Multidimensional Coding of Multimodal Languaging in Multi-Party Settings

Christophe Parisse[1], Marion Blondel[2], Stéphanie Caët[3], Claire Danet[4],
Coralie Vincent[2], Aliyah Morgenstern[5]
[1]INSERM, Modyco, CNRS & Paris Nanterre University, Nanterre, France ;
[2]SFL, CNRS & Paris8, France
[3]STL, CNRS & Lille University, Lille, France
[4]Dylis, Normandie University, Rouen, France
[5]PRISMES, Sorbonne Nouvelle University, Paris, France
cparisse@parisnanterre.fr

Abstract

In natural language settings, many interactions include more than two speakers, and real-life interpretation is based on all types of information available in all modalities. This constitutes a challenge for corpus-based analyses because the information in the audio and visual channels must be included in the coding.

The goal of the DINLANG project is to tackle that challenge and analyze spontaneous interactions in family dinner settings (two adults and two to three children). The families use either French, or LSF (French sign language). Our aim is to compare how participants share language across the range of modalities found in vocal and visual *languaging* in coordination with dining. In order to pinpoint similarities and differences, we had to find a common coding tool for all situations (variations from one family to another) and modalities.

Our coding procedure incorporates the use of the ELAN software. We created a template organized around participants, situations, and modalities, rather than around language forms. Spoken language transcription can be integrated, when it exists, but it is not mandatory. Data that has been created with another software can be injected in ELAN files if it is linked using time stamps. Analyses performed with the coded files rely on ELAN's structured search functionalities, which allow to achieve fine-grained temporal analyses and which can be completed by using spreadsheets or R language.

Keywords: multimodality, interaction, spoken language, sign language, ELAN, template, coding

1. Analyzing real-life interactions

Language has long been studied out of its ecological context, first through written forms characterized by their linearity, then through invented sentences, and finally with a focus on speech, in experimental studies or semi-guided interactions. Even when gestures are integrated in the analyses, environments are most often stripped of objects or other activities whose affordances have a multitude of impacts on their use (but see Mondada, 2016). Those limits can be viewed as strengths as they have conducted to fruitful research on *the language system*. However, in order to capture the full complexity of *language use*, new approaches and methods are needed in which all our semiotic resources can be analyzed as they are deployed in their natural habitat involving the orchestration of bodies engaged in a variety of situated activities with a diversity of artifacts.

The aim of our research is to capture language in its ecological environment in order to articulate its actional roots and its symbolic functions. Following Boutet (2018), we thus analyze the bodies of our participants as both the support (the instrument) and the substrate (that which constitutes and structures) for "languaging" (Linell, 2009). Our approach grounds language in embodied action rather than viewing it only as a code or a symbolic system (Bottineau, 2012). What we call "languaging" is not only relative to the languages and cultures a subject uses, but also to the available semiotic resources that can be coordinated and enable us to embody mental constructions. Reversely, the semiotic resources we use shape, construct and contribute to the meaning of our interactive productions.

We share Mondada's multimodal approach (2019: 47) according to which "research in multimodality – that is, the diversity of resources that participants mobilize to produce and understand social interaction as publicly intelligible actions, including language, gesture, gaze, body postures, movements, and embodied manipulations of objects – can be further expanded by considering not only embodied resources for interacting but also embodied practices for sensing the world in an intersubjective way."

This means that, in real-life interaction, understanding other people does not only rely on the language produced. Linguistic analyses, even when determined with great precision, i.e through standardized notation systems, with very good intercoder agreement, rely on a range of cues and on the context. To participate fully in an interaction, a semiotic understanding of the situation needs to be achieved. In many situations, what was linguistically produced is insufficient to understand the meaning, or is even sometimes misleading. The linguistic material is a vital part of semiotic understanding, but the context and all the actions of the participants are also crucial.

The goal of this paper is to present an analytical framework and method that will be used to annotate and study a corpus of natural multiparty interactions in family dinner settings. This corpus is collected in the context of a project funded by the ANR, "Multimodal LANGuage practices in French family DINners" (DINLANG) in which our main aim is to analyze the coordination between the co-activities of dining and languaging in French and French sign languages. The material used for this project, especially the ELAN template described in this paper and our annotation guide, is can be downloaded from the NAKALA repository

(https://doi.org/10.34847/nkl.97ff535e). This material will be updated during the duration of the project.

