Conventions for multimodal transcription
(initial version: 2001; current version: 5.0.1, Nov. 2019)

Please, when using these conventions, refer to the following web site:
https://www.lorenzamondada.net/multimodal-transcription

This file contains two versions of the conventions:
• Short version (usable in the appendix of articles and chapters)
• Long version (usable as a tutorial)

See also the following web sites:
The English version is available under two addresses:
https://www.lorenzamondada.net/multimodal-transcription
and
For the French version (not updated):

See also for a conceptual discussion of these conventions:
Multimodal transcript conventions
(short version)

Embodied actions are transcribed according to the following conventions developed by Lorenza Mondada (see Mondada 2018 for a conceptual discussion).
https://www.lorenzamondada.net/multimodal-transcription

*   *   Descriptions of embodied actions are delimited between
+   +   two identical symbols (one symbol per participant and per type of action)
∆   ∆   that are synchronized with correspondent stretches of talk or time indications.
*--->  The action described continues across subsequent lines
---->*  until the same symbol is reached.
>>    The action described begins before the excerpt’s beginning.
---->> The action described continues after the excerpt’s end.
.....  Action’s preparation.
----   Action’s apex is reached and maintained.
, , ,  Action’s retraction.
ric   Participant doing the embodied action is identified in small caps in the margin.
fig   The exact moment at which a screenshot has been taken
#    is indicated with a sign (#) showing its position within the turn/a time measure.
Multimodal transcript conventions
(long version)

https://www.lorenzamondada.net/multimodal-transcription

1. Principles

These conventions are conceived to annotate all possibly relevant embodied actions, such as gesture, gaze, body posture, movements, object manipulations, etc. that happen simultaneously to talk or during moments of absence of talk.

The convention is based on two principles:

a) *Timing*: each embodied action is precisely temporally located within the course of the multimodal activity and it is delimited by two brackets, which indicate on the left its emergence and on the right its completion.

b) *Characterization of the embodied action*: each embodied action is shortly described.

The conventions are based on economy and consistency: by putting no more annotations than necessary, and by choosing annotations that are interpretable univoquely.

2. Identification of the participants doing the embodied action

Every embodied movement is attributed to a participant, identified by his/her pseudonym and by a symbol consistently used for the same participant through the transcription.

(1) Example:

* delimits gestures done by LAU
↓ delimits gestures done by PAL
∆ delimits gestures done by BRU
+ delimits gestures done by VIV

Sometimes, it might be relevant to create several lines for different embodied actions done by one participant at the same time. In this case, different symbols will be used – if possible by choosing similar symbols in order to have a visual link between different actions of the same person.

(2) Example:

* for gestures done by LAU
* for gaze by LAU
+ for gestures done by VIV
† for gaze by VIV
etc.

If the embodied action is done by the current speaker, then its description is *not* preceded by her/his identification in the margins;

(3a) Example:

1 CUS *EH s’il vous plaît* un bout d’gruyère
*points at gruyère*

If it is done by another participant, (s)he will be identified in the margins. It might be useful to use capitals for the identification of the speaker and the normal style for the identification of the participant doing the embodied action.
Conventions for multimodal transcription

Lorenza Mondada — 4

(3b) Example:

1  LAU  +"ah oui"  je+ sais
   vev  +grasps pen+

Note that in some cases, when there are numerous lines referring to several participants, it can be wise for the sake of clarity to identify all of the participants doing embodied actions, including the speaker:

(3c) Example:

1  LUC  mais #ceux-là#  là φ çaζ seraitξ  dommage≠  φhein,#  de:
     but these  there it would be too bad  right, to:
  luc  ζpoints w open palm------------------ζ
  luc  looks at JEA-->
  jea  -->φlowers hand-----------------φraises H-->
  yan  ↓...turns to R-->
  eli  $turns to R-->
  fig  #fig1  #fig.2  #fig.3#

3. Timing: Delimitation of the beginning and the end of an embodied action

Every embodied action has a temporal trajectory that is delimited by two identical symbols, one indicating when the action begins and the other one when it ends. These two symbols are inserted either in the line of talk or in a measured indication of time, in order to allow a synchronization of the verbal/silent conduct and the embodied conduct. These two symbols are spatially aligned, one above the other, in order to represent their simultaneous unfolding.

