitative and quantitative analysis of data from PN 662 as a case study. First, if ‘paragraphs’ or episode boundaries are provided at all, they are provided online, as the performance unfolds. More importantly, such boundaries pertain not only to the word level, but also to a wide range of performative features clearly perceptible in the audio recording. This is why the general terms ‘macro-’ and ‘micro-segmentation’ are preferred over ‘clauses’ and ‘paragraphs’. The present oral song in fact features multimodal segmentation (Bonifazi and Elmer 2012a). Macro-segments, in particular, are suggested on the word level through changes of spatial and/or temporal settings, special formulas introducing new characters (Elmer 2009), shifts between third-person narration and reported speech, and the insertion of evaluative comments. The altered tempo and melodies of special lines, the repeated rhythm of line blocks, the lowering of the base tone, and the insertion of falsettos at line starts are considered as well.

The paper concludes by drawing conclusions from a graph overviewing the frequency and the distribution of all these features.

### References

- Bonifazi, Anna and David F. Elmer. 2012a. Composing lines, performing acts: clauses, discourse acts, and melodic units in a South Slavic epic song. In *Orality, Literacy and Performance in the Ancient World*, E. Minchin (ed.), 89-109. Leiden and Boston: Brill.
- Elmer, David F. 2009. Presentation Formulas in South Slavic Epic Song. *Oral Tradition* 24.1: 41-59.
- Hinds, John. 1977. Paragraph structure and pronominalization. *Papers in Linguistics* 10: 77-99.
- Hofmann, Thomas. 1989. Paragraphs & anaphora. *Journal of Pragmatics* 13, 239-250.
- Longacre, Robert E. 1979. The paragraph as a grammatical unit. In: Givon, Talmy (ed.), *Syntax and Semantics: Discourse and Syntax*, vol. 12, 115-134. Academic Press, New York.
- Van Dijk, Teun. 1982. Episodes as units of discourse analysis. In: Tannen, Deborah (ed.), *Analyzing Discourse: Text and Talk*, 177-195. Georgetown University Press, Georgetown.

## Detangling the function of the most frequent discourse particle in an understudied language

Natalia Cáceres Arandia (SeDyL\_CNRS)

The particle *yaawö* of Ye'kwana (ISO 639-3:mch, Cariban, Venezuela) is never given a spontaneous translation into the contact language, Spanish. In isolation, speakers of the language will accept *entonces* 'then' as a possible translation. However, the particle is so frequent that even non-Ye'kwana speakers can distinguish the form in connected public speech and repeat it back to Ye'kwana speakers to ask about its meaning. This paper discusses the possible functions of *yaawö* in discourse arguing it allows the speaker to signal further elaboration, at least in some of its occurrences.

In a Ye'kwana corpus of 5158 text units of oral interactions, a third (n=1734) of the units contain at least one occurrence of *yaawö* (2068 tokens). This is by far the most frequent form in the corpus: the second represents half of the occurrences (*ke*, 935 tokens, 'propietive', 'instrumental' and 'quotative') and the third form a fourth of the occurrences (*mödö*, 536 tokens, 'inanimate medial demonstrative' and at least two other undefined functions).

There have been no studies about discourse particles in Ye'kwana other than a non-comprehensive list of particles and their generic meaning (Cáceres Arandia 2011). The particle studied here is likely a lexicalized form of the combination of the anaphoric inanimate pronoun *iyö* and the general locative postposition =*awö*., but the contracted form *yaawö* is never found with a locative meaning. I show in this paper that *yaawö* can be found in at least three different functions: at the beginning of a clause (1e) as a temporal or situational connective (the least frequent function); at the end of clauses containing a pronominal mention of an argument (1c) or no mention at all (2) which is given in nominal form right after *yaawö*; and at the end of clauses which are connected by the event being described (1a-b, 1d-e), allowing the speaker to signal that something more is coming related to what they have just uttered.

- (1a) *Iyö tü-w-entatü-tojo=mma kün-en'-aakö tüwü yaawö.*  
 3IN:INV 3S:REFL-INTR-mouth.rinse-INSTR=EXCL 3/3:DIS-drink-PDI 3SG then  
 '(Agouti) was drinking the mouthwash.' {CtoAbjPic.053:FcM}
- (1b) "Ünjü" *ke kün-ö'dü-aakö, "a-w-entatü-tojo mödö yaawö.*  
 no QUOT 3S:DIS-say-PDI 2S-INTR-rinse.mouth-INSTR 3IN:MD then  
 '“No” said (Bee), “that one is your mouthwash.”' {CtoAbjPic.054:FcM}
- (1c) "[*Edö-öne*] *iyö o-wokü-dü yaawö [eese naa=dü=üne].*"  
 3IN:PX-INTS 3IN:ANA 2-drink-POS then this.side 3S:COP=REL=INTS  
 '“[This one]<sub>i</sub> is your drink, [the one that is on this side]<sub>i</sub>.”' {CtoAbjPic.055:FcM}
- (1d) "Aaa," *yööje kün-entatü-jötü-i yaawö.*  
 ah thus 3S:DIS-rinse.mouth-PLAC-PRP then  
 '“Ah,” (Agouti) rinsed her mouth thus.' {CtoAbjPic.056:FcM}
- (1e) *Yaawö=öne iyö yookü-dü kun-tu-i yaawö.*  
 then=INTS 3IN:INV 3:drink-POS 3/3:DIS-give-PRP then  
 'At that moment, (Bee) gave (Agouti) her drink.' {CtoAbjPic.057:FcM}
- (2) *Yööje kun-ummicha-i yaawö, tü-uwa-dü.*  
 thus 3/3:DIS-let.go-PRP then 3:REFL-basket-POS  
 'She let it go thus, her basket.' {CtoAbjPic.040:FcM}

## All together now: Identifying units above the clause in Ngarinyin

Stef Spronck (Utrecht University/University of Helsinki)

Due to the influence of tagmemics (Pike 1964; Longacre 1970) and Systemic Functional Grammar (Halliday and Hasan 1976) on early Australianist linguistics, descriptive grammars of Australian Aboriginal languages from the 1960-1980s often contain at least some comments on discourse features. This includes analyses of narratives, information structure and emphatic marking, but also suggestions about units above clause level, such as thematic units and paragraphs, along with their own typical markers.

