

Segmentation

Thomas Hanke, Sung-Eun Hong, Susanne König, Reiner Konrad, Gabriele Langer, Silke Matthes, Rie Nishio, Anja Regen

1. Version: July 2011

*Revision 1: November 2012: Supplemented by rules for doubtful cases (changes of handshape and orientation; see text in **bold face**); modification in “Prolonged Signs”*

It is not a simple matter to determine where a sign begins and ends. It can be presumed that competent signers of DGS are able to intuitively identify signs within the natural flow of the language. Nevertheless, it is reasonable to define guidelines for segmentation to ensure the highest consistency possible for the treatment of tokens in the annotation. With these guidelines we specify the segmentation rules that have been used at the IDGS for other projects more precisely and make them explicit. In contrast to what has been done in other projects¹, we have chosen the approach of a narrow segmentation in which transitional movements between signs do not count as part of either the preceding or the following sign. Following this approach the number of boundaries between segments that need to be annotated is almost doubled, as the end of one sign usually does not accord with the beginning of the following sign. However, the advantage of narrow segmentation is that the forms of the tokens matched to one type look more similar to each other since differences resulting from transitions particularly at the end of a sign are omitted. This makes the task of lemma revision much easier. Lemma revision is a quality control measure during which all tokens of one type are inspected one after the other to check whether token-type-matching is correct. Also future automated comparisons of tokens forms will profit from a narrow segmentation, because it reduces differences of form between token segments (i.e. visual noise) that would require manual postprocessing.

In the following, first we provide the guidelines for segmentation that are used in the context of the DGS-Corpus and that have also been used in the Dicta-Sign project. Following this, additional conventions and rules for specific types of signs or particular phenomena of articulation are presented.

¹ Among others about Australian Sign Language (Auslan): *Endangered Languages Documentation Project* (URL: <http://www.auslan.org.au/about/corpus/>; Annotation guidelines: Johnston, Trevor. 2011: Guidelines for annotation of the video data in the Auslan Corpus. [manuscript constantly revised, as of 06/19/2011; current version online; URL: <http://www.auslan.org.au/about/annotations/>], about Sign Language of the Netherlands (NGT): *NGT Corpus Project* (URL: <http://www.ru.nl/corpusngt/>; Annotation guidelines: Crasborn, Onno / Zwitserlood, Inge. 2008: Annotation of the video data in the Corpus NGT. Nijmegen: Radboud University. [current version online; URL: http://www.ru.nl/aspx/download.aspx?File=/contents/pages/515436/corpusngt_annotationconventions.pdf]), about British Sign Language (BSL): *BSL Corpus Project* (URL: <http://www.bsllcorpusproject.org/>; Guidelines Annotation guidelines: s. Johnston 2011).

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 Liddell/Johnson (1989)²: 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.*
 - ***If the hand is placed at the starting location of the sign without the handshape and/or orientation being fully established (yet), cut as soon as the movement starts.***
- *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*

² Liddell, Scott K. / Johnson, Robert E. 1989. American Sign Language. The phonological base. In: *Sign Language Studies* 18 (64), pp. 195-277.

Application of the model to Auslan in: Johnston, Trevor / Schembri, Adam. 2006: Australian sign language (Auslan): An introduction to sign language linguistics. Cambridge: Cambridge University Press, pp. 111-116.

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.*
 - ***This also applies if, throughout the movement, there's a change of handshape which is not a proper part of the sign; e.g. if the handshape of the subsequent sign is pre-emptively formed.***
- *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).*

Generally, all parameters are treated equally, as it is assumed that at a certain point (at the beginning as well as the end of a sign) all parameters (except movement) with relation to the sign are established. If this is not the case, meaning if e.g. the corresponding handshape is only established after the movement has already begun, or if during the movement the handshape of the subsequent sign is pre-emptively formed (assimilation), the movement is considered as the main indicator. The cut will then occur at the beginning or the end of the movement.

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.

Additional Conventions

- *Indexical Signs*

Indexical Signs as well as occurrences of conventional signs that are based on pointing (indexing image producing technique) are always considered movement-hold-units, which entails that the movement—ending with or without contact—is part of the sign proper.

- *Fingerspelling*

Letters that are signed in one fluent succession most likely refer to one word or to one proper noun and are segmented as one token. Discontinuations and pauses that coincide with the end of a unit of meaning lead to annotation with more than one token-tag. The same applies whenever the person signing hesitates to think or considerably changes the location of the hand(s). In case of doubt, fewer rather than more token-tags should be created.

- *Numbers*

In contrast to previous projects, the DGS-Corpus considers numbers no longer as whole units. Instead, following formational aspects numbers are divided into different groups of which the elements can be combined systematically. Through this particular approach, the various realisations of the group's elements can easily be accessed. The groups are:

- *Numbers from 0-12*
- *Numbers from 13-19*
- *Tens from 20 onward (20, 30, ... 90)*
- *Hundreds (100, 200, ... 900)*
- *Thousands (1000, 2000, ... 9000)*

The signed number 1925 in DGS is split into four tokens; it either is segmented as 19+100+5+20 (when used for designating a year) or as 1000+900+5+20.

- *Prolonged Signs (disruptions of the flow of signs)*

Signs that are held significantly longer in their end position than normal are also tagged as one token. Disruptions occur, for example, because the person signing pauses to reflect—often indicated by a change of direction of sight towards the monitor, the interlocutor or any other direction,—or to indicate that he/she wants to continue with

his/her turn or wants to add something he/she cannot remember at that moment. The token tag of a prolonged sign ends when the sign form dissolves (conf. rules of segmentation).

- *Repeated, offset movements*

Especially when segmenting so-called directional verbs difficulties might occur, for example, if a sign such as GIVE is executed repeatedly while being moved laterally. Here, the movement is the determining factor for meaning as well as segmentation: If the sign is executed in its entirety each time, meaning that the hand returns to its original starting position over and over again, it is cut as several tokens. Each token bears the meaning of “giving something to one (particular) person”. In the case of one slurred movement, in which the hand does not return to its starting point, the occurrence should be cut as only one token, as it carries the meaning of “giving something successively to several (different) people” (distributive).

- *Back-channel feedback*

The term back-channel behaviour or back-channel signals involves manual signs and gestures, facial expressions, mouthing, mouth gestures and gestures through which the interlocutor lets the speaker know that he/she is following what is being signed. Typical back-channel signs are, for instance, YES or RIGHT. Expressions that can be assigned to back-channel behaviour are not covered by the basic annotation.

However, signs and mimic-gestural behaviour should be segmented if they

- *result in a change of speaker (turn-taking),*
- *cause the speaker’s reaction or answer (with regards to content),*
- *induce a dialogue that gives rise to a new subject.*

- *Mouthing*

The criteria of segmentation suggested above include signs only, that is manual signs. A simultaneously articulated mouthing does not affect the segmentation. When the articulation of a mouthing takes longer than the execution of a sign or when it starts before the beginning of a sign, the segment’s boundaries are not changed but follow the manual sign’s properties only.

In the context of the basic annotation, mouthing is merely included due to its relevance concerning the meaning of signs and of sign sequences. We assume that mouthing is mostly simultaneously articulated to the respective sign, knowing full well that it is nothing unusual that beginning and end of sign and mouthing do not align perfectly.