LREC 2022 Marseille

SLTAT 2022

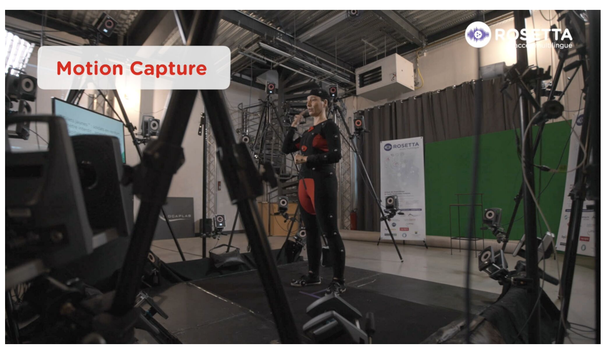
Example-Based Multilinear Sign Language Generation from a Hierarchical Representation

Boris Dauriac - Mocaplab, Aubervilliers, France – boris.dauriac@mocaplab.com
Annelies Braffort – LISN, Paris-Saclay Univ., CNRS, Orsay, France – annelies.braffort@lisn.upsaclay.fr
Élise Bertin-Lemée - SYSTRAN, Paris, France – elise.bertinlemee@systrangroup.com

Introduction

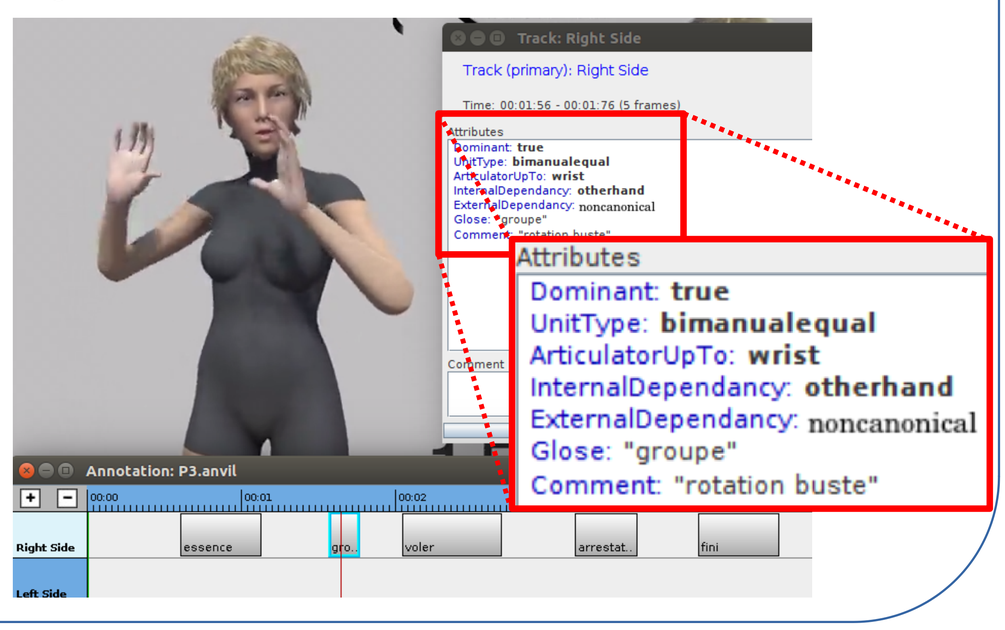
- **Rosetta** : French project on audiovisual content accessibility → Automatic translation from text to French Sign Language, displayed through avatar animation
- 3 main contributions :
  - Rosetta-LSF corpus : aligned corpus of text, SL captured using a mocap system and AZee descriptions
  - Translation system from text to AZee (a representation of SL content)
  - **System allowing to generate virtual signer animations from AZee input**
- Chosen approches
  - Generation of basic animations by rendering **motion capture** as video clips to ensure a high level of bio-realism
  - Generation of final animation using a **hybrid method** : multitrack, smart concatenation, edition of animation blocks to ensure contextual adaptation

