UNIVERSITÀ
degli STUDI
di CATANIA

DIPARTIMENTO DI SCIENZE UMANISTICHE

DOTTORATO DI RICERCA IN SCIENZE DELL'INTERPRETAZIONE

XXXVI CICLO

Gaia Caligiore

CODIFYING THE BODY: EXPLORING THE COGNITIVE AND SOCIO-SEMIOTIC FRAMEWORK IN BUILDING A MULTIMODAL ITALIAN SIGN LANGUAGE (LIS) DATASET

TESI DI DOTTORATO

Coordinatore:
Chiar.mo Prof. Antonino Sichera

Tutor:
Chiar.ma Prof.ssa Sabina Fontana

Anno Accademico 2022/2023

Abstract

Codifying the body and describing its movements during the act of signing is a task that has long been discussed in the field of sign language studies, with enquiries on how to capture and label signed discourse being established since the very beginning of contemporary sign language research (Stokoe, 1960). Starting from this premise, and laying its ground on an interdisciplinary perspective, the present work aims to present a multimodal dataset in Italian Sign Language (LIS), describing a process of data collection and computational annotation. Traditionally, the field of sign language research has been characterized by a dichotomy between perspectives of contrast and continuity within features and attributes associated to vocal languages. In this regard, this thesis proposes a challenging middle-ground viewpoint, integrating characteristic insights from formalist and functionalist approaches in the study of sign languages. The goal is to draw aspects and methodologies from these perspectives that can be advantageously integrated to promote an interdisciplinary, current, and accessible collection and description of LIS.

Positioning itself theoretically within the cognitive and socio-semiotic framework of sign language description (Volterra *et al.,* 2022), this work aims to suggest and apply guidelines and best practices for the collection and computational annotation of video data in LIS by taking into account other corpora collections and annotation experiences. It ponders the peculiarities of LIS, such as its multimodal and multilinear nature, while also considering the daily use made of it by the Italian signing community. In this perspective, language cannot be separated from its sociocultural and historical context (Russo Cardona, 2004b), nor from its embodied nature (Lakoff & Johnson, 1980), as it is through the body that we humans live and describe our experiences. This assertion becomes even more important when the object of study is a sign language mediated through the visual-gestural channel, where the body itself constructs meaning.

Building on these assumptions, this thesis introduces a video collection method for LIS. This method is grounded in the use of multimodal and synchronized RADAR sensor and camera video capturing tools, allowing for detailed information gathering on all body parts involved in the act of signing, whether referring to manual or body elements. The subsequent annotations, developed using the ELAN software, occur on different levels, employing vocal labels in Italian and English, along with a specific annotation system for sign languages: Typannot (Bianchini, 2023). The interdisciplinary nature of the project is also reflected in the accessibility of annotations, enabling various users to access the information contained in the recorded videos, regardless of familiarity with LIS.

This thesis thus demonstrates that the suggested system of data collection, employing RADAR sensor and camera video capturing tools, alongside integrated multimodal

corpora annotation techniques, accurately reflects the intricate nature of a signed language. This approach extends beyond the current state of the art by offering a pathway for the advancement of sign language recognition systems and more effective automatic translation tools.

Codificare il corpo e descrivere i suoi movimenti durante l'atto di segnare è un tema ampiamente discusso nel campo degli studi sulle lingue dei segni. Infatti, fin dagli albori della ricerca contemporanea sulle lingue dei segni (Stokoe, 1960), sono stati intrapresi studi mirati a comprendere come catturare e categorizzare in modo efficace il discorso segnato. Il presente lavoro, a partire da questa premessa e applicando una prospettiva interdisciplinare, ha come obiettivo la presentazione di un dataset multimodale in Lingua dei Segni Italiana (LIS), descrivendo i processi di raccolta e annotazione computazionale dei dati. Nella ricerca sulle lingue dei segni, la dicotomia tra prospettive di contrasto e continuità rispetto alle lingue vocali ha storicamente caratterizzato il campo. In quest'ottica, la tesi propone un approccio sfidante e mediano, integrando punti caratteristici di prospettive formaliste e funzionaliste nello studio delle lingue dei segni. L'obiettivo è trarre da queste prospettive aspetti e approcci che possano essere vantaggiosamente integrati per favorire una raccolta e descrizione interdisciplinare, attuale e accessibile della LIS.

Posizionandosi teoricamente nel quadro cognitivo e socio-semiotico della descrizione delle lingue dei segni (Volterra et al., 2022), questo lavoro mira a proporre linee guida e buone pratiche per la raccolta e annotazione computazionale di materiali video in LIS, tenendo conto delle peculiarità della lingua stessa, come la sua multimodalità e multilinearità, e considerando anche l'uso quotidiano che la comunità segnante italiana fa di essa. In questa prospettiva, la lingua non può essere separata dal suo contesto socioculturale e storico (Russo Cardona, 2004b), né dalla sua natura embodied (Lakoff & Johnson, 1980), poiché è attraverso il corpo che noi esseri umani viviamo e descriviamo le nostre esperienze. Tale affermazione assume ulteriore importanza quando l'oggetto di studio è una lingua mediata dal canale visivo-gestuale, in cui è il corpo stesso a costruire significato.

A partire da queste premesse, questa tesi introduce un metodo di raccolta video per la LIS. Il metodo presentato si basa sull'uso di strumenti di acquisizione video multimodali e sincronizzati, come sensori RADAR e videocamere, consentendo una raccolta dettagliata di informazioni su tutte le parti del corpo coinvolte nell'atto di segnare, che si tratti di elementi manuali o corporei La successiva annotazione, sviluppata utilizzando il software ELAN, avviene su diversi livelli, adottando etichette vocali in italiano e inglese, insieme a un sistema di annotazione specifico per le lingue dei segni: Typannot (Bianchini, 2023). La natura interdisciplinare del progetto si riflette inoltre nell'accessibilità delle annotazioni, consentendo a diversi utenti di accedere alle informazioni contenute nei video in LIS, indipendentemente dalla familiarità con la LIS.

Questa tesi dimostra dunque che il sistema proposto di raccolta dati, utilizzando sensori RADAR e strumenti di acquisizione video delle telecamere, insieme a tecniche integrate di annotazione di corpora multimodali, è in grado di riflettere con precisione la complessa natura di una lingua dei segni. Questo approccio va oltre lo stato attuale dell'arte offrendo un percorso per lo sviluppo di sistemi di riconoscimento delle lingue dei segni e per la realizzazione di strumenti di traduzione automatica più efficaci.

Index

Explanation of Capitalization Choices and Terminology

In this study, certain instances of the word *sign* and *deaf* are capitalized. The lowercase *deaf* denotes individuals with varying degrees of hearing loss. Given the diversity in backgrounds, identities, and ages of sign language acquisition among individuals, the decision was made to capitalize *Deaf* when referring to the collective group sharing language, identity, and culture. A group that primarily uses sign language for communication (or is proficient in signing) and identifies with a set of norms and beliefs that connect them both to the Deaf community and more broadly, to society. The timing of hearing loss, whether a Deaf person was born into a signing family, or when sign language was acquired, was not taken into consideration.

Similarly, the term *sign* may or may not be capitalized. Tommaso Russo Cardona's perspective (2004b) has been adopted in this regard. Therefore, the word *sign*, when used to refer to the main object of study in general linguistics, will be written in lower letters. On the other hand, when *Sign* is capitalized, it will be used to refer to the units that make up signed discourse, i.e., the joining of manual and body components that occur in a sign language and that have gone either through a process of standardization or present characteristics outlined by Volterra *et al*. (2022).

In this work, the terminology *vocal language labels* will be used instead of the more conventional term *gloss,* typically employed to define the fundamental meaning of a sign in vocal languages. This choice aligns with the proposal put forth by Antinoro Pizzuto *et al.* (2010), who argue that the terms *transcription* and *gloss* have often been employed inaccurately when referring to translations of the fundamental meaning of a Sign or sign language utterance. Consequently, the tradition of using caps lock to write glosses has been abandoned in favor of a translational approach, where the vocal language label simply stands as a written rendering of the general meaning of the Sign in a spoken language (either Italian or English). For this reason, in accordance with the strategies employed by Volterra *et al.* (2022), this work will attempt, whenever possible, to avoid the use of vocal language labels altogether, opting for the inclusion of videos or images to avoid the distortion of the sign language data presented.

To further reduce the sole reliance on vocal language labels, this thesis includes QR codes in the print version and hyperlinks in the digital edition. These are placed next to images or vocal language labels, allowing readers to access video representations of the Signs or utterances discussed. Whenever possible, examples are sourced from SpreadTheSign. If not available or too specific, a Deaf signer was consulted to perform the Sign or utterance.

Chapter 1
Mapping the landscape: relating contrast and continuity in sign language research

From the days of ancient Greece, sign languages have captured the attention of scholars and thinkers alike, going as far back as Plato's *Cratylus*, where he asked the question: «If we had no faculty of speech, how should we communicate with one another? Should we not use signs, like the deaf and dumb?[1]» (Plato, ca. 360 B.C.E./2018, p. 20). It was through this question that Plato not only acknowledged the existence of gestural codes employed by deaf individuals in ancient Greece, but also displayed ad acknowledgment of the inherent communicative force residing within these codes.

Historical records provide glimpses of standardized gestural codes that endured through time, representing the existence of communicative systems rooted in a visual-gestural modality. The efforts of historical luminaries spanning from the 16th Century onwards— such as Pedro Ponce de León, Charles-Michel de L'Épée, Roch-Ambroise Cucurron Sicard, Laurent Clerc, Jean Massieu, and Thomas H. Gallaudet— have left an indelible mark. These figures not only embarked on the journey to describe the «[…] natural and spontaneous language of signs» (Gallaudet, 1848, p. 81) but also contributed to the education of the deaf youth. Yet, it was only much later that a shift occurred, with the publication *Sign Language Structure* by William Stokoe (1960). Stokoe's work revealed American Sign Language (ASL) as a language in its own right, a strong departure from the predominant notion that sign languages were more similar to pantomime and theatrical gestuality (Volterra *et al*., 2022).

---

[1] This passage is taken from Benjamin Jowett's 1892 translation of a Greek text, as featured in its 2018 edition. In this specific section, the terms 'deaf and dumb' are used solely for the purpose of adhering to the language found in that historical translation. It is crucial, however, to emphasize that this terminology is now considered outdated and carries negative connotations. It is well-recognized that language evolves over time, reflecting changing perspectives and values of societies. Therefore, this note is included to acknowledge the inaccuracy of the translation by modern standards and to advocate for the use of more respectful and inclusive language. It's important to recognize that terminology like 'deaf and dumb' can stem from certain historical ideologies and may have long-lasting social and cultural impacts.

Language[2] plays a central role in transforming a group of individuals into a cohesive community with shared culture and history. Through languages, people engage in the intricate process of describing both the world around them and their own identities using arbitrary systems of semiotic representation. Keller (1995) refers to languages as *phenomena of the third kind*: cumulative and unintended outputs of the communicative actions of humans.

In the progression of sign language studies, a clear pattern emerges, revealing increased awareness concerning the nature of these languages. This culminates, from a sociopolitical standpoint, with an ongoing streak of local sign language recognitions by nations worldwide[3]. Within this discourse, an ongoing stimulation that continues to fuel the interest of scholars towards sign languages can be witnessed, looking at them from different, complex and, sometimes, opposed perspectives, leading to a continuous reciprocal enrichment.

1.1. Connecting pathways: LIS research, recognition and societal dynamics in Italy

Over the course of more than half a century, significant changes have occurred in the field of sign language linguistics, reshaping the perception of sign languages from pantomimic gestural codes to universally acknowledged and legally recognized natural languages in numerous countries throughout the world.

The 1880 *International Congress of Education for the Deaf* (ICED) in Milan left a lasting impact that reverberated through the centuries, profoundly shaping not only sign

---

[2] In the present thesis, the term 'language' is used equivalently to the Italian translation *'lingua'* and is not meant to refer to the faculty of language, which corresponds to *'linguaggio'* in Italian. Instead, 'language' is conceptualized here as the tangible expression of communication, whether in vocal or signed form.
[3] https://wfdeaf.org/news/the-legal-recognition-of-national-sign-languages/ (Accessed on August 8th, 2023).

language studies, but also the broader perception of sign language within both Deaf and hearing communities. This congress produced eight resolutions, all unified under one the principle: «[t]he pure oral method ought to be preferred» (Gallaudet, 1881, p. 6). The repercussions of these resolutions had a profound effect on the trajectory of LIS by solidifying its development, recognition, and standardization in Italy.

The declaration '*Il gesto uccide la parola*' (Gesture kills speech) characterized a century-long heritage of oralism that was imposed on the deaf people of Italy. This approach suffocated an entire group of people, obstructing the progression of awareness regarding the linguistic nature of Italian Sign Language (henceforth LIS-*Lingua dei Segni Italiana*), as well as strongly influencing the communication methods used by deaf individuals. Furthermore, the aftermath of the Milan Convention obstructed the progress and acknowledgment of LIS as a legitimate language, impacting both sociopolitical and linguistic perspectives. This legacy preserved a negative perception of LIS and extended to other sign languages, leading researchers to position and define them in a state of continuity with vocal languages, to validate their linguistic status.

In the early 1980s, the exploration of signed communication in Italy began with the involvement of hearing and Deaf researchers. Initially the language now known as LIS lacked an official name and was often referred to as 'mime' or 'gesture' by both signers and individuals external to the signing community (Volterra *et al.,* 2022). This naming reflected in the Sign used to describe it (see Fontana *et al.,* 2015, p. 4).

The emergence of the first publications on this topic (Montanini *et al.,* 1979; Volterra, 1981) and the dedicated efforts of researchers like Elena Antinoro Pizzuto, along with productive collaborations between Deaf and hearing researchers, sparked what could be termed as a 'revolution' in the field of sign language studies in Italy (Fontana *et al.,* 2015, p. 9). Central to the evolution of these groundbreaking perspectives was the influence of factors *external* to language, including meta-linguistic awareness, social attitudes within the signing community, and the visibility of the community and its language to a wider audience. Antinoro Pizzuto was, in fact, the first Italian National Research Council (henceforth CNR, *Consiglio Nazionale delle Ricerche*) researcher to carry out studies that were entirely centered on the role played by iconicity in LIS. Her work paved the way for significant theories on describing sign language through iconicity (Antinoro Pizzuto *et al.,* 2006; Cuxac & Antinoro Pizzuto, 2010), exploring sociolinguistic variations (Volterra & Antinoro Pizzuto, 2002), and dealing with sign language video data annotation and processing (Antinoro Pizzuto *et al.,* 2008, Antinoro Pizzuto *et al.,* 2010).

The visibility of sign language in Italy, especially in media representation, has undergone significant changes due to technological advancements, aligning with global trends. Regarding content accessibility in Italy, the late 1980s witnessed the introduction of subtitles in movies on Italian television. However, the progress of subtitling on public television channels has been since inconsistent. In fact, according to Volterra *et al.* (2022), less than half of the programs on national Italian television channels are currently subtitled. The importance of media accessibility in terms of subtitle availability or LIS interpreting, for ensuring equitable access to critical information to deaf individuals, became especially evident during the Covid-19 pandemic in Italy. In fact, the mobilization and protests led by the Italian National Agency for the D/deaf (henceforth ENS- *Ente Nazionale Sordi*) at a national level encouraged governmental authorities, including the Prime Minister, to prioritize the translation of their press releases into LIS (Tomasuolo *et al.,* 2021). This recognition highlighted the crucial role of LIS in providing accessible information to the Deaf population during challenging times.

The pandemic also highlighted the challenges faced by D/deaf people when face-to-face communication became complex. This heightened visibility, together with the advocacy efforts of the signing community, contributed to the official recognition of

LIS and Tactile Italian Sign Language (LISt) by the Italian Government in May 2021[4]. Indeed, this historical milestone once again reflects the collective efforts of a community that has gradually gained awareness of its identity and status, acknowledging LIS as a language deserving proper representation and recognition.

The delayed but significant official acknowledgment of LIS in Italy has played a central role in challenging conventional perceptions of language for the wider public. This acknowledgment, although occurring later than in some other contexts (Busatta, 2022), has contributed to widen the public understanding of languages as systems of communication expanding beyond auditory codes. The official recognition of LIS denotes a significant landmark to be located within an ongoing process of shifting perceptions, both within and beyond the Italian signing community, spanning several years. Additionally, the evolution of technology since the 1980s has drastically altered communication for sign language users. Multimedia platforms such as social networks, YouTube and video conferencing tools (e.g., Zoom, Google Meet), have revolutionized information exchange and communication with a specific impact on sign language users and, as a direct consequence, on their languages, which were influenced in terms of use and standardization (Fontana, 2022a). Alongside these advancements, the post-pandemic integration of sign language content on public Italian television channels has significantly amplified the visibility of signing. The combined efforts of advocacy by national signing communities, supported by organizations like ENS, alongside the widespread use of social media and technology, have accelerated discussions surrounding the standardization, dissemination and documentation of LIS. This progress reflects the growth of extensive social and academic efforts (Fontana *et al.,* 2015) that have mutually shaped and propelled each other's advancement.

---

[4] See: Article 34 ter. https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=2021-05-21&atto.codiceRedazionale=21A03181&elenco30giorni=false (Accessed on August 25th, 2023).

Considering these societal, cultural, technological, linguistic, and academic advancements collectively, it becomes evident that these processes are bound within a co-constructive framework of mutual influence. Within this holistic framework, both internal and external factors to the language and the community evolve hand in hand, on a parallel trajectory.

Within this evolving and multifaced framework, language collection is among one of the issues that arises. A fundamental pursuit within linguistics, language collection involves capturing and preserving the unique linguistic features and cultural nuances of a language. In the context of sign languages, comprehensive collections are still needed, and technology has proven essential in this task, allowing linguists to employ raw and annotated video recordings to discuss the qualities of sign languages, based on qualiquantitative evidence. Within this framework, computational processing and annotation are essential companions to language data collections, as they involve segmenting and labeling Signs within recorded data, laying the foundations for subsequent analyses. By using computational tools, linguists can process sign language data with information pertaining not only to manual components, but also related to body components (see Volterra *et al.,* 2022), facilitating systematic exploration and even comparisons across different sign languages. In summary, complementary dynamics such as the legal recognition of sign languages and the intensification in use of technology and multimedia tools, have led linguists focus on sign language collection, followed by possible computational processing of the data. As technology advances and interdisciplinary collaborations flourish, these efforts will not only enrich the understanding of sign languages but will also result in concrete and practical outputs that could foster greater inclusivity and communication accessibility for national and international signing communities.

In conclusion, the historical trajectory of LIS and the Italian signing community over the past forty years mirrors a broader narrative of interconnected dynamics where societal evolution, linguistic inclusivity and technological integration walk hand in hand with research. In this context, it is possible to put forward the hypothesis that the relationship between factors traditionally considered external to language and linguistic research becomes central, encouraging a reciprocal influence.

1.2. On continuity and contrast in the description of sign languages

Moving beyond sociocultural factors strictly related to historical and political events into the context of linguistic research, it is found that in the field of sign language linguistics, continuity and contrast are two complementary research strategies that have been used (and, to an extent, still are) to describe gesture, signed and vocal languages by means of models that, depending on the historical and methodological research framework, follow paths that highlight patterns of continuity and contrast.

1.2.1. Shaping contemporary linguistic theories: contrast between formalism and functionalism

In the prominent language theories of the 1960s, structuralism, as introduced by de Saussure (1916), held a central position. Ferdinand de Saussure is acknowledged as the father of modern studies in linguistics and the structuralist approach. This approach, in turn, plays a crucial role in informing functionalist approaches to linguistics, more specifically.

De Saussure described languages as systems comprising regulations, words, morphemes, and phonemes shared by societies (*langue*) and shaped by users in their daily speech (*parole*) (De Mauro, 2019). Discussing the Swiss linguist's hypotheses, De Mauro (1991) emphasizes that the signified (*signification*) aspect of a sign (*signe*) is, consequently, determined by the context in which it is situated, therefore the circumstances of *parole*. De Saussure stressed the importance of studying languages comprehensively. He viewed languages as the ever-evolving outcomes of sociocultural elements (*masse parlante*) and historical circumstances (*le temps*) (De Mauro, 1991). This perspective prompted an examination of language in synchronic, diachronic, and panchronic frameworks. Consequently, any relationship established between things and signs occurs within specific time and space, leading to a continual shift in meaning over time (also see Tullio De Mauro's introduction to his translation of de Saussure's Course, 2005).

This description establishes an approach to language studies that perceives language as a structured code, while also acknowledging it as a social action influenced by time. In fact, according to de Saussure (1916), language never exists outside of society, and its inherently social nature is a fundamental characteristic. In the USA, Saussurean structuralism drew considerable attention during the initial phase of Noam Chomsky's studies. However, Chomsky later distanced himself from direct reference to de Saussure (De Mauro, 2005) by adopting certain aspects, while opposing others. Whereas both scholars explore the relationship between language structure and its use, their description of human knowledge encompasses individual and social categories to varying degrees.

Chomsky's concepts of *competence* and *performance* (1965) differ from de Saussure's langue and parole in significant ways. While *langue* comprises a semiological and socially shared system, competence refers to an individual's knowledge of a language, devoid of social influences. Additionally, whereas de Saussure's *parole* is inseparable from *langue*, representing how speakers (or signers) apply their language knowledge, Chomsky's competence theory sees human language as a system allowing free expression of thought, largely independent of external control, need-satisfaction, or practical purposes. In doing so, Chomsky dismisses the notion that communication is a primary function of language, labeling it a vulgar and instrumental view, thereby relegating the act of using language (i.e., performance) to a residual category (van Valin, 2003).

In Syntactic *Structures* (2002b) Chomsky states that he considers language as «[…] a set of sentences, each finite in length and constructed out of a finite set of elements» (p. 13). From a linguistic perspective, within this approach, all languages (at least, spoken ones) possess a finite set of phonemes that combine to create sentences. Sentences will then constitute languages. This combination process is possible thanks to grammar, i.e. a fixed set of combinatory rules that help us identify *grammatical* and *ungrammatical* sentences. To explain these concepts, Chomsky presents the sentence '*colorless green ideas sleep furiously*', to explain that the grammaticality of a sentence does not depend upon its semantic meaning, or the credibility of certain word combinations, but rather on the syntactic correctness of a sequence of words. This sentence, in fact, «[…] though nonsensical, is grammatical» (Chomsky, 2002b, p. 16).

De Saussure and Chomsky's works have led linguistic research down two distinct paths. Chomsky's path envisions language as composed of a finite set of rules, free from socio-cultural and temporal influences. It seeks to define the structures of language, interpreting utterances as symbols within a larger equation (De Mauro, 1991). On the other hand, in the Saussurean structuralist framework, language is perceived as an outcome of collective theoretical and practical knowledge that manifests within a specific moment in time. This division within linguistic research is mirrored in sign language studies which exhibit a dual nature characterized by two complementary, yet somewhat contrasting, methodologies: *formalist* and *functionalist* approaches, two paradigms that significantly influence and guide the course of research within the field.