Against this background, I present fieldwork and archival corpus data from the Ngarinyin language (Australia; Worrوران), in which several discourse marking strategies can be distinguished. After introducing a second position epistemic clitic, which can be used as a reliable indicator for determining clausehood, I present three different kinds of discourse markers *above* the clause, the first of which is a ‘paragraph marker’ (also cf. Rumsey 1987), as in (1).

- (1) *liny ngay-a-ngarri*                      *ho:: wularnburr*  
 see 1pl.INC-go.PRS-SUB              oh! things
- babilij*    *burr-ma*            *buno*    *ngala*    *jiriki*    *nyonya*  
 lift.up    3pl-do.PRS            3pl.DIST animal    bird            other
- jina*                      *liny-ju*                      *nyirr-yi-nyi-ngarri*            *aw!*  
 msg.PROX    see-**PARAGRAPH** 1pl.EXC-be-PST-SUB            ah  
 ‘Then we saw: “Oh, it is full of things!”’  
 ‘This bird [is there]...’  
 ‘And then we suddenly saw another one, oh!’ (100903-19NGUN, 2:00-2:04)

The paragraph marker, in bold in (1), signals both that the event so marked is unexpected and introduces a new episode in the narrative. I argue that examples marked as in (1) display structural episode marking.

The two other discourse marking strategies I discuss are a set of (partially combining) pronouns, which may serve to signal topic (dis)continuity and an epistemic authority marker, which signals that the current speaker assumes that the addressee does not yet know the identity of the subject referent. I argue that both of these strategies create discourse structures as well, but that they are less clearly identifiable than episodes marked by the paragraph marker.

I conclude with a brief theoretical discussion about the methodological divide that, I would suggest, exists in the study of language between the treatment of units within/below and above the clause. I claim that the fundamental differences in approaching these linguistic levels is what leads to confusion about the status of units above clause level, since it produces inconsistent analyses of the three components language consists of: representation, cohesion and stance (cf. Spronck 2020). An integrated approach is needed.

### References

- Halliday, M. A. K., and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.  
 Longacre, Robert E. 1970. “Sentence Structure as a Statement Calculus.” *Language* 46 (4): 783–815.



- Pike, Kenneth L. 1964. "Discourse Analysis and Tagmeme Matrices." *Oceanic Linguistics* 3 (1): 5–25.
- Rumsey, Alan. 1987. "Lative and Translative in Ungarinyin." In *A World of Language: Papers Presented to Professor s.a. Wurm on His 65th Birthday*, edited by Donald C Laycock and Werner Winter, 603–11. Canberra: Pacific Linguistics.
- Spronck, Stef. 2020. "The Representation-Cohesion-Stance Hypothesis." In *Signs of Life: Changes and Continuity in Language, Thought and Identity*, edited by Vera DaSilva Sinha, Ana Moreno-Núñez, and Zhen Tian, 75–109. Amsterdam/Philadelphia: John Benjamins.

## Larger discourse units in Beja (North-Cushitic, Afroasiatic)

Martine Vanhove (LLACAN, CNRS, INaLCO, EPHE)

Ends of episodes (i.e. larger discourse units) in Beja folktales, are easily recognizable thanks to a segmental cue, the discourse marker *e:n*, which grammaticalized from the Perfective 3<sup>rd</sup> person plural of the quotative verb *di* ‘say’. In an oral corpus of 32 folktales, counting 7919 words, they amount to 474 tokens, i.e. 6% of the total number of words. Episodes are of various length, may consist of one utterance or several utterances, matrix and subordinate clauses, or coordinated clauses, and dialogues. Conversely, in other genres (procedural texts, humorous stories and jokes, anecdotes, personal narratives, and conversations), *e:n* is much less frequent, or even absent in descriptive texts such as the *Pear story* (Chafe 1980). Intonation contours, i.e. prosodic cues, thus play a more important role. In addition, the use of various TAM, also plays a role in identifying the beginning and the end of an episode, especially in folktales.

This presentation will discuss the interplay between the three parameters, comparing one tale, one conversation and one anecdote. Several typical situations occur:

1. While *e:n* occurs after any syntactic type of utterance in folktales and anecdotes, its use in conversations seems to be limited to the end of dialogues, and when reporting a situation, a clear trace of its quotative origin.
2. Final vowel lengthening is the main cue to identify that identify turn maintenance, and that an episode is not finished, but falling contours may mark the end of a turn taking or an episode, but not necessarily.
3. Regarding TAM, the Perfective typically occurs at the beginning of an episode when starting with an independent or a coordinated clause, and at the end of a tale or any other narrative genre, while the Imperfective typically occurs at the end of a series of events in an episode, even when the preceding verbs are in the Perfective or the Aorist as in the following example.

*hi:da:b*    *ʔi:ba:b-ja:n*    *ʔi:ba:b-e:*    *ha:f=i:b*  
 together    travel-PFV.3PL    travel-CVB.SMLT    country=LOC.SG

*i-ganif-na=e:b*    *o:=do:r*    *jha:m*    *dha:j*  
 3M-kneel\MID.PFV-PL=REL.M    DEF.SG.M.ACC=time    leopard    DIR

*jʔ-i=t*    *e:=kam*  
 come-AOR.3SG.M=COORD    DEF.PL.M.ACC=camel\PL

*ji=i:-biri-n=e:=na:=je:b*    *ka:m ho:j danri*    *e:n*  
 REL.PL.M=3-have\AOR-PL=REL=thing=LOC.PL    camel    LOC    kill\IPFV.[3SG.M]    DM

‘They traveled (PFV) together. While travelling, when they stopped (PFV) in a country, a leopard came (AOR) towards them and killed them (IPFV) a camel that they had (AOR).’

### References

Chafe, Wallace (ed.). 1980. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, New Jersey: Ablex.

