# Eye Gaze Annotation Practices: Description *vs.* Interpretation

**Annelies Braffort**
LIMSI-CNRS
Campus d'Orsay, bat 508, BP 133, F-91403 Orsay cx, France
E-mail: annelies.braffort@limsi.fr

**Abstract**

If sharing best practices and conventions for annotation of Sign Language corpora is a growing activity, less attention has been given to the annotation of non-manual activity. This paper focuses on annotation of eye gaze. The aim is to report some of the practices, and begin a discussion on this topic, to be continued during the workshop. After having presented and discussed the nature of the annotation values in several projects, and explain our own practices, we examine the level of interpretation in the annotation process, and how the design of annotation conventions can be motivated by limitations in the available annotation tools.

**Keywords:** eye gaze annotation, sign language corpora

## 1. Introduction

For Sign Languages (SL), reception of linguistic information is primarily conducted through the eyes. The addressee usually fixates his gaze at the signer's face, particularly the area around the signer's eyes. Eyes are also one of the body components that convey linguistic information, together with other non-manual and manual ones. Several eye aspects can be considered: blinking, eye aperture, and eye gaze. In this paper, we focus on eye gaze.

Eye gaze has a number of different linguistic functions. Some of these functions have been pointed out in the literature (Engberg-Pedersen, 1999): Certain lexical signs may require a specific gaze direction, some iconic constructions require a gaze directed at the hands or at signing space, and gaze is also involved in role shifts. In some theoretical models (Cuxac, 2000), gaze have a semiotic function, allowing distinguishing between two modes of expression: without or with an illustrative process (gaze toward addressee or not).

Analysing SL corpora, by looking at the eye gaze values, their durations, and the co-occurring or surrounding events conveyed by other manual and/or non-manual components, can provide evidences for the definition of formal descriptions linked to functional categories.

SL corpus linguistics is a recent field, and if some practices begins to be promoted and shared, e.g. the use of a database for the lexical signs for annotation consistence and reliability (Hanke, 2008; Johnston, 2008), or even some kind of standardisation (Shembri, 2010; Crasborn, 2012), less attention has been given to the annotation of non-manual activity.

The aim of this paper is to report some of the practices related to eye gaze annotation (section 2), including ours (section 3), and begin a discussion on this topic, to be continued during the workshop (section 4).

## 2. Eye gaze annotation practices

This section reports the practices used to annotate eye gaze in five projects, for Auslan, ASL and various European SLs. They have been selected to illustrate the various practices.

### 2.1. Annotation conventions for the Auslan corpus

The Auslan corpus annotation guidelines (Johnston, 2013), designed using the Elan annotation software, is regularly updated as the annotations progress.

The current version of the annotation scheme includes a tier to code eye gaze with four possible values: **a** for "addressee", **t** for "target", **o** for "other", and **z** for "cannot be coded".

These four values code the target of eye gaze.

### 2.2. Annotation conventions for the ASL Linguistic Research Project in Boston

The ASLLRP project includes the development of annotation software (SignStream) and documentations on the conventions used for the annotation.

The annotation scheme includes an eye gaze tier, with the following values:

- Direction of eye gaze: **left**, **right**, **up**, **down**.
- These values can be combined: **up/lf**, **up/rt**, **dn/lf**, **dn/rt**.
- **addressee** is used to code when eye gaze is directed toward the addressee.
- **track-hand** is used to code when eye gaze follows the hand.
- It is also possible to code eye gaze directed at a specific location, such as **i** (**i** is an index for a given location), or **"under table"**, or even **indef** in the case of an indefinite reference.

In this scheme, the eight first values give the direction of eye gaze from the perspective of the signer. The other values give the target of eye gaze.

### 2.3. Annotation conventions for the ECHO project

The ECHO European project included a case study devoted to SL. A comparable corpus of three European SL was constituted (the SL of Sweden, United Kingdom, and the Netherlands), together with a common annotation scheme. This annotation scheme includes a gaze tier, with the following values:

- **l-90**: left close to 90 d° (of midsaggital plane)
- **l**: left, close to 45 d° (of midsaggital plane)
- **r-90**: right, close to 90 d° (of midsaggital plane)
- **r**: right, close to 45 d° (of midsaggital plane)
- **u**: upward
- **d**: downward
- Combinations are possible, e.g. **ru** (right and upward)
- **lh**: to the left hand
- **rh**: to the right hand
- **bh**: to both hands
- **p**: toward a person present
- **c**: toward the camera

In this system, the six first values and the associated combinations code the eye gaze direction, while the others code the target of eye gaze.
For the direction values, a different granularity is used depending on the plane: The horizontal plane is segmented into four values, the vertical one into two values.