Formalist approaches in sign language studies apply the framework of generative research. Drawing from tools and ontologies originally developed for the formal analysis of spoken languages, formalist studies seek to contribute to the understanding of sign languages by establishing general patterns that facilitate interconnections between advancements in sign language research, and research conducted on vocal languages. More broadly, formalist directions in linguistic research emphasize internal systematicity, while often excluding consideration of external explanatory factors (McElvenny, 2019).

On the contrasting side, functionalist approaches within sign language studies establish distinct investigative frameworks, offering an alternative to formalist theories. These functionalist perspectives diverge from theories that emphasize a division between language structures and functions, disregarding the communicative constraints inherent in communication. More precisely, functionalist approaches in the present work are intended following Bates *et al.*'s (1991) definition of the framework, characterized by the belief that natural languages are created, used, and governed by communicative functions, prioritizing the functional aspects of language in its creation and usage. These approaches deliberately detach from existing preconceptions tied to vocal languages concerning sign languages:

> *Functionalism* can be defined as the belief that the forms of natural languages are created, governed, constrained, acquired and used in the service of communicative functions. So defined, functionalism is the natural alternative to theories of language that postulate a strict separation between structure and function, and/or theories that attempt to describe and explain structural facts *sui generis,* without reference to the constraints on form that are imposed by the goals of communication and the capabilities and limitations of human information processing.
> (Bates *et al.,* 1991, p. 134)

1.2.2. Continuity with structuralist paradigms in early sign language studies

In relation to the relevance of structuralist theories for language studies, de Saussure's theorizations found an audience in Charles F. Hockett, who outlined thirteen design features to delineate languages (1960). These features became a reference point in the early stages of sign language studies, coinciding with the emergence of Stokoe's article on American Sign Language (ASL) (1960).

Hockett's first design feature endorsed a new description of language, by reciting as follows: «The auditory-oral channel is perhaps the most obvious design feature. There are systems of communication that use other channels, for example, gesture […]» (1960, p. 90). Hockett's suggestion that a communication system could operate on an alternative channel such as the visual-manual one, opened the possibility of excluding the auditory-oral channel as a mandatory element in defining language. This proposition allowed sign languages to potentially meet several design features, aligning with the prevailing structuralist theoretical framework. Consequently, sign language researchers sought to establish that sign languages fulfilled additional parameters. In particular, two parameters emerged within this context: arbitrariness and duality of patterning (Volterra *et al.*, 2022).

In his pioneering studies on ASL conducted at Gallaudet University in the late 1950s, William Stokoe, initially an English scholar, proposed an unconventional theory asserting that ASL constituted a fully-fledged language. What was once perceived as a fragmented and imprecise amalgamation of gestures underwent a paradigm shift in perception. Stokoe (1960) demonstrated, through the application of structural linguistic paradigms, as highlighted by Volterra *et al.* (2022), that ASL Signs could be broken down into structures corresponding to the phonological level of vocal languages. He termed these structural units 'cheremes', which corresponded to hand shape, movement and spatial location, to which he later added palm orientation (Stokoe *et al.,* 1965). Stokoe himself defined cheremes as analogous to phonemes (1960, p. 69).

This theoretical framework found fertile ground in the Italian context, particularly through its reception by Virginia Volterra who, at the time, worked as a researcher at the CNR where she adapted Stokoe's parameters to LIS (Volterra, 1987). Her objective was to demonstrate that, like vocal languages, it was indeed possible to establish the existence of double articulation in LIS (Volterra *et al.,* 2022).

Another influential structuralist paradigm that influenced early LIS research involved attempts to demonstrate that, in continuity with vocal languages, signed languages presented degrees of arbitrariness. Ferdinand de Saussure's foundational principle, outlined in *Course in General Linguistics*, stated that: «[t]he bonds between the signifier and the signified is arbitrary […] the linguistic sign is arbitrary» (1916, p. 67). The Swiss linguist further added that «Onomatopoeia might be used to prove that the choice of the signifier is not always arbitrary. But onomatopoeic formations are *never* organic elements of a linguistic system» (*Ivi*, p. 69). These affirmations led early LIS researchers to highlight arbitrary traits. In fact, «[…] [i]n the preliminary stages of sign language research, it was important to emphasize the arbitrary elements and marginalize iconicity» (Volterra *et al.,* 2022) as it was perceived as a characteristic deriving from gestuality which, to some extent, was seen as somewhat of a lessening of the linguistic nature of sign languages. This perspective was evident in Stokoe's (1960, p. 13) definition of «[…] the language of signs in general use among the American deaf […]» as a language that is as arbitrary as any. Certainly, these efforts highlight the prevailing linguistic notions deeply rooted in the dogmatic beliefs that a language lacking the principles of arbitrariness and duality of patterning would fall short of meeting two fundamental criteria necessary for its classification as a 'true' language.

1.2.3. Rethinking the relationship between action, gestures and Signs in a framework of continuity

The initial studies on sign languages marked the emergence of a new field of research that extended beyond mere linguistic signs. It revealed pathways and patterns in human communication that encompassed not just language but also actions and gestures, opening new dimensions for exploration and understanding. Bates *et al.* (1975) introduced the ontogenetic centrality of gesture, affirming that the interplay between gestures and speech lays the foundation for language development. However, this interplay must take into consideration the interconnection among actions, gestures and language, within the framework of a theory of evolutional continuity.

Armstrong, Stokoe and Wilcox discussed the importance of gesture in the development of linguistic competence. The authors put forward a theory, suggesting that syntax originates from gestures. They suggested that gestures, especially iconic and visible ones, facilitate the transition from disorganized sequences of vocal components and gestures to hierarchically organized structures (Armstrong *et al.,* 1995, p. 198). This progression toward complex gestural arrangement can be attributed to the development of mirror neuron systems, enabling hominids to master intricate hierarchical structures (Rizzolati & Arbib, 1998; Arbib, 2003). Michael Arbib further expands on this notion (2000; 2005), proposing that as gestures evolved into complex imitation-driven forms, they became *protosigns*, thereby contributing to the processes that facilitated the emergence of *protospeech*.

Regarding language acquisition, LIS and ASL researchers theorized a developmental path for gesture and first words, similar to that of embodied cognition theorists (see section 2.2.3). They suggested that words initially link to acting on prototypical objects and tools within stereotyped routines or scripts, gradually becoming decontextualized (Volterra *et al.*, 2017; Volterra *et al.,* 2022). Observations on hearing monolingual infants revealed that their production of words and gestures is related to objects they could act on, indicating the significant role of gestures in language development (Capirci *et al.,* 1996). Additionally, caregiver input influences a child's preference for a specific language modality (Fontana & Mignosi, 2012). Remarkably, deaf children acquire sign language at a pace similar to hearing children's acquisition of vocal language (Bates, 1979).

David McNeill's perspective on gestures aligns with these notions, positing that gestures exist alongside words and sentences in speech, but possess qualitative differences from them (1992). McNeill characterizes gestures as catalysts for thought and speech (2005) by observing five speakers describing a scenario of a person climbing a building. Despite employing different gestures and speech, a common thread emerged: an iconic upward gesture representing upward movement, albeit with varying hand configurations, remained consistent across all descriptions (McNeill, 1992, p. 109)[5]. In presenting these observations, McNeill proposes an embodied process of action representation that reflects in gestures. He suggests that gestures play a central role in shaping thought processes, by embodying and reflecting information perceived through the body. McNeill further posits that gestures are not merely supportive of speech but are an integral component of it. At the most extreme level, gestures contribute meaning independently, filling informational gaps that cannot be expressed through speech (McNeill, 1992, p. 128).

These postulations pave the way for the hypothesis of the concept of action, gesture, and Sign continuity, which emerges from challenging the traditional dichotomy that had seen gestures and Signs as separated processes, as mentioned in the previous section. On this topic, Adam Kendon, a prominent researcher in gestural studies, advocated for transcending the division between gestures and Signs in favor of a comparative analysis of their semiotic attributes (2008). Kendon *et al.* (1981) noted that sign languages represent a specialized elaboration of gestural expression and aligned with McNeill's theories on the relation between gesture and sign, by advocating for moving beyond this divide (Kendon, 2008). Kendon built upon McNeill's concept of *Growth Point*, originally introduced in 2005. McNeill coined this term to indicate a unit where imagery and language intersect in a framework where metaphors hold significance in aiding in the conceptualization of ideas or actions, following a pragmatic and cognitive approach (Fontana, 2009). Kendon (2008) further highlights

[5] In reference to early studies exploring LIS and the correlation between Sign order among Deaf signers and gesture use among non-signing individuals, refer to the works of Volterra *et al.* (1984) and Laudanna & Volterra (1991).

that both Deaf signers and hearing speakers use gestures. However, a Deaf signer possesses the ability to saturate gestures with a distinct layer of meaning, by integrating them within a signing context that is part of their sociocultural repertoire.

McNeill expanded on the concept of continuity by translating it into a scheme, referred to as *Gesture Continuum*[6] (2012). This scheme delineates five types of gestures, as labelled by Kendon (2004). McNeill states that the Gesture Continuum is based on three dimensions: the necessity of speech accompanying the gesture, the degree to which the gesture resembles language, and the level of conventionalization in its form. He highlights that, as one moves along the continuum from gesticulation to Sign, the relationship between gesture and speech undergoes changes, based on these dimensions (McNeill, 2012). The continuum outlined by McNeill (2006; 2012) starts with gesticulation, characterized by hand and arm movements accompanying speech, lacking language properties but synchronizing with speech patterns about 90% of the time. Moving along, Language-Slotted gestures fill grammatical gaps within sentences. Distinct from gesticulation, they do not synchronize with speech but become integrated into the sentence structure. After that, Pantomimes are gestures or sequences describing actions or objects, separate from speech yet not fully expressive and can be detached from speech entirely. Emblems, such as signaling ‘OK’ by extending middle, ring and pinky finger while forming a circular shape by joining thumb and index finger, or bringing the index finger to the mouth in a request for silence, are conventionalized and shared within communities, carrying specific meanings.

Lastly, sign languages constitute fully expressive, standardized and recurrent Signs (comprised of manual and body components) forming independent, socially shared linguistic codes. Interestingly, the simultaneous coordination of Sign and speech in sign

[6] Also referred to as *Kendon Continuum*.

languages can lead to errors in both codes, showcasing their independence from speech[7] [8].

McNeill's Gesture Continuum, delineating various types of gestures from gesticulation to sign languages, reveals the nuanced relationship between gestural and vocal communication. This continuum, ranging from gestures accompanying speech to fully expressive sign languages, serves as a central framework in understanding how gestures are linked with speech patterns and evolve into independent signed codes. By illustrating these diverse forms of communication within the continuum, the integration of gestures into language structures, challenging the conventional separation of gestures and Signs as auspicated by Kendon (2008), can be witnessed.

While Kendon and McNeill's work has significantly contributed to understanding the symbiotic relationship between action, gesture, and the continuity between gestures and sign languages, emerging perspectives widened the lens to perceive language as a *multimodal* entity within a holistic framework, where language is looked at as a complex interplay of multiple communication modalities. A paradigm shift of this kind, views language not merely as an outcome of gestural or vocal codes, but as a multifaceted system comprising components that synergistically convey meaning in a complementary manner. This approach leads to an exploration of languages, both visual-gestural and auditory-vocal, as comprehensive, dynamic, and multimodal

---

[7] The insights presented in this section draw from McNeill's work, including his article *Gesture: A Psycholinguistic Approach* (McNeill, 2006) and his book *How Language Began* (2012).

[8] McNeill's Continuum provides a schematic representation of distinct types of gestures. The release of *How Language Began* (2012) coincided with that of *Segnare, parlare, intendersi* (Fontana & Mignosi, 2012), a book in which the authors discuss the gesture-Sign continuity hypothesis from an evolutionary perspective. There seems to be a connection between actions, gestures, and language: a model that accompanies hearing and deaf children into a process where «[…] actions lead to language through gestures» (Fontana & Mignosi, 2012, p. 127). It seems that children from 9 to 13 months start interacting with others through gestures, vocalizations, and words. Fundamental in this process is deixis: a pointing gesture, entirely dependent on the context, performed to draw attention to objects, events or people. Gestures go through a process of decontextualization, becoming increasingly independent from a certain referent to the point of developing into symbolic acts (Fontana & Mignosi, 2012, p. 115). By using these acts, children can refer to an object, event, or person without the need for it to be present.

systems. If the contemporary reevaluations of the relationship between gesture and language is taken into consideration, a coherent narrative emerges highlighting the profound influence of gestures on the emergence and evolution of language, challenging traditional dichotomies between actions, gestures and Signs.

This journey through empirical studies, theoretical frameworks, and paradigm shifts reveals a lively landscape where actions, gesture, speech and sign languages are intricately intertwined. Kendon's proposal of moving beyond the juxtaposition of the terms 'gesture' and 'sign' (2008), and his terminology proposal of 'utterance visible action', describing language as systems where gesture and words are integrated, leads to the application of a multimodal view to language, intended here as signed and vocal. By considering language as a multimodal entity (Fontana, 2009) and, therefore, reframing *what* constitutes language, new considerations on both the paralinguistic and the extralinguistic emerge.

1.2.4. Embracing the multimodal perspective on language

The path that has been described up to this point reflects a general shift in sign language studies, elevating their status from unrecognized to objects of a novel level of enquiry, encouraging a reconsideration of the very essence of language.

It has been discussed how, starting in the 1960s, sign languages arose to the status of linguistic entities by being aligned vocal languages, emphasizing shared attributes. However, a reconsideration on the nature of gesture and Sign revealed their intrinsic relation to actions and the bodily perceptions of the world, based on human experience[9]. This connection resonates with Kendon's (2014) description of language as a system encompassing both signed and spoken language, thereby advocating for a

[9] Tallerman and Gibson (2012) write that language is not a monolithic entity. Instead, it constitutes a multifaceted construct, a composite of traits that likely emerged over long periods of time, some of which might have been present in species preceded humans.

multimodal approach to describing ad, consequently, studying language. Within this framework, research in gesture studies highlights the multimodal nature of language. It reveals that language is not confined to a singular auditory-vocal or visual-gestural system. Instead, it comprises a union of complementary systems employed by both signers and speakers, based on their communicative requirements. This perspective leads to a significant implication: the very essence of a language's linguistic nature and the standards used to define it should be reconsidered (Fontana, 2009).

An interesting perspective to consider in this regard, involves a closer examination of de Saussure's perspective on arbitrariness. As related by Sallandre (2003), de Saussure introduces a distinction between signs characterized by radical arbitrariness (*radicalement arbitraire*) and relative arbitrariness (*relativement arbitraire*) (1916), describing level of arbitrariness that characterizes signs that are, to some extent, motivated. In this context, de Saussure considers that the best starting point for any language's analysis is taking into consideration the *limits* of arbitrariness. In fact, according to de Saussure (1916), a total and complete reliance of any linguistic system on the concept of arbitrariness would lead to maximum complication. This is why we are able to find order and regularity in some signs, or parts of some signs, as no language exists where motivation is absent language (but at the same time, no language exists where *everything* is motivated). These extremities constitute the spectrum where language unfolds.

Vigliocco *et al.* (2014) present a fascinating perspective within multimodal language studies by proposing a compelling 'thought experiment': what if the exploration of language had commenced with sign language as its focal point? Within this hypothetical scenario, Vigliocco *et al.* suggest that numerous dogmatic beliefs regarding language would have been overturned or, at the very least, rescaled in terms of perceived importance. One significant implication of this thought experiment is challenging the deep-rooted notion of arbitrariness as the defining quality of language's signification mechanisms. In fact, from this point of view, sign languages, with their 'visible' iconicity would render the notion of strict arbitrariness implausible. Therefore languages, whether signed or spoken, would have been initially characterized as systems comprising both arbitrary *and* iconic components. This hypothetical reimagining encourages a reconsideration of how fundamental nature of languages are conceptualized (see also Volterra *et al.,* 2022, Chapter 6). In particular, Kendon (2014) encourages research aimed at rethinking language studies within a multimodal approach, dismissing the notion that language solely comprises linguistic elements conveyed through speech or Sign. This new theoretical perspective aligns with Kendon's (2012) view on language as a system that extends beyond just spoken or signed forms, highlighting the challenge the original structural linguistics framework faces in accommodating the prevalent iconicity observed in sign languages.

Thus, the multimodal approach profoundly influences the comprehension and depiction of language, breaking away from its confinement to the written form, and perceiving it as a face-to-face event. Embracing this perspective, a multimodal language model will then include paralinguistic aspects like prosody and gesture in spoken languages, along with body components in signed languages (often referred to as 'non-manual components') (Perniss, 2018). Additionally, it will consider extralinguistic elements, including sociocultural influences. In fact, if a multimodal perception of language is recovered, then, language, as a face-to-face event (not considering its signed or vocal nature) will have to be considered as such: an event. Undividable from the sociocultural context in which it is taking place.

1.3. The present thesis

Provided this background, the present thesis aims to offer a novel perspective on the collection and processing of a sign language within a framework that acknowledges LIS as a multilinear, multimodal, and sociocultural entity. This approach will take into consideration the ‘rediscovered’ paralinguistic and extralinguistic elements by developing within the functionalist Cognitive and Sociosemiotic Framework (henceforth COS-S), as adapted to LIS by Volterra *et al.* (2022), whose theoretical nature and implications will be explored in the upcoming chapters. Simultaneously, this study will attempt to establish a connection between the COS-S Framework and formalist sign language studies. Despite often being seen as opposing approaches (see section 1.2.1.), the present thesis seeks to discuss and problematize these two methodologies with the aim of identifying common ground and balance. The goal is to establish a framework that encourages joint and reciprocal contributions, allowing each approach to offer partial methodological applications to LIS data collection and processing.

Hence, this study occupies an intersegmental space, reimagining frameworks that have previously progressed on parallel paths within sign language research. This exploration is possible due to the continuous evolution of sign language studies. Initially aimed at affirming sign languages as comprehensive linguistic systems, the field has since expanded beyond mere investigations into their linguistic essence, prompting the development of various methodological approaches, defining how research on sign languages should be managed and conducted. Within this perspective, methodologies specific to language and functionalism adopted in this work are employed within the COS-S Approach. This approach dedicates efforts to elevating semantics as the key ‘engine’ of language. Furthermore, a central aspect of the functionalist approach lays in the description of languages as tools employed by humans for communication and interaction with their surroundings, both through signing and speaking. For this reason, the pragmatic and semantic objectives of communication acquire central roles, as focus element of this linguistic theory (Volterra *et al.,* 2022).

Expanding on the existing formalist and functionalist paradigms, the methodology employed in this study finds partial alignment with both approaches, leaning more towards functionalism. Therefore, the exploration, collection, computational representation and processing of LIS data presented in this work will be conducted through tools and methodologies defined as ‘language-specific’, i.e., specifically developed for studying sign languages.

Having acknowledged the prominence of the functionalist approach in this study, it is also vital to recognize a challenge in processing sign language data using computational tools. In this context, formalist perspectives offer language modeling methodologies that can facilitate the type of processing discussed in this work. Presently, the landscape of processing tools for sign language data largely relies on linguistic resources, annotation strategies, and methodologies initially designed for vocal languages. This reliance unsurprisingly stems from the historical predominance of investigations into vocal language and resulting developments of approaches to sign language studies. In its attempt to navigate this complex relation among formalist and functionalist paradigms, this study aims at harmonizing a COS-S approach to LIS studies with the pragmatic need to employ existing resources and methodologies for LIS data collection and processing.

As a consequence, the descriptions of LIS presented in this thesis will encompass its historical evolution and its role as a language representing a minority culture in Italy. Moreover, as a language used in face-to-face interactions, this work intends to evaluate how its elements can be formally collected and processed within an interdisciplinary approach that involves the use of multimodal recording tools to document the language and proposes a multi-level annotation system. This system aims to depict LIS devoid of influences stemming from paradigms and coding methods developed for spoken languages. Simultaneously, it seeks to create a multimodal LIS resource accessible to both signers and non-signers. The discussion in this work intends to provide a comprehensive understanding of the potential methods and approaches for studying LIS from a holistic perspective.

1.3.1. Research questions

As discussed in the previous section, this study aims to investigate the application of the functionalist COS-S framework to sign language collection and processing within an interdisciplinary research context. To achieve this, the present thesis addresses the following research questions (RQ):

RQ1. Is it possible to balance structure and variation in the collection, description and analysis of LIS?

RQ2. Can sign language resources be developed to facilitate interdisciplinary collaboration among researchers, providing both human-readable and machine-readable data through language-specific annotation methods? How can LIS-specific multimodal linguistic features be effectively represented during the annotation of LIS data without compromising the language's integrity for machine processing?

RQ3. Are there tools available for the collection of LIS data that can effectively capture its unique characteristics and multimodal linguistic features while ensuring accurate, high-quality, and comprehensive data collection for research and analysis purposes?

RQ4. Which guidelines should be established in selecting tools and developing annotation methodologies within a COS-S framework?

In this study, the response to the research questions (RQ1-RQ4) will be derived from the outcomes of LIS data collection and processing tasks. In relation to RQ1, as described in sections 1.2. and 1.3., two main paths have developed in the context of linguistic enquiry: one focused on formal structures of language, applied to sign languages, and another favoring a functional approach to language description. These two approaches will be further discussed in Chapter 2, where sections 2.1. and 2.2. will focus on a discussion of the two distinct approaches in linguistic inquiry. Section 2.1. will focus on one path, examining the emphasis on formal structures of language, especially when applied to sign languages. In contrast, Section 2.2. will concentrate on an alternative approach favoring a functional perspective in language description. This exploration will lead to Section 2.3., where a proposal for achieving a balance between these methodologies will be advanced. This proposal will be framed within a practical context, aiming to establish balance between the two approaches. In doing so, this section will lay the foundation for continuing to address RQ1, by setting the stage for a comprehensive understanding of the interplay between these methodologies and their potential synthesis.

In Chapter 3, the discussion will begin by exploring the challenges of sign language processing, with a focus on the Italian context. Section 3.2.2. will provide insights into the most significant sign language corpora collections, offering a comprehensive understanding of their relevance. Additionally, section 3.2.3. will include a description multimodal LIS data collection initiatives conducted in Italy, offering a localized perspective on data collection efforts. Transitioning to computational linguistics tasks, section 3.4. and 3.5. will focus on the contemporary landscape of sign language processing, highlighting both its advancements and challenges.