Family dinners are shared moments of everyday life which present a perfect opportunity to study how language and interactive practices are transmitted to and used by children in order for them to construct meaning (Morgenstern et al., 2021). Because the subtle interweaving of these practices while eating fully engages the body, our project highlights the semiotic differences between participants using a spoken language, French, and a sign language, Langue des Signes Française (LSF), including children at different ages.

The project includes recordings from families primarily using a spoken language and from families primarily using a sign language (families where at least one member is deaf). This means that the analytical framework cannot use a written language form of transcription as the underlying structure of coding features, contrarily to what is most often found in language corpora, because it does not exist for French Sign Language and is not sufficiently relevant to code various multimodal forms of language and actions.

Moreover, using written language for semiotic analyses is dangerous as the main features of written language tend to hide the real properties of language interactions. Indeed, a body of literature (see Harris, 1990; Linell, 2005; Love, 2017) has demonstrated how written language forms have led to misunderstandings on what language actually is.

Using a phonetic transcription as the basic structure of the corpus contents could be possible, as we could conduct the required analyses for both vocal *languaging* (using mouth movement description) and visual *languaging* (using hand movement description). But these analyses would not be sufficient to achieve our goal of providing a semiotic analysis, and in both vocal and visual *languaging* it would be difficult and would not be desirable, according to our perspective, to draw a semiotic division between vocal productions, signs and gestures.

To avoid drawing a dividing line between actions, gestures, speech and sign, and conducting interpretations that suffer from the "written language bias" we use an analysis based on modalities and on interactions between participants at the highest structural level, and transcription or symbolic coding at the lowest structural level. This does not mean that our coding will not contain spoken language transcriptions when they are possible, or written descriptions and translations when they are useful. But these elements are not the ideal theoretical representation of language (see Harris 1990, Linell 2005, Love 2017).

2. Interaction and modalities

Our theoretical framework combines language socialization, cognitive grammar, interactionist and multimodal approaches to *languaging*. We borrow the term *languaging* to refer to multimodal language use – "linguistic actions and activities in actual communication and thinking" (Linell, 2009: 274) expanding the term to include speaking, gesturing and signing. We study how children's socialization to a variety of modes of expression in their daily experiencing (Ochs, 2012)

through dinners shapes the development of their language use.

Interaction is a powerful theoretical framework for the analysis of semiotics (see Linell, 2009; Mondada, 2008). So as to avoid the pitfalls described above, we have based our analysis on two main features: interaction and modalities. Who is participating, and in which modality, are our topmost levels of analysis. We also include collective activities (that cannot be ascribed to one participant only) at our top-level analysis.

3. Recording set-up

Our goal of gathering and coding real-life interactions in dinner settings also has consequences on the equipment used to record the interactions. Our recording set-up is designed to collect as much information as possible without being a hindrance to the dinner participants.

As the location of the dinner is fixed, always around a table (of any shape), the recording equipment is also fixed. We have three recording points where a camera and a sound recorder are placed. A 360° camera is placed above the table on a boom stand. A 360° sound recorder is also positioned on the boom stand (see Figure 1). Although the 360° provides very good information about what the participants are doing and what is happening at
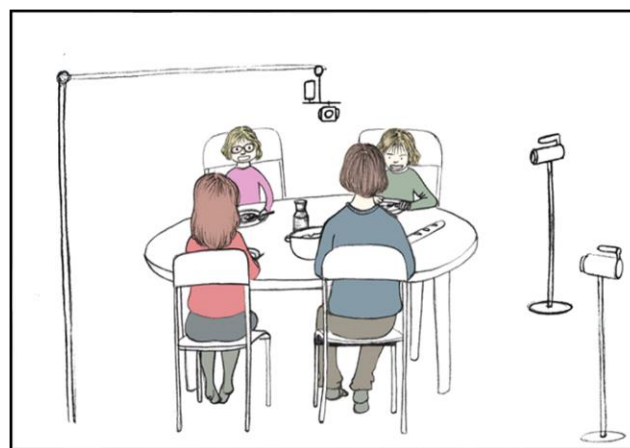

Figure 1: Recording set-up - @Claire Carpentier

the table, it is not perfect for coding gaze direction and the posture of the participants. Additional recordings are done with two classic wide-angle cameras located at two different sides of the table. Each camera provides a frontal/side view of half of the participants and a rear-view of the other half. The cameras also record sound, and we have added better quality microphones.

