Proceedings of the 10th Workshop on the Representation and Processing of Sign Languages (sign-lang@LREC 2022), pages 42–47
Language Resources and Evaluation Conference (LREC 2022), Marseille, 20-25 June 2022

Applying the Transcription System Typannot to Mouth Gestures

Claire Danet[1], Chloé Thomas[2], Adrien Contesse[1], Morgane Rébulard[1], Claudia S. Bianchini[3], Léa Chevrefils[2], Patrick Doan[1]
[1] Ecole d'Art et Design d'Amiens, 40 rue des Teinturiers, 80080 Amiens, France
{claire.danet, morgane.rebulard, adriencontesse, pdoan.atelier} @gmail.com
[2] Dynamique du langage in situ, Université Rouen-Normandie, 17 rue Lavoisier, 76130 Mont-Saint-Aignan, France
{thomaschloe2103, leachevrefils}@gmail.com
[3] Laboratoire Formes et Représentations en Linguistique, Littérature et dans les arts de l'Image et de la Scène, Université de Poitiers, 5 rue Lefebvre, 86073 Poitiers, France
claudia.savina.bianchini@univ-poitiers.fr

Abstract
Research on sign languages (SLs) requires dedicated, efficient and comprehensive transcription systems to analyze and compare the sign parameters; at present, many transcription systems focus on manual parameters, relegating the non-manual component to a lesser role. This article presents Typannot, a formal transcription system, and in particular its application to mouth gestures: 1) first, exposing its kinesiological approach, i.e. an intrinsic articulatory description anchored in the body; 2) then, showing its conception to integrate linguistic, graphic and technical aspects within a typeface; 3) finally, presenting its application to a corpus in French Sign Language (LSF) recorded with motion capture.

Keywords: Typannot, transcription system, mouth gestures

1. Introduction

Typannot is a transcription system designed to annotate every signed language (SLs), which takes into account all the SLs components, i.e. the manual parameters (Handshape, Initial location of the upper limb, and Movement) as well as the non-manual parameters (Mouth Action, Eye Action, Head, and Bust). It stands out from other current transcription systems, like HamNoSys (Hanke, 2004) or SignWriting (Bianchini, 2012), by adopting a descriptive model based on the articulatory possibilities of the body rather than the visuo-spatial characteristics of SLs gestures. This novel descriptive perspective is essential if we want to study the role of the body in the structuration of SLs next to the observations allowed by existing transcription systems.

2. State of the Art: Mouth Gesture

In SLs literature, the role of mouth has been the most studied among facial expressions: indeed, the lower part of the face plays one of the most important functions. Studies have reported a distinction between mouth movements, mouthings and mouth gestures. The mouthings would be the result of an oral education and/or a situation of contact with the hearing community and are labializations which resemble the surrounding vocal languages (Crasborn, 2006); moreover, mouthings generally tend to reproduce the most relevant phonetic part of a lemma of the spoken language. Conversely, mouth gestures are mouth movements specific to SLs (Crasborn *et al.*, 2008; Woll, 2001). It is generally recognized that the mouth assumes different roles, ranging from lexical to morphemic function (adjectival or adverbial). An example of the lexical role is given by the minimal pair [TOO BAD] (facial expression: frowned eyebrows, lips corners down) and [WIN] (facial expressions: wide eyes, eyebrows up, lips corners up) in French Sign Language (LSF), where both signs are textbook homonyms that are partially disambiguated by mouth gesture. In many signs, the mouth plays an important part and may be the only parameter in action, such as to express boredom in a story, i.e. the addition of puffy cheeks with outward airflow without any hand signs (Boyes Braem and Sutton-Spence, 2001).

These different studies show the importance of mouth movements on SLs research. To study these various movements within the corpus, it is necessary to have a complete and efficient transcription system. To date, there are already systems for annotating mouth movements, such as HamNoSys (Hanke, 2004), Vogt-Svendsen notation (2001) or Hohenberger and Happ notation (2001). The typographic system Typannot offers a complementary point of view based on the body articulatory possibilities to describe and note the movements made by the mouth, regardless of its function (mouth movements, mouthings or mouth gestures): in this paper we will focus on mouth gestures.

