

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)

```
2 LUC pξeu- *r'gardez #+%le:#ξ* .hβh+h #le papi%βllon +bleu là:,*
      may- look the .hhh the butterfly blue there
luc ξ.....ξpoints twd insect----->
eli *one step fwd----*another step fwd-----*
yan %looks-----%pivots-->
jea +looks-----+pivots-->>
fig βturns H back---βpivots-->
      fig.2# #fig.3 #fig.4
```



See also the following web sites:

The English version is available under two addresses:

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

and

https://franzoesistik.philhist.unibas.ch/fileadmin/user_upload/franzoesistik/mondada_multimodal_conventions.pdf

For the French version (not updated):

http://icar.univ-lyon2.fr/projets/corinte/documents/convention_transcription_multimodale.pdf

See also for a conceptual discussion of these conventions:

Mondada, L. (2018). Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality, *Research on Language and Social Interaction*, 51:1, 85-106.

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)
Δ Δ	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 screen shot 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
- Δ 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.

(3b) Example:

1 LAU +°ah oui° je+ sais
viv +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 ξmais #ceux-là# là φ çaζ seraitξ↓ dommage\$ φhein,# de:
but these there it would be too bad right, to:
luc ξpoints w open palm-----ξ
luc ζlooks 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 Δ se retrouve ici:,Δ
Δpoints to doc-----Δ

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°
viv +points-->
2 (0.5)
3 VIV mais alors+ i- il y va pour les dé- pour les sortir
-->+

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)
viv +points-->
3 VIV mais alors+ i- il y va pour les dé- pour les sortir
-->+

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 Δeuh cette attribution de: de terrain est bien moins claire.
  bru Δlooks at PAL-->>
```

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,
  >>points-----@
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,
  >>points-->
2 euh cette @attribution 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 Δcas@ de figure de: (.) des +causses, @en général,
  >>points-----@ @palm open-->
  bru Δlooks at VIV-->
  viv +points-----+
2 euh cette@ attribution de: de terrainΔ est bien moins Claire
  -->@
  bru -->Δ
```

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 Ale so[leil Δca]talan,
           HE IS, he is from Barcelona, he brings back the catalan sun,
           *moves-----*

2 ELI                                           [le soΔleil]
                                           [the sun]
eli                                           Δpivots fwd--Δ

3 YAN      exacte[ment.]
           exact[ly.]

4 JEA      [la me]:r, le *bleu:, voilà.
           [the seje:, the blue:, that's it.
           *begins to walk slowly-->

5 ELI      mhm.
6          (0.3)
7 JEA      voilà. (.) c'est *tout.
           that's it. (.) that's all.
           ---->*walks away-->>
```

(13b) Example (more granularity):

```
1 JEA      IL EST, *il est* de bar*celo:ne, i ramène Ale so[leil Δca]talan,
           HE IS, he is from Barcelona, he brings back the catalan sun,
           *Lfoot-*Rfoot---*

2 ELI                                           [le soΔleil]
                                           [the sun]
eli                                           Δpivots fwd-Δ

3 YAN      exacte[ment.]
           exact[ly.]

4 JEA      [la me]:r, le *bleu:, voi*là.
           [the seje:, the blue:, that's it.
           *1 step fwd*1 more step-->

5 ELI      mhm.
6          (0.3)*
jea      -->*1 step fwd-->
7 JEA      voilà. (.) c'est *tout.
           that's it. (.) that's all.
           -->*walks away-->>
```

If a participant does multiple relevant embodied movements at the same time, they will be described in different lines:

(14) Example:

```
1 PAL .hh [ c'est des [ c'est le leV schéma euh: Vordinaire, .hh
  lau [looks at PAL[
  lau V.....Vpoints-->>
```

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---->
  fig #fig.1 #fig.2
```



This enables the conventions to transcribe silent embodied actions, done by humans as well as by animals (Mondada, 2018/ROLSI).

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 -->+holds money twd SEL-----+,,,,,....+takes drink->
  b sel ->•gz money-->
  c sel *takes to bc*bar-c*back on c*,,,,,....*grasps money*
  d SEL |--°merci°---|

4 (0.6)+(0.2)+(2.9)*(0.2)*•(0.1)+±(0.2)| (0.4)±(0.2)| (0.2)*+(0.3)|
  a sel *hands over the change-----*,,,,->>
  b sel -->•gz CUS----->>
  c SEL |danke schön|
  d cus ->+,,,,+holds RH twd SEL--+extends RH-----+grasps change->>
  e cus tgz money----+turns away-->>
  f CUS |-danke au--|
```

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-ill+ 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-----*
```

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

7. Screen shots

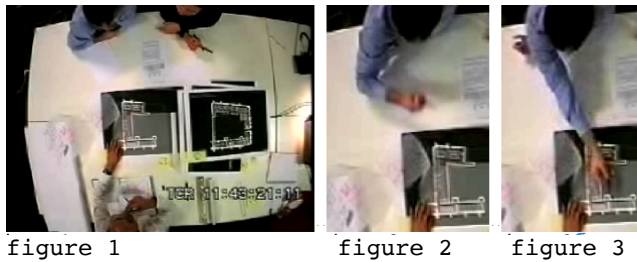
Multimodal transcripts are hybrid objects that rely on text and images, which are integrated in the transcription.