### Abbreviations

ACC accusative; AOR aorist; COORD coordination; CVB.SMLT simultaneity converb; DEF definite; DIR directional; DM discourse marker; IPFV imperfective; LOC locative; M masculine; MID middle; PFV perfective; PL plural; REL relator; SG singular.

## Split ergativity and larger discourse units in Tima (Niger-Congo, Sudan)

Birgit Hellwig and Gertrud Schneider-Blum (University of Cologne)

This presentation focuses on the Niger-Congo language Tima, which is spoken in the Nuba Mountains of Sudan. Tima is a language with split ergativity, where ergative case marking interacts with constituent order. In unmarked SVO order, the A argument remains unmarked for case. In a number of other word orders, however, and especially in the case of OVA order, the A argument receives ergative case.

Stimuli-based research has given us a good idea of the semantic, referential and discourse properties that favour the use of one structure over another. In a series of publications (Becker & Schneider-Blum 2020; Casaretto et al. 2020; Schneider-Blum 2023; Schneider-Blum & Hellwig 2018; Schneider-Blum et al. 2022), it was shown that ergative structures tend to be used in contexts where there is a prominence conflict between the two arguments of a verb: the A argument is less prominent than the non-A argument on at least one prominence hierarchy, e.g., on the animacy hierarchy (e.g., when a non-human agent acts upon a human patient), the referentiality hierarchy (e.g. when the agent is not identifiable as an individual), and/or the givenness hierarchy (e.g. when the agent is new, while the patient is given).

While such prominence conflicts robustly favour ergative structures in our controlled data set, they do not obligatorily trigger them, and speakers can always choose to use non-ergative structures instead. We hypothesized that discourse factors show responsible in this respect, and we therefore conducted a corpus analysis of case marking in natural discourse, with a focus on narrative monologues. This analysis shows that ergative structures indeed play a role for structuring discourse. They occur, for instance, after narrative boundaries where they (re-) introduce an agentive referent in subject function at the start of a new referential chain, establishing it as the most prominent referent. This is illustrated with the following example, where unit (1) ii features such a boundary. The preceding discourse unit in (1) i. elaborates on the activities of one protagonist, the mother (expressed in non-ergative pronouns and zeros). In (1) ii, a new episode starts and the attention shifts to another protagonist, the children, who will remain prominent throughout the upcoming episode.

- (1) i. ‘After that she (i.e. the mother) left the matter of stirring and she left angrily the Tima mountain; she went to the Abjunuuk area; then she went to live there in Abjunuuk, because she<sub>ZERO</sub> had got angry, she went there; she stayed until she<sub>ZERO</sub> gave birth; she<sub>ZERO</sub> got children, five children.
- ii. And then the **children**<sub>ERG</sub> said [...] they<sub>LOG</sub> won’t stay here, that they<sub>LOG</sub> come back to their mountains.

In addition to marking boundaries, ergative structures are also observed to occur in elaboration contexts. In these contexts, attention is temporarily shifted to the O argument, allowing a narrator to elaborate on it before turning attention back to the A argument.

### References

- Becker, Laura & Gertrud Schneider-Blum. 2020. Morphological marking of contrast in Tima. *Glossa: A Journal of General Linguistics* 5(1). 125–1.

- Casaretto, Antje, Gerrit J. Dimmendaal, Birgit Hellwig, Uta Reinöhl & Gertrud Schneider-Blum. 2020. Roots of ergativity in Africa (and beyond). *Studies in African Linguistics* 49(1). 111–140.
- Schneider-Blum, Gertrud. 2023. The impact of attentional centering on ergative marking in Tima. *STUF - Language Typology and Universals* 76(1): 87–112.
- Schneider-Blum, Gertrud & Birgit Hellwig. 2018. Reference tracking in Tima and its interplay with ergative marking. *Studies in Language* 42:4, 970–993.
- Schneider-Blum, Gertrud, Sonja Riesberg, Birgit Hellwig & Nikolaus P. Himmelmann. 2022. On resolving prominence conflicts – exceptional case marking in Tima and Yali. In Gianollo, Chiara, Łukasz Jędrzejowski & Sofiana I. Lindemann (eds.), *Paths through meaning and form*, 211–215. Cologne: USB Monographs.

**In search of episodic units in traditional narratives:  
A corpus-based study of Wan and Chuvash**  
Tatiana Nikitina (CNRS-LACITO)

This study aims to develop a methodology for identifying morphosyntactic phenomena that are sensitive to larger-than-sentence units in narratives. I start from the assumption that, if means for signaling such units exist, their use should correspond overall to units defined by meaning or function in narrative, and that such units should be readily identifiable by applying quantitative methods of analysis to a sufficiently large sample of data. Ideally, such methods should be cross-linguistically applicable and based on objective criteria and form-independent definitions.

To prepare the data for quantitative analysis, I annotated a selection of traditional stories from the Wan corpus (Mande, Côte d'Ivoire; Nikitina 2023) for episodic units, relying on a combination of criteria for narrative analysis from Labov (1972) and criteria for thematic unity from Dijk (1982). The annotation was performed at the clause level and made it possible to isolate narrative event sequences, to identify clauses appearing at their boundaries and to contrast them with clauses appearing sequence-internally.

I then examined the frequency of use of the major grammatical markers and discourse particles in sequence-internal clauses and compared it to their frequency at sequence boundaries. I looked specifically for asymmetries in the distributions that would suggest that a particular marker is sensitive to segmentation into episodic units. Theoretically, the method allows me to identify markers that signal segment boundaries as well as markers that signal continuity within segments (if any such markers exist in the language). I discuss the results of the comparison for the phenomena commonly associated with narrative-structuring functions in other studies; such common suspects include temporal conjunctions and discourse particles (e.g., Bakker 1993), lexical expressions such as 'it happened that' (e.g., Brinton 1993), "backgrounding" devices such as nominalization (e.g., Mithun 2002), the choice of referential expressions (pronouns vs. nouns, e.g., Fox 1987), and the choice of verbal categories such as tense and aspect (e.g., Schiffrin 1981).