3. The Transcription System Typannot

The parameters for the description and study of SLs have gradually been established based on the work of Stokoe (1960). They include: the shape of the hand, its position and orientation, movement, and facial expression. Together they allow the description of language structure at a sub-lexical level. This categorization is found in the different types of representation systems, whether phonological (i.e., Stokoe) or phonetic (i.e., HamNoSys). In both cases, the transcription systems mainly rely on a **visuo-spatial** conception of these parameters. Indeed, those categorizations refer to an observation of gestural phenomena from a visual and spatial perspective: the hand has a **shape** and is in one **place**, is oriented in one **direction** and will follow a **trajectory**, the face has an **expression**. This mode of representation shows the gestures from an external point of view (visible) without seeking to precisely explain the bodily organization which partially underlies the forms / locations / orientations / trajectories / expressions (however, HamNoSys uses articulatory principles to represent manual shapes). Without contesting

the strengths and merits of this approach, the fact is that currently it is not possible to systematically inform the way in which these forms are produced at a bodily level and consequently the role of the body in the language structure cannot be questioned. In view of the many works on embodied cognition (Varela *et al.*, 1991), the postulate is that the body is at least the vector of these forms, and at most the environment in which they **occur**, **articulate**, and **transform**. Being able to characterize SLs through a **specific body description model** would allow researchers to distinguish two levels of structuration that appear intrinsically linked: 1) a bodily level describing the way in which the articulatory possibilities of the body are dynamically organized; 2) a linguistic level describing how these bodily organizations can form meaningful structures within the language.

3.1 A Description Rooted in the Body: the Kinesiological Approach

The objective of Typannot is to propose a **phonetic** transcription system based on a **body articulatory** model. To do this, it follows a **kinesiological** perspective (Boutet, 2018; Chevrefils *et al.*, 2021), which makes it possible to understand the principles and mechanisms of movement at an anatomical and biomechanical level. The system adopts two registers of information referring to: 1) the articulatory structure; and 2) the mode of activation of the latter. The register of the articulatory structure is divided into three parameters: hand (Doan *et al.*, 2019), upper limbs (Bianchini *et al.*, 2018), and face; each of them has distinct parts (e.g., arms, forearms, hands) that can be arranged according to different degrees of freedom of their own (e.g., the upper limbs have seven degrees of freedom). The second register makes it possible to describe a specific body organization to which activation principles are associated (e.g., impulse, tension, amplitude). Together, these two registers allow investigating the dynamics of transformation of the gesture and questioning the processes of constitution and modulation of its meaning.

3.2 Appropriating a Bodily Perspective: a Grapholinguistic Reflection

At a grapholinguistic level, the design of Typannot, taken as a typographic transcription tool, poses several challenges related to the model and its use. Thanks to the involvement of typographic and UX/ UI[1] designers, it is possible to question how to typographically implement this model and help users to appropriate it. Once finalized, Typannot shall consist of a family of characters and an input interface covering the information registers. While existing transcription systems have traditionally to choose between a linear representation (linked to the decomposition and queryability of data) and a readable graphical synthesis of the sign (as in the case of SignWriting), Typannot is a system capable of combining the two (Fig. 1)

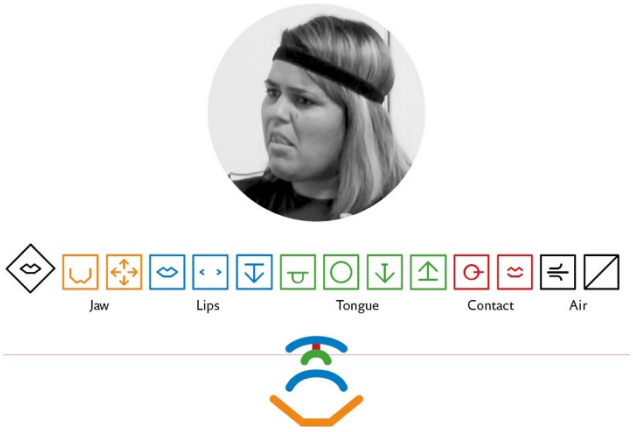

Figure 1: Typannot: description of a mouth gesture articulation in generic (middle) and in "composed glyph" (bottom).

Indeed, by exploiting the automatic ligature functionalities allowed by the OpenType[2] technology, the transcriptions in Typannot can be displayed either in a so-called "generic glyph", where the description has the form of a sequence of queryable characters, or in a so-called "composed glyph", which displays a thumbnail representing the articulatory subsystem (hands, upper limbs, face) in a simplified and visually explicit form. The purpose of this functionality is to be able to change the "focal point" of the observation, according to the needs and the context of use, without losing information. To succeed in producing the very large quantity of thumbnails corresponding to the possible combinations, a program to generate them automatically was created (Fig. 2).

