

General Project Aim

- 1) Create a **standardized procedure** for annotating non-manual markers (NMMs) in sign language and gesture data
- 2) Create a **standard method** to **quantify inter-annotator agreement**

Stage 1: Initial Guidelines

- We present a first version of the annotation guidelines in Oomen et al. (2023)¹
- 12 tiers of NMMs

Stage 2: First Annotation Round

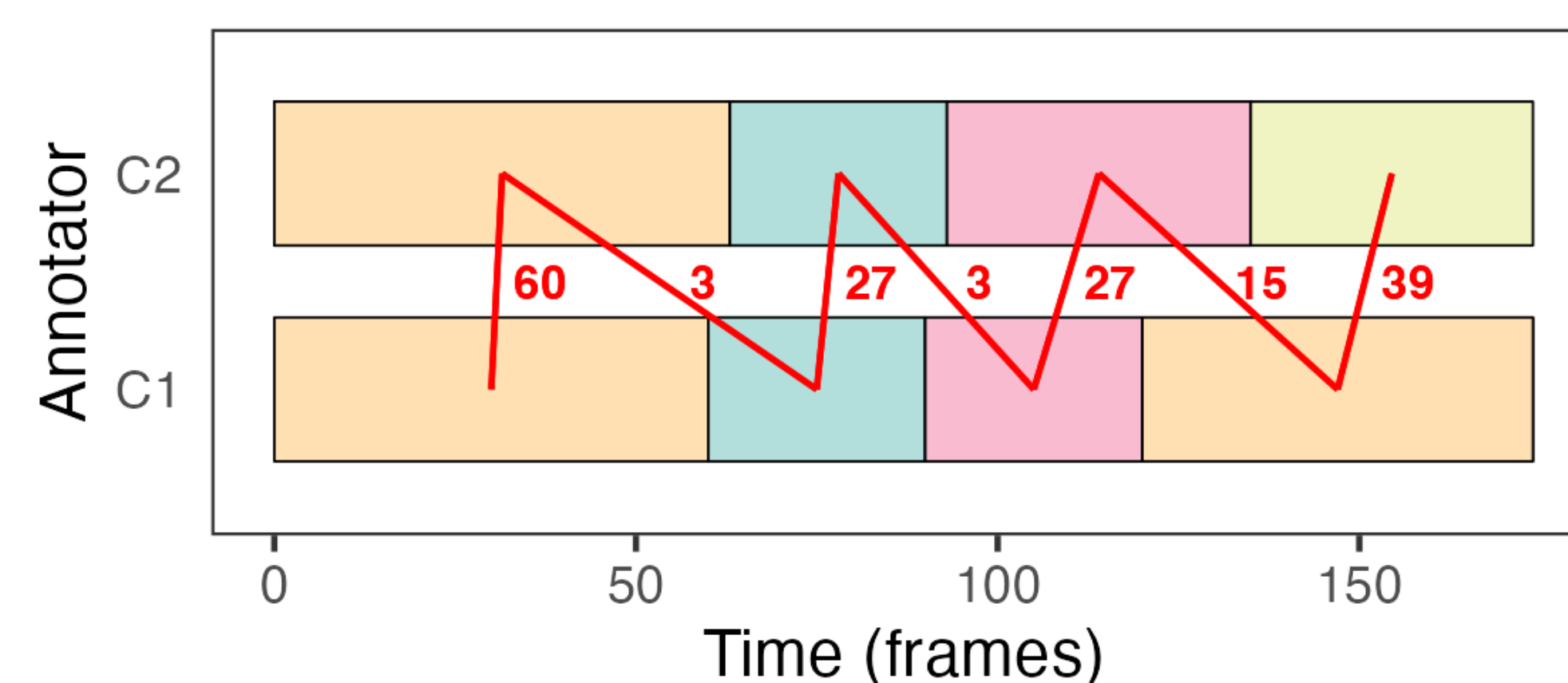
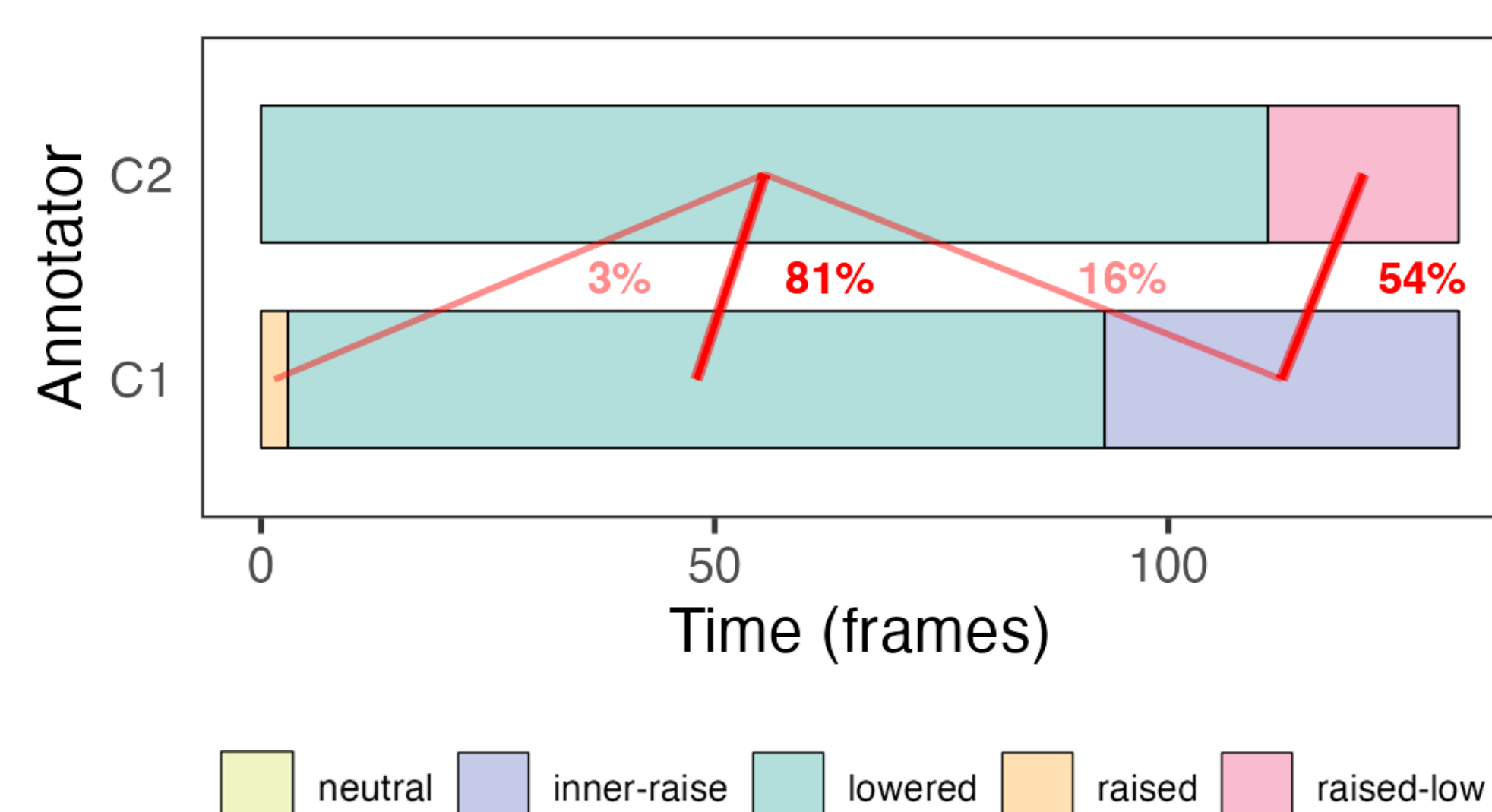
- Test set of 60 interrogative sentences in Sign Language of the Netherlands (NGT)
- Two coders produced annotations in ELAN
- Coder 1: 585 events
- Coder 2: 564 events