The description of the action is inserted between these two symbols.

(4) Example:

1  BRU  la maison qui A se retrouve ici; A
       Αpoints to doc-----A

If an embodied action begins on a line and continues either the next line or some lines later, its description is followed by an arrow pointing to the direction of the next symbol/landmark that indicates its end. In this way, the arrow works as an instruction for the reader to search, in the following lines, for the next arrow pointing at the same symbol, closing that annotation.

(5) Example:

1  LAU  +"ah oui"+
   vev  +points-->
2   (0.5)
3  VIV  mais alors+ i- il y va pour les dé- pour les sortir
       --2+

If an embodied action begins in the middle of a pause, then the pause has to be segmented into smaller temporal fragments in order to insert the landmark.

(6) Example:

1  LAU  "ah oui"
2   (0.3) + (0.2)
   vev  #points-->
3  VIV  mais alors+ i- il y va pour les dé- pour les sortir
       --2+

Note that if an embodied action is synchronized with(in) a pause, there is always an identification of the participant doing it in the margin.

In some cases, when the end of the embodied action is located several lines after its beginning, it is possible to help the reader to find when it ends by indicating a line number.
after the first arrow. Nonetheless, be careful in not over-using this type of indication: if it is used too much, it can blur the readability of the transcript by adding too many redundant indications.

(7) Example:
1  LAU  "ah oui"
    viv  +points--->1.9
  ((transcription continues))
9  VIV  on garde une agnelle+ pour: cinq brebis euh
        --->

If an embodied action continues until the end of the excerpt and afterwards, its description is followed by a double arrow. In this case, there will be no second landmark closing the temporal span of the action (and this is indicated by the double arrow):

(8) Example:
4  PAL  par rapport au cas de figure de:: (.) des causses, en général,
5  bru  euh cette attribution de: de terrain est bien moins claire.

Bruno looks at Pal from line 5 on, after the end of the excerpt.

If an embodied action begins before the beginning of the excerpt, this is indicated by an initial double arrow, like in the example below:

(9a) Example:
1  PAL  par rapport au cas@ de figure de:: (.) des causses, en général,
2  euh cette@ attribution de: de terrain est bien moins claire

(9b) Example:
1  PAL  par rapport au cas@ de figure de:: (.) des causses, en général,
2  euh cette attribuiton de: de terrain est bien moins Claire

If embodied actions of more than one participant are described, the first line is generally devoted to the speaker, and the following ones to the other co-participants. In this case, the identifier of the doer of the embodied action is not repeated in the margin (the absence of identifier makes it clear that the doer is the speaker):

(10) Example:
1  PAL  par rapport +au @cas@ de figure de:: (.) des +causses, @en général,
2  euh cette@ attribution de: de terrain@ est bien moins Claire

In case of overlapping turns, the annotations of embodied conducts are placed after the overlapped talk:

(11) Example:
1  LAU  "ah [oui,] ici c’est vrai."
2  VIV  [mais] al’ors i- il y va p’our les dé- pour les sortir
      lau  +points--->
      viv  +nods

In this case it is good to indicate all identifiers of the speakers/doers of embodied action.
4. Trajectories of embodied actions

Embodied actions have a temporal trajectory, which can be roughly described by
distinguishing a) a preparatory phase, b) a recognizable shape of the action, c) a retraction or
withdrawal phase. Their annotation is inspired by conventions used by Kendon (1990) for
gestures and Goodwin (1981) for gaze. For all embodied movements, the trajectory is
indicated in the following way:
..... small dots indicate that the embodied action is emerging,
..... commas indicate that the embodied action is withdrawing, retracting.
The embodied action itself is described when it has reached its recognizable shape, which can
also be maintained for some time.
(12) Example:
1 VIV sur des terres assez *euh ass- a:*- assez *bonnes*
 ....... *points---*\