Once these topics are covered, Chapter 4 will shift its focus to the crucial aspect of annotation in sign language research. Annotation stands as a central element in any inquiry involving sign languages, calling for consideration and exploration. Chapter 4 will offer practical solutions and examples to address RQ2, aiming to highlight effective approaches to annotation, within the scope of sign language research. Chapter 5 will then lay down the foundation for the practical efforts outlined in the following chapter. In Chapter 5, experiences in multimodal sign language collections will be examined. This retroactive analysis will aid in determining the optimal tools and methods for gathering high-quality, multimodal sign language data, ultimately contributing to addressing RQ3. Lastly, RQ4 will find answer in Chapter 6, where comprehensive guidelines for the efficient collection and annotation of LIS data within the COS-S framework will be presented. These guidelines will be put into practice through the description of three distinct data collection efforts, referred to as the

MultiMedaLIS Datasets (MULTImodal MEDicAl LIS Datasets) throughout this work, elaborating on the creation process of the third version of a multimodal dataset (the MultiMEdaLIS_3 Dataset) which, at the present time, to the author's knowledge, is the biggest RADAR/camera dataset for a sign language. For clarity, throughout this thesis, the three data collections will be sequentially numbered based on their chronological development: the first being MultiMedaLIS_1 (Winter 2022), followed by MultiMedaLIS_2 (Spring 2022) and, finally, MultiMedaLIS_3 (Summer 2022).The examination of the annotations developed for the third version of the Dataset will reconnect with RQ1, offering a practical demonstration of the theoretical notions discussed in Chapter 2.

Therefore, the present thesis seeks to go beyond current literature by not only outlining a theoretical framework that employs functional COS-S methodologies for a systematic analysis of LIS intended for computational processing within an interdisciplinary applicative context, but also by presenting a concrete application of this framework. This is achieved through a detailed description of the data collection process, annotation efforts, and subsequent analyses, outlined in Chapter 6.

Chapter 2
Navigating linguistic theories: formalism, functionalism and sign language

2.1. The core characteristics of formalist theories

In opening this section, it is useful to begin by introducing the concept of 'form'. Influential over linguistic analyses carried out since the late 19th century, studies on 'form' in language focus on overtly describable aspects of language, often placing semantics and pragmatics in secondary positions (McElvenny, 2019). As highlighted in the previous chapter, early studies conducted on sign languages were predominantly framed within structuralist theories. Indeed, it has been argued that initial findings from ASL linguistics research provided validation for these language theories[10] (Armstrong *et al.,* 1995). In addition, it has been discussed that structuralist theories initially impacted Chomskian linguistics (De Mauro, 2005), although Chomsky eventually moved towards excluding extralinguistic elements entirely. For the purpose of the present work, before delving into more detail into Chomsky's theories, it might be beneficial to outline the fundamental characteristics of formalist theories.

Formalism, as expressed by Taylor (2003, p. 22), fundamentally rests upon the notion of predefined and immutable categories, exhibiting defining traits. Firstly, linguistic entities are classified within a category based on their possession of specific and essential features. This criterion acts as an 'entrance', allowing linguistic entities to enter a category only if prerequisites are met. Secondly, the features that delineate categories operate in a binary manner. Entities either fulfill the criteria, entering the

[10] To this regard the following statement by Langacker can be considered: «What one finds in language depends in large measure on what one expects to find» and «If one expects language to be organized into discrete 'modules', abundant evidence for modularity will be forthcoming […]» (Langacker, 1987, p. 11).

category, or remain outside of it. Thirdly, categories exhibit clear and well-defined boundaries. Lastly, within each category, all members are regarded as equal.

Chomsky's prominence in contemporary formalist theories, as evidenced by his writings, highlights the notion that grammar is independent of meaning and context. On this theme, Givón lists three postulates (Givón in Wilcox & Wilcox, 1995, p. 135). Firstly, language functions as a distinct module within the mind/brain, separate from general cognitive processes. Secondly, language structures can be examined *independently* of their communicative purpose. Lastly, the relationship between the linguistic code and its mental referent is arbitrary, divergent from the evident iconicity seen in pre-human communication. Consequently, language is regarded as *disembodied*. Mental linguistic units are not aligned with articulatory or physical characteristics. Within the confines of the formalist viewpoint, there is no mutual influence between the mind and language. Lastly, a cornerstone of formalist and Chomskian theories is the concept of universality. A fundamental postulate proposed by Chomsky is that languages share a set of universal features. For instance, all languages have nouns and verbs, subjects, objects, or noun phrases that are combined according to specific sets of syntactic rules.

2.1.1. Chomsky's generativism and the formalist perspective: exploring language as grammar

In Chapter 1, Chomsky stood out as a central figure in contemporary formalist linguistic studies. In fact, he is credited with his pioneering work on various theories, particularly his ideas about the Poverty of Stimulus argument, language's innate competences and the model of Universal Grammar. These theories are key in shaping modern formalist approaches to studying language.

Chomsky believes that the theory of Universal Grammar (UG) is supported by his Poverty of Stimulus hypothesis (1959), which suggests that children possess an innate understanding of grammar principles that guide their language learning process. This theory was crafted to provide a biologically motivated contrast to behaviorist accounts of language acquisition[11].In particular, Chomsky suggests that children are born with an inherent cognitive framework, referred to as the *innate language faculty* (Chomsky, 1965; 1975; 2002a). This innate system encompasses principles and parameters constituting UG, defined as the genetically determined language capability (1986). These principles and parameters are believed to equip children with the essential tools to acquire any spoken language.

In the course of his research, Chomsky embarked on a quest to answer fundamental questions, such as: What constitutes knowledge of language? How is this knowledge acquired? It was within this pursuit of answers that UG emerged. As the understanding of language acquisition and the innate cognitive faculties of humans unfolded, generativist investigations progressed to describe and explore the structural attributes of languages. Let's begin by addressing a question. What makes Generative Grammar 'generative'? During his early studies, Chomsky defined Generative Grammar to describe the intentionally unrealistic ideal speaker/listener's 'pure' innate competence, within a perfectly explicit grammatical context (1965; 1986). In this context, grammar is not dependent on humans' abilities to produce and/or understand language (Chomsky, 1965). To this regard, Chomsky further elaborates on the object of linguistic enquiry by distinguishing between *E-language* (Externalized language) and *I-language* (Internalized language), another cornerstone of his framework. E-language

[11] In behaviorist theories, linguists must work to recognize similarities and regularities in human linguistic behavior. Among its developers is B. F. Skinner (1938; 1957), who developed a methodical analysis of the behavior of organisms, both human and animal. Skinner wrote that «[…] the ultimate direction and control (of advanced systems of organism behavior) have been assigned to entities placed within the organism and called psychic or mental» (1938, p. 3). The American Psychologist also focused on how humans act and interact with their surroundings to obtain a response (1957, p. 53). Thus, within behaviorism, language is acquired by children as any other skill is. It takes place through a process of response to linguistic inputs, followed by reinforcements and 'punishments' (Shormani, 2014). The Behaviorist approach was criticized by Chomsky who disagreed with the assumption that language acquisition is the result of a response to stimuli but rather an innate ability of humans (1959).

encompasses observable linguistic behaviors across various languages, traditionally serving as the object of linguistic inquiry. However, Chomsky posits that these concrete manifestations of E-language rely on a system of representations of grammatical rules and linguistic structures innate in individuals, known as I-language (Chomsky, 1986). As it constitutes the internal mental representation of language, I-language then becomes the proper subject for linguistic study (Tallerman & Gibson, 2012). This proposal aims to shift the focus of the study of language, abandoning its view as an externalized object, a change that, according to Chomsky, was «[…] very much in order» (1986, p. 26), as grammar is best approached as a self-contained study independent of semantics (Chomsky, 2002b). In fact, the foundational assumptions that motivated this shift had been motivated by Chomsky's claim that obscure and intuition-bound notions lack utility in linguistic enquiry. In fact, he highlighted the productivity of rigorously formulating and applying linguistic theories to material without resorting to adjustments or vague formulations (1959). The focus on I-Language, intended as grammar, as the main topic of linguistic enquiry, aligns with Chomsky's identification of a hierarchical organization of the elements within sentences. This hierarchy starts from the biggest constituent of the sentence and then progresses towards smaller constituents until individual elements are reached (Chomsky, 1970; Cook & Newson, 1996). For instance, hierarchical organization might be *sentence* > *phrase* > *individual elements*. Syntactic structures hold, therefore, a fundamental position within formalist studies.

The fundamental principles of Chomsky's early formalist theories have undergone substantial transformations over the years. More recently, Chomsky has highlighted the importance of transcending the confines of a specific theoretical framework that could restrict the comprehension of language generation. Instead, he urges for a contemplation of the notion that the principles governing the construction of structured expressions may not be as exclusively tied to language as once believed, potentially extending their relevance to diverse domains. This point of view suggests the potential to uncover principles that support various manifestations of structured expression, extending beyond the boundaries of I-Language (Chomsky in Haspelmath, 2018).

2.1.1.1. The formalist approach to sign language studies

Through the progression of the previous sections, a crucial question emerges in relation to the suitability of employing formalist approaches in the study of sign languages as unique forms of expression that are based on the visual-manual modality. In fact, for researchers working with sign language, can formalist approaches, which find their core in the need for structured rules often conveyed through written language, be effectively adapted to the study of sign languages? The branch following formalist principles, aims to adapt and extend formalist concepts, such as phonemes, morphemes and syntax to the unique characteristics of sign languages. The driving force behind this adaptation is the belief that shared features are more revealing than dissimilarities.

In the previous chapter, it was mentioned that following the publication of Stokoe's work in 1960, a paradigm for analyzing sign languages emerged, characterized by a structuralist approach. This reflected in the first description of LIS published in the 1987 book, edited by Virginia Volterra[12], titled *La lingua italiana dei segni. La comunicazione visivo-gestuale dei sordi* (Italian Sign Language: The visual-gestural communication of the deaf), a book marked by an attempt to define the elements of LIS within a framework of continuity with vocal languages, a natural consequence of Stokoe's early formulations on ASL. However, descriptions of sign languages based on formalist frameworks have been carried out continuously by linguists like Sandler and Lillo-Martin, who embarked on a journey based on this framework, to redefine traditional linguistic terms to suit the specifics of sign languages like ASL and BSL (British Sign Language). For instance, the term *phonology*, typically associated with

[12] The book features contributions from Serena Corazza, Maria Luisa Franchi, Alessandro Laudanna, Elena Antinoro Pizzuto, Elena Radutzky, Benedetto Santarelli and Maria Luisa Verdirosi.

the study of sound combinations in vocal languages, undergoes a transformation in the context of sign languages. Rather than referring to combinations of sounds, it takes on a novel meaning as «[…] a finite set of meaningless contrastive units that combine in constrained ways to form meaningful morphemes and words» (Sandler & Lillo-Martin, 2006, p. xv). This adaptation aligns with the distinctive nature of sign languages and their visual-gestural phonological structure. Additionally, Sandler and Lillo-Martin take initial structuralist approaches to sign language studies a step further, by positing that the generativist approach to sign languages syntax can provide valuable models for identifying non-canonical word orders within sign languages (Sandler & Lillo-Martin, 2006, p. 12), approaching the interplay between universal linguistic principles and sign languages with optimism. Sandler and Lillo-Martin thus perceive this interplay as a facilitator for the evolution of new theories applicable to both signed and vocal languages. However, while they appreciate the persuasiveness of rigorous theories that unveil cross-modality commonalities, they also recognize that discrepancies offer insights into the diverse modalities of language systems, acknowledging that Chomskian principles come with limitations in accommodating variation[13]. Therefore, they state that their objective is to leverage general theories of linguistic structure to analyze sign languages' morphology, phonology, and syntax, culminating in a collection of properties that hold true *across* languages. To ensure that their research aligned with a universal framework Sandler and Lillo-Martin prefer to focus on contexts where sign languages adhere to universal principles or fall within an allowed range of variation[14] (Sandler & Lillo-Martin, 2006, p. 285). This approach seeks to strike a balance between uncovering shared linguistic traits and acknowledging the unique characteristics of sign languages, contributing to the ongoing dialogue between generativism and sign language studies.

[13] For instance, vocal languages can be categorized according to the location of the object of the verb or the WH- element (Chomsky, 2002b, p. 17).
[14] «[…] we stand to advance our understanding of the phenomenon of human language by continuing to elaborate the description and analysis of languages in each modality, without ever losing sight of the properties of mind that correspond between the two» (Sandler & Lillo-Martin, 2006, p. 510).

Within the framework of generative principles, researchers have attempted at describing and *transcribing* sign language utterances. On the topic of syntactic structures and LIS, Branchini and Geraci (2011) conducted a study on the *Corpus LIS* to identify how Subject (S), Object (O) and Verb (V) combine in LIS utterances. This description looks at LIS phrases to identify the combination among lexical elements and phrase head. As a result of this research, the authors found that within the Corpus, two utterance orders are found: SOV (54%) and SVO (46%), leading to the definition of LIS unmarked utterance order of as SOV, making it a head-final language (see also Geraci, 2004). The formalist approach to language studies is particularly evident in the presented examination of LIS structures and showcases a *fil rouge* of formalist investigation that is the identification of common, concrete, practical and scientifically articulated rules. This type of approach comes with advantages and disadvantages. On one hand, efforts towards delimiting the scope of investigation, focused and tangible outcomes in describing and analyzing sign languages within a universal framework, can give valuable results that contribute to a broader understanding of linguistic principles that transcend modalities. By adhering to a set range of parameters, researchers can establish a solid foundation for comparative analysis which, in turn, may uncover shared linguistic patterns and cross-modal similarities. On the other hand, setting boundaries can potentially complicate the exploration of language-specific characteristics, thereby limiting the comprehensive understanding of a sign languages. This approach might seem limiting, particularly if it results in the exclusion of unique linguistic features from theoretical considerations.

Transitioning from the exploration of generative approaches to (sign) language research to the domain of cognitive studies, particularly the modularity theory, offers a shift from structural ivestigations to a broader examitation of cognitive faculties and their compartmentalized nature. This transition broadens the focus from linguistic structures to a more encompassing understanding of how cognitive processes have been described as operating within modular and encapsulated systems.

2.1.2. Cognitive studies: the modularity theory and its developments

In Fodor's expansion of Chomsky's innateness model, he introduced the *modularity theory,* proposing that cognitive faculties are compartmentalized into independent vertical modules, each dedicated to specific tasks, such as language or visual reception. According to Fodor, external information undergoes transformation into data processable by each module. These modules are encapsulated, autonomous and innate, operating rapidly and automatically through specialized processing mechanisms. They are pre-tuned to a database, meaning they can process specific types of data while automatically ignoring other inputs. Additionally, modules can handle specific problems with precise structures, producing outputs without relying on other modules. The produced outputs are 'codified' in a suitable manner for central processing, which is influenced by the beliefs of the organism. In establishing this, Fodor draws a separation between modules and central processing. This modular organization of the mind profoundly shapes how we process information and execute intricate mental tasks. For instance, the visual module can process visual information independently, allowing rapid and intuitive object recognition (Fodor, 1983). Fodor's modularity theory thus implies a disembodied mind model, where information is received and processed independently within distinct modules. There is no alteration, addition, or removal of the perceived information as it is processed within encapsulated cognitive modules.

Fodor's inflexible model of modularity of cognition is taken into consideration by Karmiloff-Smith, who argues for going beyond modularity, challenging Fodor's dichotomy between modules and central processing. Karmiloff-Smith introduces the concept of *modularization* versus modularity, by arguing that the development of predispositions within the cognitive system is influenced by environmental inputs, suggesting that cognitive processes do not strictly conform to fixed modules. Karmiloff-Smith's Representational Redescriptional (RR) model suggests that our innate predispositions only fully manifest through interaction with the environment, emphasizing an epigenetic nature to cognitive development. This model describes how implicit information becomes explicit knowledge, initially within specific domains (intra-domain) and then extends to create connections between different domains (inter-domain). In putting forward this model, Karmiloff-Smith proposes a union of Piaget's constructivism (Piaget, 1976; 2005), suggesting that humans lack domain-specific knowledge, and Fodor's modularity (1983).

Forty years after Fodorian modularity, new descriptions of the human brain's processing capabilities have emerged, describing the brain as a *complex network*. In this framework, complex networks have revealed similarities between social and biological networks as they both showcase segregation within themselves, while enabling integrations across their respective subnetworks. In particular, in complex networks the existence of subnetworks is acknowledged suggesting their transversal operation by integrating content across different subnetworks (Wig, 2017). In this context, the brain appears to process information in a *continuum* among plasticity and rigidity. Karmiloff-Smith's critique of Fodor's 1983 theory highlights one of its limitations regarding the lack of discussion on scenarios where a module lacks input. She points out instances, like deaf-born individuals, where the brain appears to selectively adapt to process visual-manual inputs linguistically (Karmiloff-Smith, 1995, p. 10). This adaptation seems to appear as a result of the absence of a sensory input, a feature unaccounted for by Fodor's theory (Fontana, 2009). More recently, studies in cognitive science have presented findings in favor of cross-modal plasticity of the brain (first theorized by Bavelier & Neville, 2000), particularly observed in deaf and blind individuals. These individuals exhibit brain plasticity that is connected with their experience (Benetti *et al.,* 2017). This suggests that the brain could serve as a system that is capable of adapting based on experiences, able to execute specific tasks regardless of the sensory systems transmitting the stimuli on which the tasks are based (Piazza & Pavani, 2022).

Acknowledgments of the centrality of sensorimotor experience and studies on the cross-modal plasticity of the brain lead back to the topics of contrast and continuity, a leading theme in the present thesis. The concept of contrastive modules finds renewed relevance in contemporary cognitive sciences, now positioned within a revised notion of continuity, suggesting that segmented brain areas can operate in a *continuum*, adapting the hypothesis to allow for their interconnected functionality. Furthermore, plastic descriptions of the brain align with studies on embodiment, finding empirical support in their correlation to embodied experiences as outlined by Lakoff and Johnson (1980) and in the discovery of mirror neurons (Gallese *et al.,* 1996; Rizzolati *et al.,* 1996).

The mentioned empirical foundation lies in the identification of mirror neurons as a source of knowledge that enables the human brain to gather sensorimotor information, encompassing both who and what we interact with (Gallese *et al.,* 2022). The activity of mirror neurons appears to allow for a direct understanding of actions, whether observed, heard or implied, through a mechanism of embodied simulation (Gallese, 2005; Gallese *et al.,* 2008) that is then adapted for thought and language (Gallese, 2008) and appears to be multimodal (Ferrari & Coudé, 2018). Beyond strict embodiment, mirror neurons seem to help humans process our and others' actions and intentions through social-emotional learning processes (Butera & Aziz-Zadeh, 2022), reconnecting to the relevance of external stimuli described by Karmiloff-Smith (1995). The concept of embodiment, aided by mirror neurons— introduced in this section, and further discussed in section 2.2.3.— extends beyond assimilating sensorimotor information. It also plays a central role in comprehending social interactions, discerning intentions, and interpreting emotions conveyed through language. This linkage between embodied simulation, facilitated by mirror neurons, and the understanding of social cues and intentions acts as a critical bridge in understanding the social dimensions inherent to language and communication.

2.2. The Social Dimensions of Language and Communication

Karmiloff-Smith's modularization introduces a dynamic layer to our understanding of language acquisition and use. By emphasizing the interaction between innate predisposition, external stimuli, and the broader sociocultural environment, modularization sheds light on the multifaceted nature of language development. This aspect holds relevance for functionalist approaches to human cognition and language, as illustrated in the present section. According to the RR model, external inputs play a decisive role in preventing innate predispositions from remaining latent (Karmiloff-Smith, 1995). Therefore, the sensory and motor perceptions through which we process external stimuli shape our cognitive abilities. Consequently, it is essential to consider how we experience and assimilate external inputs and, in turn, their significance in both understanding and expressing language.

Functionalist approaches assert that the fundamental focus of language studies should extend *beyond* grammar and syntax. Within this framework, grammar acquires a secondary role, serving as a means for signers and speakers to crystallize their experiences as beings existing in the world, employing organized codes to facilitate creative expression (Duranti, 2007, p. 120). The shift from language as grammar (Chomskian I-Language) redirects attention towards comprehending how humans act using. From this point of view, language is studies *as* a social action, emphasizing its definition through embodied perceptions of our surrounding world.

2.2.1. Language as a social action

By recognizing the corporeal foundation of thought and language, cognitive linguistics offers a renewed interpretation of the very foundations of linguistic systems, operating at the levels of lexicon, semantics, and grammar. In adopting this interpretation, Tommaso Russo Cardona made a significant contribution through his works, such as *La mappa poggiata sull'isola* (The map placed on the island) (2004b), marking a transition within sign language studies the Italian context. This shift involved a departure from an abstract and theoretical exploration of (sign) language studies, in favor of a more tangible and pragmatic investigation of language's everyday application within the Deaf and signing community, highlighting the unique processes of signification inherent to Deaf Culture, as manifested by Deaf signers[15]. Russo Cardona's perspective is closely linked to Charles Sanders Peirce's (1903), who carried out his research on semiotics during the early 20th century, becoming one of the most influential scholars in the field. Peirce challenged de Saussure's dual vision of meaning (signifier/signified), preferring a triadic sign relation in different contexts. For Peirce, communication takes place through signs, where a sign is anything that stands for something (object), to somebody (interpretant), in some respect (context) (Mick, 1986; Peirce in Marmaridou, 2000). Central to Peirce's framework is the concept of the *interpretant,* denoting an interpretation of one sign through another. This recursive nature allows any initial meaning to be continuously re-interpreted, setting off an infinite chain of interpretants. The interpretant plays a dual role, serving as both the interpreted sign and the interpreting sign, thus enabling a continuous process of semiosis (Mick, 1986, p. 199). Peirce's framework of semiotics, with its emphasis on the triadic nature of signs and, in particular, the concept of interpretant, resonates with Hymes' perspective on communication within specific social and cultural contexts. In fact, Peirce's recognition of the intricate relationships between signs, objects and interpretants falls in line with Hymes' emphasis on understanding communication as a product of the cultural values, social institutions, and roles within a community.

Dell Hymes, an American sociolinguist and anthropologist, discussed the importance of the community and social setting in which communication takes place, in shaping communication itself. Hymes criticized the impersonal and formal view of linguistics

[15] Moreover, starting from in-use signed language, Russo Cardona explored the processes of iconicity and productivity in LIS on several registers, from everyday conversations to poetry and theatre performances (Russo Cardona *et al*., 2001; Russo Cardona 2004a; 2004b; 2005).

that was popular at the time (late 1970s) and stated that a distinction must be made between the 'infiniteness' and 'equivalence' of languages according to formal theories, and the actual use made of languages by their speakers and signers, that makes them 'finite' and 'unequal' (1974, p. 72). According to Hymes linguistics should be framed within ethnography: i.e., the communicative, social and cultural context where language takes place. In fact, he had already suggested, ten years before, that:

> The same linguistic means may be organized to quite different communicative ends; the same communicative ends may be served by organization of, or by focus upon, quite varied means. Facets of the cultural values and beliefs, social institutions and forms, roles and personalities, history and ecology of a community must be examined together in relation with communicative events and patterns […].
> (Hymes, 1964, p. 3)

Hymes' approach aligns within the COS-S perspective in the context of sign languages (Volterra *et al.,* 2022), as it takes into consideration the influence and weight played by communities and cultures on languages. The unique circumstances of Deaf communities have, following Hymes' hypotheses, unquestionably shaped their languages. Take, for example, the bimodal bilingualism of Deaf signers (Fontana, 2013; Volterra *et al.,* 2022): the constant interplay between signed and vocal languages in everyday life places Deaf individuals in a social context where the majority hearing population employs acoustic and auditory-based languages. Such an interaction has led to the adoption of components from vocal languages, such as mouthings, into sign. languages[16], a characteristic that stands as the concrete realization of this interlingusitc and intermodal interplay.