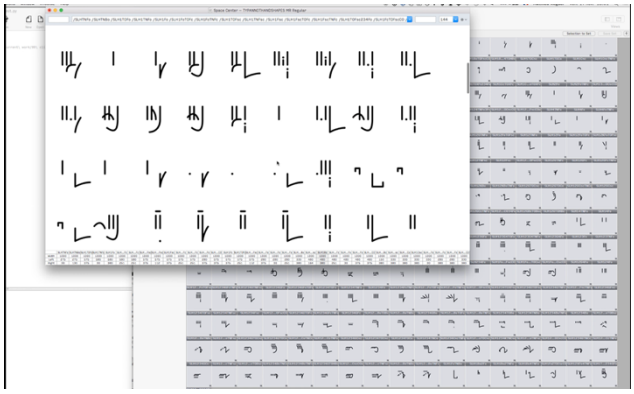

Figure 2: Automatic generation of "composed glyphs" for handshapes.

Alongside the actual typographical issues, a UX/UI approach favoring the assimilation of the model has also been developed for the input interface, named Typannot Keyboard. Indeed, by crossing the interactions allowed by the interface and visual feedbacks, the user can intuitively understand to what a variable corresponds (Fig. 3). This digital interface offering several input devices to adapt to a wide spectrum of transcription approaches and to allow

[1] UX/UI : User eXperience Design & User Interface.

[2] OpenType is a vectorial font format that allows encoding any character associated with Unicode, regardless of the platform (Mac, Windows, Android, etc.); OpenType fonts can have advanced typographic features that handle complex scripts and typographic effects like ligatures.

easy integration of the articulatory principles offered by the system; moreover, it allows easy access to composed glyphs without having to know the composition logic beforehand.
This interface is in progress and does not yet include mouth gestures. Despite this, the mouth gesture typeface with generic glyphs (see section 5) already exists and it is possible to use it on any software supporting OpenType (e.g. Word, Excel, ELAN, etc.).

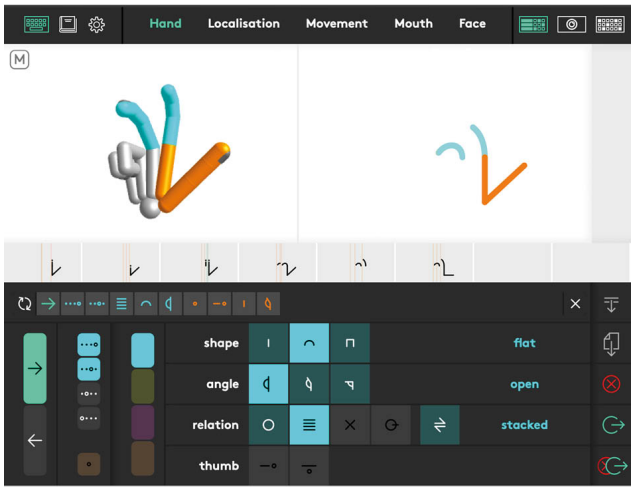

Figure 3: Typannot Keyboard for handshapes.

For a detailed description of the OpenType functionalities and of Typannot Keyboard, cfr. Danet *et al.* (2021) and Chevrefils *et al.* (2021).

4. The Case of the Mouth Gesture

In the Typannot system, mouth gestures are part of one of the two sub-parts of the face, i.e. Facial Action, which includes 1) Eye Action which concerns the upper part of the face and 2) Mouth Action (MA) which takes into account all the possibilities of the lower part and which corresponds to mouth gestures. In order to determine the description characteristics of MA, the work started with the observation of the existing formal descriptions in the literature (Boyes-Braem and Sutton-Spence, 2011), to understand the issues and the specific needs necessary for the realization of the mouth gestures in SLs (what is perceived by the interlocutor). Then, descriptions were reconsidered according to the principle of the articulatory approach and according to specific criteria of transcription (viewable, transferable, and searchable textual data) and design (genericity, readability, modularity, and hand-writability). This method allows a consistent coordination and unification of the typographic and graphic system for the different body parts (Danet *et al.*, 2021).

