Annotation of Nonmanual Signals for Automatic Sign Language Generation

Rosalee Wolfe, John McDonald, Melissa Bialek
School of Computing
DePaul University
Chicago, IL

Outline

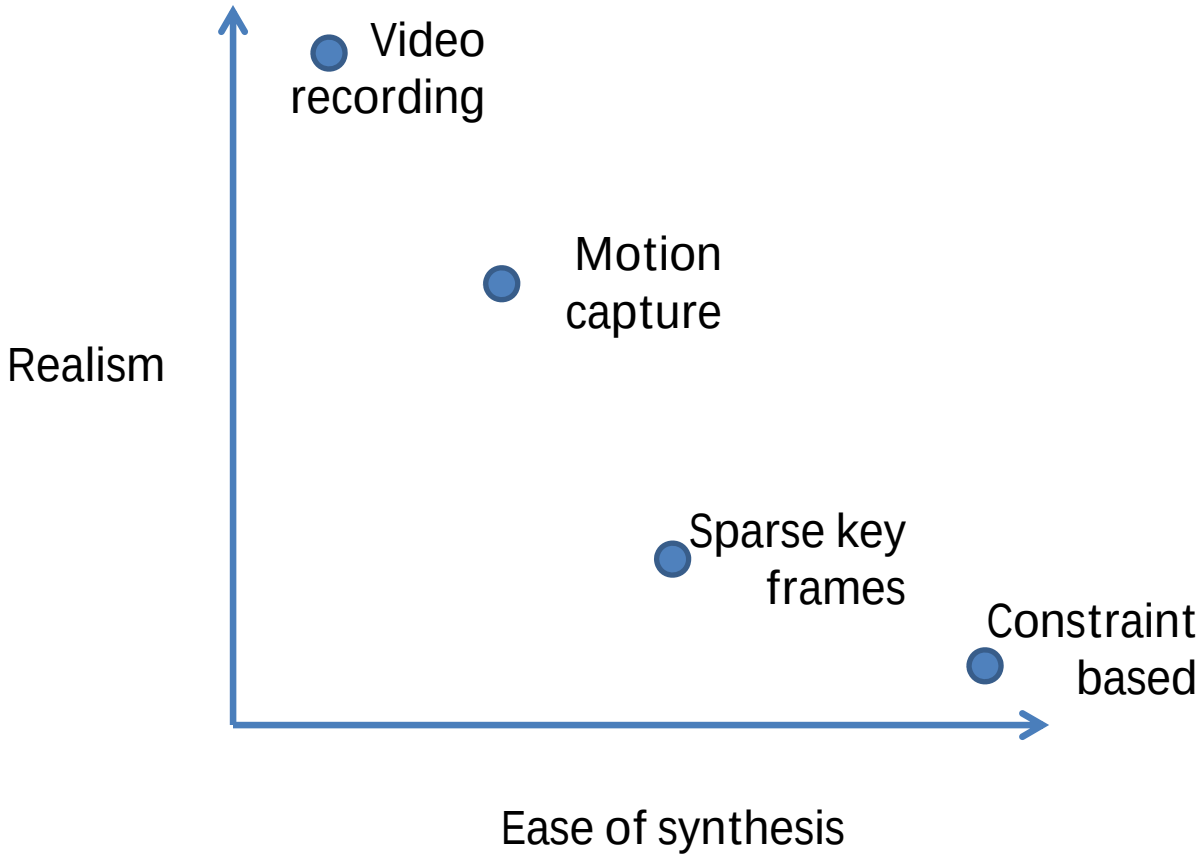
- Basic technologies of sign language generation
- Current annotation practice
- Challenges in current annotation approaches
- Open questions and possible alternatives

Language generation

- Required in any automatic translation system
- Spoken to signed language
  - Must portray human motion

Alternatives

Can be combined!



Generation and Annotation

- Video, mocap
  - Linear time-based data
  - ELAN, iLex, ANVIL
- Sparse key frame, constraint-based
  - Annotations drive these systems.

Current annotation practice

- Transcription systems
  - HamNoSys
  - Sign Writing
- Guidelines
  - ECHO conventions
  - Auslan corpus
  - ASLLRP
- Theory neutral or theory dependent?

