

Segmentation

There are two basic approaches how to segment continuous signing into individual signs:

- A sign starts where the preceding one ends – fluent signing means there are no gaps between signs
- Transitional movements between signs do not count as part of either sign. Therefore, there are gaps between two signs during which the articulators move from the end of one sign to the beginning of the next. (In some cases where there is no transition, there might not be a gap either.)

Both approaches have their pros and cons. However, in the context of Dicta-Sign the second approach offers advantages for the subsequent processing:

- A token tag then represents that part of the signal that is described by HamNoSys.
- Variation between tokens is much lower than if the transition would be part of the sign.

Most researchers following this approach have some rules on paper, but mostly rely on their intuition where to cut. To minimise differences between the different groups, we try to be as explicit as possible:

One-handed signs

The easy case here are signs with a HMH structure in the sense of L&J: The sign starts at the beginning of the initial hold and ends at the end of the final hold. (Please note that this means that segmentation is not strictly bottom-up i.e. data-driven, since we use our knowledge about the type to cut/tag the token.)

Sign Starts:

- *The sign starts as soon as its handshape has been formed and is placed in the right orientation at the starting location of the sign.*
- *In cases where two signs share a hold (i.e. one sign ends in a hold, and by chance the next sign is beginning with a hold at exactly the same location with the same handshape and orientation), cut the hold in the middle. (Here it is obvious that there cannot be a gap between the two tags.)*
- *In case of signs without a specific starting location, look for a discontinuity in the movement (e.g. sudden change in direction) from the end of the previous sign and the end of the target sign. That point is then the starting point.*
- *In case of a continuous movement from the beginning of a sign to the end of the next sign (e.g. THINK YOU in lax signing), cut in the middle/at the peak of that movement. (This is then also the end of the previous sign, i.e. there is no gap in-between the two signs.)*

Sign Ends:

- *If the sign finishes with a hold it ends at the end of the hold (just before the first change of one of the parameters).*
- *If the sign finishes with a movement, then cut just before a change of movement direction.*
- *If there is no change of movement, a change of handshape or orientation marks the end of the sign.*
- *In case there's no change of handshape or orientation but a continuous movement from the previous to the following sign, the sign ends in the middle /at the peak of that movement (see above).*

Two-handed signs

In principle, you can apply the criteria for one-handed signs to both hands individually. The result is that in some cases you have different timings for the two hands. If you use two tiers for the transcription, you could easily accommodate that – at the expense of more time needed to cut things and with unclear implications wrt a cognitive point of view what a sign is. For one tier, and that also holds for cutting the video itself, which is what counts for image processing, you have to come up with a combined criterion. Interestingly, the easiest and most consistent definition to cut both hands in parallel is to just concentrate on the dominant/active hand and ignore the other:

For two-handed signs, follow the above definition applied to the dominant/active hand.