4.1 Articulatory Description of Features

To do so, gesture is deconstructed into discrete elements that can be divided into four levels of information:

- level 1. the articulatory parameter that the transcription refers to;
- level 2. the different parts that compose the parameter;
- level 3. the different variables associated with each parts;
- level 4. the values assigned to those variables.

Each of those levels have a limited set of characteristics that defines them like individual bricks of information.
After several iterations and thinking to optimize and organize these bricks, the XYZ axes were taken as the common referent: these allow to imagine the MA (parameter) composed of different face elements (parts) as fixed elements, having activable zones (variables) that carry transformations on these 3 axes (values).

PARTS

- ***Jaw, Lips*[3] *(i.e. corners, tubercles and vermilion borders), Tongue, Air***

VARIABLES and **values**

- ***CONVERGENCE*:** indicates the approximation, one towards the other, of the two elements constituting the part in question (e.g., Lips Convergence = coming together of the lips)
- ***DIVERGENCE*:** indicates the moving away, one opposite the other, of the two elements constituting the part in question (e.g., Jaw Divergence = opening of the mouth)
- ***CONTACT*:** Alveolus, Dental arc, Vermilion, Corner, Cheek
- ***SELECTION*:** Upper, Lower, Both Vermilion(s); Left, Right, Both Corner(s)
- ***POSITION*:** Up, Down, Down+, Left, Right, Fore, Fore+, Back
- ***SHAPE*:** Flat, Round, Tip, Blade
- ***CHANNEL*:** Outward, Inward
- ***STREAM*:** Obstructed, Restricted

Thereafter, it is necessary to order these elements in a robust syntax. The descriptive order was motivated by the logic of transformation and by the frequency of the activated elements. In this way, the Jaw comes first because it directly influences the openness of the lips. The lips may appear to diverge from each other when in reality they inherit the position of the jaw, they have not been activated and therefore remain in a "neutral" state. Thus, the graphematic formula of MA takes into account all the bricks, their levels of description and the logic of the transformation.

4.2 The Double Graphic Representation

Within the Typannot system, there are different graphic representations: generic glyphs and composed glyphs. The generic glyphs allow a detailed representation of each position of the articulators of the face, i.e. for the mouth

[3] The term Lips incorporates corners, tubercles and vermilion borders (the last two parts being refereed together as "vermilion").

gestures: Jaw, Lips, Tongue, Air; conversely, composed glyphs are an arrangement of the articulators position.

4.2.1 Generic Glyphs

Once defined, those characteristics form the generic components of the Typannot transcription system called generic glyphs. Graphic symbols can be assigned to them and later encoded into a font to perform like letters (Fig. 4).

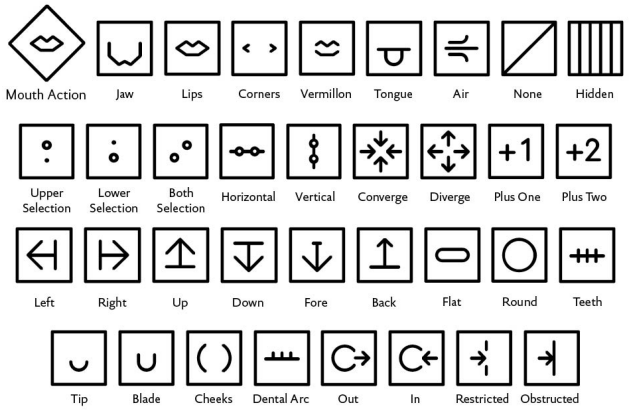

Figure 4: Table of generic glyphs for Mouth Action.

With these few generic components, it is thus possible to generate an infinite number of mouth gesture combinations (Fig. 5). The systematic organization of information into four levels supplemented by a syntax makes it possible to produce a manipulable and queryable transcription. Finally, thanks to the principle of genericity, Typannot allows annotators to query and compare data through different levels of analysis, from a single attribute to a combination of features.