The description of action and its segmentation is an analytical decision, depending on the
precision and the granularity of the transcription (its detail).
For example, the following transcripts offer two contrasted versions: the first transcribes
Jean’s movements (his walk) more globally than the second (where different steps are
distinguished, showing the coordination between the steps and the organization of the
progression of his talk):
(13a) Example (less granularity):
1 JEA IL EST, *il est de bar*celo:ne, i ramène *le so[leil ^ca]talan,

2 ELI [le so^leil]

3 YAN exacte[ment.]

4 JEA [la me]:r, le *bleu:, voilà.

5 ELI mhm.

6 (0.3)

7 JEA voilà. (. ) c’est *tou.

(13b) Example (more granularity):
1 JEA IL EST, *il est* de bar*celo:ne, i ramène *le so[leil ^ca]talan,

2 ELI [le so^leil]

3 YAN exacte[ment.]

4 JEA [la me]:r, le *bleu:, voilà.

5 ELI mhm.

6 (0.3)*

7 JEA voilà. (. ) c’est *tou.

If a participant does multiple relevant embodied movements at the same time, they will be
described in different lines:
5. Timing and transcription of silent actions

Silent embodied actions are transcribed by reference to their temporality, indicated on the numbered line, in fragments of seconds.

(15) Example: (two baboons come closer to each other and enter in interaction)

1. (1.4) + # (1.8)** (0.2) + # (0.3) * (0.1)
   Ava >>walks twd BIN*pivots, posterior twd BIN-->
   Ava + looks back at BIN*
   Bin + looks at AVA----+ looks at AVA’s posterior--->

In some cases, silent embodied actions can be represented in such a way that they constitute the main course of action, with talk having a secondary line of action:

(16) Example: (the customer gives money to the seller who gives back some change, both thanking each other)

3. * (0.7) +*(0.1)+ (1.0) ** (0.5) +*(0.2)+| (0.2)+*(0.3) |
   a cus +*hands over the change------------------------*,++,-->>
   b sel *grasps money*
   c sel + turns away--->
   d SEL |danke schön|

What constitutes the numbered line—the turn-at-talk or the timing—is an analytical decision (see Mondada 2019/SocInt).

6. Description of embodied actions

The description of embodied actions is not straightforward and relies on analytical/conceptual decisions.

It is important to avoid physicalist descriptions (e.g. ‘arm moves 45° down to the floor’) as well as intentionalistic descriptions (e.g. ‘wants to grasp the bottle’).
Some categories tend to standardize action (e.g. ‘points’, ‘gazes’), some tend to be very
general (e.g. ‘gesticulates’), some can be very specific (e.g. ‘points with the top of the pen’),
some require to stay generic (e.g. ‘extends arm’). This depends on the granularity of the
analysis, as well as on the local relevancies governing the action.

(17) Example:

1  LAU     "ah [oui, *ici c’est vrai."*
2  VIV     [mais *alors i-il+ y va p*our+ les *dé- pour les sortir
             lau   *extends H with pen*,*,*,*,*,*
             viv   +3 big nods--+

Instead of describing the movement as a ‘pointing’ gesture, a decision has been made here to
treat the movement—which is then retracted, and then abandoned— as a simple extension of
the hand (which can project pointing but also writing) and not yet as a something more
precise, like a gesture.

The description of the movement can be more or less precise:

(18a) Example:

1  MAR     ici [c’est *les vestiai*re*s?*
2  ROB     [eh         [oui mais
             mar   *points--------*

(18b) Example:

1  MAR     ici [c’est *les vestiai*re*s?*
2  ROB     [eh         [oui mais
             mar   *LH points w pen*

Although ‘points’ could be a good enough annotation, further expanded in the analytical text,
in some cases ‘points with his pen’, ‘points with open horizontal palm’ etc. might be more
accurate.