To assess the cross-linguistic applicability of the proposed methodology, I also used it to annotate and analyze a sample of stories from a narrative corpus of Chuvash (Turkic, Russia; Nikitina 2022). The morphosyntactic profile of Chuvash is strikingly different from that of Wan, yet the methodology yields comparable results and allows me to formulate specific hypotheses for further testing. I speculate on possible sources of the differences in the marking of episodic units in the two languages and discuss general implications of my findings.

## References

- 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.
- Brinton, Laurel J. 1993. Episode boundary markers in Old English discourse. Jeffers, R. J. & H. Aertsen (eds.) *Historical Linguistics 1989: Papers from the 9th International Conference on Historical Linguistics, Rutgers University, 14-18 August 1989*. Amsterdam: John Benjamins, pp. 73-89.
- Fox, Barbara. 1987. Anaphora in popular written English narratives. Russell S. Tomlin (ed.) *Coherence and Grounding in Discourse: Outcome of a Symposium, Eugene, Oregon, June 1984*. Amsterdam: John Benjamins, pp. 157-174.

- Labov, William. 1972. The transformation of experience in narrative syntax. Chapter 9 from W. Labov. *Language in the Inner City: Studies in the Black English Vernacular*. Philadelphia: University of Pennsylvania Press, pp. 354-405.
- Mithun, M. 2002. Rhetorical nominalization in Barbareño Chumash. *Survey Reports, Survey of California and Other Indian Languages* 12: 55-63.
- Nikitina, Tatiana. 2022. A narrative corpus of Chuvash. In Nikitina, Tatiana, Ekaterina Aplonova, Izabela Jordanoska, Ekaterina Biteeva, Abbie Hantgan-Sonko, Guillaume Guitang, Olga Kuznetsova, Rebecca Paterson, Elena Perekhvalskaya & Lacina Silué (eds.) *The SpeechReporting Corpus: Discourse Reporting in Storytelling*. CNRS-LLACAN & LACITO, <http://discoursereporting.huma-num.fr/index.html>
- Nikitina, Tatiana. 2023. A narrative corpus of Wan. In Nikitina, Tatiana, Ekaterina Aplonova, Izabela Jordanoska, Ekaterina Biteeva, Abbie Hantgan-Sonko, Guillaume Guitang, Olga Kuznetsova, Rebecca Paterson, Elena Perekhvalskaya & Lacina Silué (eds.) *The SpeechReporting Corpus: Discourse Reporting in Storytelling*. CNRS-LLACAN & LACITO, <http://discoursereporting.huma-num.fr/index.html>
- Schiffirin, Deborah. 1981. Tense variation in narrative. *Language* 57(1): 45-62.

### Episode structure in A'ingae

Kees Hengeveld (ACLC, University of Amsterdam)

A'ingae is a language isolate spoken in the border region of Colombia and Ecuador. A sketch of the grammar of the language can be found in Fischer & Hengeveld (2023). The language exhibits a number of grammatical features that demonstrate that the semantic category of Episode is highly relevant for its analysis. This paper gives an overview of these features and their interaction. I use the theory of Functional Discourse Grammar (Hengeveld & Mackenzie 2008) to develop the description of the system, as this framework posits the episode as one of its layers of semantic analysis.

Episodes can be defined as combinations of events or situations that are thematically coherent, in the sense that they show unity or continuity of participants, time, location, and reality domain. Episodes defined in this way are formally reflected in the grammar of A'ingae in at least two different ways: (i) in the marking of new and contrastive topics, and (ii) in cosubordinating clause chaining constructions in combination with tail-head linkage.

The new and contrastive topic markers are shown in (1) and (2):

- (1) Khaki=**ta**                      jaja='fa                      dûtsi'ye=ndekhû.  
 next\_day=NEW    go=PLS                      child=APL  
 'The next day the young men went out.' (BC03.001)
- (2) Kaje=ni=**ja**                      me'in.  
 downriver=LOC=CONTR    NEG  
 'Downriver it's not like that.' (BC24.056)

Topic markers mark the element that is responsible for there being a new episode. This may reflect a change in time (1), location (2), but also a change of participants or reality domain. The latter means that conditional clauses are marked with a topic marker as well.

Turning now to clause chaining, consider the combination of three clause chains in combination with tail-head linkage illustrated in (3):

- (3) a Paña=mba    anthe=pa    bûtu=in    ji                      tsa=ni.  
 hear=SS    stop=SS    run=SIM    come                      house=LOC  
 '(The mother) heard it, stopped work and came running to the house.' (BC08.023)
- b Ji=pa                      ichuru'chu=ma    khase                      isû.  
 come=SS    gourd\_bowl=ACC1    again                      take  
 'Coming she took out the gourd bowl again.' (BC08.024)
- c Isû=pa    khase                      khusha='kan.  
 take=SS    again                      drum=SIMIL  
 'Taking it she again tried sort of drumming on it.' (BC08.025)

The last verb in (1a), *ji* 'come', is repeated as the first verb in (1b). Similarly, the last verb in (1b), *isû* 'take', is repeated as the first one in (1c). The non-final verbs in (1a-c) are all marked for same subject, showing subject continuity.

As shown in this example, not every individual clause chain corresponds to a semantic episode. Unambiguous marking of a new episode occurs when the head in a tail-head construction carries one of the topic markers discussed above, as in (4).

- (4) a. Tsa='ma=tsû ju=ni Pasto=ni=nda=te kan'jen.  
 ANA=FRT=3 DIST=LOC Pasto=LOC=NEW=RPT live  
 'But it is said that she lives in Pasto.'
- b. Kan'jen=mba=**ta** dû'shû khi=chu=khu tsûi ja-yi.  
 live=SS=NEW child thus=CLF:RND=AUG walk go-PROSP  
 'A child that lives there, this big, was walking along.' (BC26.122-123)

In (4b) the topic marker =*ta* attaches to the head, a phenomenon that is only observed with new episodes.

The two phenomena discussed thus crucially interact in signalling episode boundaries.