Hymes' criticism towards the 1970s impersonal and formal view of linguistics finds applications in language collection as it can be adopted as a starting point for considerations on the collection of natural versus artificial sign language data, and how the nature of the data influences formalization processes.

[16] For a discussion of this phenomenon is LIS, see Fontana (2008) and Fontana and Raniolo (2015).

Following the path established by Hymes, Duranti approaches language analyses by defining them «[…] *as the study of language as a cultural resource and speaking as a cultural practice*» (1997). In particular, Duranti adopts the ethnopragmatic perspective, according to which language is best studied and understood as a social action, from a culture-internal perspective, thus considering culture-internal values, beliefs, attitudes and emotions (Goddard, 2006). In adopting the ethnopragmatic perspective, it becomes evident that Signs and words possess a distinctive 'force' within everyday interactions (Duranti, 2007), a force that transcends the mode of communication employed, be it written, spoken, or signed. This inherent force, embedded in Signs and words during daily interactions, accentuates the crucial significance of direct face-to-face communication in molding identities.

As the subject of embodied cognition will be explored in the following section, the groundwork laid by Peirce, Russo Cardona, Hymes, and Duranti, highlighting the inseparable connection among language, culture, and cognition, will be further examined and taken into consideration, but from a predominantly cognitive standpoint, with the aim of examining how our physical experiences and interactions contribute to shaping cognitive linguistic processes.

2.2.2. Language as embodied cognition

Cognitive linguistics is grounded in three fundamental hypotheses. Firstly, it rejects the notion that language is an autonomous cognitive faculty in humans. Instead, language is viewed as an integral component of human cognition, playing an essential role in various social and cognitive skills. This perspective suggests that processes involving language are not distinct from other cognitive abilities used in non-linguistic domains, such as visual or sensorimotor skills. Secondly, it asserts that sounds and utterances are both inputs and outputs of cognitive processes governing speaking and understanding. Additionally, within the theoretical context of cognitive linguistics, knowledge of language emerges from language use in a bottom-up process, where categories are constructed from cognition, leading to the development of specific utterances for particular contexts. Consequently, words are considered units that can only be analyzed after having understood the mental processes that give rise to their meaning. This leads into the third principle, highlighting the importance of semantics. Words are not viewed merely as grammatical units, but their semantic value must be continually considered since meaning is the outcome of a cognitive process and should not be treated independently (Croft & Cruse, 2004).

The grounding of cognitive linguistics in the understanding that language is intertwined with human cognition forms the basis for the concept of embodiment. Within cognitive linguistics theoretical context, the notion of embodiment is an additional hypothesis that contributes to disrupting the formal view of language as a purely abstract system. Instead, it suggests that humans engage similar neural structures to both experience sensory-motor reality and comprehend language describing these experiences. Indeed, when considering language both as a social action (section 2.2.1.) and as embodied (section 2.2.2.) a convergence is found in the understanding that language surpasses mere systems of symbols or internalized abstract rules, emerging from contextual interactions. The perspective of language as a social act recognizes its molding within social contexts, interactions, and communicative intentions. Similarly, the embodied viewpoint highlights the grounding of language in our physical engagements and interactions with the world. Language as a social act highlights its role in constructing social realities (Goddard, 2006; Duranti, 2007), while the embodied perspective explores how our bodily experiences and interactions actively shape the meanings and cognitive aspects of language.

2.2.3. Embodiment, image schemas and the role of metaphors

As discussed above, the concept of embodiment challenges the notion that language is an entirely formal, abstract, and arbitrary system (Fontana & Mignosi, 2012). In particular, the paradigm suggests that humans employ the same neural structures to experience the sensorimotor aspects of reality and to understand verbs, names and sentences that describe those same sensorimotor experiences (Buccino & Mezzadri, 2013). Shapiro states that:

> The emphasis on embodiment is intended to draw attention to the role an organism's body plays in performing actions that influence how the brain responds to the world while at the same time influencing how the world presents itself to the brain. Similarly, the focus on situatedness is meant to reveal how the world's structure imposes constraints and opportunities relative to the type of body an organism has, and thus determines as well the nature of stimulation the brain receives. Within embodied and situated agents, the brain remains a crucial organ of cognition, but it must accept a downgrading in status from star to co-star, an equal partner in the creation of cognition alongside body and world.
> (Shapiro, 2011, p. 173)

Shapiro (2011, p. 197) organizes the discussion on embodied cognition around three themes: conceptualization, replacement, and constitution, commonly shared among embodied theories. The fist theme is conceptualization: the acquisition of concepts is limited or constrained by the properties of a body. The organism will always rely on its body to understand its surroundings. This implies that different organisms with different bodies will have a different understanding of the world. The second theme is replacement: cognitive science is based on a computational framework that does not represent cognitive processes correctly. Computationalism, i.e., a discrete description of cognitive processes, should be surpassed with tools that allow a description of cognition as continuous and dynamic. The third and final theme is constitution: the body and the world are constituting elements of cognition. They are not mere causal influences on cognition. but play a fundamental role in structuring cognitive processes (extended cognition).

The centrality of metaphor in cognitive processes and the production of language was introduced by Lakoff and Johnson (1980), who advocated for a reexamination of the weight of metaphors in language, theorizing that most of human's conceptual systems is of metaphorical nature. Traditionally, metaphors were regarded as peripheral elements but, according to Lakoff and Johnson, they play a central role in comprehending the world we perceive:

> The concepts that govern our thought are not just matters of the intellect. They also govern our everyday functioning, down to the most mundane details. Our concepts structure what we perceive, how we get around in the world, and how we relate to other people. Our conceptual system thus plays a central role in defining our everyday realities. If we are right in suggesting that our conceptual system is largely metaphorical, then the way we think, what we experience, and what we do every day is very much a matter of metaphor.
> (Lakoff & Johnson, 1980, p. 4)

In stating this, Lakoff and Johnson assert that metaphors are, first and foremost, a cognitive mechanism that helps us structure conceptual systems and, as a secondary consequence, a linguistic phenomenon (Fontana & Cuccio, 2013). In this framework, metaphors are no longer just a matter of language as intended in formalist studies (therefore as a matter of words) but rather a matter of human thought, shaping how we perceive and understand the world (Lakoff & Johnson, 1980). Lakoff and Johnson go on to state that «[…] most fundamental concepts are organized in terms of one or more spatialization metaphors»[17] (1980, p. 50), as concepts become more complex, metaphors do so as well. The authors go on to list a series of metaphor typologies such as: entity and substance metaphors, container metaphors, metaphors related to the visual field, events, activities, states, personification and so on, thus showing how permeating they can be to our everyday life and activities.

Cuccio (2015) writes that both language production and comprehension trigger the experience of bodily feelings, «[…] in virtue of the mechanism of Embodied Simulation, the comprehension of bodily-based metaphors determines the experience of a particular bodily state […] allowing us to build cross-domain mappings» (p. 105).

[17] In their book *Metaphors We Live By* (1980), the authors provide several examples of space-oriented metaphors deeply rooted in the physical and cultural experience. For instance, sadness and happiness are related to the down/up spatial contrast, think of the idioms of 'feeling down' as opposed to 'feeling up'.

However, how can embodied simulation facilitate the human brain's transition into the abstract realm of signification? If a metaphor is considered conventional, this implies that it is deeply rooted within a culture and is no longer seen as a cross-domain comparison: «[t]he more conventional and less deliberate the metaphor, the less motor activation will be observed in the brain» (*Ivi*. p. 100). Therefore, metaphors are a successful source of language creation and comprehension since they ease us into understanding both bodily and abstract concepts by harmonizing our brain and body. Thus, the concrete or abstract nature of a word (or Sign) does not imply any type of difference, since they are all ground in perception and action (Borghi *et al*., 2014).

Another process related to embodiment in cognitive linguistics is metonymy. Whereas metaphors are strategies used to understand something in terms of something else, metonymy is employed to use one entity to refer to another. If a metonymy represents part of a whole, many parts could stand for *that* whole. The part we choose determines the aspect of the whole we are focusing on (Lakoff & Johnson, 1980, p. 80). When it comes to understanding, metonymies and metaphors use different kinds of paths. They both serve the same purpose, yet metonymies «[…] allow us to focus more specifically on certain aspects of what is being referred to» (*Ivi,* p. 81).

Examining the connection between linguistically meaningful units and brain responses, Meister *et al.* (2007) prove that the ability to discriminate between a noise and a phoneme is caused by the posterior perisylvian region— that is, the region of the human brain involved in language and speech that includes Broca's and Wernicke's areas, respectively the area associated to the production of speech and the interpretation of sounds— and the left premotor cortex. This suggests a common cortical architecture for the development of motor abilities and discrimination of complex articulatory sound patterns. Friston (2012) thus asserts that our acquisition of information is intrinsically linked to movement, and he highlights that the brain's interface with the surrounding environment is exclusively mediated by the body. Indeed, essential activities such as speaking, walking, and observing depend on the body's integral role. This intricate interplay between the body and the brain's engagement with the environment aligns with the findings of Pulvermüller, who published a study stating that certain speech stimuli «[…] are perceived as a bare noise if placed out of speech context […] However, they are perceived as phonemes if presented in appropriate speech context. […]» (Pulvermüller *et al*., 2009, p. 10).

2.2.3.1. The biological anchoring of embodiment

Buccino and Mezzadri (2013) elaborate that the mechanism driving the recruitment of motor representations, particularly in words expressing actions and verbs, may reside within the mirror neuron system. Mirror neurons (Gallese *et al.,* 1996; Rizzolati *et al*., 1996) are located in the premotor and parietal cortex of apes and exhibit activity both when an animal executes an action on an object, and when it observes or recognizes a similar action performed by another individual, or even when the same action is heard[18]. Therefore, mirror neurons establish a connection between the 'observer' and the 'actor'[19] (Rizzolatti & Arbib, 1998, p. 188). If human language is rooted in our bodily experience and our interactions with others are filtered by the mirrored production and recognition of actions, then it is plausible that human communication took place initially through action recognition and thus gestural elements. On this topic, Rizzolati and Arbib write:

> Our proposal is that the development of the human lateral speech circuit is a consequence of the fact that the precursor of Broca's area was endowed, before speech appearance, with a mechanism for recognizing actions made by others. This mechanism was the neural prerequisite for the development of inter-individual

---

[18] «*Il meccanismo attraverso il quale le parole che esprimono azioni, i verbi, possono determinare il reclutamento di rappresentazioni motorie coinvolte anche nell'esecuzione di quelle stesse azioni potrebbe risiedere nel sistema dei neuroni specchio. I neuroni specchio sono neuroni trovati nella corteccia premotoria e parietale della scimmia, che sono attivi sia quando l'animale esegue un'azione diretta su un oggetto, sia quando osserva e riconosce la stessa azione o un'azione simile eseguita da un altro individuo o quando la stessa azione viene ascoltata»* (Buccino & Mezzadri, 2013, p. 7).

[19] This discovery is of central importance not only for embodied theories, but also for sign language studies and is, therefore, relevant for the development of the data collection presented in this work (refer to Chapter 4, 5 and 6).

> communication and finally of speech. We thus view language in a more general setting than one that sees speech as its complete basis.
> (1998, p. 190)

Mirror neurons collectively establish a ‘vocabulary’ of motor actions, hosting representations of movements that can be accessed via perceived stimuli, a hypothesis initially introduced in Rizzolati *et al.* (1988). This vocabulary consists of ‘words’ encoding the objectives of actions or the specific manners in which these actions can be executed. These ‘words’ embody actions represented by groups of mirror neurons that are activated upon the observation of ongoing actions upon certain objects (Rizzolati & Luppino, 2001; Fogassi, 2008). Research on mirror neurons has shown that actions are encoded in abstract terms, facilitating the understanding of the relationship between the agent and the object, which is seen as a demonstration of intentionality. This process involves the brain creating conceptual representations based on a model that represents the perceived world's space as occupied by meaningful objects (Fontana, 2009).

The exploration of the embodied theory’s biological basis and the existence of a cognitive ‘vocabulary’, summarizes the relation between action representations and motor movements. This theorization resonates with Johnson’s theorization of *image schemas* (1992) as combinations of schemas and images: conceptual structures that contribute to the creation of elaborate and abstract concepts through metaphorical projection (Langacker, 2008). In other words, an image schema is a synthetization of bodily and sensory experiences that results in structures that help humans in their daily interaction with their surroundings. Image schemas are generalization of sensory experiences, extrapolated from their original context and reenacted in similar ones. They combine with each other allowing humans to perform mundane tasks, such as getting a glass of water or going to the bank.

By looking at the two elements that comprise the term, it can be found that *schemas* are a recurring term in philosophy and cognitive science, originating from Greek rhetoric, describing a series of strategies used by rhetoricians to create embellished arguments (Oakley, 2007). The concept of *schema* is present in cognitive science's lexicon as well, described as links between general patterns. In particular, «[…] a cognitive representation comprising a generalization over perceived similarities among instances of usage » (Langacker, 1984, p. 23). On the other hand, images are abstract patterns rendered schematically, that help individuals frame their experiences (Oakley, 2007).

2.2.4. Sign languages as embodied action

It has been observed how neurobiological characteristics support the structuring our experiences and discussed theorizations on the role of metaphors in language. A question now arises: how do metaphors align with this sign language studies?

If Langacker's statements (1984) are taken into consideration, it becomes apparent that metaphors act as 'bridges' between distinct domains of understanding. In fact, Langacker states that metaphors are a primary means of enhancing and constructing our mental world, where basic organizational features of one conceptual domain, that are usually more directly grounded in bodily experience, are projected onto another (2008). Therefore, metaphors enable us to grasp abstract or complex concepts by relying on our familiarity with more concrete, embodied experiences. For instance, one might describe time as 'flowing', linking the abstract concept of time to the tangible experience of water moving (Radden, 2003).

The integration of metaphors within the framework of neurobiology highlights the significance of cognitive processes in shaping our perceptions. Metaphorical mappings between domains can influence not only how we understand but also how we interact with and navigate the world around us. If this integration is considered while studying sign languages, Lakoff and Johnson's propositions (as presented in section 2.2.3) will hold great significance, stemming from the very foundation of sign languages, which are rooted in signers' experiences of the world, a world they not only perceive but also *describe* through the visual-manual system.

Furthermore, in terms of the influence of extralinguistic societal factors in constructing metaphors, Russo Cardona (2005) theorizes that not only bodily experiences influence language metaphors, but so does the shared knowledge of Deaf communities. In particular, he describes that shared cultural knowledge is required to understand Deaf signers' metaphors, highlighting the significance attributed by Deaf individuals to vision. In LIS metaphors, the iconic quality of Signs allows for embodied and expressive representations within a merged mental space formed through cross-domain mapping. Understanding these metaphors calls for a deep cultural understanding linked to the Deaf experience, enabling the creative interaction of specific features from each domain in the blend. Additionally, within LIS metaphors, vision assumes a notably crucial role as it is conceptualized as a multifaceted instrument for both elaborating upon and reshaping knowledge, demonstrating its involvement in the cognitive and expressive contexts of sign language communication (Russo Cardona, 2005). The examples provided by Russo Cardona regarding the heightened importance of vision within LIS metaphors highlight the deeper layers of meaning that arise from the integration of embodied experiences and cultural insights in shaping sign language expression and understanding.

The focal question becomes, once again, *what is* language? This reevaluation of language-defining concepts in contemporary linguistic enquiry, where action, gesture, and Sign exhibit continuity, positions iconicity as a fundamental aspect of linguistic exploration, offering a renewed, embodied perspective on languages. In functionalist theories, language, both as an entity and in its constitutive elements, cannot be perceived as formal, rigid, or fixed. There is not a moment in language use where bodily perceptions and contextual information are excluded. As such, a disembodied perspective of language cannot coexist with languages that are crafted through the human body. By asserting this point, a connection with the definition of functionalism embraced in this study is recovered (Bates *et al.,* 1991), connecting and making sense of linguistic enquiry as based on languages that created, governed, constrained, acquired and used in the service of embodied communicative functions.

Moreover, the emphasis put by cognitive linguistics on the relevance of metaphors provides a precious framework for investigating the conceptual structures of languages, revealing how iconicity is manifested in both vocal and signed linguistic expressions through embodied mappings. In asserting so, we find that iconicity is a quality of language that manifests at varying degrees depending on the language modality, whether vocal or signed, in more overt or covert ways.

Christian Cuxac, a French linguist, significantly contributed to the process of redefinition and reevaluation of iconicity in sign languages, considering it a fundamental aspect of signed communication. In particular, Cuxac developed the *semiogenetic approach* (Cuxac, 2000) for LSF (*Langue des Signes Française*) and categorized iconic productions in LSF as Transfer Units (*Unités de Transfert*— TU*)* and classifies them into three categories: Transfer of Size and Form, Transfer of Situation, and Transfer of Person (Cuxac & Sallandre, 2002; 2007). Elena Antinoro Pizzuto later contributed to prove that Cuxac's model is able to account for the description of all sign languages, and not just LSF (see Cuxac & Antinoro Pizzuto, 2010).

The categorization of these structures offers a distinction between various iconic strategies employed in sign languages to convey meaning: telling *without* showing and telling *through* showing. Telling without showing involves using Lexical Units (*Unités Lexematiques*— LU), while telling through showing employs TU. Cuxac's model highlights the significance of body components in signed discourse, particularly the direction of eye gaze, an element that distinguishes standard Signs from a TU, with the signer directing their gaze toward the interlocutor in the former case, and toward their hands or assuming the eye gaze of the represented entity in a TU (Volterra *et al.,* 2022). Regarding the relationship between semiotics and iconicity, Cuxac suggests that the manual and body components comprising TUs carry inherent meaning (Cuxac & Antinoro Pizzuto, 2010). Cuxac theorizes that the existence of similar Signs across different sign languages emerges from signers' everyday experiences. Signs evolve as signers select relevant information pertaining to a referent. In this process, signers identify pertinent attributes of a referent, aligning with the norms of the linguistic system they employ.

Forty years of research on language as a social act and language as embodied, perhaps find a contemporary unification in Annelise Kusters' enquiries, encompassing embodied cognition, metaphor studies and the examination of language as a multimodal action. Kusters carries out ethnographic studies on multimodality and semiotic repertoires, encompassing both spoken and signed languages. She writes that:

> […] all human interactions, and linguistic repertoires, are (and always have been) multimodal. Language in use, whether spoken, signed or text, is always and inevitably constructed across multiple modes of communication and through 'contextual' phenomena such as the use of the surrounding physical spaces.
> (Kusters *et al*., 2017, p. 2)

The re-introduction of the concept of language as multimodal takes us back, once again, to the journey of re-evaluation of para and extralinguistic elements that linguistics has gone through in the past six decades. Kusters aims at going beyond the notion of languages existing as isolated and unique systems, thus shifting the discourse towards the multimodal and multilingual approach. This perspective acknowledges repertoires— i.e., the tools and means available to language users (Kusters *et al*., 2017, p. 4)— as: «[…] a heteroglossic realm of embodied potentialities and constraints» (Kusters *et al*., 2017, p. 10). Specifically, the repertoire of Deaf signers interested Kusters due to its peculiar communicative context as it has been observed that signers can use several linguistic varieties (both signed and spoken) thanks to the constant Sign/speech, Sign/writing and Sign/Sign contact (Kusters *et al*., 2017, p. 6).

2.2.5. The functionalist approach to sign language studies: in favor of a cognitive and socio-semiotic approach

In describing the qualities of the contemporary formalist approach to sign language studies, Volterra *et al.* (2022) outline a set of fundamental components characterizing the COS-S model as employed the Italian context. A notable shift is highlighted, stating that «[…] whereas during early-stage research on LIS we reiterated the difference between Signs and gestures, nowadays we find ourselves in a very different theoretical perspective […]» (Volterra *et al.*, 2022). Indeed, through the COS-S paradigm, sign languages are examined by exploring into how meaning is constructed by signers through their communicative practices (Fontana & Roccaforte, 2019). Additionally, both production and perception are profoundly influenced by individual users' ideologies and awareness (Fontana, 2022a).

The present section will focus on the discussion of the core traits of this theoretical paradigm applied to sign languages, drawing a continuous line with sections 2.2.3. and 2.2.4. of this thesis work. In doing so, it will be highlighted how functionalist models are pertinent to sign languages, while also identifying their methodological distinctions from earlier approaches in sign language studies.

Firstly, following an embodied perspective, body components play a crucial role in constructing meaning within sign languages. Therefore, aspects like facial expressions, posture, eye gaze and oral components ought to be given significant consideration when studying sign languages. For instance, in LIS, the recipient can distinguish between a question and a statement by observing whether the signer raises their eyebrows or not. This exemplifies how body components hold a similar significance to the way prosody functions in vocal languages, entering the context of the paralinguistic. The postulation that manual and body components are all fundamental aspects of signed productions finds its concretization in the description of the mechanisms used by LIS signers to create meaning. In the scheme developed by Volterra *et al*. (2022) Cuxac's model (2000) is adapted (see section 2.2.4.) by identifying the mechanisms of: pointing (using Deictic Units), telling (using Lexical Units) and showing (using Transfer Units).

Moving forward, iconicity is not solely confined to shaping Signs, but also holds a pivotal role in the formulation of utterances. Moreover, the approach goes beyond Stokoe's description of sign language components as cheremes[20], which are no longer viewed as isolated, meaningless units within a sign, but are instead seen as intrinsic components of a continuous signing flow, where manual and body components acquire equivalent significance. Importantly, the scope of iconicity extends past sign languages, as it is also observed in vocal languages. The concept of iconicity, a dynamic interplay between form and meaning, has broadened its influence across various linguistic modalities[21]:

> […] human beings can express themselves by means of pointing, describing and depicting. Through pointing, it is possible to locate and recall a referent in space; through words or signs of the lexicon, it is possible to describe or label meanings; through illustration or depiction, it is possible to show immediately what people are talking about. (Volterra *et al.,* 2022, p. 29)

Additionally, the functionalist COS-S paradigm rejects the application of investigative methodologies originally developed for vocal languages to the study of sign languages. Historically, researchers attempted to identify structures inherent to vocal languages to depict signed languages. However, this perspective diverges from this standpoint, acknowledging that each human language possesses its own distinctive qualities.

[20] Stokoe (1960) Stokoe *et al.* (1965) characterized cheremes as the building blocks of ASL signs. He compared them to the phonemes in vocal language: minimal units without intrinsic meaning, which can be combined to form signs.

[21] Think of onomatopoeic sounds.

Treating all languages with uniform analytical tools, particularly those designed for languages of differing modalities, is no longer considered entirely suitable.