We thus obtain three different video-recordings and four different audio-recordings. The 360° video can be converted to two 180° videos to make it easier to use with ELAN (2021)[1]. All the video and audio files are synchronized with a classical video clap. The recording is analyzed from the first call to the table and is stopped when all the participants have left the table.

---

[1] https://archive.mpi.nl/tla/elan - Max Planck Institute for Psycholinguistics, The Language Archive, Nijmegen, The Netherlands

4. ELAN features

ELAN (Brugman & Russel, 2004) has many powerful features which can be used to organize both audible and visible data (Boutet & Blondel, 2016; Vincent, 2020). The two main features are the template and the searching module, and each feature can be set up according to a temporal or a structural organization.

4.1 Template, tier structure, vocabularies

The template is the most well-known feature of the ELAN software. It is a static organization that is decided at the beginning of the coding process and where later on changes are limited in scope. A template organizes all the tiers used in ELAN, and especially the relations between the tiers.

At the highest level of organization in a template, we created tiers (that we called the main tiers) with a temporal organization. The main tiers are all independent and they contain annotations that are characterized by their beginning and their end boundaries.

The other tiers (that we call dependent tiers) are organized according to their relationship with elements of the tier called the "parent" tier. The parent tier can be one of the main tiers, or any of the other dependent tiers, producing a tree-like organization of the data. There are several modes of organization defined in ELAN, but in our current project, we only use the "Time Subdivision" mode. In this mode, a dependent tier contains a description of a temporal division of the annotations in the parent tier. Coding the dependent tier is not mandatory, but if coded, the dependent tier cannot contain blanks and must have the same span as the parent tier.

When coding data, it is possible to restrict the annotations to a specific (controlled) vocabulary, so as to avoid using erroneous or undescribed annotations. We plan to use this feature whenever it is possible, once our inventories of all categories found in the data seem sufficiently comprehensive, especially to code gaze properly.

4.2 Searching module

The first function of the searching module is to find elements in the annotation on the basis of any string or keyword. ELAN presents many powerful search features, including features that search words or strings of words in as many files as needed, and features that allow the user to replace targeted elements by new items.

One of the searching tools of ELAN is the "Structured search of multiple EAF". As indicated by its name, this searching tool can structure data and process as many files as necessary.

A very interesting feature of the structured search is that it allows the user to improve on the functionalities offered by the template. A major reason for using templates is that they enable us to organize our data and to improve the quality of our annotations. But they also make it easy to extract information from the transcriptions and export them to a statistical software or a spreadsheet for further analysis.

However, the template organization of ELAN, although powerful, is limited because it accepts neither temporal overlaps nor distant temporal relationships between the dependent tier and the main tier. Dependent tiers have to be included within the temporal boundaries of the parent tier. If two annotations have only partial overlap, or do not overlap, the template tool cannot be used, as the extraction and exportation of those annotations is not possible.

It is possible to go beyond this limitation thanks to the "Multiple Layer Search" option of the "Structured Search of multiple EAF" searching tool. With this option, we can search for co-occurrences across multiple tiers (layers). A co-occurrence can be specified by a temporal or a hierarchical relation. Results can be exported in a tabular format (CSV). For each line, the export contains all the information for one hit, thus with this format, it can be further exploited in any spreadsheet processor or statistic tool. For example, we can find all the gazes of one specific participant that occur before the gazes of another participant and overlap with them. The time elapsed between the gaze of the first participant and the gaze of the second participant can be controlled to avoid having a first gaze that ends too much time in advance (by setting a maximal value, for example two seconds between the beginning of the first and the second gaze) and a first gaze that begins nearly at the same time (by setting a minimal value, for example 100 milliseconds between the beginning of the first and the second gaze).

The Multiple Layer Search can find any temporal relation, even when a long time has elapsed between the occurrences in each tier. It can also find sequences of annotations for the same participant, or any combination of sequences and overlaps. The number of overlaps and the number of annotations in a sequence can be more than two annotations long. Finally, the relationship between annotations can be described using all the possible temporal and structural settings of ELAN, but also using sets of participants, or sets of coded tiers, etc. The searching tool can be used to locate elements in the files, as well as to export the results. Finally, the queries can be saved, so complex queries can be reused easily whenever new data is available.