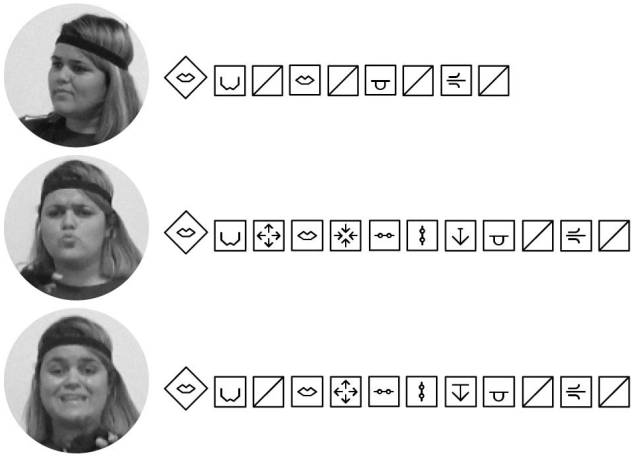

Figure 5: Mouth Action examples, with pictures of the face and the corresponding generic glyphs, according to the established syntax.

4.2.2 Composed Glyphs

The decomposition into generic glyphs allows the generation of a multitude of mouth gestures combinations and the technical capabilities to analyze them. As they are arranged linearly (Fig. 5), the reader/transcriber must make an effort to visuo-spatially reconstruct the mouth gestures, to see them as units.

This is why it is important to propose the second graphic form of the Typannot system, which allows to "read" intuitively and quickly what is transcribed. This consists of producing a logographic representation (composed, unique and recognizable), which depicts the desired mouth gesture while retaining all the information bricks. Recent advances in font encoding technologies (e.g., OpenType properties) and typographical features (e.g., contextual ligatures) allow designing a system that gives users the ability to transparently display one shape or the other while maintaining data integrity.

However, due to a large number of strokes to be graphically represented in a small space, typographic choices were necessary. Typannot uses the modular design approach to be able to compose mouth gestures: each characteristic is symbolized by a graphic module, which can vary according to its transformation on the face and on the neighboring characteristics (Fig. 6). These modules are organized and transformed by respecting a grid and rules of composition. A specific graphical formula has been defined, which translates the generic element of information into a unified and visually explicit glyph (Fig. 7).

LIPS	Jaw Neutral			
	Horizontal / Vertical	Converge	Neutral	Diverge
Neutral	Converge			
	Neutre			
	Diverge			
Fore	Neutre			
	Diverge			
Back	Neutre			

Figure 6: Modules and "composed glyphs" variations for Mouth Action.

Thanks to this modular framework and a scriptable font design environment (i.e., RoboFont[4]), it is possible to automate the modules composition in order to generate all the composed forms that users need.

[4] RoboFont is a software for typeface creation that can automatically generate contextual ligatures from graphic modules and layout instructions. For the development of Typannot, Frederik Berlaen, creator of RoboFont (https://robofont.com), has kindly provided GestualScript with a license to use his software.

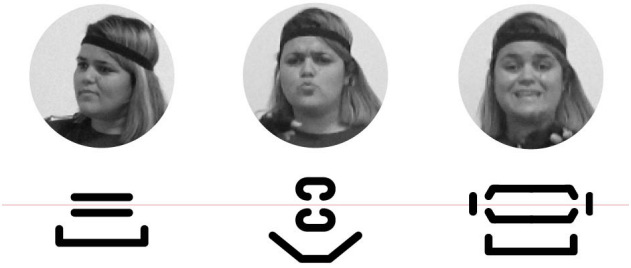

Figure 7: Mouth Action examples, with pictures of the face and the corresponding composed glyphs.

5. Corpus Application

Thomas' thesis (in progress) identifies non-manual patterns within interrogative utterances in LSF. A corpus has been recorded, using different means such as Motion Capture (MoCap) through hardware solutions like *Perception Neuron*® and software solutions like *OpenFace*®, as well as three 4K and HD cameras. It originates from 6 deaf signers, 18 to 25 years-old, using LSF as their main language, and has a total duration of 1h43[5]. In order to conduct the transcription of this corpus, Thomas uses the Facial Action transcription system of Typannot in its entirety (MouthAction and EyeAction).

To transcribe this corpus, Thomas is using the multimodal software ELAN, developed by the Max Planck Institute for Psycholinguistics of Nijmegen, in the Netherland[6]. Two kinds of transcription have been used in this corpus: in French glosses for the translation and with the Typannot generic glyphs for the facial expressions. The use of generic glyphs implies having to annotate each articulator separately (lips, jaw, tongue, eyebrows, etc.). Indeed, when extracting data from a spreadsheet, each value of the articulators must be requested individually. If this had been transcribed in the same line, it would not be possible to know if, for example, the requested value "down" relates to the lips or the jaw; this problem arises from the system economicity, the same generic glyph being used for different articulators: composed glyphs solve this issue and simplify the annotation scheme.