Equally significant is the spotlight cast on community dynamics. As languages function within specific communities, they exercise a profound influence on how and whether languages evolve. Hence, an imperative arises to refrain from analyzing sign languages (or any languages) in isolation from their corresponding communities. Recognizing the symbiotic relationship between language and community unveils a more comprehensive understanding of the interplay between linguistic evolution and sociocultural contexts (Volterra *et al.,* 2022; Fontana & Roccaforte, 2019; Fontana, 2022a).

Having described the theoretical framework of the functional COS-S model as applied to sign language studies, an observation emerges. The model's applicability extends, in fact, to both signed and vocal languages (Volterra *et al.,* 2022). Hence, it becomes evident that this perspective works bidirectionally. On the one hand, it draws from a theoretical background established for languages by reviewing the described theories within the context of sign language studies. On the other hand, it leads to a methodology that that results in a new and fresh perspective for studying all languages, signed and vocal alike. This dual nature of the approach showcases its adaptability and versatility, reworking existing linguistic theories while also generating fresh insights that enrich our understanding of human language in its entirety (Cuxac, 2000; Russo Cardona, 2004b; Volterra *et al.,* 2022). Therefore, scholars adapting in the COS-S approach direct their focus towards the unique features that define sign languages. This choice is guided by the understanding that certain aspects, inherent to sign languages, can find explanations when examined through the lens of fundamental cognitive linguistic principles. These foundational principles will be described in the following paragraphs.

2.2.5.1. On iconicity

As mentioned in Chapter 1, in early sign language research, iconicity was seen as an element inherited from gestuality and, therefore, as a characteristic that had to be minimized to prove sign languages' status as natural languages, in a context dominated by a phonocentric majority, even to the point of being seen as an 'embarrassment' (Pietrandrea, 2012, p. 270). In the contemporary COS-S framework however, as discussed by Cuxac (2000) (see section 2.2.4), iconicity is valued as a fundamental consequence of semiotic strategies of Sign formation.

Within the Italian context, an effort to define iconicity is LIS was made, once again, by Tommaso Russo Cardona (2004a) in a paper where he described two types of iconicity in sign languages: *frozen* and *dynamic*. The former, frozen iconicity influences the relationships between form and meaning, while the latter, dynamic iconicity, affects the correlations established between forms and meanings within specific utterances, textual contexts, and situational settings (Russo Cardona, 2004a, p. 167). Russo Cardona further illustrated dynamic iconicity, proposing that forms and meanings can establish various types of relations by exploiting the order of Signs in an utterance, as well as variation in manual and body components. He conducted an analysis on a small corpus of poetry, dramatized narratives, and conferences delivered by native LIS signers (see section 3.2.3.). This examination led him to identify distinct forms of dynamic iconicity. Firstly, recast iconicity, which involves the use of the same handshapes to denote different relations between a Sign and its referent, with the handshape's meaning being recast based on the context. Secondly, iconicity of parameters in discourse, which encompasses iconic features discernible in the location, movement, and orientation of a Sign, which mirror the relationships between two represented referents. Thirdly, iconic reverberation: which, in the context of signed poetry, denotes a pattern that employs similarities and differences in the manual components of Signs to convey specific semantic connotations. Russo Cardona (2004a) found that, within the analyzed corpus, dynamic iconicity was present at different levels in all the analyzed registers (although at different percentages). His findings contribute to a broader understanding of how signers employ both iconic and conventional components to construct meaning in signed utterances, by clarifying the strategies that signers employ to convey subtle nuances, create semantic connections and enrich utterances through the interplay of form and meaning at different registers. Moreover, Russo Cardona's exploration of iconicity challenges traditional assumptions and contributes to a paradigm shift in the field as he states that the interrelation among arbitrariness and iconicity in sign languages is even more evident, as sign language linguistics is embodied, due to the visual-gestural modality of signed linguistic systems.

Consequently, iconicity and the embodied perspective are able to explain why Signs referring to the same entity may be similar among sign languages. It is precisely due to their visual-manual modality that sign languages can imitate reality. Hence, despite their cultural differences and due to their shared biological features, human beings will experience reality in similar ways. As Volterra writes: «The world around us is crawling with visible objects. Thanks to the visual-gestural modality it is possible to discuss 'visible things' using visible Signs, and this increases the similarity between signifier and signified» (Volterra in Elmetti, 2010). Due to their visual-manual modality, sign languages are believed to be more visually iconic than spoken ones.

On the topic of iconicity in vocal languages, Winter *et al*. (2017) write that even if these languages are superficially less iconic than signed ones, a balance between iconicity and arbitrariness is found. Additionally, iconic English words seem to belong to specific semantic fields– a trait shared with iconic Signs as well– namely the auditory (to hiss), tactile (to crash), olfactory (to sniff), gustatory (to suck) and visual domain. Therefore, the experience of the surrounding world will influence the strategies used to describe it with language, regardless of its modality. This can be stated both for abstract and concrete concepts, such as dreaming or eating[22] both articulated over the neck in the upper or lower part of the signer's head, due to a process that led to the identification of pertinent features of a referent based on the signer's experience.

Now, thanks to the extensive research on the topic of language acquisition in deaf and hearing children, the central role played by gestures in language acquisition has been noted, whether it be vocal or signed. A particular focus is now placed on deictic gestures. Deixis is not only a central means used by children in the preliminary stages of language development, but also one of the iconic strategies used in sign languages to create significance (Antinoro Pizzuto *et al.,* 2006; Volterra *et al.,* 2022). Using a semiotic perspective, two studies were published in the early 2000s. As mentioned in previous sections, as well as in the present one, Russo Cardona reflected on the productivity of iconicity in sign languages (2004a) and Cristian Cuxac described a string of TUs (2000), that is multilinear structures comprised of manual and body components that are strictly linked to the context in which they are found. As Elena Antinoro Pizzuto and colleagues write, TUs are used to introduce a referent in discourse for the first time through deixis, or to refer to something or someone that has already been mentioned through anaphora. TUs are fundamental to the construction of a cohesive signed text:

> Thus, to introduce a referent in discourse, a signer can produce a standard (or 'frozen') manual sign for the referent, and deictically mark it in the signing space via a manual and/or a gaze pointing, and/or also via morphological alteration of the sign's place of articulation (which may or may not be accompanied by visual indexes), thereby establishing a position in space (or 'locus') for the referent symbolized. Anaphoric reference is then made re-indexing roughly the same point in space via visual or manual pointing.
> (Antinoro Pizzuto *et al*., 2006, p. 477)

[22] See the signs for 'eat' and 'dream' in different sign languages on the website SpreadTheSign (https://www.spreadthesign.com/it.it/search/, accessed on October 12th, 2023). These examples are taken from Sabina Fontana's lecture (as part of the *Italian Sign Language Linguistics* 3 course) on May 21st, 2021

Regarding deixis as an iconic strategy with an anaphoric function, Inoue (2006) highlights that signers have the ability to use the physical location associated with both the current individual and an absent, mentally conjured referent that has been introduced in the preceding context. This phenomenon exemplifies the concept of embodied cognition, as the use of preloaded mental representations can potentially alleviate cognitive demands by leveraging the surrounding environment. Beyond their productivity, TUs are embodied strategies that highlight a peculiarity of sign languages, that is the ability to include the context in a signed discourse. Thanks to these strategies that the context acquires linguistic significance by becoming part of the co-text.

In conclusion, the interplay between embodied cognition, iconicity and semiotic strategies highlights the intricate connection between language, the body, and the world we perceive. The conventional belief that puts shape and meaning in a strictly arbitrary correlation is giving ground to a new understanding that acknowledges the coexistence of both arbitrary and iconic components in all languages, as Signs and words may be both motivated and conventional by having an iconic motivation which is also the result of an arbitrary cultural choice (Volterra *et al.,* 2022). Ultimately, sign languages exemplify how the embodied nature of communication shapes linguistic structures and fosters a deeper integration of context within linguistic expression.

2.3. Balancing structure and variation in the description and analysis of sign languages

In this concluding section aims at navigating through the two predominant branches of linguistic inquiry prevalent in contemporary research and their implications when applied to the study of sign languages.

The formal perspective on grammar, as discussed in section 2.1., depicts its operation as a 'vacuum' that systematically organizes input from the environment to generate language output. This rationale originates from Chomsky's aspiration to consider linguistics as a 'hard science' and compare it to chemistry, where abstract concepts are not dwelled upon (McElvenny, 2019). Indeed, adopting a strict and formal approach in describing and analyzing sign languages presents both advantages and disadvantages, largely due to its inherent rigidity. The rigorous nature of this approach eradicates ambiguity and variation, offering a precise portrayal of sign languages, which can significantly benefit linguistic inquiries. Müller de Quadros and Lillo-Martin (2010) support this notion, emphasizing that the formalism inherent in this approach enables the formulation of specific, detailed inquiries and explicit proposals. Their observations and generalizations are firmly grounded in empirical evidence, making them pertinent not only to linguists following a similar approach, but also to those employing different methodologies. The advantage thus lies in the precision and clarity that such formalism offers, allowing for detailed exploration of specific inquiries into sign languages.

This methodology responds to two objectives: firstly, it aims to provide a brief depiction of the fundamental characteristics of a sign language within the framework of generative theory. Secondly, it highlights the essential role of computers and technology in sign language research by facilitating thorough and systematic analyses of signed productions (Chesi & Geraci, 2009). In essence, adopting formalist approach involves describing sign languages using existing tools and methodologies, thereby confining languages within delimited yet productive frameworks.

With relation to cognition, Chomsky's formalist approach resonated with Fodor's modularity (1983), which proposed that the human mind operates through specialized modules, a concept that influenced Karmiloff-Smith's development of modularization as a constructive process, highlighting the intrinsic connection between how the human mind functions and the visual sensorimotor inputs received by the body and suggesting that action, language, and cognition share common neural systems (Fontana, 2009).

In this framework Fodorian modularity and the idea that the information received by the brain is processed through exact modules in a vertical manner is overturned and reimagined (Piazza & Pavani, 2022).

Beyond modularization, cognitive linguistics theorists and functionalists, in general, may argue that the lack of focus on context, semantics, variations and other modalities may be an element of far too great importance to be dismissed in favor of universality. Shapiro (2011) maintains that one of the main weaknesses of formalism is, in fact, the division between brain and body. If the perspective is extensively 'brain-centric', then it automatically excludes the body and the world that are, according to functionalism, fundamental parts of any cognitive processing system.

Langacker argues that as language is neither self-contained nor well-defined, then a complete formal description is impossible in principle (2008). The impossibility of applying generative linguistics as an all-encompassing framework is the common thread for functionalists who see formal UG's principles as *tendencies* rather than strict universals whose diffusion is motivated by a colonizing ethnocentric approach that does not consider variation. Consequently, differences between languages are, according to functional theories, not merely superficial, and cannot be resolved by identifying deeper structures (Evans & Levinson, 2009).

As the goal of this thesis is to present a unified and comprehensive approach to both the collection and processing of sign language data, the methodology discussed and employed for the collection and processing of a multimodal LIS Dataset (see Chapter 6) will be derived from a combined employment of formalist and functionalist methodologies, specifically tailored for studying sign languages. In particular, the present work seeks to merge these methodologies to find balance between rigidity and flexibility.

The discussion in section 2.2. highlighted the reconsideration of para and extralinguistic elements, advocating for a multimodal, cognitive, and semiotic approach in linguistic investigation. This approach stands in contrast to the formalist viewpoint discussed earlier in section 2.1. While these viewpoints have been depicted as oppositional in this chapter, following conventional descriptions, this portrayal within the present work is deliberate. It serves the purpose of offering a broad overview of these perspectives in linguistic inquiry. However, this presentation of opposition is not intended to solidify their divergence, but rather to open the door for a subsequent joint application to LIS. In particular, the goal of the present work is to transcend this dichotomy by introducing a methodological approach that integrates elements from both perspectives. By combining aspects from both methodologies, the intention is to create a comprehensive framework specifically tailored for the collection and processing of LIS data. This approach aims to harmonize the strengths of both formalist and functionalist methodologies, thereby offering a more holistic and inclusive method for studying sign languages.

2.3.1. Methodologies and constraints in implementing the proposed balanced approach

Now that the theoretical foundation for the present thesis has been laid out, and the argument for a balanced perspective between formalist and functionalist views has been made, the following section will focus on explaining how this inter-theoretical framework and proposed approach are pertinent to this specific work.

Ad discussed in Chapter 1, the primary practical goal outlined in this thesis revolves around the collection and processing of LIS data. Expanding upon this goal, there is an additional layer to this statement: it involves seeking practical applications for functionalist theories, specifically within the framework of COS-S approach to LIS. With the aim of reaching this goal, the approach adopted in this work maintains aspects of formalist theories, mainly due to constraints faced by those involved in sign language processing and annotation. These constraints often include the necessity for alphabetic annotation or enabling data to be processed further within contexts of automatic recognition and translation. The present section will therefore provide an answer to following question: How will these goals be achieved?

Functionalism, within the context of this thesis, is intended as characterized by its focus on the interconnectedness of language, the human body, and contextual inputs. It perceives language as a dynamic system embedded in social interaction, necessitating a holistic approach that extends beyond grammar and syntax. Specifically, ethnopragmatic theories find application in the three collection processes undertaken in this thesis, leading to the creation of the three MultiMedaLIS Datasets, with detailed exploration of their collection processes related in the following chapters (see section 6.3.1). It has been seen that ethnopragmatics highlights the importance of studying language as an event occurring within specific sociocultural and historical contexts (Hymes, 1974; Goddard, 2006; Duranti, 2007). The theoretical and practical pragmatic approach to LIS research and collection, as advocated by Russo Cardona (2004a; 2004b), significantly shaped the data collection processes in this work, heavily dictating the selection of the recording tools used during the data collection efforts.

In regard to data collection modality, it is obvious that LIS, as a visual-manual language must be documented through video. Beyond stating this, the selection of portable and multimodal tools was performed by embracing functionalist perspectives. The portability of the recording tools aligns, once again, with ethnopragmatic views. It enables the recording of LIS in spontaneous, natural, and even domestic contexts[23],

[23]To this regard, it is important to highlight, as will be elaborated in the following chapters, that these statements are made with the perspective of repeating the data collection. In fact, the three data collections outlined in this work were not gathered within natural contexts. This motivates their description as 'preliminary' Datasets.

allowing for its capture within settings where extralinguistic elements are genuinely perceived. These chosen tools are both portable and adaptable, facilitating the capture of LIS in spontaneous signing contexts. This approach aims to preserve the authenticity of LIS in its natural settings, giving value to the functionalist emphasis on language as a dynamic, socially embedded phenomenon shaped by real-life interactions.

On the other hand, in relation to the strict multimodality of the data collection tools, these were selected following the COS-S model to sign language studies and its emphasis on sign language multimodality and multilinearity (Volterra *et al.,* 2022) along with their potentials and constraints (Kusters *et al.,* 2017). Understanding that sign language encompasses not only manual elements but also equally significant body aspects, within this framework, is imperative to envision a language collection process that equally captures multimodality and multilinearity. Furthermore, the tools were chosen to capture LIS in a three-dimensional manner.

In terms of annotation of the third version of the preliminary Dataset, the approach involved the use of language-specific annotation tools. This decision stemmed from the functionalist concept of describing languages within distinct frameworks. By employing tools tailored specifically for LIS annotation, the goal was to capture the unique characteristics and structures of LIS, while accounting for its visibly embodied nature.

However, any annotation process comes with constraints, primarily due to the very nature of the task of annotating. In fact, this process involves adding detailed information to signing, carefully weighing considerations related to factors such as the oral nature of LIS, the alphanumerical constraints imposed by the processing systems themselves and, crucially, potential applications of the annotated data. Chapter 2 highlighted how various linguists have pursued the identification of linguistic universals in sign languages. They conducted formal analyses of Signs and signed utterances (Geraci, 2004; Sandler & Lillo-Martin, 2006; Branchini & Geraci, 2011; Mazzei *et al.,* 2013), obtaining tangible outcomes in formal descriptions of sign languages that transcend modality. This aspect is crucial when handling sign language data, as the ability to generate concrete, annotated and readable outputs is a significant advantage of this methodology. Furthermore, it sets the groundwork for developing automatic translation tools, whether in textual or visual formats (see Geraci *et al.,* 2014). Consequently, for those aiming to produce annotated data that is both machine and human-readable (as in Research Question 3 of this thesis), these ideal outputs become essential considerations. Hence, the presence of some form of alphanumerical representation becomes necessary, leading to the employment of vocal language labels. This approach not only aids in making annotations accessible to non-signers coming from different linguistic backgrounds, but also aligns with the goal of creating machine-readable data by positing a sign/word non-contextual equivalence.

The annotation process, as detailed in upcoming chapters, will undergo some adaptations to accommodate the multilinearity inherent in LIS, by employing multilinear annotations within the used annotation software. Through the combined use of vocal language labels, and a language-specific typographic system, multilayered annotations will be generated. These annotations will facilitate formal analyses are able to identify commonalities and differences among Signs, focusing on the codification of manual and body elements' positions and movements, rather than on the transcription of the Sign. In employing computational tools capable of quantitatively analyzing data in terms of recurring patterns and sign structures, this approach aims to demonstrate the feasibility of offering a detailed breakdown of linguistic elements within LIS through language-specific annotation tools.

Lastly, it is important to note the consistent use of the term 'Sign' in its singular form throughout this section. Since the preliminary Datasets were developed to test suitable tools and methodologies for collecting and processing LIS, Signs were recorded in isolation (hence, their preliminary nature). This approach allows for formal inquiries by constituting isolated units, avoiding the complexity of balancing formalism and functionalism in syntax description, given the differing emphasis on syntax in these models. Future expansions of the data collection including utterances might prompt discussions on syntax integration.

Chapter 3
Collecting and processing sign language data

Having introduced contrasting approaches to sign language studies and proposing a connection between them in the previous chapter, this chapter redirects its focus on the discussion of various aspects related to sign language collection and processing. The term ‘sign language processing’ encompasses a wide array of tools and models used to collect and analyze sign language data. In the context of the present work, given that the tangible output will be a multimodal collection of LIS Signs gathered in isolation, the creation of a semantic framework to depict LIS in a computationally viable manner (Schubert, 2020) is of interest. Specifically, Signs within the third version of the Dataset (see section 6.3.) underwent initial processing steps, including segmentation and labeling of Sign tokens (i.e., individual units), with the aim of facilitating further analyses and ultimately producing data suitable for machine translation tasks between Italian and LIS.

Consequently, this chapter aims to explore the evolution of computational linguistics, exploring language collection for processing. The establishment of its historical and contemporary context will pave the way to understand the creation and processing of the presented preliminary Datasets. Beginning with an exploration of the fundamental tasks in Natural Language Processing, the focus will then be shifted towards the application of these tasks within the context of sign languages. Emphasis will be placed on the significance of corpus linguistics methodologies in multimodal data collections and processing efforts. Additionally, considering the central role of video tools in sign language data collection, a discussion on computer vision will be included.

3.1. Sign language processing: challenges and innovations

From a historical standpoint, computational linguistics emerged significantly during and after World War II, when the concept of using electronic computers for cross-linguistic communication was introduced (Weaver, 1955). Nowadays, computational linguistics is defined as an interdisciplinary field concerned with the study of natural languages from a computational perspective. In particular, it is a subarea of computer science which deals with the production and understanding of language (Hausser, 2001) or, more generally, with the transmission of language, and is mainly interested in Natural Language Processing (NLP) and artificial intelligence (Church & Liberman, 2021). For this reason, the general goal of computational linguistics is «[…] to reproduce the natural transmission of information by modeling speaker's production and the hearer's interpretation on a suitable type of computer»[24] (Hausser, 2001). As mentioned, among the practical goals of computational linguistics is NLP which specifically focuses on the development and application of computational techniques to analyze, understand, and generate language data. Broadly, NLP aims to help computers understand language as well as we do (Donges, 2021). Large data collections play a significant role in NLP, as its applications are built based on structured collections of data are fundamental in training algorithms for tasks like language understanding and recognition. These large data collections can come in the form of corpora. More specifically, a corpus is an organized and computerized collection of authentic language obtained from various communicative scenarios that is used for linguistic analyses (Thanaki, 2017). Any kind of linguistic investigation that relies on the data derived from a corpus can be referred to as corpus linguistics (Stefanowitsch, 2020). To this regard, it is important to mention that in the context of this work, corpora collection and annotation methodologies are of central importance, as they lay the ground for the practical collection and annotation of LIS. However, since the preliminary data collections presented in Chapter 6 are not authentic, in that they collect Signs in controlled contexts, a decision has been made to refer to the

[24] Hausser's (2001) focus on language transmission through hearing and speaking can be viewed as indication of a historical lack of inclusion when it comes to the recognition of sign languages in this context. This omission highlights a broader trend of neglecting sign languages in scholarly discourse until more recent times.

outputs of the data collection efforts as ‘datasets’, intended as a collection of isolated Signs specifically related to the medical semantic domain.

Typically, a corpus is required for carrying out Part of Speech (POS) tagging. POS tagging consists of assigning different labels (POS tags) to the words of an utterance to provide information on the part of speech of each word (or token). It is usually a step that precedes parsing (Sharma, 2020) which, in turn, consists of assigning a structure, often represented in the form of a tree, to a sentence. Parsed sentences can be collected into treebanks, i.e., parsed corpora of syntactically annotated sentences. Different tagsets have been developed for the definition of constituents when analyzing natural languages. The most famous treebank based on constituency relations is the Penn Treebank (Taylor *et al.,* 2003). Another successful taxonomy of dependency relations has been developed as part of the Universal Dependencies (UD) framework. The UD project aims at developing universal annotation strategies to «[…] support comparative evaluation and cross-lingual learning […]» and to «[…] facilitate multilingual NLP and enable comparative linguistics studies […]» (Nivre *et al.,* 2016, p. 1659). Within this framework, sentences are first analyzed at the level of words (NOUN, VERB, ADJ, ADV, DET) and then dependency relations are identified from a tagsets of forty possible dependency relations (Petrov *et al.,* 2012).

When referring to corpora as large collections of language or using corpus methods to process linguistic data, the conventional reference is to written or spoken corpora (Adamou, 2019) which primarily encompass languages with a longer tradition of metalinguistic considerations. In fact, Crawford and Eniko (2016) state that corpus linguistics studies traditional models of language production: written and spoken. Modalities that, as arleady clarified, can be easily applied to spoken/written languages, but require language-specific processing for sign languages. However, when considering sign languages, it is crucial to highlight the initial challenges faced in constructing such a corpus due to the unique characteristics of these languages such as, for instance, the need to include visual data. This challenges not only highlight the need for innovative methodologies in sign language corpus construction but could also prompt a broader conversation within corpus and computational linguistics in accommodating diverse modalities and representation systems.

In the context of the present thesis, analyzing corpus applications to sign language studies is a fundamental step, as it will provide a background for establishing good practices and guidelines in processes of sign language data collection and annotation, a central topic of this work.