### Abbreviations

3	third person
ACC.REAL	realis accusative
ANA	anaphoric reference to entities or events
APL	animate plural
AUG	augmentative
CLF:RND	classifier round and small object
CONTR	contrastive topic
FRT	frustrative
LOC	locative
MANN.CV	manner converb
NEGP	negative predicate
NEW	new topic
PLS	plural subject
PROSP	prospective
SIMIL	similative
SS	same subject

### References

- Fischer, Rafael & Hengeveld, Kees. 2023. A'ingae (Cofán). In: Patience Epps & Lev Michael (eds), *Amazonian languages: An International Handbook. Language Isolates I: Aikanã to Kandozi-Chapra*. Berlin: de Gruyter Mouton, 65-123.
- Hengeveld, Kees & Mackenzie, J. Lachlan. 2018. *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: Oxford University Press.



### **Annotating boundary cues to determine paragraphs in spoken language**

Katharina Gayler<sup>1</sup>, Maria Bardaji<sup>1</sup>, Sonja Riesberg<sup>1,2</sup>, Nikolaus P. Himmelmann<sup>1</sup>  
(<sup>1</sup>Universität zu Köln, <sup>2</sup>CNRS-LaCiTO)

‘Episodes’ or ‘paragraphs’ as structuring units in discourse have been defined as “coherent sequences of sentences of a discourse, linguistically marked for beginning and/or end, and further defined in terms of some kind of ‘thematic unity’ – for instance, in terms of identical participants, time, location or global event or action” (van Dijk 1981: 177). While this definition sounds conclusive at first sight, determining the boundaries of such units in spoken language is not at all a straight-forward task, and prone to circularity.

In this talk, we explore a bottom-up method for determining the boundaries of larger discourse units. In a pilot corpus consisting of 10 re-tellings of the Pear Movie (Chafe 1980) from two different languages – Totoli (Western Malayo-Polynesian) and Yali (Trans-New Guinea) –, we annotated the following syntactic and prosodic boundary marking cues:

- (filled) pauses
- prosodic final falls
- temporal expressions
- existential/presentational constructions
- discourse markers
- tail-head linkage
- (beginning & end of clause chaining)

Our general hypothesis is that where these features cluster, speakers intended to mark a discourse boundary.

But our corpus study also shows that:

- cues do not necessarily have to cluster
- there are smaller and larger clusters
- there are language-specific prototypical clusters
- there are speaker-specific clusters

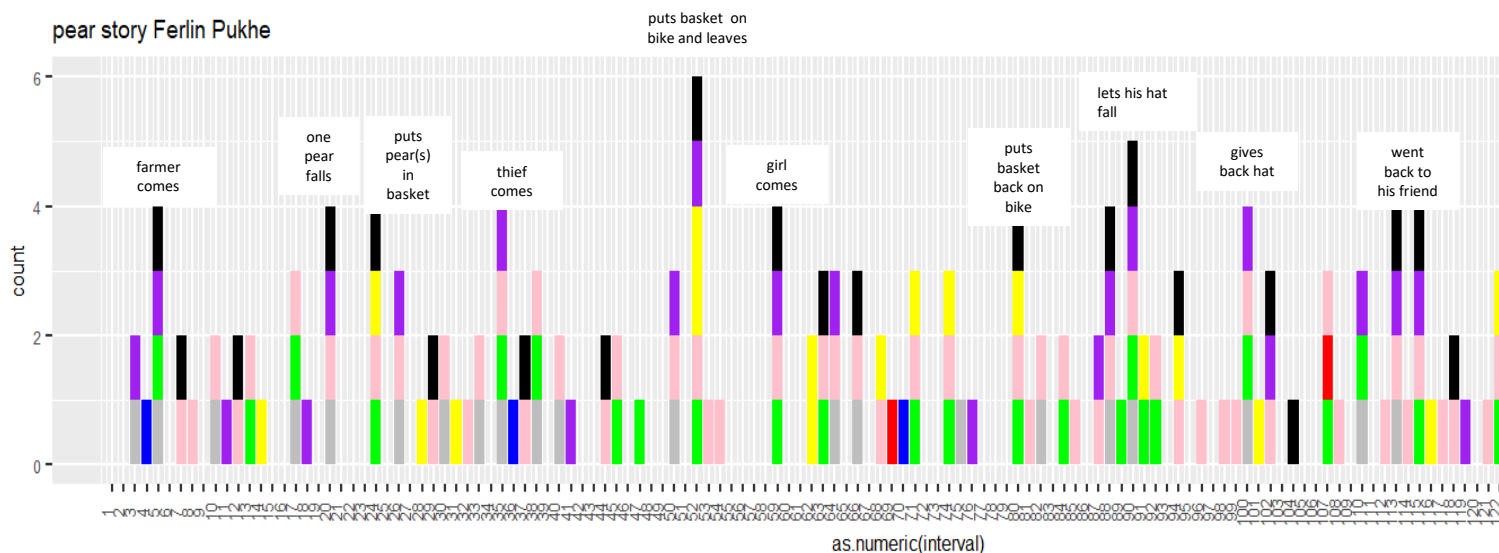
The first two points have relevant theoretical consequences: The fact that features do not necessarily cluster shows that none of them, on their own, can be claimed to have a boundary marking function *per se*. And the fact that we observe smaller and larger clusters implies that we can determine weaker and stronger boundaries in discourse. The Figure on the following page shows clusters of various size in one Yali narrative (with the respective scene in the Pear Movie).

When comparing the Totoli data with the Yali data, we find that the semantic units demarcated by feature clustering largely overlap. But we will also show that there are certain differences in the way Totoli and Yali speakers in our sample structure their narratives. We will furthermore present some language- and speaker-specific boundary marking strategies.

### **References**

Chafe, Wallace. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.

van Dijk, Teun A. 1981. Episodes as units of Discourse Analysis. In Deborah Tannen (ed.), *Analyzing Discourse: Text and Talk*, 177–195. Georgetown: Georgetown University Press.