Facial anatomy: Brows

- ECHO conventions:
  - r (raised) f (furrowed)
- Auslan guidelines:
  - raised brows / lowered brows
- Signstream:
  - Rudimentary envelope
    - S “start”
    - Steady state
      - -lowered / lowered / +lowered
      - -raised / raised / +raised
    - e “end”
- SignWriting: seven settings

What, but not why

- Phonetic / phonemic annotations are descriptive.
- What the brows are doing, but not why
- SL generation requires knowledge linguistic processes that cause the brow movement.
  - Must make decisions on how to control the movement
  - Set priorities for the contribution of each process

One example: Up, up and more up!

- English:

  Have you succeeded at last? ☺

- ASL:

  _____happy
  _____y/nq

  PAH

Three processes

1. Affect (extra linguistic)

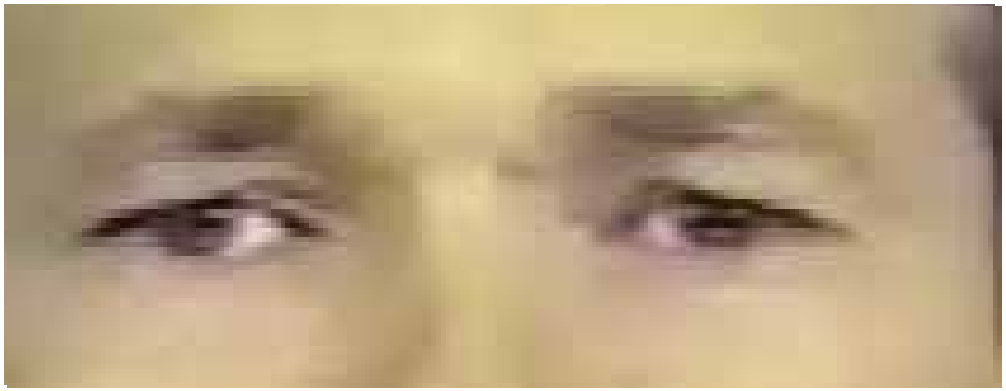
Bridges and Metzger (1996) *Deaf Tend Your.*

Three processes

2. Syntax

Three processes

3. Lexical



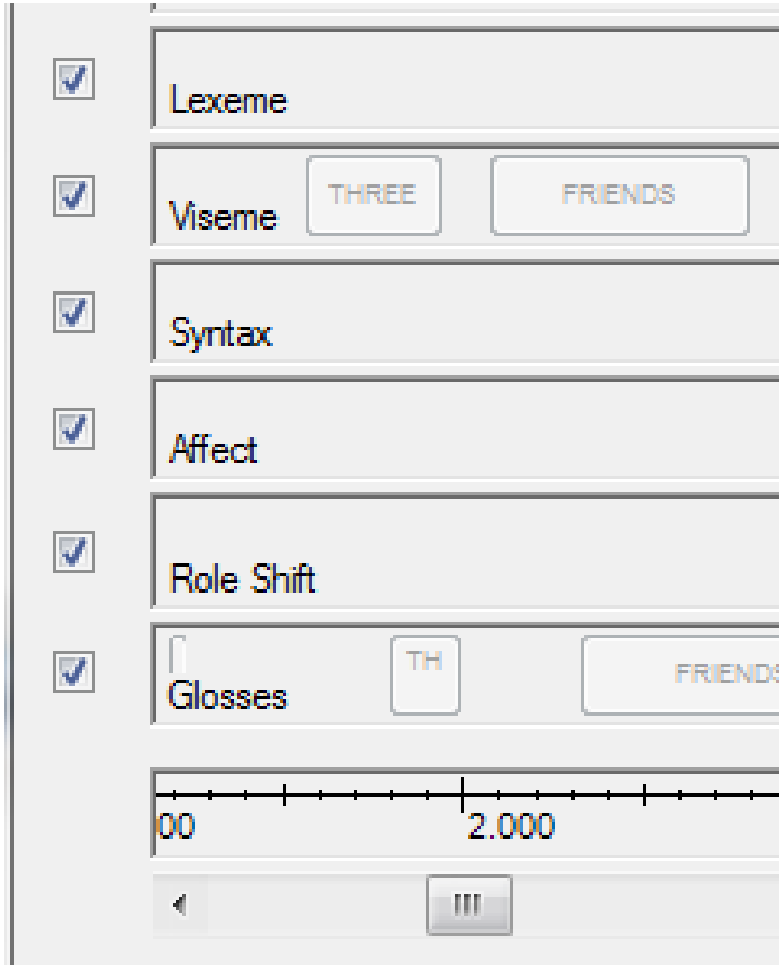
AH

Reinforcing vs. competing

- Reinforcing
  - Per example: All contributions are upward exclusively, or downward exclusively
  - Another example: A WH-question signed in an angry manner.
- Competing
  - Some contributions are upward; others downward
  - Examples
    - WH-question asked in a happy or surprised manner
    - Yes/no question asked in an angry manner