5. Coding features for a multimodal analysis

Coding features are organized hierarchically. We placed the participants and collective tiers (for example conversational topics) at the main level. Collective tiers link several participants together and are rarely attributed to a single participant (there could be for example occurrences of the youngest child producing a vocal or gestural monologue while the other members of the family are conversing together). Coordinated tiers also include the participation framework (Goffman, 1981), e.g. the group of participants that are involved in an interaction. Even if some of them do not actually produce vocal or visual language and are not either the main speaker/signer or the main addressee, they form a 'framework' in which each member has a participation status. We thus code the cues that enable us to assess that status (as speaker/signer, addressee or overhearer for instance).

For each participant, the coding describes a set of dimensions. A dimension can be a resource, in the sense of Mondada (2019) and Feyaerts et al. (2022), which can

correspond to language (mouth and hand), gaze, body postures, actions, manipulations of objects, … Dimensions can also be analyses of the situation, for example the theme of the conversation, the participation framework, the discourse topic ...

As a rule, coding indicates that the participant is doing something, either producing sound or movement (with hands, arms, head, bust, gaze or facial articulators), thus the production is not necessarily verbal or with conventional meaning.

6. Implementation in the current project

6.1 Organization of the ELAN template

The ELAN template includes a series of main tiers and dependent tiers. On the main level, we placed the participants as individuals, and the participants as groups. Each of these main tiers is divided into all the modalities necessary for our analysis. Modalities are logical subdivisions of participants and groups, but we did not use the structural mechanisms of ELAN to organize the division into modalities, so as to keep a maximal flexibility in the use of ELAN (structural properties cannot be used to describe overlaps). Queries about the relationships between modalities (as presented in 4.2.2) use the temporal organization of the transcription rather than its structure.

6.2 Participants

There are four to five participants in our recordings. Two parents, typically but not necessarily the mother and the father, and two or three children. Each participant is associated to a unique tag. Each participant has the same set of dimensions coded.

6.3 Groups

Groups refer to situations where it is necessary to code something that is shared between several participants. In our work, this includes:

Presence: Who is present in the situation, including participants that are not producing anything at the target moment, but who could participate (in real life you can talk to several people, but not all of them will necessarily answer you. Nonetheless, your discourse will take their presence into account, so it is necessary to code this information). There can be presence in the audio channel, the visual channel or both.

Themes: What is happening in general? What is the situation about? What is the topic of the conversation? Several topics can co-exist. Themes are often deduced from the situated *languaging*, or from the semantics of non-verbal actions.

Participants: Who is actually involved in the participation framework?

6.4 Other dimensions

For all participants, several dimensions can be analyzed. The four main dimensions are: presence, gaze, audible production, visible production. These four dimensions are structurally independent from each other. In all dimensions, all events have a beginning and an end. But there is no specific structural relation between them. This is not because such a relation cannot exist, but because we cannot use structural constraints on these relations, as structural constraints in ELAN limit the possible temporal constraints (it becomes impossible to indicate overlap, precedence, or succession). However, analyses of these temporal relations can be conducted thanks to the searching system described below.

The main level of analysis simply indicates:

- A participant is present in the room, is not actually in the room but it is possible to hear her or him, or is not present at all.
- A participant produces something visible.
- A participant produces something audible.
- A participant gazes at one or several persons or objects.

A controlled vocabulary is used in the tier referring to presence: in-camera field, off-camera field, out of-room-can-be-heard, out of-room-cannot-be-heard. For gaze, the annotations contain information about the participants or inanimate elements (objects) the gaze is directed at. For an audible or visual event, we can describe both the actions and the *languaging* that occur during the event. The main transcription level allows us to segment and tag the events, but without specifying their nature. The information is included in the dependent action and *languaging* tiers. These annotations are temporally organized (with beginning and ending boundaries) using ELAN's "Time subdivision". The coding system for *languaging* is not the same in the vocal and visual modalities.

6.4.1 Descriptions of actions

All actions are first coded as dining related or not. Further description can be included in a comment or in a sub-tier in natural language, whether they may carry a communicative value or not. If they have a communicative value, the description of the action includes the symbol "§". Actions can be quite automatic and non-intentional, or intentional. Most actions are coded in the visual part of the template, but audible actions can also be coded.