- chain end = follows the end of a clause chaining construction
- chain start = constitutes the beginning of clause chaining construction
- DM = contains a discourse marker
- final fall = follows a final fall
- FP = follows a filled pause
- pause = follows a long pause
- presentational = contains a presentational construction
- THL = constitutes the head of a tail-head linkage construction

**Difficulty in identifying the discourse boundaries in Burmese**  
San San HNIN TUN (INALCO/Lacito-CNRS)

Subject and object are usually overtly marked in many western (or Indo-European) languages whereas it is not the case in Burmese. Theoretically, there are a subject marker /ka<sub>1</sub>/ and an object marker /ko<sub>2</sub>/ in Burmese, which are postpositionally attached to the NP, allowing the word order to be flexible [e.g. (1)-(2), producing identical English (translation)].

- (1) /cə.ma<sub>1</sub>    **ka<sub>1</sub>**        θu<sub>1</sub>    **ko<sub>2</sub>**        me<sub>3</sub>    dɛ<sub>2</sub>/  
1Sg.F.    **Sub.Mr**    3Sg.    **Obj.Mr**    ask    SFM  
*I asked him/her.*
- (2) /θu<sub>1</sub>        **ko<sub>2</sub>**        cə.ma<sub>1</sub>        **ka<sub>1</sub>**        me<sub>3</sub>    dɛ<sub>2</sub>/  
3Sg.        **Obj.Mr**    1Sg.F.        **Sub.Mr**    ask    SFM  
*I asked him/her.*

However, it is common practice that the (grammatical) subject (or even the object) is often dropped in a naturally occurring discourse in Burmese, as long as the speaker feels that the information is retrievable in the (situational) context. Furthermore, object pronouns such as *it* are not systematically used in Burmese [e.g. (3) where the object pronoun *it* is to be retrieved in the context]. Moreover, address terms such as *Father*, *Daughter* or proper names are often used as pronouns: e.g. when one hears *daughter* in a discourse, it could be *I*, *You* or *She*, depending on the situational context [e.g. in (4.a) 'daughter' means *I* since it is a woman speaking to the father but in (4.b) 'daughter' means *you*, since it is the father speaking to the daughter – both sentences taken from a spoken discourse]. To my experience (as a language teacher & researcher of the Burmese language), besides the fact that when new information is introduced, the subject (or object or their markers) are overtly marked, there are no other morphosyntactic clues but only the semantic content that plays a key role in identifying the discourse boundaries.

- (3) /θi<sub>1</sub>        dɛ<sub>2</sub>/  
*know*        SFM  
*I know/knew (it).*
- (4.a) /θə.mi<sub>3</sub>    you<sup>n</sup><sub>3</sub>        ka<sub>1</sub>    pya<sup>n</sup><sub>2</sub>        yau?        pa<sub>2</sub>        bi<sub>2</sub>/  
*daughter office*    *from back*    *arrive*        *Ptcl(polite)*    SFM  
*I'm now back from the office.*
- (4.b) /θə.mi<sub>3</sub>    ko<sub>2</sub>        sau<sup>n</sup><sub>1</sub>    ne<sub>2</sub>        da<sub>2</sub>/  
*daughter* **Obj.mr**        *wait*    **Progressive**        SFM  
*I'm waiting for you.*        [cf. 2 excerpts from the spoken corpus 'RP.PePe']

Similarly, my experience as a language instructor (of Burmese as a foreign language) suggests that one way to tell someone's linguistic competence is to see whether the person is producing isolated sentences or if his/her sentences are well-knitted: i.e. whether a learner is capable of developing a topic in a well-thought-out, sequential sentences, through the use of appropriate transition words such as (equivalents of) *Therefore* (/di<sub>2</sub> dɔ<sub>1</sub>/), *And then* (/pi<sub>3</sub> dɔ<sub>1</sub>/), etc. That said, we are under the impression that Burmese discourse – spoken or written – seems arbitrary: when a discourse is divided into paragraphs, the dividing lines may vary from a speaker to another.

In this presentation, I plan to discuss the difficulties in defining discourse boundaries in Burmese through examples taken from excerpts from four different spoken Burmese corpus texts (a short story, an interview, a narration & a radio play). It is our hope that we can gain some more linguistic insights from this workshop to define discourse boundaries in Burmese.

### Abbreviations

- 1     *First person*
- 3     *Third person*
- F     *feminine*
- Mr   *marker (=particles with grammatical functions)*

NP *Noun Phrase*

Obj *object*

Ptcl *particles (=when a lexical item does not express grammatical functions)*

SFM *Sentence Final Marker (=lexical item indicating the end of a sentence)*

Sg *singular*

Subj *subject*

### **Bibliography**

Chafe, Wallace, 1994, *Discourse, consciousness & time. The flow and displacement of conscious experience in speaking & writing*. Chicago: The University of Chicago press.

Dept. of Myanmar Language Commission. 1994. *Myanmar-English Dictionary*.

ⓂN\MAAGÇLIP\ABIDAN\ Yangon, Myanmar: Ministry of Education, Union of Myanmar

Eggs, Suzanne & Diana Slade. 1997. *Analyzing casual conversation*. London & Washington: Cassell.

Harris, Zellig S. 1952. 'Discourse analysis' in *Language* Vol 28, N° 1, Linguistic Society of America, pp. 1-30 [JSTOR - <https://doi.org/10.2307/409987>]

Hnin Tun, San San & Mathias Jenny, 2013, 'Differential subject marking without ergativity: the case of colloquial Burmese', in *Studies in Language*: Vol. 37, Issue 4, Amsterdam, Holland: John Benjamins Publishing Co.

McCarthy, Michael & Ronald Carter, 1994, *Language as Discourse: perspectives for language teaching*, Harlow: Longman.

Okell, John. 1969. *A Reference Grammar of Colloquial Burmese*, 2 vols. London: Oxford University Press.

Okell, John. & Anna Allott. 2001. *Burmese/Myanmar Dictionary of Grammatical Forms*. Surrey, UK: Curzon.

Ozerov, Pavel, 2014, *The system of information packaging in colloquial Burmese*, Thesis, Australia: La Trobe university.