During the transcription of this corpus, the transcription system had the advantage of being able to be implemented in the ELAN software as well as in a spreadsheet as a font, which allows to make numerous enquiries and analyses. The system – easy to learn and use[7] – is readable, thus allowing to quickly know what is annotated; moreover, it has the advantage of being useful to transcribe the different elements of the LSF.

A segment of annotated corpus follows (Fig. 8):

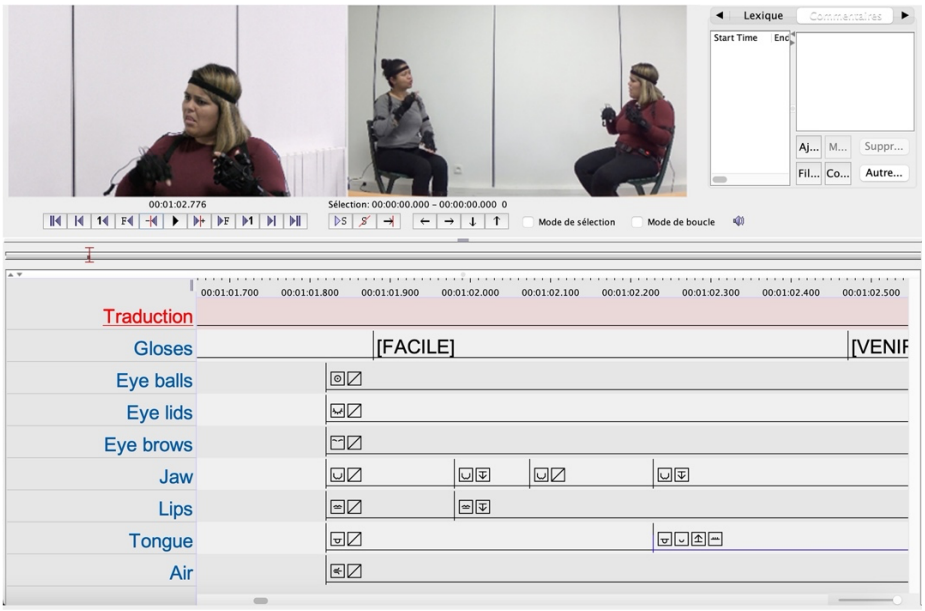

Figure 8: Frames from the visual corpus of gestures and a segment of its annotation with "generic glyphs" on ELAN.

6. Outlook

Typannot has a fine and precise grid for all the face articulators allowing to link the transcriptions to various data captured by MoCap systems such as the *OpenFace*® software. One of the objectives of Thomas' thesis (in progress) is to define thresholds between Typannot values and MoCap data. For example, for all the "lips diverge" within the corpus, limens will have to be issued on the 3 axes. Establishing these limens has the advantage of allowing (semi)-automatic transcriptions, resulting in reduced time in respect of manual transcriptions and/or the

[5] The corpus is composed of two types of elicited dialogues: 1) obtained by asking speakers to talk about the issue of accessibility (to culture, transport, health, etc.); 2) based on a questions&answers game (WH-questions and polar questions).

[6] https://www.mpi.nl/corpus/html/ELAN_ug/index.html

[7] Since 2019, students of the License "Sciences du Langage et Langue des Signes Française" at the University of Poitiers follow a 4-hour course explaining the basic principles of Typannot and the practical use of Typannot Handshape: they are then able to use the system in their linguistic analyses.

possibility of creating larger corpus to analyze. This limen principle has been experimented for the recognition of Typannot Handshape with *Leap Motion Controller*®[8] technology. However, for mouth gestures, the gestural phenomena to be recognized are limited to a smaller surface and therefore require a finer approach (Brumm and Grigat, 2020).

7. Acknowledgements

The project is funded by Ecole Supérieure d'Art et de Design (ESAD) de Amiens, the Hauts-de-France Region and the DGCA of the French Ministry of Culture.
For the corpus: Joëlle Mallet.

