This is a description of how the program extracts data from the ELAN files.

For each ELAN file, do the following:

Find the colour tiers (named "Simulus\_1","Simulus\_2","Stimulus\_1","Stimulus\_2", "Color(Stimulus)\_1","Color(Stimulus)\_2", "Stimulus\_11", "Stimulus\_22")

This tier has annotations that mark the start and end of each trial (one picture).

For each annotation in the colour tiers.

Find all the annotations between the start and end times of the colour annotation (+ or - 100 milliseconds), for the following tiers:

SIGN TIERS:

"Formal category/Sign\_1" (participant 1)

"Formal category/Sign\_2" (participant 2)

These annotations are signs produced by the signers

For each of these signs:

Get the label of the sign annotation: \*This is the variant name\*

Get the try marking value:

Find annotations in the try marker tier (named "Try-marker\_1") that occur in the times of the sign annotation (+ or - 50 milliseconds).

If there is an annotation, and it is "Yes" then \*this sign is try-marked\*

Get the T-1 value:

Find annotations in the T-1 tier (named "T-1\_Repair") that occur in the times of the sign annotation (+ or - 50 milliseconds).

If there is an annotation, and it is "Yes" then \*this sign is a problem source\*

Get the teaching value:

Find annotations in the teaching tier ("Teach Sign\_1", "Teach Sign\_2") that occur within the times of the sign annotation (+ or - 50 milliseconds).

Get the first of these annotations.

If there is an annotation, and the annotation label is "Yes", then \*this is a teaching context\*

All of these signs are \*not T0Check signs\*

Take all of this information and make one line in the results file spreadsheet for it.

For each annotation in the check tier ("T0/Check\_1", "repair initiation\_1","Repair initiation\_1","T0/check\_1","T0/Check\_2", "repair initiation\_2","Repair initiation\_2","T0/check\_2" )(within this trial):

All of these signs \*are T0Check signs\*

Get candidate understanding value:

Find annotations in the candidate understanding tier ("T0\_1\_candidate(understand)”, “T0\_2\_candidate(understand)") that occur within the times of the sign annotation (+ or - 50 milliseconds).

If there is an annotation, and the annotation label is "Yes", then \*this is a candidate understanding form\*

Take all of this information and make one line in the results file spreadsheet for it.

The output of the program is a spreadsheet called "variants.csv" where each line is a particular occurrence of a sign, and each column contains information on that sign (source file, trial number, start time, end time, iconicity, indexicality, try-marking, teaching, candidate understanding etc.