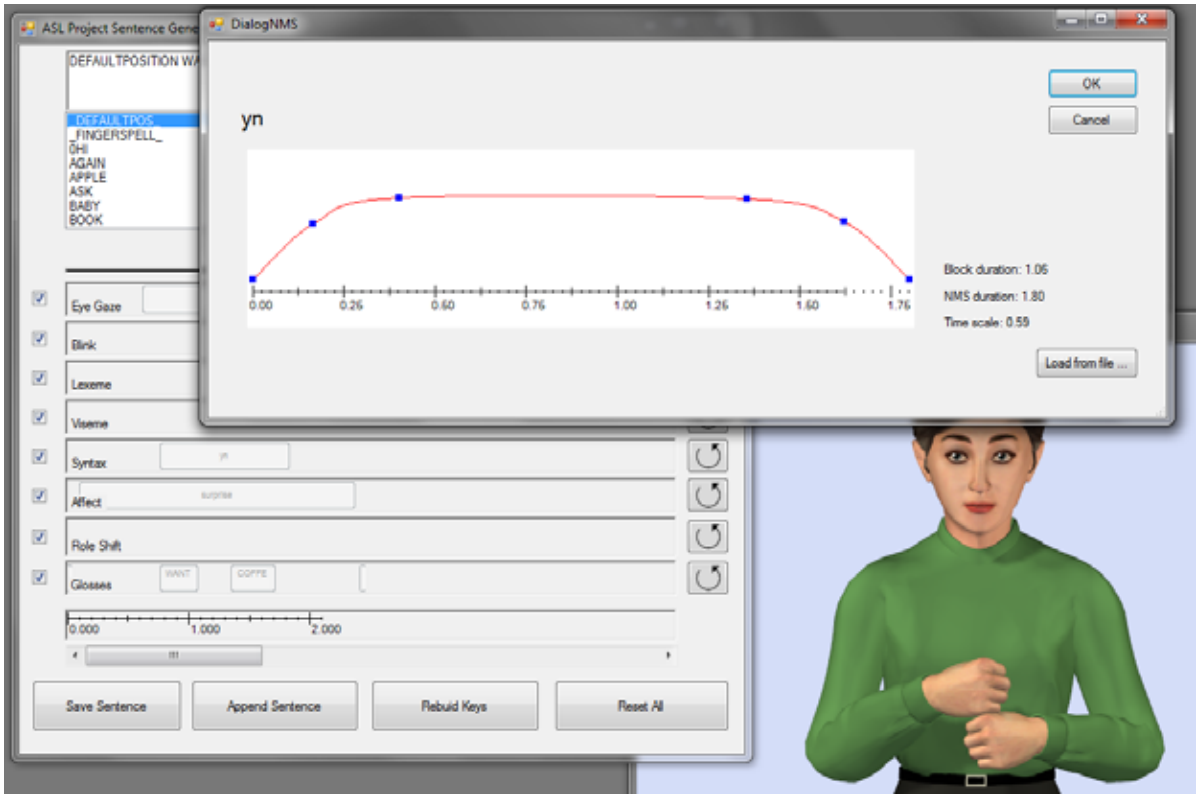
Our system

- Facial nonmanual signals organized by linguistic / extra-linguistic process.
  - Affect
  - Syntax
  - Viseme: for mouthing
  - Lexeme: for nonmanual signals that part of lexical items
- Within a tier, data is organized into blocks.



Block structure

- Pose or poses
- Intensity envelope



Coffee

Toy Car

Questions

- For generation, annotations need to be theory dependent. How practical is it to add these?
- For generation, contrastive annotation tags such as “raised”/”lowered” for brows carry incomplete data.
  - How should this data be acquired?
  - Or should it be the responsibility of the generator program to reconstruct it?

Possibilities …

- Have separate theory-dependent tracks for generation
- The generation could serve as an alternative for testing the theory.
- Possible explorations
  - What constitutes the minimal requirements for communicating a role shift?
  - Pragmatic vs. syntax vs. lexical processes – which dominate each facial feature?

Questions?

Thank you!

http://asl.cs.depaul.edu