8. Bibliographical References

Bianchini, C.S. (2012). *Analyse métalinguistique de l'émergence d'un système d'écriture des Langues des Signes : SignWriting et son application à la Langue des Signes Italienne (LIS)*. Thèse de doctorat, Université de Paris 8 & Università degli Studi di Perugia. https://doi.org/10.13140/RG.2.1.3817.4563

Bianchini, C.S., Chevrefils, L., Danet, C., Doan, P., Rébulard, M, Contesse, A. and Dauphin, J.F. (2018). Coding movement in Sign Languages: the Typannot approach. In *Proceedings of the Fifth International Conference on Movement and Computing (MoCo'18), sect. 1(9), ACM*, pp. 1-8. https://doi.org/10.1145/3212721.3212808

Boutet, D. (2018). *Pour une approche kinésiologique de la gestualité*. Habilitation à diriger des recherches, Université de Rouen-Normandie.

Boyes-Braem, P. and Sutton-Spence, R. (Eds) (2001). *The hands are the head of the mouth: the mouth as articulator in sign language*. Hamburg: Signum Press.

Brumm, M., and Grigat, R.R. (2020). Optimised preprocessing for automatic mouth gesture classification. In Proceedings of the nineth Workshop on the Representation and Processing of Sign Languages *(LREC 2020)*, ELRA, pp. 27-32.

Chevrefils, L., Danet, C., Doan, P., Thomas, C., Rébulard, M., Contesse, A., Dauphin, J.-F., and Bianchini, C.S., (2021). The body between meaning and form: kinesiological analysis and typographical representation of movement in Sign Languages. Languages and Modalities, 1:49-63. https://doi.org/10.3897/lamo.1.68149

Crasborn, O.A. (2006). Nonmanual structures in sign language. In K. Brown (Ed.), *Encyclopedia of language & linguistics 8*. Oxford: Elsevier, pp. 668-672.

Crasborn, O.A., van der Kooij, E., Waters, D., Woll, B. and Mesch, J. (2008). Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language & Linguistics*, 11(1):45-67. https://doi.org/10.1075/sll.11.1.04cra

Danet, C., Boutet, D., Doan, P., Bianchini, C.S., Contesse, A., Chevrefils, L., Rébulard, M., Thomas, C. and Dauphin, J.-F. (2021). Transcribing sign languages with Typannot: a typographic system which retains and displays layers of information. *Grapholinguistics and its Applications*, 5(2):1009-1037. https://doi.org/10.36824/020-graf-dane

Doan, P., Boutet, D., Bianchini, C.S., Danet, C., Rébulard, M., Dauphin, J.F., Chevrefils, L., Thomas, C. and Réguer, M. (2019). Handling sign language handshapes annotation with the Typannot typefont. *CogniTextes (AFLiCo)*, 19:1-24. https://doi.org/10.4000/cognitextes.1401

Hanke, T. (2004). HamNoSys: representing sign language data in language resources and language processing contexts. *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, ELRA, pp. 1-6.

Hohenberger, A. and Happ, D. (2001). The linguistic primacy of signs and mouth gestures over mouthing: evidence from language production in German Sign Language (DGS). In P. Boyes-Braem, R. Sutton-Spence (Eds), *The hands are the head of the mouth: the mouth as articulator in sign language*. Hamburg: Signum Press, pp. 153-189.

Stokoe, W. (1960). Sign Language structure, an outline of the visual communications systems of American deaf. *Studies in Linguistics Occasional Paper*, 8. Reprinted as 2005. *Journal of Deaf Studies and Deaf Education*, 10(1):3-37. https://doi.org/10.1093/deafed/eni001

Thomas, C. (in progress). *Étude des paramètres non-manuels en LSF au sein d'énoncés interrogatifs : entre transcriptions manuelles et capture de mouvement*. Thèse de doctorat, Université de Rouen-Normandie.

Varela, F.J., Thompson, E. and Rosch, E. (1991). *The embodied mind: cognitive science and human experience*. Cambridge MA: MIT Press.

Vogt-Svendsen, M. (2001). A comparison of mouth gestures and mouthings in Norwegian Sign Language (NSL). In P. Boyes-Braem, R. Sutton-Spence (Eds), *The hands are the head of the mouth: the mouth as articulator in sign language*. Hamburg: Signum Press, pp. 9-40.

Woll, B. (2001). The sign that dares to speak its name: echo phonology in British Sign Language (BSL). In P. Boyes-Braem, R. Sutton-Spence (Eds), *The hands are the head of the mouth: the mouth as articulator in sign language*. Hamburg: Signum Press, pp. 87-98.

[8] Leap Motion Controller is an optical hand-tracking module that captures the hand movements of your hands with unparalleled accuracy.