3.2. Corpus Linguistics: from foundations to sign language analyses

Corpus linguistics originates in the pioneering works of Father Roberto Busa: a Jesuit priest who worked on the creation of an index of all of St. Thomas Aquinas works, known as the *Index Thomisticus*[25]. Busa's work began in the year 1950, which marks the beginning of computational work on texts (Tognini Bonelli, 2001). Ten years later, in the 1960s, the first electronic corpus of written American variety of the English language was published as the *Brown Corpus*, compiled at Brown University, containing around a million words of printed prose in the US, during the year 1961 (Francis & Kucera, 1964). Thanks to the release of tape recorders, the University of Edinburgh collected an electronic corpus of vocal language, comprising more that 150,000 words taken from informal conversations in English (Tognini Bonelli, 2001). Therefore, in the 1960s, linguists started building corpora of both written and spoken texts. Since the beginning of corpus linguistics, this subject has been anchored to the theoretical paradigm of the empiricist approach, based on the observation of naturally-occurring and authentic data (Assunção & Araújo, 2019). Moreover, corpus linguistics has flourished in the recent years becoming a lively subject with multilingual (and more

[25] https://www.corpusthomisticum.org/it/index.age (Accessed on November 18th, 2023).

recently, multimodal) collections of languages and language varieties that are not used solely to study the grammar of a language, but to study and collect the language itself (McEnery & Wilson, 1996). Coincidentally, corpus linguistics has shown a shift from the description of language as *langue* rather than *parole,* an aspect which has been criticised by Chomsky, who stood for the description of language as competence rather than performance, thus favouring the I-Language rather than its use, as performance is often not representative of competence. Moreover, Chomsky has argued that corpora could not be considered as evidence of how a language works and, consequently, could not be sees as tools fit for linguistic enquiry (Chomsky in McEnery & Wilson, 1996, p. 10). In contrast, Tognini Bonelli states that the assumption, when it comes to the definition of corpus linguistics, is that the collection is gathered according to specific design criteria and with a specific purpose in mind. Moreover, *corpora* allow those who access them to observe facts in languages through concrete examples, leading to the possibility of generalisation of one or more phenomena, based on the repetition of patterns (2001).

One topic that needs discussing is the application of corpus linguistics to sign language collection and analysis. Crawford and Eniko (2016) provide an example of corpus linguistic applications in their introduction to the subject. They present a comparison in the use of the English words *equal* and *identical* within the *COCA Corpus* (*The Corpus of Contemporary American English*), comprised of more than one billion words of text collected from 1990 to 2019. Making such a comparison, i.e., between two synonimic 'words' in sign languages, requires muli-layered considerations and analyses, as this comparison might not translate to sign languages easily. In fact, in comparing between *equal* and *identical*, Crawford and Eniko (2016) rely on a well-stablished written representation of English words, with an analysis conducted on a corpus that spans almost thirty years. However, sign languages lack a universally accepted writing system and, historically, researchers have relied on vocal language labels and annotation attempts to represent Signs in written forms, strategies which do not capture their multimodal and multilinear richness. Additionally, the documentation of sign languages is often not as extensive as that of spoken ones, as many sign languages lack large-scale corpora. In conclusion, a simple comparison among English words from *COCA Corpus* opens the discussion on the challenges that arise when applying similar methodologies to sign languages.

3.2.1. An overview of spoken Italian Corpora

The concept of corpus encompasses some key characteristics. In its broadest sense, in the context of sign language linguistics, a corpus has been described as a collection of text, whether signed, spoken, or written, that serves as the basis for linguistic analysis (Fenlon & Hochgesang, 2022). Moreover, a corpus should be machine-readable, meaning that the data it contains can be readily searched, retrieved, and organized (McEnery & Wilson, 1996). Based on the subject's name, it can be stated that corpus linguistics *is* a branch of linguistics. However, it must be noted that this belonging is of tricky nature. Corpus linguistics cannot be classified as a branch of linguistics as syntax, morphology, semantics and so on. As can be inferred, all these branches are concerned with the description of one or more aspects of language, whereas corpus linguistics is a methodology that can be applied in most areas of linguistic research, to the point that it can allow us to differenciate between approches taken to the study of language (McEnery & Wilson, 1996). In today's landscape, the union of technology and linguistics has allowed reaching new frontiers of linguistic investigation. Moreover, the effectiveness of language processing tools strongly depends on the quality and volume of language-related data available.

Back in the late 1980s, researchers made an effort to document recurring linguistic patterns in both written and spoken Italian. This was partly inspired by Tullio De Mauro's work on understanding how people use language in their daily lives. As a result of these efforts, corpora of spoken and written Italian were developed to document linguistic phenomena including diamesic varieties, and partially accounting for the historical evolution of the language (Barbera, 2013).

In regard to written Italian[26], one key resource is the *Contemporary Written Italian Corpus* (CORIS/CODIS— *Corpus di Italiano Scritto Contemporaneo*), which was established in 1998 at the University of Bologna. This corpus was designed to be a representative collection of Italian, similar to the well-known *Brown Corpus*. Over a period of two decades, from the 1980s to the 1990s, this corpus was updated every three years. New material was added through the inclusion of real electronic texts that were chosen because they reflected spoken Italian well. In total, over 100 million words were collected from sources like the press, fiction, academic writing, legal documents, and miscellaneous texts[27]. Another resource is the Lexicon and Frequency *Corpus of Written Italian* (CoLFIS— *Corpus e Lessico di Frequenza dell'Italiano Scritto),* which comprises 3,798,1275 words from books, newspapers, magazines, and even data from the Italian National Institute of Statistics (ISTAT— *Istituto Nazionale di Statistica*), all published between 1992 and 1994[28]. For studying the historical progression of Italian, the *Diachronic Corpus of Written Italian* (DiaCORIS— *Corpus Diacronico di Italiano Scritto Contemporaneo)* includes texts written between 1861 and 1945, offering insights into how the language has changed over time (Onelli *et al.,* 2006).

Turning to vocal language, the *Linguistic Corpora for Spoken and Written Italian* (CLIPS - *Corpora Linguistici per l'Italiano Parlato e Scritto*)[29] is a significant resource. Collected between 1999 and 2004 as explained by Albano Leoni (2007), this corpus is substantial, featuring 100 hours of audio recordings. It stands as one of the most important Italian language resources (Morlicchio *et al*., 2021) with the aim of

[26]In this section, the focus is exclusively on Italian corpora, as they offer the closest parallel to sign language corpora in the Italian context.
[27] https://corpora.ficlit.unibo.it/coris_eng.html (Accessed on September 3rd, 2023).
[28] https://www.istc.cnr.it/en/grouppage/colfiseng (Accessed on September 3rd, 2023).
[29] http://www.clips.unina.it/it/ (Accessed on September 3rd, 2023).

capturing various regional, social, stylistic, and individual language variations. This corpus includes both male and female voices and is organized into distinct categories. It is partially transcribed, segmented, and phonetically annotated, enhancing its usability. The corpus is further divided into diatopic and diaphasic varieties. Diatopic covers different geographic locations spanning from Milan to Palermo, while diaphasic includes radio talks, spontaneous dialogues, and phone conversations, encompassing a wide range of vocal language situations. Collectively, these corpora encompass only a fraction of the available resources for spoken Italian. Nonetheless, their existence constitutes a contribution towards enhancing our comprehension of both spoken and written Italian, attesting for its diamesic and diachronic variations[30].

In the past years, linguists engaged in the collection of spoken Italian data have been confronted with the necessity of developing tools and methodologies for analyzing spoken material across various dimensions. This includes the creation of frameworks to transcribe vocal language, and suggesting standardized coding systems (Morlicchio *et al.*, 2021)[31]. Furthermore, this interest in vocal language corpora has given rise to collaborative efforts between linguists and computer scientists, thus promoting interdisciplinary partnership. Despite their different objectives, these groups share a common goal: the systematic collection of languages. This shared interest serves a twofold purpose. On one hand, it promotes linguistic research aimed at understanding vocal languages. On the other hand, it serves as a stimulus for advancements in

[30] For an updated list of data banks, corpora end archives, see the list of *Accademia della Crusca*, https://accademiadellacrusca.it/it/contenuti/banche-dati-corpora-e-archivi-testuali/6228 (Accessed on September 3rd, 2023).

[31] «In order to carry out annotations of the spoken material, initially, there were no applications capable of satisfying all the needs of segmental, prosodic, and textual level analysis of the samples. […] Another problem arose in a later phase when it was realized that annotations on various levels of analysis were not consistent, as a significant variability was observed among transcriptions and annotations. It was noted that even transcriptions and annotations produced by the same person could exhibit some inconsistencies among themselves. This prompted the researchers involved in the project to engage in a series of collaborative efforts to find transcription frameworks and propose standardized coding schemes to be used for transcribing and annotating the recorded data» (Morlicchio *et al.,* 2021, p. 243).

automated language processing, including language recognition, language generation, and machine translation.

Within this landscape of progress, sign language corpora have gained traction and attention, increasing in size and in number of participants (Schembri & Cormier, 2022). Within the Italian context, however, the availability of LIS corpora in the past ten to twenty years has improved, yet not significantly. This can be traced back to different factors: on the one hand, LIS is a minority language, a characteristic which makes its documentation more complex. At the same time, existing LIS corpora are often not accessible to all. In fact, many corpora are primarily used by their creators (Barbera, 2013) and may not be widely accessible or might have limited public availability.

At the present time, the challenges encountered by experts dealing with vocal language corpora, including the establishment of transcription standards and coding systems or the scarcity of open-source corpora, are similar to the challenges being presently encountered by sign language linguists. In fact, the absence of comprehensive documentations of LIS is an issue that presents a significant challenge for all those interested in corpus-based research. On this topic, two of the scholars that have made pioneering contribution to LIS studies (and sign languages research, more broadly), namely Virginia Volterra and Elena Antinoro Pizzuto, have been writing since 2002 in favor of the need to collect more LIS corpora, particularly with a focus on sociolinguistic variations (Volterra & Antinoro Pizzuto, 2002; Antinoro Pizzuto *et al.,* 2008; Antinoro Pizzuto *et al.,* 2010). With, the *Corpus LIS* (see section 3.2.3.1.) being collected a few years after (Geraci *et al.,* 2011) the wish made by Volterra and Antinoro Pizzuto seem to have encountered some sort of completion, however, despite being an excellent resource, the *Corpus LIS* is not yet available to the wider public[32].

[32] Although the *Corpus LIS* is not currently accessible, its developers have expressed their willingness to share it with anyone interested. After having reached out to both the Venice and Milan branches of

3.2.2. Sign language corpora

Much like spoken language corpora, sign language corpora consist of collections of preferably fully annotated data following shared conventions (Hodge & Crasborn, 2022). The development of sign language corpora is still in its early stages compared to spoken language, with a 15-years-long and and continuously evolving tradition, and with researchers anticipating its growth as an indispensable aspect of sign language research (Schembri & Cormier, 2022). Sign language corpora allow researchers to analyze language patterns without solely relying on introspection by signers. As Fenlon and Hochgesang point out, «[s]igned language corpora have the potential to be revolutionary for our understanding of signed language linguistics, leading us to reflect on what has been reported to date, and to formulate new theories and ideas regarding the nature of signed languages» (2022, p. 2). Consequently, the importance of such documentation spans various contexts, from language documentation and advancing linguistic theories on sign languages, to poromoting discussions on sign language resource annotation and post-processing for applications like machine translation, which is come of the long-term goals in the NLP domain (Thanaki, 2017).

Traditionally, sign language corpora collections have relied on setups with multiple cameras in well-lit studios, as seen in the NGT (*Nederlandse Gebarentaal*) and SSL (Swedish Sign Language) corpora (Hanke & Fenlon, 2022). However, there are instances of corpora being collected in more domestic or familiar settings, like the *Corpus LIS* (Geraci *et al.,* 2011), or by recording online conversations through video chat, as in the *Gianfreda Corpus* (2011). These corpora can include data from native

---

the group, they demonstrated their openness to providing access to the corpus. However, some challenges have arisen in this process. The *Corpus LIS* exists in a limited number of copies, stored on hard disks, and the process of digitally transferring the content is complex, as the videos within the corpus are substantial in size, for a total of hundreds of gigabytes, which implies a time-consuming transfer procedure. Furthermore, it is important to mention that the authors are working on making the complete corpus available online in the near future.

signers (such as in the instances mentioned above) or hearing learners (as in La Grassa, 2016). Another collection of everyday communication data that falls into the domestic category is the *Signes en famille* corpus[33]. This corpus gathers spontaneous interactions occurring in the daily lives of French families with at least one Deaf child, where LSF serves as the primary means of communication (Morgenstern & Caët, 2021). Regardless of the context, sign language corpora should share a common trait: high data quality achieved through the use of equipment designed to meet the expectations of future research (Hanke & Fenlon, 2022).

Drawing from past experiences in corpus collection, it appears clear that the development and use of sign language corpora are rapidly growing in significance, especially in contexts like the Italian one, where such collections are not yet common. In fact, sign language corpora, much like spoken language corpora, offer a unique avenue for studying signing communities and their languages. As highlighted by Hodge & Crasborn (2022) and Schembri & Cormier (2022), the creation and enhancement of sign language corpora can play a central role in the documentation of these visual-gestural languages and can significantly benefit broader linguistic research involving them.

3.2.3. Multimodal data collections of LIS

Despite data collection efforts being primarily centered on spoken Italian, as highlighted in the previous section, sign language researchers in Italy have actively worked on the development of LIS corpora and datasets, a task which, in turn, calls for the development or adaptation of annotation standards[34], tailored for the codification of manual and body components. Consequently, it involves a multifaced process of data collection and annotation as sign language dataset may include varying content

[33] https://www.ortolang.fr/market/corpora/signes-en-famille
[34] Both from standards created for other sign languages or vocal languages.

that mainly depends on the purpose of the data collection itself. Within this context, sign language datasets can be divided into two main categories: datasets that include videos portraying continuous signing or datasets that include videos portraying Signs in isolation.

Systematic efforts to compile and document LIS have been carried out, leading to the development of small-scale continuous LIS corpora by researchers, typically for individual research needs. These initiatives can be attributed to the scarcity of readily accessible collections of LIS data, encouraging researchers to create their own resources. For instance, to develop a small treebank for LIS of around 250 utterances (Caligiore *et al*., 2021), two videos were selected (one featuring a Deaf signer and the other a hearing LIS interpreter) which were then manually segmented and annotated. If it had been possible to access existing LIS corpora at that time, the resulting in research could have been more robust, based on genuine and natural LIS data by signers, and open to cross-comparisons with similar analyses.

The scarcity of available LIS data collections has led researchers to embark on the creation of their own resources. As a result, numerous limited-size LIS corpora have been established independently, showcasing the efforts made to document LIS, each serving different purposes, depending on the type of data collected. The methodologies employed for collecting LIS data display a wide range of approaches. In fact, researchers use techniques ranging from naming tasks to semi-structured and spontaneous interviews with Deaf signers and videorecording sessions with hearing learners of LIS as a second language (L2) second modality (M2) (Cardinaletti & Mantovan, 2022). These documentations serve different purposes, from the documentation of the language itself to the creation of tools for automatic translation.
For instance, scholars like Perea Costa and Russo Cardona, have contributed to this field by developing LIS corpora through distinct data collection methodologies: direct data collection involving interviews with Deaf participants (as exemplified by Perea

Costa), and indirect data collection where videos available online are gathered and organized according to specific research objectives (as exemplified by Russo Cardona)[35]. These examples provide insights into the diverse approaches that researchers have undertaken to address the challenge of limited LIS data resources.

The *Corpus Perea Costa* was gathered by Maria de la Luz Perea Costa from 1998 to 2001. The primary goal of this corpus was to investigate similarities and differences between LIS and LSE (Lengua de Signos Española), with a specific focus on Signs related to food and the eating habits of Deaf people in Italy. The *Corpus Perea Costa* comprises a collection of 2.143 Signs and 68 minutes of video recordings (Roccaforte, 2016). To ensure a diverse representation of the Deaf population in Italy, the data collection process involved 30 Deaf participants who each had at least one D/deaf parent. These participants were carefully selected and equally divided into groups of six, from the cities of Genoa, Gorizia, Rome, Foggia and Caltanissetta. The data collection methodology employed involved two main tasks. First, the participants performed a naming task based on 108 images depicting different foods. Subsequently, they were asked to describe what they had eaten on the previous day, to collect spontaneously elicited LIS data (Roccaforte, 2015). The strategic combination of these two tasks within the Corpus, allowed to capture both isolated instances of 30 Deaf people's lexicon, as well as spontaneous utterances related to food.
Another notable corpus[36] collected among LIS signers is the one compiled by Russo Cardona, which comprises «[…] 823 different manual Signs and/or signed constructions (a total of 1,491 Sign tokens) produced by native LIS signers» (2004a). Russo Cardona's corpus is divided into three distinct sub-corpora, each of which represents a different register. The first sub-corpus consists of five LIS poems, the

[35] The following paragraphs contain additional examples of LIS data collection efforts. However, Perea Costa, Russo Cardona and Gianfreda have defined their data collections as *corpora*, making them particularly relevant to mention in this section of this text.
[36] An official name does not seem to have been given to the corpus by the author.

second sub-corpus consists of three lectures on various topics, the third sub-corpus features three dramatized narratives. Collectively, these sub-corpora provide approximately 25 minutes of recorded video material.

Lastly, a corpus developed by Gabriele Gianfreda presents a distinctive feature: it was collected through video chats. This method of data collection (i.e., the recording of video chats on ooVoo) is not only different from the previous examples but also holds particular significance as in their lives Deaf individuals frequently employ video chats in to communicate with other signers (Gianfreda, 2011). This corpus is comprised of six conversations among Deaf signers, varying in age, family background, geographical origin, and age of acquisition of LIS. The participants were invited to video call with whomever they chose and discuss a topic of their preference, without any time limits. While the exact duration in minutes of this corpus is not specified by the author, it provides valuable insights into the spontaneous interactions of Deaf signers in real-life video chat scenarios. Indeed, as can be seen, despite sharing a common interest in Deaf signers' production, and using video recordings, these corpora diverge in their data collection methodologies and research objectives, providing three different examples of strategies that can be employed when approaching LIS collection.

In addition to the corpora mentioned above, there are numerous other independent data collections conducted by videorecording Deaf signers. For instance, during the 4th National LIS Conference, held in Rome in 2018, several studies presented findings from data collections, as exemplified by the presentations of researchers such as Ricci *et al.* (2018), Fornasiero (2018), Di Renzo & Slonimska (2018), Conte (2018). Additionally, Fontana & Raniolo (2020), Cuccio *et al.* (2022), and Fontana (2022a), to name just a few, have all undertaken data collection efforts. Certainly, numerous additional data collections exist although, regrettably, they lack extensive documentation as the existence of these collections is often mentioned in the articles and conference presentations authored by various researchers. These examples represent a merely subset of the landscape of independent data collections carried out on Deaf signers in the Italian context.

In recent years, with increased interest in sign languages and LIS, their documentation has reached new areas of enquiry. For instance, researchers have been carrying out documentation and investigation processes of LIS as a L2 for hearing learners. One corpus in this context is the *LISAU Corpus* (*LIS di Adulti Udenti*— LIS of Hearing Adults) (La Grassa, 2016). This corpus stands out from the previoulsy-mentioned corpora due to its different focus. In fact, the *LISAU Corpus* does not collect LIS data from native signers, but from hearing adults who are larning LIS as their L2 M2. This corpus is comprised of approximately 35 minutes of videorecordings, involving six LIS learners at different proficiency levels (ranging from students of LIS I, II and III levels). Additionally, two Deaf signers were included in the corpus, serving as a control group.

The documented collections, for which authors have provided information on length and token numbers, are compared in Table 1 below. This comparison highlights that these corpora are typically less than one hour long and contain fewer than 2,200 tokens.

Developers	Number of tokens	Total length	Participants
Tommaso Russo Cardona	1,491 tokens	25 minutes	Deaf
Maria de la Luz Perea Costa	2.143 tokens	68 minutes	Deaf
Matteo La Grassa	n.a.	35 minutes	Deaf and hearing

Table 1. Comparison of Russo Cardona, Perea Costa and La Grassa's corpora.

In an additional research project, following a naming experiment with hearing bimodal bilinguals (Pretato *et al.,* 2017), Peressotti & Navarrete (2019) developed a LIS corpus comprised of 234 isolated Signs taken from SpreadTheSign[37], with the aim of linking iconicity with different psycholinguistic variables by asking hearing spoken Italian speakers to provide a rating of the iconicity of a sign. These examples collectively

[37] https://www.spreadthesign.com/it.it/search/ (Accessed on September 3rd, 2023).

illustrate the expanding scope of research and the ongoing efforts made by researchers to compile and document LIS (with its varying registers and more or less proficient Deaf and hearing users) through structured processes of data collection.

As mentioned in the introduction to this section, LIS data collection extends to projects that go beyond language documentation. An example is the LIS4ALL project (realized within the context of the ATLAS[38] project), developed for the automatic translation from spoken Italian to LIS in Turin's Porta Nuova rail station (Geraci *et al.,* 2014; Geraci & Mazzei, 2014). Among the tasks carried out for this project, an artificial corpus[39] of 50 LIS utterances was included (Mazzei, 2015). Additionally, the ATLAS project also been employed by another application through Rai (RAI—*Radiotelevisione Italiana*) Virtual LIS: a platform intended to show 3D signing avatars[40]. It is also worth mentioning that a substantial but private database for LIS is comprised of all the videos that have been broadcasted as part of the TG LIS of the Italian national television.

Despite most corpora collections being private, an exception to this accessibility challenge is the online dictionary SpreadTheSign, a project that traces back to 2004. While initially intended as a dictionary for sign languages, SpreadTheSign also serves as a versatile resource for language documentation[41]. SpreadTheSign was established with the goal of developing a multilingual dictionary of sign languages for Sign comparison and learning (Hilzensauer & Krammer, 2015). Since its creation, SpreadTheSign has expanded significantly, becoming the world's largest online

---

[38] The ATLAS Project, co-funded by the Piedmont Region, aimed to develop services to automatic translation services from written Italian to LIS. This translation was displayed using a computer-generated avatar, http://www.crit.rai.it/CritPortal/progetti/?p=297&lang=en (Accessed on September 3rd, 2023).
[39] Term used by Mazzei (2015).
[40] http://www.crit.rai.it/CritPortal/progetti/?p=2566 (Accessed on October 10th, 2023).
[41] It should be noted that SpreadTheSign features signs and utterances performed in a controlled environment, and not in natural or spontaneous contexts.

dictionary for sign languages, making it possible for D/deaf and hearing people to look up or learn Signs from their national sign language, or other sign languages. At the moment, SpreadTheSign is comprised of more than 60,000 videos in 44 different sign languages[42] including Signs in isolation, as well as utterances. The University of Venice Ca' Foscari played a central role in providing LIS translations of the available sings (Cardinaletti, 2016). Moreover, the website's Sign-matching feature makes it particularly suitable for cross-linguistic comparisons among multiple sign languages.
Having provided an overview of existing public and private resources for LIS documentation, the following sections will focus on the *Corpus LIS*, the largest collection of spontaneous, semi-structured, and structured LIS videos by Deaf signers.