## 2.4. ViSiCAST European project

The ViSiCAST European project didn't include a task on corpus annotation and the design of an associated annotation scheme, but some work has been done on a computing representation of signed utterances. An XML system called SiGML, based on HamNoSys, has been designed. This is a timed multi-tier representation where each tier encodes one of the parallel information channels. One of the tiers is used to represent eye gaze, with the following values:

- **AD**: toward addressee
- **FR**: far
- **HD**: towards the signer's own hands
- **HI**: Towards the signer's own dominant hand
- **HC**: Towards the signer's own non-dominant hand
- **UP, DN, LE or RI**: up, down, left or right
- **NO**: no target, unfocussed
- **RO**: rolling eyes

Here also, there is a mix between directional type and target type values. Moreover, a new kind of value is used, which is dynamic: "rolling eyes".
Another particularity of this system is that it is considered that head movement and eye gaze can be linked. This is represented in the head tier, not in the eye gaze one, and here also, this is a dynamic value.

## 2.5. Intersign network

The Intersign European network[1] aimed at developing standards and guidelines for the study of European SL. Six SLs were represented.
One of the contributions was related to eye gaze in Danish SL with considerations about notations issues for forms and functions (Engberg-Pedersen, 1999). In this contribution, three levels of interpretation in the notations are proposed, from pure formal to pure functional:

1. At the formal level:
   - eye contact with the receiver;
   - some other direction than the receiver;
   - eye blink.
2. At an intermediate level, when eye gaze is directed at the signing space:
   - Are instances of eye gaze in some other direction than the receiver in a meaningful direction or not?
   - If the direction is meaningful, is it in the direction of a locus or in the direction of a configuration?
3. At the functional level, where there are five categories, based on a distinction between two types of signing depending on who the signer's locus represents: the signer as sender of the current utterance (i.e. the sender level) or one of the individuals talked about (i.e. the character level). The following category definitions are extracted from (Engberg-Pedersen, 1999):
   - the narrator's eye contact with the receiver (sender level),
   - avoidance of eye contact at major boundaries by blinking or by looking away in no particular direction (sender level),
   - reference-tracking eye gaze in the direction of a locus just before a predicate or with a topical nominal or a resumptive pronoun (sender level),
   - imitative eye gaze with constructed action, thoughts or dialogue, imitates the holder of the point of view or the quoted person (character level),
   - configurational eye gaze with polymorphemic predicates (it can be the sender or the character level).

In all of these levels, the values code the eye gaze target.
Something particular in this system is the presence of eye blink, which is not a target value. In other annotation schemes, eye blink is considered as one of the possible values of eye lid or eye aperture tiers, or even as a specific tier (Braffort & Chételat, 2011).

## 2.6. Main trends and particularities

From this report, we can notice that:

- three of these five projects propose as annotation values a combination of directional and target values, and two of them only target values;
- the directional values, based on a segmentation of the signing space from the perspective of the signer, are more or less the same, with in one case a different segmentation of the signing space (more than two values in one plane);
- the target values are quite different; only the "addressee" value is common to all the annotation scheme; some values are more or less detailed, some are present only in one scheme;
- four schemes includes additional values with no equivalent in the other studies: a parameterised value when the gaze is directed toward the signing space, and two dynamic values that doesn't code a direction (blink and rolling eyes).