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 u:ne 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
```

4 LUC *non:, on# a on a*# touj- ah oui, (.) parce qu- (.) parce
 *.....*points--->>
 fig #fig.2 #fig.3



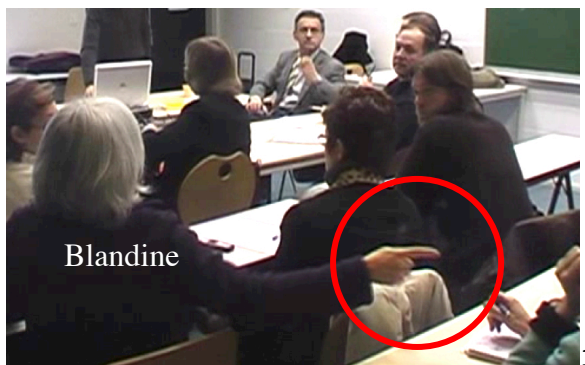
5 que là i bloquerait cette issue
 6 GEO ben oui.

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-->
 2 (0.3) % (1.9)
 bla -->%extends her arm backwards-->
 3 BLA anne ram%beau# is [ehm the person who:
 -->%points back-->>
 fig #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

Bibliography

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

- Mondada, L. (2019). Transcribing silent actions: a multimodal approach of sequence organization. *Social Interaction. Video-Based Studies of Human Sociality*, 2:2. DOI: 10.7146/si.v2i1.110964 (<https://tidsskrift.dk/socialinteraction/article/view/110964/160083>).
- Mondada, L. (2018). Transcription in linguistics. In L. Litosseliti (ed.) *Research Methods in Linguistics* London: Bloomsbury (2d ed.), 85-115.
- Mondada, L. (2018). Visual practices: video studies, multimodality and multisensoriality. In D. Favareau (ed.). *Co-operative Engagements in Intertwined Semiosis: Essays in Honour of Charles Goodwin*, Tartu: University of Tartu Press, 304-325.
- Mondada, L. (2018). The multimodal interactional organization of tasting: Practices of tasting cheese in gourmet shops, *Discourse Studies*, 20(6) 743–769. DOI:10.177/1461445618793439
- Mondada, L. (2018). Questions on the move. The ecology and temporality of question/answers in mobility settings. In Deppermann, A. & Streeck, J., eds, *Modalities and temporalities*. Amsterdam: Benjamins, 161-202.
- Mondada, L. (2018). Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality, *Research on Language and Social Interaction*, 51:1, 85-106.
- Mondada, L. (2017). Walking and talking together: Questions/answers and mobile participation in guided visits, *Soc Sc Information, S.I. on Human motion and Social Context*, 56:2, 1-34.
- Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20:2, 2-32.
- Mondada, L. (2015). Multimodal completions. In : Deppermann, A., Günthner, S. (eds.), *Temporality in Interaction*. Amsterdam: Benjamins, 267-307.
- Mondada, L. (2014). The local constitution of multimodal resources for social interaction. *Journal of Pragmatics*, 65, 137-156.
- Mondada, L. (2014). Pointing, talk and the bodies: Reference and joint attention as embodied interactional achievements. In Seyfeddinipur, M. & Gullberg, M. (eds.). *From Gesture in Conversation to Visible Utterance in Action*, Amsterdam: Benjamins, 95-124.
- Mondada, L. (2014). Bodies in action: multimodal analysis of walking and talking, *Language and Dialogue* 4:3, 357–403.
- Mondada, L. (éd). (2014). *Corps en Interaction. Participation, spatialité, mobilité*, Lyon: Editions ENS.
- Mondada, L. (2013). Embodied and spatial resources for turn-taking in institutional multi-party interactions. *Journal of Pragmatics*, 46, 39-68.
- Mondada, L. (2012). Video analysis and the temporality of inscriptions within social interaction: the case of architects at work. *Qualitative Research*, 12, 3, 304-333.
- Mondada, L. (2012). The conversation analytic approach to data collection. J. Sidnell, T. Stivers (Eds.). *The Handbook of Conversation Analysis*, Blackwell-Wiley, 32-56.
- Mondada, L. (2011). Understanding as a embodied, situated and sequential achievement in interaction. *Journal of Pragmatics*, 43, 542-552.
- Mondada, L. (2009). Emergent focused interactions in public places: A systematic analysis of the multimodal achievement of a common interactional space. *Journal of Pragmatics*, 41, 1977-1997.
- Mondada, L. (2009). The Embodied and Negotiated Production of Assessments in Instructed Actions, *Research on Language and Social Interaction*, 42 : 4, 329–361.
- Mondada, L. (2008). Documenter l’articulation des ressources multimodales dans le temps: la transcription d’enregistrements vidéos d’interactions. In: M. Bilger (éd). *Donnees orales, les enjeux de la transcription*. Perpignan: PEP, 127-155.
- Mondada, L. (2007). Transcript variations and the indexicality of transcribing practices. *Discourse Studies*, 9/6, 809-821.

- Mondada, L. (2007). Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, 9 :2, 195-226.
- Mondada, L. (2006) 'La pertinenza del dettaglio: registrazione e trascrizione di dati video per la linguistica interazionale', in Y. Bürki and E. de Stefani (eds). *Trascrivere la lingua. Dalla filologia all'analisi conversazionale*, pp. 313–44. Bern: Lang.
- Mondada, L. (2005). La constitution de l'origo déictique comme travail interactionnel des participants : une approche praxéologique de la spatialité. *Intellectica*. 2/3, 41-42, 75-100.
- Mondada, L. (2004), Temporalité, séquentialité et multimodalité au fondement de l'organisation de l'interaction: Le pointage comme pratique de prise du tour. *Cahiers de Linguistique Française*, 26, 169-192.
- Mondada, L. (2003). Working with video: how surgeons produce video records of their actions, *Visual Studies*, 18/1, 58-72.