6.4.2 Descriptions of audible *languaging*

Audible languaging contains the name of the language used in the main tier line. In a dependent 'script' tier, there are orthographic transcriptions of what is said. They can be completed by symbolic codes to indicate intonation, onomatopoeia, laughter, etc. They follow the principles used in classic spoken language transcriptions. More specifically they follow the convention of the CHILDES system (CHAT: MacWhinney, 2000) as the audible *languaging* is first transcribed using the CHAT software independently for the rest of the coding system. When it is finished, a conversion is performed using a specific tool (TEICORPO: Parisse et al., 2022) which produces an ELAN file with the transcription in the correct ELAN tier.

6.4.3 Descriptions of visual *languaging*

These descriptions target all symbolic gestures (arms, faces, torso). In our theoretical approach, we consider that sign language and gestures are in the same continuum because gestures have symbolic meaning in sign languages and are used productively in languaging.

Moreover, there is ambiguity in sign languages between what is a gesture specific to LSF and a shared gesture with the surrounding cultural community (that signers and non-signers may use, such as pointing, shrugs or headshakes), as these shared gestures are often fully incorporated (or grammaticalized) in sign language.

So, we do not separate sign language and gesture in our template. Visual languaging can also be produced by hearing speakers when they produce symbolic gestures.

French Sign Language (LSF) is annotated in the sub-tier named "script" and using ID-Gloss (consistent labels in the written surrounding language, including codes for non or semi-lexical unit, see Johnston, 2014). A free translation is provided on another dependent tier.

More fine-grained analyses of the hand-movements are not currently included in our work, but will be conducted later in the project.

6.4.4   Addressee(s)

The description can be completed by a dependent description of the addressee(s) of the *languaging*. This line, called "interloc" contains only a controlled vocabulary with the codes of the participants. It is possible to include more than one addressee at the same time.

6.5   Example for a French speaking family

F	visible in camera		
act-aud-F	---		
lng-aud-F	ENG		
script-aud-F	so did you do the science exam?		and what was it like?
interloc-lng-aud-F	1-Ca		1-Ca
act-vis-F	kisses the boys		
Ca	visible in camera		
lng-aud-Ca	ENG		
script-lng-aud-Ca		yes.	
interloc-lng-aud-Ca		1-F	

Table 1: simplified version of speaking family annotation

Table 1 shows a simplified version of part of the coding for a speaking family. The notations in _'- - -' delimit the duration of the father's visible or audible production. Code F is for Father, Ca for elder child. Aud is for Audible, Lng for *Languaging*, Interloc is the addressee (so for example interloc-lng-aud-Ca corresponds to addressee(s) of the elder child when producing *languaging* in audible form). 1-Ca corresponds to *languaging* directed at the elder child only, 1-F to *languaging* directed at the father only.

6.6   Example for a LSF signing family

M	visible in camera		
lng-vis-M	Sign Language		
script-lng-vis-	gest-hand	gest-hand 'wait'	THEME MASK, PT2 READY IDEA

M	'wait'		CLOTH WHAT
interloc-lng-vis-M	1-Cb		1-Cb
reg-M	1-Ca	closed eyes	1-Ca
theme1	control Cb	disguise	

Table 2: simplified version of visual language annotation

Table 2 shows a simplified version for a signing family. M is the mother, Cb is the younger child. The *languaging*

Figure 2 : query for some *languaging* followed by at least 300ms by an action for the mother

part is in the lng-vis-M tier instead of the lng-aud-F tier. An example of theme is given.

6.7   Queries

As presented in part 4.2.2, queries will be very useful to structure the data available in our corpus. Indeed, having a *languaging* and semiotic approach means that there is no preset organization of the data such as what can be imagined in a theory based on the primacy of conventional speech or sign. In real-life, gestures can take on symbolic meaning for both the speaker/signer and their addressee(s) on the spur of the moment, language forms can be used in a repetitive manner just to tease someone or emphasize a situation, there can be a gaze before or after either a word or a gesture, etc. Data organization is not stable.