Rosetta-LSF aligned corpus

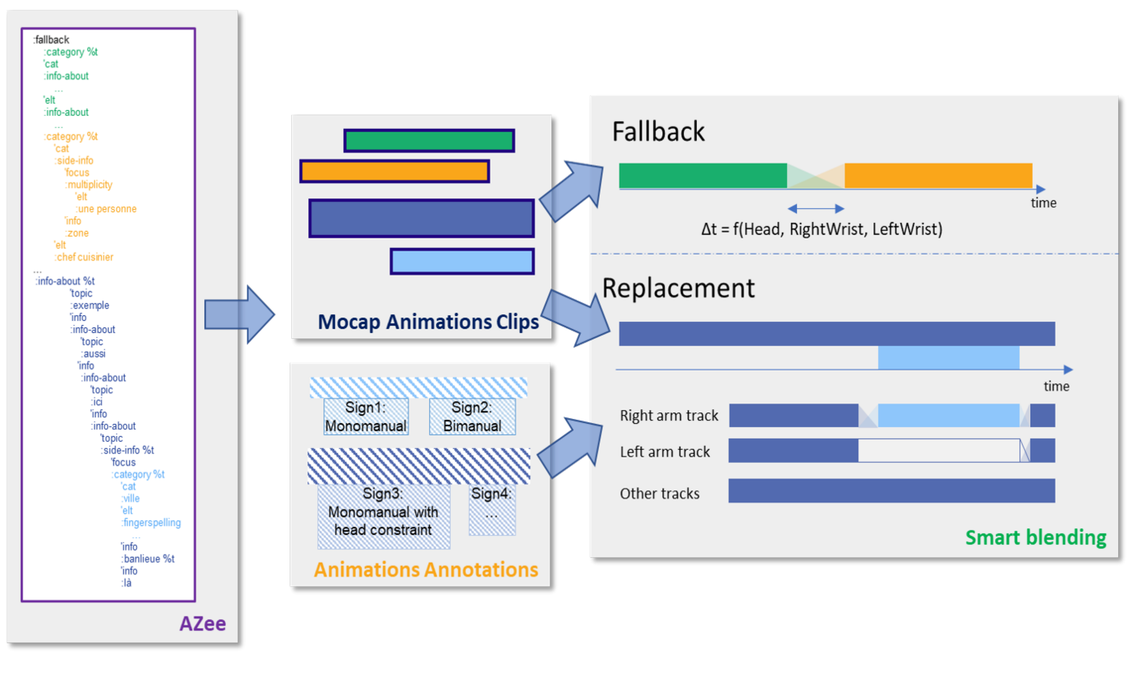


- Description of the corpus constitution in [Bertin-Lemée et al. 2022] (LREC main conf.)
- Task 1: translation into LSF of 194 news in French, for instance:
  L'Everest menacé de réchauffement climatique
  Everest threatened by global warming
- Rendering on a 3D avatar with the same body proportions as the signer
- Annotation of 4 types of internal and external constraints, designed to allow extraction of relevant elements for generation

- **UnitType**: between the 2 hands in bimanual signs
- **ArticulatoryUpTo**: from fingers to elbow (more than just handshape)
- **InternalDependency**: between the hand and other parts of the body
- **ExternalDependency**: with respect to the signing space
- Example: sign GROUP



Generation Methodology



- Input : **AZee** expression describing a SL utterance
  Tree-structured expression, each node → portion of the utterance
  Including information on the corresponding animation **block**
- Basic animations : Mocap animation clips
  Multitrack approach : animations split in several tracks (~ anatomical)
- Final animation using **non-linear blending** approach with 2 cases :
  1) **Fallback**, when two blocks must be linked on all tracks.
     - Calculation of the distance covered by the end-effectors with the highest dynamics (head, right and left wrists) to estimate a biorealistic duration.
     - Proportional calculation on the distance covered
  2) **Replacement**, when a sub-block must be replaced on some tracks only.
     In many cases: replacement of arm animations while keeping the other tracks unchanged to keep the AZee structure.
     Use of the UnitType and InternalDependancy attributes.
     - New duration : inserted block duration → other tracks stretched or squeezed
     - Replacement bimanual to monomanual (or reverse)
       → non-dominant tracks emptied (or filled)
     + computation of the transitions in all cases

Results & Prospects

- Test of the functionnality of the system on 7 sentences (see videos)
  Positive feedback on the fluidity of the animations
- Aspects to be addressed
  - Nonmanual: Not yet annotated, but methodology already allows for managing these elements
  - Signing space : Not yet integrated in the generation methodology, but annotation already provides information (ExternalDependancy attribute)
  - Evaluation : design of metrics to evaluate linguistic fluency and "body fluency"

Acknowledgement:
This work has been funded by the Bpifrance investment project "Grands défis du numérique" as part of the **Rosetta** project.