[1] http://www.sign-lang.uni-hamburg.de/intersign/intersign.html

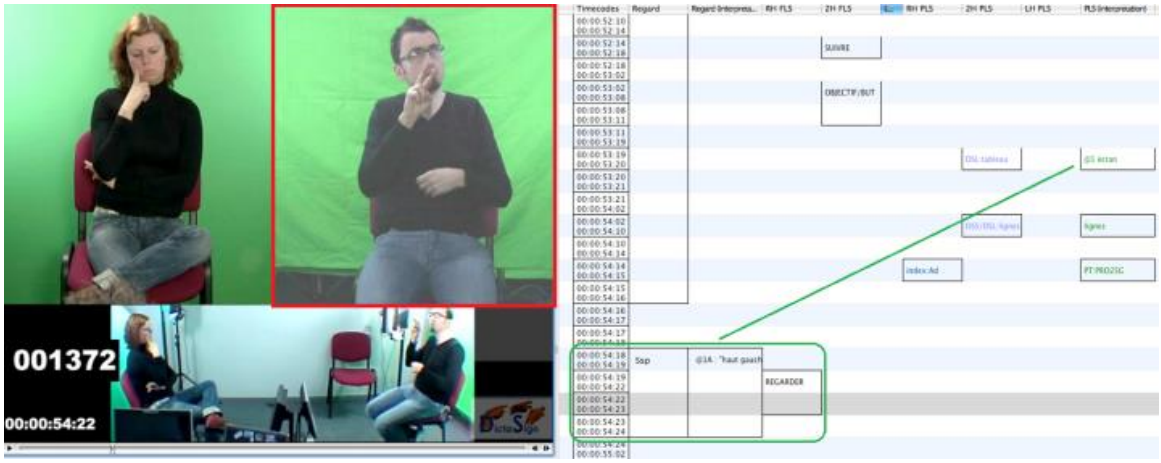


Figure 1: Extract from the annotation of the LSF part of the Dicta-Sign corpus

## 3. Gaze annotation in the French Sign Language part of the Dicta-Sign corpus

This section reports the practices used to annotate eye gaze in the French Sign Language (LSF) part of the Dicta-Sign corpus (Matthes et al, 2010), which was a comparable corpus created during a European project including studies on four SLs (German, Greek, English, and French).

### 3.1. Annotation scheme

First, we used only one type of values, in order to facilitate the design of analysis requests that could be more complex in case of mixed values.

Then we used target values, because this allows saving time for analysis. Moreover, this avoids using an arbitrary segmentation of the signing space. We based our controlled vocabulary on the Auslan annotation guideline, with additional details for the cases where eye gaze is directed toward a target, being virtual or real.

Finally, we distinguished two levels of annotation, a more formal one, to code the target kind, and a more interpretative one, to code the supposed target itself in case of target in the signing space. For that, we used two tiers, called Gaze and Gaze interpretation.

The tier "Gaze" allows identifying the target, with the following values:

- **ad**: addressee
- **ssp**: signing space
- **hd**: hand or part of hand
- **real**: real object (e.g. elicitation material, such as paperboard and computer screen) or other person than the addressee
- **x**: far (e.g. the signer is thinking or is looking away without a given target)
- **?**: cannot be coded

The tier "Gaze interpretation" is used to code more information in the case of a *ssp*, *hd* or *real* value in the Gaze tier, with the following values:

- ***@code***: associated with a *hd* value; *code* refers to hands or fingers, identified more or less precisely using a code (e.g. @I_PAD(r) means the index pad of the right hand)
- ***@id:txt***: associated with a *ssp* value; *id* refers to a previously annotated entity located in the signing space, and *txt* id a textual description of the referred entity (e.g. "@2" refers to the localised entity number 2)
- ***code:txt***: associated with a *hd* value; *code* can take one of the three values hands, hand(r) or hand(l), and *txt* described the referred entity hold by the hand(s) (e.g. "hand(r):billet" refers to the right hand holding a ticket (*billet*))
- ***txt***: associated with a *hd* or *real* value; *txt* is a textual description of the referred entity (e.g. "Bottle", or "top right corner of the screen"), or the real object (e.g. "screen") or person (e.g. "moderator").

With this organisation, we can have detailed information on the way eye gaze is used in constructions requiring a gaze directed at the hands or at the signing space.

### 3.2. Detailed example

Figure 1 illustrates an example of eye gaze annotation with our annotation scheme. The annotation software used is iLex (Hanke, 2008). In this view, time flows from top to bottom, and tiers are vertical. In this example, we have annotated the eye gaze that is associated with the lexical sign REGARDER that means "to look at":

- The first two tiers in the figure are used for eye gaze.
- The third tier is used for the lexical signs performed by the right hand, here **REGARDER**.
- Notice also the eighth tier, which is used to add interpreted information in case of depicting signs. In our example, it has been used to attribute an index to an entity that has been located in the signing space: **@1 écran** means entity number 1, which is a screen (*écran*).
- The value for the Gaze tier (Regard) is **ssp** for signing space.
- The value for the gaze interpretation tier is **@1A: "haut gauche"**. This means that the target is a sub-part of the entity number 1, this sub-part being interpreted as the top left corner of the screen, from the perspective of the signer.

By using this method, we can design requests that allow us to automatically link values related to spatial annotation in different tiers.

## 4. Discussion

This section proposes thinking about the various practices, their pros and cons, as a start for more interactive discussion during the workshop.

### 4.1. Description vs. interpretation

A first point is the identification of the level of interpretation in the annotation process, and all the possible biases that annotators do not realize that they have, as they will have common knowledge on the grammar of written language.