3.2.3.1. The Corpus LIS project[43]

The *Corpus LIS* is an extensive collection of LIS data that accounts for sociolinguistic variation among Deaf signers of LIS in Italy. This initiative drew inspiration from previous experiences of collecting sign language corpora abroad, particularly on the experiences provided by researchers who had collected ASL and Australian Sign Language (AUSLAN) corpora. The authors' primary goals were twofold: collecting a large amount of data, suitable for quantitative analysis, and to create a representative corpus of LIS usage in Italy (Cecchetto *et al.,* 2011). To achieve these objectives, interviews were conducted in various cities and towns, spanning from the North to the South of Italy, including Sicily. The locations include Turin, Milan, Brescia, Bologna, Salerno, Bari, Catanzaro, and Ragusa. A total of 165 Deaf participants, approximately 18 from each city or town, took part in these interviews. Gender distribution was nearly equal, with 41% women and 59% men. The participants were further categorized into three age groups: young (10-30), middle-aged (31-54), and older (55+).

[42] https://www.spreadthesign.com/it.it/about/statistics/ (Accessed on October 13th, 2023).
[43] I would like to express my gratitude to Prof. Carlo Checchetto for kindly sharing part of the *Corpus LIS*.

The recruitment of participants was assisted by local contacts with the Deaf community in each area. The involvement of these local contacts was important, similarly to the approach used in the *Corpus Perea Costa*. In fact, during the interviews, the role of the local contact was to ensure that the interviewees felt as comfortable and spontaneous as possible. Moreover, the only member of the research team present at the beginning of the sessions was Deaf. Each interview session involved participants from the same geographical area and age group. In the room with the participants, three cameras were placed: two focused on each signer, and one to capture both participants together. Video recordings were collected in high quality and saved in mpg2 format.

During the interview, participants were asked to perform four different tasks. Two free tasks, free conversation, and individual narration, and two semi-structured tasks, question-elicitation session, and a picture-naming task (Geraci *et al.,* 2011). The free conversations include video recordings of participants discussing a topic of their choice for around 40 minutes or less. The individual narration session included 5-minutes-long narrations of personal experiences. During the question-elicitation session, participants worked in pairs. Signer 1 described a picture that Signer 2 could not see. Signer 1 had to provide information about the picture, while Signer 2 noted the details provided by Signer 1 by writing them in a form provided by the contact person. This task aimed to target linguistic constructions, particularly polar and WH- questions (Geraci *et al.,* 2011*,* p. 536). Lastly, the picture-naming task featured 42 pictures shown to the participants, who were asked to elicit the corresponding Signs. The pictures covered various categories, including colors, family members, months of the year, objects for which LIS Signs were unknown (e.g., flamingo or basil), initialized Signs, fingerspelled Signs, and Signs with potential variations based on the age and geographical origin of the signer (Geraci *et al.,* 2011, pp. 540- 541). Given the number of participants and the length of the video recordings, at the present time the *Corpus LIS* represents the largest and most comprehensive resource of its kind for LIS.

After collecting the *Corpus LIS*, the researchers faced the task of annotating this resource. Two Deaf researchers, Santoro and Poletti (2011), provide insights into the process of partially annotating the *Corpus LIS*. It is important to note that as of 2011 (i.e., the year of the publication of the volume in which Santoro and Poletti's chapter is contained), the complete annotation of the *Corpus LIS* had not yet been completed. In fact, the authors outline general guidelines for future annotation, acknowledging that annotating the entire *Corpus LIS* horizontally could take several years. In doing so, they offered a general annotation procedure[44], which was also adopted by other authors in the same book. Santoro and Poletti highlight the use of ELAN as the preferred and widely used tool for editing and annotating sign languages (see section 4.3.). The annotation procedure described consisted of various levels within ELAN. The main level is aimed to unambiguously identify the signer's gender, age, and geographical origin. Dependent levels are created to investigate specific elements of signed discourse that were relevant to the researchers' particular research objectives. These levels cover aspects such as body components (referred to as non-manual markers), WH- questions (Geraci & Bayley, 2011) and eyebrow raising (Conte *et al.,* 2011). Additionally, two levels were designated for annotating vocal language labels and parts of speech. Lastly, three levels of annotation included a literal translation of the utterance into Italian, as well as literal and free translations into English.

In conclusion, the *Corpus LIS*, with its rich collection of LIS data, could serve as a valuable resource for various linguistic analyses and applications but also to simply capture the sociolinguistic and regional variations of LIS across Italy. More broadly, the annotation of LIS corpora, plays a crucial role in linguistic research for several

[44] This is not uncommon in corpora annotation processes as often, due to lack of funding and/or resources, annotators are not able to proceed with the annotation of their entire corpus. However, a solution that accounts for the annotation methodology and the theoretical framework behind it, is to provide a sample of annotation and describe annotation guidelines (Hodge & Crasborn, 2022). It is worth mentioning that the manual annotation process of this extensive corpus is a complex and time-consuming task, and the full annotation of the Corpus LIS seems to have remained a work in progress at the time of this discussion.

reasons. In fact, the annotation process aids in making LIS data accessible and interpretable by both researchers (both signing and not), and automated systems.

3.3. Strategic integration of computer vision tools in sign language data collection

Studies examining sign language processing encompass a broad spectrum, examining multiple sign languages (Bragg *et al.,* 2019) as well as specifically focusing on LIS (Fontana & Caligiore, 2021). Within this landscape, datasets and corpora serve as foundational pillars for any processing effort. Annotations accompanying these sets of data are commonly developed using ELAN (Wittenburg *et al.,* 2006; Crasborn & Sloetjes, 2008), a software specifically tailored to work on multimodal data. One main characteristic to consider in the context of sign language processing, is that since sign languages do not have a standard written form, data collections will not exist without the inclusion of videos captured through various recording devices that are chosen based on the specific objectives of data collection.

If machine translation stands as one of the primary objectives in data collection endeavors, the application of computer vision emerges as a valuable asset. Computer vision, in essence, revolves around using cameras to analyze and comprehend real-life scenarios. In their work, Bragg *et al.* (2019) delineate a spectrum of methodologies in sign language recognition, showing a preference for non-intrusive systems. These systems not only aim for minimal intrusion but also strive to capture the three-dimensional essence of Signs through the employment of depth cameras.

For the data collection processes detailed in this thesis, the strategic integration of computer vision played a central role in the selection of recording tools. Particularly, a decision was made to incorporate depth cameras, infrared cameras, and other visual sensors. These tools, collectively, will facilitate a comprehensive approach to sign language analysis by enabling the simultaneous capture of vital visual and spatial dimensions, crucial for accurate interpretation. The employment of various recording technologies serves a dual function for the recorded data. Firstly, it aids in sign language recognition processes. Secondly, when combined with other recording tools like RGB cameras, it generates high-quality data suitable for processing through sign language corpus linguistic methodologies, such as multilayered annotation. This dual approach enhances the usability and versatility of the recorded data across different analytical methods and applications.

3.4. Navigating sign language machine translation

As machine translation is one of the long-term goals in the NLP domain (Thanaki, 2017), when it comes to sign language processing from signed to spoken languages[45], the journey typically begins with sign language recognition, including finger-spelling and Sign classification. The information extracted from this recognition phase is then used to translate the source sign language into a target spoken or sign language (De Coster *et al.,* 2023). Machine Translation's objective is to provide rapid translations, albeit with limitations in literary contexts (Mitkov, 2003; Hudecová, 2020; Raheem Jabbar & Bani Madhi, 2021). Jurafsky and Martin (2020) emphasize machine translation's core role in information accessibility, a vital aspect for sign language data processing and language documentation. In relation to challenges, they encompass morphological, syntactic, and lexical disparities among languages, including multiple word meanings and lexical gaps.

Not considering processing models relying on data captured through invasive tools like gloves or wearables (e.g., bracelets or motion capture devices applied to the body) that

[45] The translation aspect is relevant in this work because the Dataset's initial processing (see section 6.5.) involves recognizing LIS signs and establishing consistent bidirectional translations, both from LIS to spoken Italian and from spoken Italian to LIS of signs in isolation. It is important to note this to specify that the Dataset is not only for translating LIS into spoken Italian for non-signers.

fail to capture facial expressions, approaches based on video recordings are favored, as they can be built upon existing video data and have the advantage of capturing both bodily and manual components (Bragg *et al.,* 2019). In their review on the state of the art for sign language processing, De Coster *et al.* (2023) emphasize a fundamental difference between spoken and sign language machine translation: the input. While spoken languages are typically input as text, sign languages require video input. However, regardless of the input modality, the output in both sign language to spoken language and spoken language to spoken language translation is typically in the form of text. An essential commonality in these approaches is the use of vocal language labels. In fact, both text to video models (Camgöz *et al.,* 2018) and video to text models (Camgöz *et al.,* 2020) directly or indirectly incorporate spoken language labels (De Coster *et al.,* 2023).

When it comes to evaluating sign language machine translation models, it is argued that Deaf signers should be involved in the evaluation process (De Coster *et al.,* 2023, p. 17). This aspect is crucial and extends to the steps preceding the final output of translated text or video. In fact, it is important that the data on which translation is based are assessed by signers to ensure both its acceptability and acceptance. In cases involving non-spontaneous signing, steps can be taken to involve a Deaf control group that can provide valuable feedback on the data. However, it is also important to acknowledge that limitations in data collections, complexities in sign language annotation, and the current state of machine translation present challenges that prevent the development of all-encompassing tools (Müller *et al.,* 2022). In this context, from a linguistic perspective, discussing challenges related to the later stages of sign language processing, such as machine translation, is complex and perhaps at this stage fo this work, premature. First and foremost, it is crucial for the signing and sign language research communities, both at national and international levels, to collaborate in the development of fully annotated resources that are openly accessible. Without this foundational step, especially within the context of LIS discussions around sign language machine translation may remain largely theoretical or limited to better-documented sign languages such as German or Dutch.

3.5. Applying and developing sign languages processing systems: an open discussion

Applying and developing processing systems for sign languages poses significantly more intricate challenges, compared to spoken languages. These challenges stem from the characteristic of sign languages, giving rise to both compositional and sociological complexities in linguistic analysis and translation tasks. From the sociological standpoint, it is important to mention that sign languages have often been the unaware object of sign language recognition and translation projects that were aimed at translating from a sign language to a spoken language. It is also worth noting that some of these resources were created by researchers who had limited awareness of the sociolinguistic context of their respective national Deaf and sign language communities. An example of this limitation is the use of invasive tools such as gloves, which can hinder the naturalness of signers and, more importantly, fail to capture the essential body components. This can significantly impair even the human signer recognition abilities in relation to, for example, idiomatic Signs that involve mouth gestures, or the comprehension of the intention conveyed through a signed utterance (question vs. statement).

Traditional linguistic concepts like morphology and syntax encounter complexity when applied to sign languages, as sign languages exist within a multilinear and simultaneous context, defying linear models typically used in linguistic analysis. Attempting to view sign language parameters as equivalent to ‘morphemes’ an association initially proposed by Stokoe in 1960, or trying to impose linear traits, creates challenges. In fact, variations in the semantic and pragmatic nature of in the formational parameters can lead to the simultaneous transmission on the role of agent of an action or provide information on the object or referent (Fontana & Roccaforte, 2015). Consequently, iconicity cannot be avoided as the reliance of sign languages on the possibility of 'showing' (Volterra *et al.,* 2022) is a funding characteristic of signed communication and poses challenges in finding equivalent translation in a target language.

The concept of applying syntax of sign languages adds another layer of complexity. Identifying fixed and non-marked orders of components in LIS is intricate, with studies revealing a theoretical SOV (Subject-Object-Verb) order for non-marked LIS utterances being contrasted by research on practical usage that proves variations due to diatopic and diachronic factors (Branchini & Mantovan, 2021). Not to mention fingerspelling, where each hand shape corresponds to a letter in the Roman alphabet (Pugeault & Bowden, 2011; Fowley & Ventresque, 2021). As for the use of sign language corpora for machine translation tasks, as mentioned in the previous section, an immediate obstacle is encountered due to the scarcity of large corpora for sign languages, particularly LIS. An issue than cannot be easily solved, as developing such corpora demands rigorous data collection efforts and manual annotation.

This conclusive section does not want to present definitive solutions, but to highlight the complexities inherent in applying NLP and machine translation to sign languages. The following chapters will explore the implications of these challenges and potential solutions, within the context of this work. Chapters 1 through 3 have primarily focused on theoretically framing this thesis, whereas Chapters 5 and 6 will explore the practical application of these frameworks in LIS data collection and annotation. Before that, Chapter 4 will focus on addressing the practical issue of the oral nature of sign languages by providing an overview of annotation systems developed for sign languages. Understanding the historical evolution of sign language codification and computational annotation systems is considered essential before moving onto the practical aspects of data collection and processing.

Chapter 4
Transcription systems and representation strategies for sign languages: from Mimographie to Typannot

As discussed in the previous chapter, NLP applications are built based on structured collections of data that undergo processes of tokenization and POS tagging (Donges, 2021). To this regard, a new question arises. How can NLP, which primarily relies on written texts, be applied to sign languages? This question opens a conversation on the strategies employed for representing sign languages and the tools that have evolved for this purpose over time. In this chapter, an exploration of various systems designed for the annotation and transcription of sign languages will be undertaken, an exploration that will encompass both spontaneous and artificial systems, as well as analogue and digital formats. Following that, the main software used for annotating sign languages, namely ELAN, along with its main features, will be introduced.

4.1. Transcription systems for sign languages

As mentioned earlier, sign languages are oral languages, in that, similarly to many vocal languages, they do not have standardized and formal writing systems and are, therefore, exclusively realized in face-to-face communicative contexts (Antinoro Pizzuto *et al.,* 2006; 2008; 2010; Peters, 2000; Garcia, 2010). Whereas, on the one hand, this oral nature allows for an observation of sign language in its multidimensional and multimodal nature, free from any influence of written representation, it can also raise certain challenges when it comes to framing the language itself, together with its intricacies and phenomena (Fontana, 2014). This oral nature presents a significant challenge not only for NLP, in fact, it is a well-recognized issue in the field of sign language research, to the point that Deaf and hearing members of international signing communities, at different levels of integration within the community itself, have devised solutions to address this challenge over the years. In fact, it is not true that sign languages are *never* written. On the contrary, different methodologies have been developed[46] over the years. Among these strategies, Mimographie, Stokoe Notation, HamNoSys, SignWriting and Typannot are found.

Mimographie, introduced in 1825 by Bébian, a French-Caribbean educator, was the first system devised with the practical goal of enabling the rapid transcription of sign language, similar to the speed of writing vocal language (Bébian, 1825). The Mimographie system comprises 187 characters that describe body movements, facial expressions, locations of execution in the signing space and handshapes (Bianchini, 2016), as depicted in Figure 1.

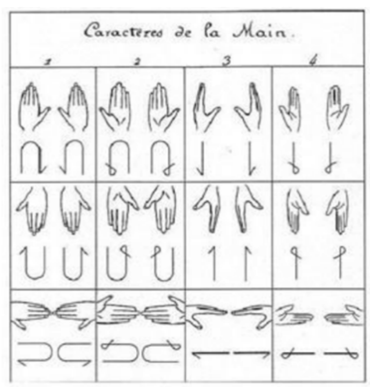


Figure 1. Mimographie hand shape symbols (Celo, 2015).

Bébian's attempt at providing a tool for the annotation of Signs was reprised on several instances in the 1900s, but for different purposes. William Stokoe, for example, developed the Stokoe Notation system in order to study ASL Signs from a linguistic perspective (1960). Therefore, his original intention in developing this system was not that of creating a system that Deaf ASL users would adopt in their daily lives. In order to do so, he divided Signs into three aspects which he deemed relevant and named

[46] It appears that no language-specific writing system has organically developed and been embraced by Deaf signers within their communities. This is because, as bimodal bilinguals, they often lean on the alphabet of the vocal language that forms part of their linguistic repertoire.

cheremes[47]: configuration, position or location, movement and palm orientation (Stokoe, 2005). Stokoe Notation symbols are written from left to right, in horizontal linear direction, as shown in Figure 2, with the function of each symbol being clarified in Table 2.

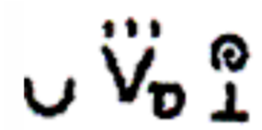


Figure 2. Stokoe Notation of the ASL Sign 'snake' from Martin (2003).

Symbol	Chereme type	Description
	Chereme of Position	Location of execution of the Sign is the lower face.
	Chereme of configuration	Index and ring finger extended and spread.
	Chereme of motion	Twisting, pronative movement.
	Chereme of motion	Three sharp, staccato movements.
	Chereme of motion	Away from the signer.
	Chereme of motion	Circular motion.

Table 2. Subdivision of the cheremes the constitute the ASL Sign 'snake'.

Another prominent notation system developed for sign languages, originating in Germany, is the Hamburg Notation System (HamNoSys). It was first introduced in 1987 and further refined with its 2.0 version in 1989 by Prillwitz. HamNoSys, while rooted in Stokoe Notation, was conceived with broader objectives extending beyond sign language description. These objectives included the integration of iconicity through the creation of new symbols, rather that reliance of the Roman alphabet, the interest in efficiency, compatibility with standard computer tools, and lastly, extensibility, permitting the incorporation of new symbols: a valuable asset for ongoing developments in the field (Hanke, 2004). As Hanke notes, «[a] HamNoSys notation for

[47] The definition of chereme, as provided by Stokoe is: «[…] that set of positions, configurations, or motions which function identically in the language; the structure point of sign language (analogous to ‘phoneme’)» (2005, p. 33).

a single Sign consists of a description of the initial posture, which includes nonmanual features, handshape, hand orientation, and location, along with the actions that alter this posture, either sequentially or in parallel» (Hanke, 2004, p. 1).

HamNoSys is, once again, to be used linearly. It introduces additional features compared to Stokoe Notation, with the most notable being the incorporation, although limited, of body components, referred to as nonmanual features in Hanke's terminology (body components, in the COS-S theoretical framework adopted by the present work) (see Figure 3). This aspect is particularly relevant, as the evolution of sign language studies has increasingly emphasized the role of body components, a dimension that was omitted in previous notation systems like Stokoe's. In HamNoSys, developers define coding schemes, typically employing alphanumerical labels, to precisely identify movements (Hanke, 2004, p. 3).

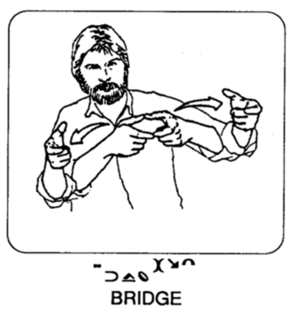


Figure 3. An instance of HamNoSys transcription accompanied by an image and English translation for the DGS Sign 'bridge' (Prillwitz, 1989).

While these notation systems have achieved varying levels of success and recognition, they have encountered challenges in gaining widespread adoption among researchers and the signing community, remaining confined to smaller groups of scholars. These challenges stem from factors such as the complexity of learning these systems and their limitations in representing multilinear and multimodal elements.

In the upcoming sections, a more detailed exploration of two notation systems will be provided, each notable for distinct reasons. SignWriting (section 4.1.1.) stands out as, perhaps, the most widely used notation system among researchers, especially in Italy.

Whose ease of memorization, usability, and long-term legibility have been demonstrated by Claudia S. Bianchini in her 2012 Ph.D. thesis. The second system, Typannot (section 4.2.) although yet to be released, holds significant promise.

4.1.1. SignWriting

SignWriting (henceforth referred to as SW) originated as an adaptation of Dance Writing, developed by dancer Valerie Sutton in 1974 primarily to describe dance movements or general physical motions. SW quickly captured the attention of sign language researchers due to the iconicity of its symbols, which appeared to be effective in conveying both manual and body components. In SW, the representation of body components is particularly noteworthy, as previous notation systems had either omitted or only partially accounted for these elements (see the section above). SW has the versatility to serve as a universal representation system for sign languages[48], however, no single sign language uses all the available SW characters (Everson *et al.,* 2012). A distinctive feature of SW is its capacity to provide precise representations of both manual and body components. It can also be effectively used for depicting Transfer Units or capturing the spatial and multilinear organization of signed discourse (Volterra *et al.,* 2022).

Within the Italian context, in 1998, Elena Antinoro Pizzuto enlisted the assistance of two Deaf researchers to evaluate this code. It was found to be easy to learn, particularly for proficient signers, and effective in its representation (see Di Renzo *et al.,* 2011), leading to its increasing adoption among researchers in Italy. Furthermore, SW lends

[48] «Sutton SignWriting can be used to write any sign language, natural or constructed. Formal SignWriting supports all sign languages without requiring the addition of new characters or updated fonts. […] SignWriting uses a closed set for characters with completed fonts that do not need to be updated» (Slevinski, 2022). For a list of all sign languages that have been written using SW, see Everson *et al.* (2012).

itself to computer processing, facilitated by software such as SignPuddle[49]. Remarkably, SW made history by becoming the first sign language writing system to be included in the Unicode Standard[50] in 2015, with a repertoire of nearly 700 characters. These characters do not comprise the entire repertoire, which amounts to more than 37,000 symbols. Additionally, an advantage that SW provides, even when compared to video data, is that it makes it possible to represent a Sign with a single graphic unit as well as provide a static image of the Sign itself (Bianchini, 2012).

4.1.1.1. A quick overview of SignWriting Symbols[51]

In SW, the units are referred to as symbols or, in the Italian context, as *glifi* (*glyphs)*[52], and unlike Stokoe Notation and HamNoSys, they are not arranged linearly but in a vertical orientation, from the top to the bottom of the page (Thiessen, 2011). The chosen perspective is an expressive one, in which Signs are represented as the signer sees them during their own performance, either in the vertical or horizontal plane (Di Renzo *et al.,* 2011).

By employing combinations of SW symbols, it is possible to represent the formational parameters of sign languages (see Figure 4), as detailed by Volterra *et al.* (2022).

[49] While SignPuddle stands out as the predominant software in this domain, it is worth noting that encoding a single sign can be a time-consuming process. Several efforts have been made to enhance the computational encoding of SignWriting symbols, including the work of Borgia (2015) and Borgia *et al*. (2012). SignPuddle can be accessed at https://www.signpuddle.org/ (Accessed on August 9th, 2023).

[50] «The Unicode Standard is a character coding system designed to support the worldwide interchange, processing, and display of the written texts of the diverse languages and technical disciplines of the modern world» (Unicode, 2017).

[51] This section provides a brief overview of the methodology and symbols used in SW to depict manual and body components. The presented overview follows the sequence of symbols' descriptions as outlined by Di Renzo and colleagues in *Scrivere la LIS con il SignWriting* (2011).