Ozerov, Pavel, 2015, 'Information structure without topic and focus: Differential Object marking in Burmese' in *Studies in Language*, Vol. 39:2 [<https://doi.org/10.1075/sl.39.2.04oze>]

Schiffirin, Deborah, Deborah Tannen & Heidi Hamilton, 2003, *The Handbook of Discourse Analysis*, Oxford, UK: Blackwell Publishers Limited

Stenström, A.-B. 1994. *An introduction to spoken interaction*. London & New York: Longman.

Van Dijk, Teun, 1981, 'Discourse studies and education', in *Applied Linguistics*, Vol II, Issue 1, pp. 1 - 26 [<https://doi.org/10.1093/applin/II.1.1>]

### **Corpus :**

1. FICT.Samsara – Fiction. *Samsara*
2. INT.KLM – Interview (Interviewee talks about her experience as a journalist in Myanmar)
3. NARR.HL – speaker talks about his experience being arrested (for political reasons)
4. RP.PePe – a radio play (audio soap opera, broadcast in Burmese in Myanmar)

## Perceptual reported speech in traditional Nganasan narratives

Denys Teptiuk (University of Tartu)

Reported speech (RS) can be used to express perception across the world's languages (Adelaar 1990; Aikhenvald & Storch 2013; de Vries 2013). In Nganasan (Uralic, Northern Siberia), perceptual RS appears predominantly in traditional folktales to express what characters perceive by vision or other senses, cf. (1)–(2), and fulfills the rhetorical functions of highlighting narrative events or participants (Larson 1978). This phenomenon is a recurring feature in storytelling by different speakers, where it encodes perception together with lexical and grammatical evidential strategies (Teptiuk & Nikitina 2023).

### (1) Vision rendered in RS

*Ou, mununtu Hal'mira T'indama, tahari'ai? əmə?*  
 INTERJ say:PRS.[3SG.S] PN now this:ADJ:NOM.PL  
*əku? huəd'ə? kontəd'əə?*  
 probably:3PL.S after:ADJREL:NOM.PL sacrifice:NPST:NOM.PL  
 'Ah, – says Halmira Tchindama, – it's probably former sacrifices.'  
 (ChND\_061025\_Haljmira\_flks.270)

### (2) Hearing rendered in RS

*Təti ni mununtu[:] Əbəj maad'a*  
 that woman say:PRS.[3SG.S] INTERJ why  
*tujmə səiŋiðə?*  
 fire:NOM.1SG.POSS shoot:INTER:3SG.R  
 'The woman says: "Why is my fire crackling?"' (KNT\_940903\_KehyLuu\_flkd.81)

Perceptual RS shows only partial similarity with other perceptual expressions in dynamic texture of perception. Lexical strategies more frequently refer to static objects/properties ('A tent is seen'), while evidentials are mainly occupied with actions ('Someone arrived, I hear'). or a dynamic action (2). Furthermore, as characteristic of RS (Spronck & Nikitina 2019), additional modal meanings are observed in the expression of perception: uncertainty (1) and surprise (2).

Although many instances of perceptual RS are framed by quotative strategies, it can also be unframed (3). Similarly to (1) and (2), unframed RS often starts with an interjection.

### (3) (Fog rose and he lost the way. His wife was alone at home. She is a Nganasan woman.)

*Ou?, kəkutu tə? n'aagimi?ə.*  
 INTERJ fog PTCL.ENIM good:TRL:EP:PF.[3SG.S]  
 'Ah, the fog has lifted.' (JDH\_00\_Njaakju\_flkd.008)

Cases like (3) can be associated with multiple perspectives: protagonist's, narrator's, or belonging to *ŋala* 'tale, mouth', an independent character following events in the tale.<sup>1</sup> On many occasions, the position of an unframed perceptual RS in the narrative structure helps to identify the perspective. For instance, the perspective of *ŋala* 'tale' is often present during the Orientation, as in (3). In contrast, the protagonist's perspective is prevalent during the Complicating action (Labov 1972).

<sup>1</sup> Cf. a characteristic feature of Northern Samoyedic (Enets, Nenets, Nganasan) folklore (Dolgix 1976: 16).

This paper analyzes perceptual events rendered in RS in the subcorpus of 22 tales from the Nganasan Spoken Language Corpus (Brykina et al. 2018). First, I focus on internal characteristics of perceptual RS and compare them with other (i.e. “non-rhetorical”) instances of RS. I check if the idea that RS adds vividness to the story (Labov 2006, Verstraete 2011) applies to Nganasan narratives and how variation between framed and unframed perceptual RS may add to this argument. Second, I compare the position of perceptual RS in the narrative structure and check if its occurrence can be associated with particular narrative episodes (e.g. introducing new characters or highlighting an important action, cf. Larson 1978).

## References

- Adelaar, Willem F.H. 1990. The role of quotations in Andean discourse. In: Pinkster, Harm (ed.), *Unity in Diversity: Papers Presented to Simon C. Dik on his 50<sup>th</sup> Birthday*, pp. 1–13. Dordrecht: Foris.
- Aikhenvald, Alexandra Y. & Anna Storch. 2013. Linguistic expression of perception and cognition: A typological glimpse. In: Aikhenvald, Alexandra Y. & Anna Storch (eds.), *Perception and Cognition in Language and Culture*, 1–46. Leiden & Boston: Brill.
- Brykina, Maria, Valentin Gusev, Sándor Szeverényi & Beáta Wagner-Nagy. 2018. *Nganasan Spoken Language Corpus (NSLC)*. Version 0.2. <http://hdl.handle.net/11022/0000-0007-C6F2-8>
- de Vries, Lourens. 2013. Seeing, Hearing and Thinking in Korowai, a Language of West Papua. In: Aikhenvald, Alexandra Y. & Anna Storch (eds.), *Perception and Cognition in Language and Culture*, 111 – 137. Leiden & Boston: Brill.
- Dolgix, Boris O. 1976. *Mifologičeskie skazki i predanija nganasanov* [Nganasan mythological fairytales and legends.] Moscow: Nauka.
- Labov, William. 1972. *Language in the Inner City*, Chapter 9. The Transformation of Experience in Narrative Syntax, 354 – 396. Philadelphia: University of Pennsylvania Press.
- Labov, William. 2006. Narrative pre-construction. *Narrative Inquiry* 16(1), 37–45.
- Larson, Mildred L. 1978. *The Functions of Reported Speech in Discourse*. Dallas: Summer Institute of Linguistics.
- Spronck, Stef & Tatiana Nikitina. 2019. Reported speech forms a dedicated syntactic domain: Typological arguments and observations. *Linguistic Typology* 23(1): 119–159.
- Teptiuk, Denys & Tatiana Nikitina. 2023. Evidential strategies in narrative discourse: A contrastive approach. Paper presented at the *56th Annual Meeting of Societas Linguistica Europaea*, University of Athens.
- Verstraete, Jean-Christophe. 2011. The Functions of Represented Speech and Thought in Umpithamu Narratives. *Australian Journal of Linguistics* 31(4). 491–517.