As much as possible, the annotations should intend to be descriptive, rather than to express particular theoretical beliefs. But coding of pure descriptive information is sometimes impossible, or even useless.

This is the case for eye gaze, where a "pure description" would be anatomical (e.g. the relative position of the iris regarding a given landmark), or, less directly, mathematical (e.g. a 3d vector). We can imagine that these data could be computed automatically, using image processing tools, providing by this way purely objective annotations. But anyway, segmentation would remain to be done, and more computation would be needed to help interpretation and analysis of the data.

Of course, we can attribute a direction value to eye gaze directly, as this has been done in some of the reported studies here. But this necessitates segmenting the signing space into arbitrary zones, because direct 3d annotation is not possible in the current annotation tools. And also here, interpretation of the target remains to be done.

Then, a more "interpreted description" for eye gaze is to code the target kinds, as we have done in our project. In this case, it is not easy to define objective criteria, and the choice relies on the subjectivity of the annotator. This saves time for the next step of annotation and analysis, at the price of the risk of more errors and less annotator agreement.

### 4.2. Dependence on the available tools

Another point to consider is that it is very difficult to design conventions that are completely independent of the limitations in the annotation tools. For example, the use of index to allow links to be established automatically during analysis between eye gaze and discourse entity that are located in the signing space is due to the fact that the used tool doesn't allow to create a list of no temporal entities with associated identifiers. This kind of process is possible with the Anvil tool, but on the other hand, Anvil doesn't allow using a lexical database such as in iLex, which is an essential part of annotation tools for SL.

It is likely that our conventions, guidelines and methods will continue to evolve in the following years, as the tools available for annotation become more sophisticated.

Ideally, and this is a call toward the image processing community, the field would greatly benefit of computed descriptions and representations associated with segmentation capabilities. Conversely, progress in the linguistic field would help automatic processing by providing more knowledge on the phenomena to be processed (Gonzalez et al, 2012).

## 5. References


Braffort, A. & E. Chételat-Pelé (2011). Analysis and description of blinking in French Sign Language for automatic generation, in Gesture and Sign Language in human-computer interaction and embodied communication, LNCS/LNAI 7206, Springer, p.173-182.

Crasborn, O. & Windhouwer, M. (2012) ISOcat data categories for signed language resources. In Gesture and Sign Language in Embodied Communication and Human-Computer Interaction, LNCS/LNAI, Springer, p.118-128.

Cuxac, C. (2000). La langue des signes française (LSF) : les voies de l'iconicité, Ophrys, Paris.

Engberg-Pedersen, E. (1999). Eye gaze in Danish Sign Language monologues: Forms, functions, notation issues. Intersign project papers, v. 3, p.33 -37. URL: http://.../engbergpedersen.html

Gonzalez, M., Filhol, M. & Collet, C. (2012). Semi-automatic Sign Language corpora annotation using lexical representations of signs, in proc. of the 8th int. conf. on Language Resources and Evaluation LREC, ELDA, p.741-745.

Hanke, T. (2002). ViSiCAST Deliverable D-: Interface Definitions. URL: http://.../ViSiCAST_D5-1v017rev2.pdf

Hanke, T. & Storz, J. (2008). iLex - A database tool integrating sign language, in proc. of the 3rd workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora, LREC conference, ELDA, p.64-67.

Johnston, T. (2008). Corpus linguistics and signed languages: no lemmata, no corpus, in proc. of the 3rd workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora, LREC conference, ELDA, p.82-87.

Johnston, T. (2013). Guidelines for the annotation of the video data in the Auslan corpus. URL: http://.../AuslanCorpusAnnotationGuidelines_Johnston.pdf

Matthes, S., T. Hanke, A. Regen, J. Storz, S. Worseck, E. Efthimiou, A.-L. Dimou, A. Braffort, J. Glauert and E. Safar (2012). Dicta-Sign - Building a Multilingual Sign Language Corpus, in proc. of the 5th workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon Proceedings LREC-2012, ELDA p.23-27.

Neidle, C. (2002). SignStream Annotation: Conventions used for the American Sign Language Linguistic Research Project. URL: http://.../asllrpr11.pdf

Nonhebel, A., Crasborn, O. & van der Kooij, E. (2004). Sign language transcription conventions for the ECHO Project, Version 9. URL: http://.../ECHO_transcr_conv.pdf

Schembri, A. & Crasborn, O. (2010) Issues in creating annotation standards for sign language description, in proc. of the 4th workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies, LREC 2010, ELDA., p.212-216.