Another constraint for these annotations is their length: it is better to choose short
descriptions, fitting within the transcript’s spatialization, in a readable and intelligible way.
Short descriptions can be then expanded in the analytical commentary.

In order to achieve shortness, annotations can be abbreviated. Different solutions are possible,
as here below:

(19a) Example:

1  MAR     ici [c’est *les vestiai*re*s?*
2  ROB     [eh         [oui mais
             mar   *LH points w pen*

(19b) Example:

1  MAR     ici [c’est *les vestiai*re*s?*
2  ROB     [eh         [oui mais
             mar           *---------------*

1: Left hand points with the upper part of the pen

7. Screen shots

Multimodal transcripts are hybrid objects that rely on text and images, which are integrated in
the transcription.

It is very important to always specify the exact moment at which the image refers and to
synchronize it with respect to the relevant line of talk and of embodied conduct. This is done
by inserting a symbol (#) both on the line of the talk/of the measured time and on the line
dedicated to the image (fig in the margins).

(20) Example:

1  GEO     alors, y a quand même une autre chose à vérifier, c’est
2  que: ce: ces chambres-là n’ont que une issue de secours
3  (0.7)# (0.3)

fig #fig:1

Conventions for multimodal transcription

Lorenza Mondada — 8
Often, it is useful to insert several images one beside the other, one consecutive to the other, contrasted enough to show the trajectory of a movement, a gesture, or an embodied action (see figures 1-2-3 above). The disposition of the images within the transcript is an analytical decision.

In order to enhance the readability of the images, it might be useful to add circles and arrows highlighting a relevant detail: images can be annotated.

(21) Example:

```
1 BLA               [yeah % i mean, %turns back-->
                     $turns back-->
2 bla (0.3) % (1.9)  -->%extends her arm backwards-->
3 BLA anne ram/beau# is %ehm the person who:
   -->%points back-->
   $points back-->
   #fig.1
```

The screen shot can be accompanied with a caption (like in example 20) or followed by a simple number (as in example 21).

8. Fonts

In order to enhance the readability of the transcripts various types of font can be used. In the example below, bold is used for talk, italics for translation and normal for embodied conducts.

(22) Example:

```
1 PAL     ben suivant le cas euh: ben on tra- on est là
```
Conventions for multimodal transcription

9. Note on the difference between multimodal transcription and description

Descriptions are often put between two double parentheses:

(23) Example:
1 PAL une unité annexe, ((coughs)) qui: sert (.) uniquement, 
2 et en continu, (0.3) à la génération de renouvellement.

This convention might be useful for some comments related to vocal conducts that are not transcribed – although its limitations have been pointed at (see Jefferson 1985 on transcribing laughter).

This convention is not useful for transcribing embodied conducts, because it is not temporally precise enough and it does not describe neither the boundaries nor the length of the embodied movement (see the principles above; see the contrast between the following two examples).

(24a) Example:
1 BRU la maison qui se retrouve ici:,((points))
   the house that is located here:

(24b) Example:
1 BRU la maison qui A se retrouveA ici:,A
   the house that is located here:
   Apoints to docA,,,,,

10. Note on the difference between multimodal transcription and coding

The principle of multimodal transcription is the relevance of embodied conducts for the organization of the course of action, within the endogenous perspective of the participants. This implies a selection of relevant details when transcribing, as well as the use of situated descriptions for the transcribed actions.

By contrast, what characterizes coding is a) the use of general and a priori defined categories for describing actions, and b) the continuous annotation of conducts across the datum (e.g. of gaze shifts or gestures of the hand along an entire fragment).

The use of the aligning software ELAN is very useful for calculating timing in multimodal transcription, but its organization in predefined lines (tiers) for predefined conducts is a feature that invites to coding rather than to transcribing.
Bibliography

Some articles in which these notations have been used/explained/discussed.

Conventions for multimodal transcription

Lorenza Mondada — 11


