

An Annotation Formalism for a French–LSF Bilingual Corpus Supporting Sign Language Generation

Objectives

1. to enable in-depth analysis of linguistic processes in French and LSF
2. to provide an intermediate language oriented toward **motor program** specification, thereby facilitating the **control of signing avatars**
3. to enable the **text-to-sign translation** using large language models

Pre-semantic space, pronouns

Singular pronouns (pointing gestures)

PRO1: I, PRO2: you, PRO3: he/she

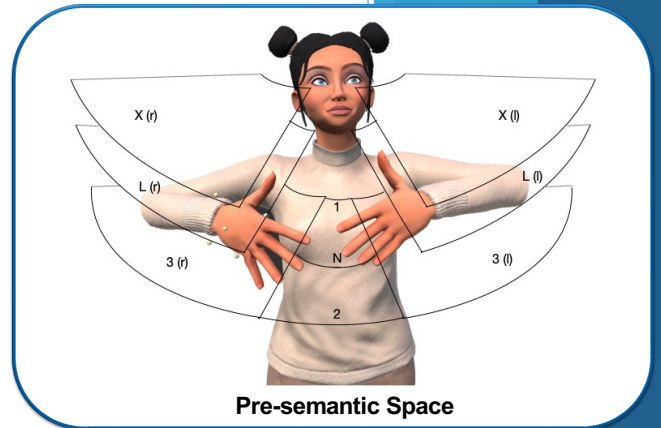
Plural pronouns: PRO4: we, PRO6: they (circular gesture)
PRO5: you (sweeping gesture)

Third-person pronouns (in trajectory verbs)

PSS1: near the signer

PSSN: neutral zone

PSSC: distant zone (objective)



Pre-semantic Space

Pointing to a location

Using the discretized space

Horizontal: L (left), N (neutral), R (right)

Amplitude: B (body), M (medium), E (extended)

Vertical: 0 (0°), 1 (30°), 2 (60°)

Paris, many persons go there [Millet, 2019]
[PARIS] [THERE_pnt:PSSL] [TOURIST] [MANY] [GO_src:N_tgt:PSSL]

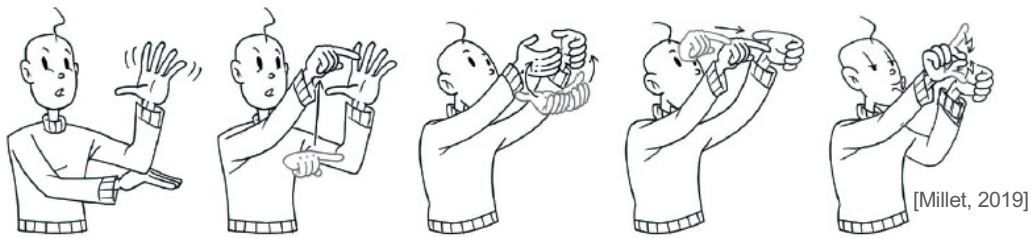
It is something too often forgotten (Le Petit Prince, Saint Exupéry)
[THAT_pnt:THING-POINTED] [THING] [FORGET] [OFTEN]

Manual proforms

1. **Verb-incorporated proforms** (trajectory verbs)
2. **Proforms representing an entity** (relative positioning)
3. **Dynamic Proforms** (persons, vehicles)

1. He gives you the book (Renart-book)
[BOOK] [mpr:prBOOK_id:RENART-BOOK]
[GIVE_src:PSS3_tgt:PSS2_mpr:RENART-BOOK]

2. The bowl (proform) is on the table (proform)
[BOWL] [TABLE] [prm:prTABLE_id:TAB1_hd:L]
[prm:prBOWL_id:BOL1_hd:R_loc:TAB1*ABOVE]



[ARBRE]

[pté-haut de l'arbre]
= loc1

[NID]

[prM-nid – pté/dedans/][prM-nid-OISEAU]
loc1-----loc-nid

There is a tree; at the top of the tree, there is a nest in which there is a bird.
[TREE_loc:L] [mpr:prTREE_id:TREE1_loc:L_hd:L]
[THERE_pnt:TREE1*TOP_hd:R_id:HIGH] [NEST_loc: HIGH]
[mpr:prNEST_id:NEST1_hd:L_loc: HIGH] [BIRD_loc:HIGH*INSIDE_hd:R]

Results

PGloss-LSF, aligned with LSF video data, is based on glosses enriched with explicit **linguistic** and **articulatory** parameters. It is applied on a **bilingual corpus** created through the SignToKids project [LREC 2024, Reverdy et al.], and evaluated quantitatively using an **automatic text-to-sign translation system** [SLTAT 2025, Reverdy et al.]