Stage 3 (present paper): Evaluating Inter-Annotator Agreement

Aim 1 (feeds into stage 4+5): evaluating the agreement between annotators

- Two methods: event-based and frame-based
- Combination of confusion matrices and agreement index (Cohen's Kappa)
- Results: **tier-specific recommendations** as well as several **general recommendations** for improvement of annotation guidelines
 - Example tier-specific recommendation: removal of the 'raised-low' label in eyebrows tier due to excess ambiguity
 - Example general recommendation: distinction should be made between **poses** and **movements** (a similar distinction has been made within the SignStream project)²

Event-based method



Frame-based method

Aim 2 (feeds into stage 6): evaluating the evaluation methods

- Previous literature advised the use of **both** frame-based and event-based methods³
- Advantage event-based method: allows for **error analysis** (determination type of error)
- Frame-based method (combined with event-based) provides information about **nature of disagreements**
- However: frame-based method is **indirect**: never provides definitive insight into disagreements
- We propose an **enriched version** of the event-based method
 - Automatic categorization of error-types
 - Keep track of additional information (e.g., duration of events with disagreement among coders)
- This yields the frame-based method **superfluous**

Stage 5: Second Annotation Round

- Collect new video data targeted at the range of labels in the updated guidelines
- Two new coders annotate the new dataset with the updated guidelines

Stage 4: New Guidelines

- Create new annotation guidelines based on tier-specific and general recommendations drawn from evaluation

Stage 6: Evaluate IAA

- Evaluate inter-annotator agreement with updated and extended evaluation procedure

Outlook

- **Community effort** to further develop guidelines as necessary
- **Divide responsibility** for specific tiers over participating research teams
- Make guidelines **accessible** in various (signed and written) languages and present them in an interactive format
- End goal: guidelines are **thoroughly validated** and adopted as a **community-wide standard**



¹ Marloes Oomen, Lyke D. Esselink, Tobias de Ronde, and Floris Roelofsen. 2023. First steps towards a procedure for annotating non-manual markers in sign languages. In NELS 53. <https://doi.org/10.21942/uva.25563453>

² Carol Neidle. 2002. SignStream annotation. <https://www.bu.edu/aslirp/aslirpr11.pdf>

³ Roger Bakeman, Vicenç Quera, and Augusto Gnisci. 2009. Observer agreement for timed-event sequential data. <https://doi.org/10.3758/BRM.41.1.137>