This is expressed in our coding by using annotations that are organized according to their temporal boundaries. We thus simply indicate within the beginning and ending boundaries of an annotation if a participant is present, if a participant is producing audible or visible *languaging* or audible or visible *acting*, if a participant is gazing at a specific person or object. A preset structure simply cannot be used. Therefore, in order to structure our data and obtain results, we must use the search options.
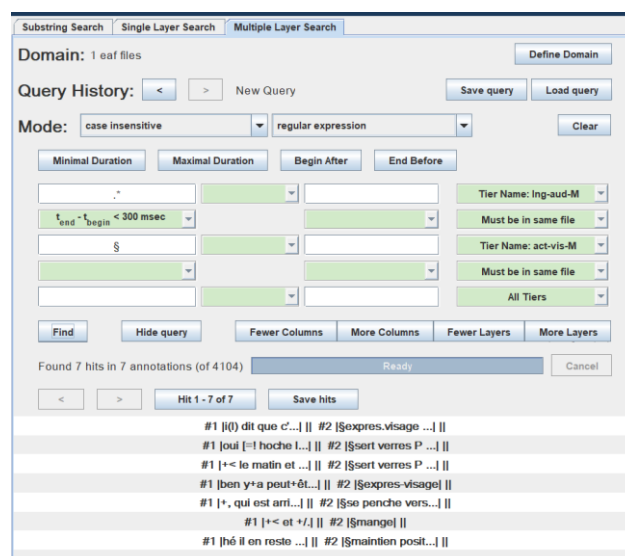
For example, if we want to know if a child's gaze precedes the mother's gesture or her spoken production, we can use a searching option to find out all the possible occurrences. Or we can find all relationships between

children's gaze and mothers' speech and conduct statistical analyses.

We can perform the same type of search within the coding for each participant. For example, we can find out what specific spoken utterance (annotated in the audible *languaging* tier) is produced by the mother before certain gestures (annotated in the visible *languaging* tier). This can be found by searching for the right overlap between speech and gesture. If we want to know only when the overlap is at least 300ms, we can add this condition (see Figure 2). There is no need to have a predefined organization of the template to analyze how speech and gesture are coordinated, as long as they are correctly coded within their temporal deployment.

If we want to find out the specific spoken interactions between the speaking participants, we can find them. If we want to cross tabulate this data with their gestures, or the absence of gesture, we can do that as well.

Queries are the perfect answer to a coding situation that is not clearly predefined, or that relies mostly on timing. ELAN is thus an excellent tool to conduct analyses on multimodal multiparty situated interactions with no preconstrued ideas on how to pair form (including action, LSF, French, gaze, and gesture) and meaning.

6.8 Limits of the implementation

There are cases where the temporal information is not sufficient to determine the degree of relationship between what is coded in the various modalities. One example is long distance temporal relations, which can be found using queries, but which could be hidden within multiple other relations that do not make sense. Another example is that things that occur at the same time might not be related, which is the case in multiparty interactions when at least two conversations occur at the same time.

These limits can be handled using structural information, which we do not use a lot because, as explained above, it is difficult to organize. Another means is to use the semantics of the values used in the coding process. For example, coding which people are engaged in a participation framework provides information about which temporal relations are meaningful or not. Using the right codes allow to make relevant queries, which are solely based on temporal information.

7. Conclusion

In this paper, we have presented the organization of an ELAN template and the use of the ELAN query module which will allow us to test our research questions. These include:

1. Because of the specialized role of gaze and of the articulators involved (mouth, hands, arms…), are there crucial differences between coordinating speaking vs. signing, and eating?
2. Will children become increasingly expert at coordinating semiotic resources and at navigating between activities?
3. Will developmental regularities according to age, cognitive, social and linguistic development as well as mode of expression (speech, sign, gesture) be identifiable despite individual and family variation?
4. Will the mode of expression and its formal components, when deployed in situated activities, have a major impact on how children construct meaning and develop language?

More research issues might be raised in the course of our project, but as of now, the multimodal nature of language has led us to develop a method to investigate how various semiotic systems such as speech, sign, gesture, posture, facial expressions and gaze but also actions and object manipulations, are simultaneously deployed, transmitted and used in the situated activities combined in family dinners. During multiparty, multimodal situated interactions in coordination with other body activities, every move, every part of the body, every object is potentially meaningful. They are deployed in a multitude of skillful variations in the collective coordination of bodies, activities and artifacts.