## **An experimental approach to episode boundaries: testing boundary cues in Western Austronesian languages**

*Maria Bardaji<sup>1</sup>; Sonja Riesberg<sup>1,2</sup>; Nikolaus P. Himmelmann<sup>1</sup>*  
(<sup>1</sup>Universität zu Köln; <sup>2</sup>CNRS-LaCiTO)

It has been proposed in the literature that narrative discourse is structured in semantic units larger than a sentence, commonly called *episodes* or *paragraphs* (Chafe 1980, Longacre 1979). These units are also demarcated by linguistic means (van Dijk 1981), such as long pauses, prosodic contours and filled pauses (Swerts 1997; 1998). In addition, it is also assumed that the boundaries of discourse units can be of different strengths. Swerts (1997; 1998), for instance, distinguishes between weak and strong discourse boundaries.

In this talk, we address the question of whether the nature of these units and their boundaries can be tested experimentally: can we prompt episode boundaries of different strength by means of visual and auditory stimuli?

We explore this topic by means of data from Totoli, a Western Austronesian language spoken in the northern part of Central Sulawesi (Indonesia). Totoli has a symmetrical voice system with two transitive constructions (*actor voice* and *undergoer voice*). Among other functions, it seems that actor voice constructions in Totoli play a role in the structuring of narrative discourse. In a corpus of retellings of the Pear Film (Chafe 1980), actor voice constructions appear to occur with increased frequency at prominent (or strong) discourse boundaries. That is, one finds more actor voice constructions at boundaries that involve a major change in the setting of a narrative (e.g. new participants, location or time) and are additionally marked by means such as discourse markers, tail-head linkage, long pauses, etc.

We hypothesize that speakers will most likely produce an actor voice construction after a strong boundary, whereas a weak boundary will be most likely followed by an undergoer voice construction. We report on an experiment that tests this hypothesis by means of a story continuation task. In this experiment, speakers first see one picture while hearing a short story segment relating to this picture and are then asked to continue the story based on a second picture. Both the picture sequence and the narrative accompanying the first picture differ with regard to whether a weak or a strong boundary is being cued at the end of the story segment. Figure 1 shows the results of a pilot of this experiment, which are promising. Speakers produced more actor voice constructions after a strong boundary (stimuli 1–4, except for stimulus 3) and more undergoer voice constructions in the weak boundary condition (stimulus 5).

While the data presented in this talk only concern Totoli, we hypothesize that the same approach can be applied to other Western Austronesian languages. The talk will present the design of a revised version of the pilot experiment (with a better control of the boundary triggers), that we plan to run with Tagalog speakers in the near future.



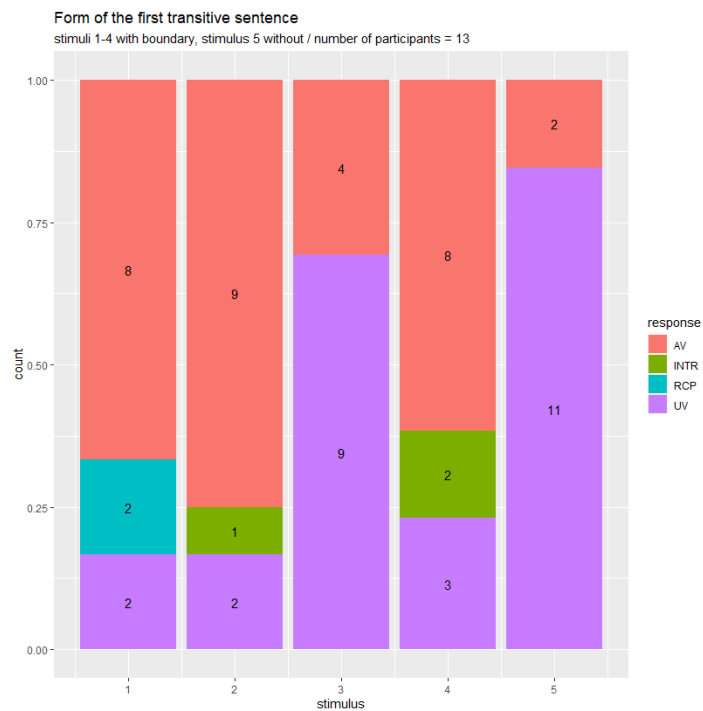


Figure 1 Voice choice of the first transitive sentence in a story continuation task

## References

- Chafe, Wallace L. 1980. *The pear stories: Cognitive, cultural, and linguistic aspects of narrative production*. Norwood, NJ: Ablex.
- Longacre, Robert E. 1979. The paragraph as a grammatical unit. In: Givón, Talmy (ed.) *Syntax and semantics 12. Discourse and syntax*, 115–134. New York: Academic Press.
- Swerts Mark. 1997. Prosodic features at discourse boundaries of different strength. *The Journal of the Acoustical Society of America*, 101(1), 514–521. <https://doi.org/10.1121/1.418114>
- Swerts, Mark. 1998. Filled pauses as markers of discourse structure. *Journal of pragmatics*, 30(4), 485-496.
- Van Dijk, Teun. A. 1982. Episodes as units of discourse analysis. *Analyzing discourse: Text and talk*, 177-195.