We thus use the affordances of the ELAN software specifically designed to annotate gesture and sign as well as other semiotic resources. ELAN can be extremely useful to analyze multiparty co-activities such as conversing and dining as it integrates temporal boundaries for the annotations and since both independent main tiers and dependent secondary tiers can be articulated in the template. Its very powerful searching options allow us to use queries in order to structure our analyses of the data. We can thus obtain results on all the possible research questions we might have concerning the orchestration of actions, gaze, signs, speech and gestures in the varying participation frameworks that occur in spontaneous ecological conversations.

8. Acknowledgements

9. Bibliographical References

Bottineau, D. (2012) Parole, corporéité, individu et société : l'embodiment dans les linguistiques cognitives. *Texto* ! vol. XVII, N°1 and 2.

Boutet, D. and Blondel, M. (2016). Les corpus de Langue des Signes Française. In Annelies Braffort (ed.) *La Langue des Signes Française (LSF) : modélisations, ressources et applications.* ISTE éditions, 47-85.

Boutet, D. (2018). *Pour une approche kinésiologique de la gestualité.* (Habilitation à diriger des recherches). Université de Rouen-Normandie. https://hal.archives-ouvertes.fr/tel-02357282

Brugman, H., Russel, A. (2004). Annotating Multimedia/Multi-modal resources with ELAN. In: Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation.

ELAN (Version 6.2) [Computer software]. (2021). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Goffman, E. 1981. *Forms of Talk*. Pennsylvania: University of Pennsylvania Press, 124-159.

Johnston, Trevor. 2001. « The lexical database of Auslan (Australian Sign Language) ». Sign Language & Linguistics 4 (1-1): 145-69. https://doi.org/10.1075/sll.4.12.11joh.

Johnston, T. 2008. Corpus linguistics and signed languages: no lemmata, no corpus. In O. Crasborn, E. Efthimiou, T. Hanke, E. D. Thoutenhoofd & I. Zwitserlood (eds) *Proceedings of the 3rd Workshop on the Representation and Processing of Sign Languages* (82-88).

Johnston, Trevor. 2014. « Auslan Corpus Annotation Guidelines ». http://new.auslan.org.au. juin 2014. http://media.auslan.org.au/attachments/Johnston_AuslanCorpusAnnotationGuidelines_14June2014.pdf.

Harris, Roy. (1990). On redefining linguistics. In *Redefining linguistics*, H. G. Davis & T. J. Taylor. (eds). London: Routledge.

Linell, P. (2005). *The Written Language Bias in Linguistics : Its Nature, Origins and Transformations* (1st edition). Routledge.

Linell, P. (2009). Rethinking Language, Mind and World dialogically: Interactional And Contextual Theories Of Human Sense-Making. Charlotte, NC: Information Age Publishing.

Love, Nigel. (2017). On Languaging and Languages. *Language Sciences* 61:113-47. doi: 10.1016/j.langsci.2017.04.001.

MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mondada, L. (2008). Contributions de la linguistique interactionnelle. Congrès Mondial de Linguistique Française, 073. https://doi.org/10.1051/cmlf08348

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Sociolinguistics*. Volume 20, issue 3: 336-366.

Mondada, L. (2019). Contemporary issues in conversation analysis: Embodiment and materiality, multimodality and multisensoriality in social interaction. *Journal of Pragmatics*, 145, 47–62. https://doi.org/10.1016/j.pragma.2019.01.016

Morgenstern, A., Caët, S., Debras, C., Beaupoil-Hourdel, P, and Le Mené, M. (2021). Children's socialization to multi-party interactive practices: Who talks to whom about what in family dinners. In Letizia Caronia (ed.) *Language and Social Interaction at Home and in School*. Amsterdam: John Benjamins, 46-85.

Ochs, Elinor (2012). "Experiencing language". *Anthropological Theory*, 12 (2): 142-160.

Parisse, C., Etienne, C. and Liégeois, L. (2020). "TEICORPO: A Conversion Tool for Spoken Language Transcription with a Pivot File in TEI", Journal of the Text Encoding Initiative [Online], Issue 13 | May 2020, DOI: https://doi.org/10.4000/jtei.3464

Vincent, C. (2020). Annotating Multimodal data of Singing and Speaking. In Frank A. Russo, Beatriz Ilari, Annabel J. Cohen (eds). *The Routledge Companion to Interdisciplinary Studies in Singing*, Vol I: Development, Routledge, https://doi.org/10.4324/9781315163734