[52]The term *glifo (glyph)* emerged according to researchers in the Italian context as a suitable description for SW symbols due to it being less associated to vocal languages. Credit for coining this term goes to Barbara Pennacchi, researcher at the ISTC-CNR in Rome (Di Renzo *et al.,* 2011).

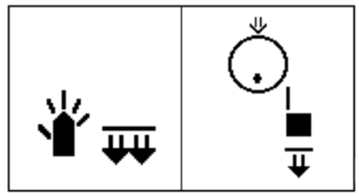


Figure 4. Two instances of SW used to represent manual components (on the left) and both manual and body components (on the right). The symbols on the left can be translated as 'five thousand', the one on the left as 'also/additionally' (Di Renzo et al., 2011).

These parameters encompass manual components, including hand shape, orientation, location of execution, and movement, as well as body components such as facial expression, oral components, movements of the bust, and eye gaze. Starting from handshape (such as closed fists, open fists, and open hands) to which lines can be added to indicate finger positions. Information regarding the palm's orientation is conveyed through the degree of shading (white and black) used to fill the symbol representing the palm. Depending on the fingers employed, SW symbols can be divided into five distinct groups to represent the positioning of each finger (Volterra *et al.,* 2022):

Group 1: one finger (index, thumb, pinky finger).

Group 2: two-finger combinations (index-middle finger combination, index-thumb combination, index-pinky finger combination, thumb-pinky finger combination).

Group 3: Three-finger combinations (thumb-index-middle finger combination, thumb-index-pinky finger combination, index-middle-ring finger combination, middle-ring-pinky finger combination).

Group 4: four-fingers combination (all fingers except the thumb, all fingers except the middle finger).

Group 5: all five fingers.

Additional symbols serve to represent movements, including various types of contact such as light touch, grabbing, rubbing, or finger insertion. Straight, curve, rotational, shaking and circular movements are also accounted for through the use of arrows, which vary depengin on the movement being vertical or horizontal (i.e., parallel to the wall or to the floor).

SW also allows for a comprehensive description of body components. The head is depicted as a circle containing symbols representing facial expressions. These expressions are further broken down into ten groups of symbols, encompassing components like the forehead, eyebrows, eyes, cheeks, nose, mouth, tongue, teeth, chin, and neck. Eye gaze is also describable, through the use of arrows that represent its direction. Movements involving the face, shoulders, and bust are conveyed through specific combinations of arrows and symbols. Type of movement is also accounted for by symbols representing if the movemen is, for instance, slow or quick. When it comes to describing mouth actions (Fontana, 2008), representing these oral components of speech necessitates reference to vocal language labels or the International Phonetic Alphabet, as outlined by Volterra *et al.* (2022). Researchers have developed symbols for punctuation as well, enabling the representation of the end of an utterance or the presence of small or extended pauses. For a table of SW gyphs see the appendix to Di Renzo *et al.* (2011), which includes both symbols used for LIS and possible variants, marked by the label VAR (variant).

4.1.2. Discussion

As previously mentioned, SW has rapidly emerged as the most widely adopted language-specific method for representing sign languages, particularly within the groups of Italian researchers adhering to the COS-S approach. However, despite numerous efforts and its innovative, language-specific nature, SW has encountered challenges in establishing itself as the primary transcription strategy for sign languages and is used within small groups of Deaf users both nationally and internationally (Hopkins, 2008). Instead, the use of vocal language labels remains prevalent in most cases. At the same time, looking at the adoption of SW symbols within the context of the formalist approach to sign language annotation, the employment of a written code distinct from the Latin alphabet poses significant challenges, primarily originating from the approach to sign language studies itself, as well as its goals. This challenge arises from several factors, including the Latin alphabet's utility in illustrating examples of universality in sign languages, its capacity to depict grammatical and syntactic Sign distribution, its ability to facilitate cross-linguistic comparisons, and its compatibility more easily with computational processing through the use of vocal language labels. The present section will not further explore the general advantages and disadvantages of using SW, nor its ease or difficulty of use, as these have been comprehensively examined and analyzed by Bianchini in her Ph.D. thesis (2012). In her work, she also provided valuable insights for those working with SW, along with suggestions for its potential future applications. Among these, shifting the focus towards the digitalization of SW, especially its integration and employment within video-processing computer programs like ELAN, is desired. However, at the present time, the challenge of effectively digitizing SW remains unresolved, as highlighted by Bianchini more than ten years ago (2012). This issue is of great significance for the present research, given its reliance on computer-based tools. As SW symbols cannot be directly incorporated into ELAN tiers, one approach is to create dedicated tiers that cover all possible symbol categories. Then, Unicode labels corresponding to each parameter[53] can be added to these tiers. However, when viewed in ELAN, these annotations will appear as sequences of Unicode labels, lacking specific information about the Sign itself and thus unable to convey their significance. Additionally, if one would need to reconvert the SW Unicode labels into SW symbols after the annotation process, such a process would require a lengthy manual restructuring effort, or the development of a specialized program for automation. The challenge of encoding SW within ELAN directly leads into the introduction of a promising and innovative transcription system: Typannot.

[53] https://unicode.org/charts/PDF/U1D800.pdf (Accessed on August 10th, 2023).

4.2. Typannot

Typannot is a typographic system designed for transcribing sign language. This system, ideated ten years ago, is currently under development by a collaborative effort involving researchers from the ESAD design school in Amiens, the FoReLLIS Lab at Poitiers University, and the DyLIS lab at Rouen University. The development approach of Typannot is based on the corporeal articulatory phonetic approach, which draws inspiration from Dominique Boutet's Kinesiological Approach (KinApp) (2018). For this reason, throughout the years Typannot has been not only the practical application of the KinApp, but also the context in which the Approach has itself developed.

4.2.1. Theoretical framework

Boutet's KinApp offers a novel perspective on studying multimodal communication, emphasizing the 'phonology of gestures' rooted in human physiology. It calls for a redefinition of the human body and its dynamics as the core driving force behind the emergence of Signs and gestures, encouraging the description of movement from an intrinsic standpoint (Chevrefils *et al.,* 2021; Morgenstern *et al.,* 2021). In the KinApp approach, the focal point is placed on movement, which acquires a new role. Rather than merely being described as a shift from one location to another, it is viewed as a holistic gestural unit with stable elements that can be described in detail (Chevrefils *et al.,* 2021). Boutet also encourages moving away from the conventional emphasis on hand shape in sign language descriptions, a direction that had already been pursued by SW developers and users. He suggests that movements are not primarily driven, for instance, by the wrist or hand, but by other body components, such as the forearm and elbow, resulting in a fluid movement pattern (Morgenstern *et al.,* 2021). This fluidity stands by 'rules' dictated by the biomechanical characteristics of the human body, which follows articulatory rules that determine various degrees of freedom of movement (see Boutet, 2018, p. 23). Therefore, bodily structures, together with their relations, potentials and limits, are taken into account (Bianchini, 2023). This aspect is what allows for the extention of this discussion on movement beyond sign languages and, more broadly, to gestures and bodily movements: «[t]his means that gesture is personally lived and understood at the level of a body that can freely transform, modulate, and interact with those cultural and linguistic forms within the limits of what is possible in terms of movement and signification» (Danet *et al.,* 2021, p. 1016).

4.2.2. Development of the Typannot transcription system

Within this framework, Typannot is a transcription system[54] being developed to represent the movements of the body in multimodal contexts (both sign languages and gestuality). Moreover, its developers (thanks to their different backgrounds) have taken into consideration the influence of computers in our daily life, which transfers of course into research. It was considered important, in fact, to take into consideration the possibility of being able to implement modern IT and typographic technologies (Bianchini, 2023).

Following the KinApp and Boutet's redefinition of the centrality of body movements from an intrinsic standpoint (2018), Danet *et al.* (2021) state that the system allows for the representation of modulations on the articulatory level, particularly the level of skeletal joints. Boutet (2018) highlights the significance of two key parameters in sign language: position and movement. Movement, in fact, naturally stems from the positioning of the signer's body, specifically the locations it occupies, which holds potentiality of movement (Bianchini, 2023). This division implies that location and movement can be transversally applied to all body parts covered by the system. These body parts are described as segments and are interconnected within a Group of Segments[55] which can be distinguished from other characters due to a diamond shape:

[54] As described by the developers on the Typannot website (https://www.typannot.com/#5-SECT, accessed on October 13th 2023), Typannot is a transcription system for linguistic and gestural studies as it is developed to enable researchers to study sign languages.
[55] Originally *Regroupement de Segments* (GRSEG).

 (hand shape),  (upper limb) and  (lips)[56]. Currently, the Typannot development team has been focused on defining the positions assumed by the following groups: HandShape[57], encompassing all the fingers, Upper Limb, consisting of the hand, forearm, arm and shoulder, Mouth, covering the jaw, lips and tongue, Eyes, involving the eyes, eyebrows, eyelashes, and nose (Bianchini, 2023). While the team has outlined more Groups of Segments, these are the ones relevant to this work, as they were used to annotate some Signs in Dataset that will be presented in the following chapters (see section 6.5). The representation of movement has not been fully developed at this point, resulting in the adoption of specific strategies for its representation in this thesis, as discussed in section 6.5.3.2.

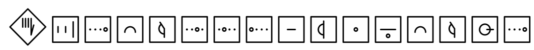
Based on each group of segments, together with the transversal movements, the delevopers of the system are designing specific characters that are grouped under the Typannot Font Family[58], which was designed following the principles of genericity, readability, modularity and inscribability (Danet *et al.,* 2021; Bianchini, 2023). Signs and movements, in a broader sense, are perceived as blocks of articulatory information that can be described in detail thanks to the adherence to the principle of genericity. This formatting ensures an unambiguous and unequivocal understanding of the information conveyed by the characters.

Moreover, the concept of readability satisfies two different applications of this principle: machine-readable and human-readable. For this reason, information encoded with Typannot can be presented in two different ways: *formule générique* (generic formula) and *glyphes composés* (composed glyphs). The first one is a linear and syntactically regulated disposition of the Typannot characters, whereas the second one

[56] These are three of the seven Groups of Segments, as the symbols for the remaining ones are currently in the process of being designed.
[57] Bianchini (2023), notes that despite the fact that the KinApp refutes the centrality of the hand, the development of Typannot started from hand shape for pragmatic reasons.
[58] In this work, the incorporation of Typannot characters was made possible thanks to the generosity of Prof. Claudia S. Bianchini and the members of the Gestual Script Team.

is an iconic and easy-to-read rendering of the codified Sign or gesture (Bianchini, 2023). It may be added that, from personal experience, despite its complex aspect at first impact, the *formule générique* of Typannot becomes easily accessible after understanding how the codification and syntax of the transcription system work. For instance, below are the two possible representations of the hand shape depicted in Figure 5.

Figure 5. Hand shape example from Bianchini (2023, p. 434).

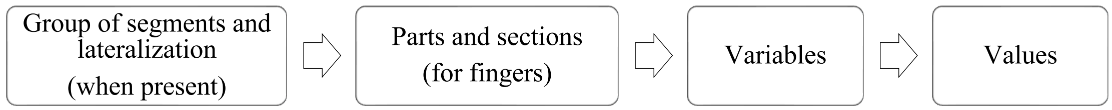
Typannot hand shape *formule générique*: 

Typannot *glyphes composés*:

The typannot font belongs to the *inventaire graphématique* (graphemic inventoire) of Typannot, which is comprised of *caractères génériques* (generic characters).

4.2.3. Syntax

As shown in the previous section, Typannot developers have provided guidance on the character order, or in other words, the syntax of Typannot. The generic formula is structured as follows: it begins with the representation of the group of segments, specifying its lateralization (left or right). Following that, the parts and selections (only for fingers) are codified. Finally, variables and their respective values are incorporated. This linear formula, as explained by Bianchini (2023) is helpful when Typannot needs to be used in contexts like Word, Excel, or, for those involved in multimodal studies, on ELAN, serving as practical guidance for transcribing with Typannot.



4.2.4. Partial description of Typannot characters

The description of the Typannot transcription system provided in this section is partial, as it only includes the components that have been used to annotate the MultiMedaLIS_3 Dataset, as described in section 6.5. For a comprehensive description of the system, please refer to Bianchini (2023), Boutet *et al.* (2016), Bianchini *et al.* (2017), Rébulard *et al.* (2018) and Danet *et al.* (2021). The subsections below are entirely adapted from Bianchini (2023).

Starting with the codification of manual components, which includes, following the KinApp model, the entire arm of the signer, the codification of the upper limb (which includes the arm, forearm, and hand) will be described, before moving on to the fingers. After that, the codification of body components will be described, specifically the mouth and eyes. Generally, the syntactic disposition of the components follows this order: group of segments, parts, variables, and values.

4.2.4.1. Upper Limb

The general segment of the upper limb includes the parts of arm, forearm and hand. After having codified the lateralization of the hand, through the left ( ) and right ( ) characters, the identification of the segment: arm ( ) forearm ( ) and hand ( ) is found. Based on the biomechanics of the human upper limb, different variables have been identified and paired into specular couples: flexion/extension ( ) abduction/adduction ( ), internal rotation/external rotation ( ), which, for the hand, corresponds to hand pronation/supination (refer to Figure 6).

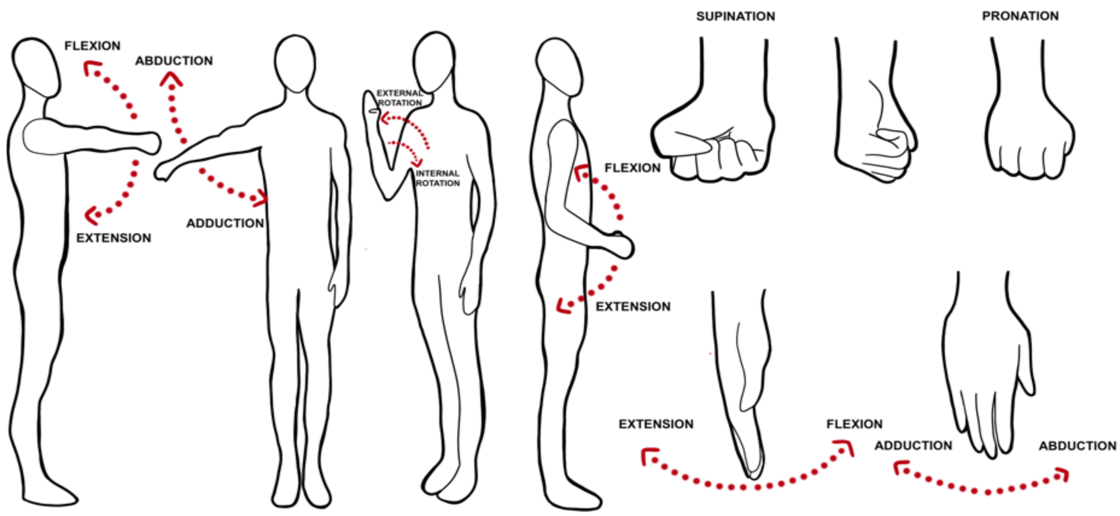


Figure 6. Upper body positions and movements (excluding information from the neck up), adapted from Narayan et al. (2021) (CC BY 4.0).

The values associated to upper limb variables are divided into two groups of four components each. These groups reconnect at the neutral position □, which is described as arms extended on the side of the torso, elbows bent at 90° and palms of the hand facing inwards[59]. From left to right, the four positive values are +4/4 □, +3/4 □, +2/4□, +1/4 □, the four negative values are −1/4 □, −2/4 □, −3/4 □, −4/4 □. It is important to note that the components of the upper limb do not have the same range of movements, therefore the degree of position/movement to attribute to the values varies depending on the section of the upper arm that is taken into consideration. For a representation of variables and values, see the Table 3 below.

Variables	Values								
Flexion/extension Abduction/adduction	+4/4	+ 3/4	+ 2/4	+1/4	0	−1/4	−2/4	−3/4	−4/4
Internal/external rotation Pronation/Supination	□	□	□	□	□	□	□	□	□

Table 3. Values of upper limb variables. Adapted from Bianchini (2023, p. 445).

[59] After discussing the neutral position with Prof. Bianchini, it was decided that, for the purpose of this work, the neutral position will be considered as arms extended on the side of the torso, elbows bent at 90°, and palms of the hand facing downwards. This decision was made because the neutral position of the dataset presented within this work is the one represented in the description above. This choice was influenced by the research carried out by Chevrefils (2022) who had already experimented with this modification of neutral position.

4.2.4.2. Fingers

Typannot allows for a codification of 300,000 different hand shapes , starting from just 22 characters. Following the established syntax, after having identified the group of segments (in this case the hand) the first characteristic that must be codified is, once again, the lateralization of the hand. After that, the position of the fingers is codified through the following characters: for the index finger, for the middle finger, for the ring finger, for the pinky finger, for the thumb.

Fingers are divided into a maximum of three selections, depending on their positioning, counting from the index fingers to the pinky finger. One selection may include one or more fingers. The thumb constitutes an additional section and is always codified separately from other fingers.

The variables related to fingers are form, angle and interaction. With regards to interaction, it is included if present and it consists in a description of how the groups of fingers are in contact with each other if present (see Table 4).

VARIABLE	VALUE	CHARACTER
Form	Extended	
	Curved	
	Bent	
Angle	Hyperextended	
	Open	
	Semi	
	Closed	
Interaction	Grouped	
	Separated	
	Crossed	
	Stacked	
	Inverted	

Table 4. The Typannot variables of form, angle and interaction for the hand shape group of segments. Adapted from Bianchini (2023, p. 446).

4.2.4.3. Mouth

The codification of the mouth includes the jaw, lips, corners of the mouth, tongue and airflow[60]. Both lips and corners of the mouth are constituted by two components (respectively upper and lower lips, and right and left corner) which can be codified together or separately. To render the movement of these components, Typannot developers have decided to describe them following a Cartesian plane (see Bianchini, 2023). Visualize it as if the jaw was divided at its center by both a vertical and horizontal axis. Consequently, the movement of the jaw, lips, and corners of the mouth follows these axes (x, y and z). Starting from the neutral position (which is the relaxed position of these components), the jaw, lips, and mouth can be located to the right, left, up, down, forwards, backwards. They can also converge and diverge. The tongue can be flat or round. It can come into contact with other components of the mouth and can be described, once again, following a Cartesian plane, based on the positions that have been described above. Lastly, the tongue can be hidden. Regarding air flow, it can be codified as neutral, an exhalation, or inhalation, with the specification of the direction in which the airflow is produced.

4.2.4.4. Eyes

The general segment eye does not have a codified character yet. However, its parts, values and variables have been codified and are nostrils, pupils, eyelashes and eyebrows. As can be inferred, the components that constitute the eye segment all belong to the upper part of the face. For eyebrows and nostrils, it is possible to codify the position of both components together, or codify them individually. The position of

[60] It is interesting to note that one of the members of the Typannot design team is Adrienne Contesse. Contesse is the creator of VocalGrammatics (https://www.vocalgrammatics.com/ accessed on August 4th, 2023), a fascinating writing system developed for beatboxing to codify how speakers manipulate the elements of the vocal tract to produce sounds.

all variables is once again codifiable using the following symbols: right , left , up , down , forwards , backwards . They can also converge and diverge .

4.2.5. Computational applications of the Typannot transcription system

As specified by its developers, one of the main functions of Typannot is its use on ELAN. In fact, the developers have «[…] already conceived an ELAN template that allows linguists to transcribe with all Typannot typefaces» (Doan *et al.,* 2019). It is possible to provide at least one instance of Typannot being used on ELAN by Léa Chevrefils, who transcribed more than 1,500 Signs of a LSF corpus (Chevrefils, 2022) also introducing controlled vocabularies directly transcribed on ELAN with the Typannot font.

Typannot represents a significant step forward in the field of sign language linguistics, building upon the past experiences and foundations laid by HamNoSys and SW (Doan *et al.,* 2019). By incorporating their language-specific attributes, codification techniques, and iconic components, Typannot creates a comprehensive system that is not only adaptable but also highly functional. Its incorporation of scriptability, genericity, modularity, and readability make it a versatile tool for various computational applications, while maintaining its iconic nature. This synthesis of past experiences and innovative approaches positions Typannot as an asset for further research and development in the realm of sign language linguistics and gestural studies. For a table comparing characteristics of HamNoSys, Typannot and SW, see Table 5 below.

System	Developed for	Extensible	Compatible with Unicode	Available on ELAN	Searchable	Iconic
HamNoSys	Research	Yes	Yes	Yes	Yes	Partially
SignWriting	Research and education	No	Yes	No	No	Yes
Typannot	Research	Yes	Yes	Yes	Yes	Yes

Table 5. Comparison HamNoSys, SignWriting and Typannot features.

4.3. ELAN

After having explored different annotation systems for sign languages, this section is devoted to the introduction of the main software employed for multimodal annotation: ELAN[61]. ELAN is a computer program for linguistic annotation, particularly useful for multimodality research. It was initially developed by Birgit Hellwig at the Max Planck Institute for Psycholinguistics in Nijmegen, Netherlands (Sloetjes, 2023), and was first released under the name ELAN in 2002. Since then, it has been widely used in multimodal research, including the annotation of sign language data, particularly for corpora annotation (Wittenburg *et al*., 2006; Crasborn & Sloetjes, 2008). As of summer 2023, ELAN is at its 6.6 version, with its manual last updated in April 2023.

The key advantage of ELAN lies in its versatility, as it allows users to develop unlimited and customizable annotation, directly bound to video frames. This is achieved through the creation of annotation layers and the subsequent segmentation of these layers into sections with associated time intervals. Annotations in ELAN are represented as Unicode text (Sloetjes, 2023).

ELAN allows annotators to tag videos frame by frame and offers the flexibility to create multiple layers of annotation called *tiers*, which can be tailored to specific research requirements (Brentari, 2010). ELAN provides the capability to establish dependencies between tiers based on the purpose of the annotation. There are two main types of tiers: independent (or *parent*) and dependent (or *child*). Parent tiers are time-alignable, directly linked to specific time intervals. In contrast, child tiers contain segments and annotations linked to their parent tiers and are typically not independently time-alignable as their time intervals depend on the parent tier. Each tier is assigned a *type* that specifies the constraints applied to it, referred to as stereotypes (Wittenburg *et al.,* 2006). A notable characteristic of ELAN is the possibility of creating *controlled vocabularies* that include, within a type, a set of annotation values presented in the form of a drop-down list that a user can select when clicking on a tier segment (Sloetjes,

[61] https://archive.mpi.nl/tla/elan (Accessed on August 26th, 2023).

2023). ELAN has played a significant role in linguistic research, particularly for annotating multimodal data and language documentation (Crasborn & Sloetjes, 2008). ELAN's versatility is especially relevant in the context of sign languages thanks to the user interface that allows for a comprehensive visualization of video data (see Figure 7), as well as tiers and segments (when present). ELAN's visual interface allows researchers to view raw video data, an essential component of any sign language collection (Antinoro Pizzuto *et al.,* 2008; Antinoro Pizzuto *et al.,* 2010), alongside the associated annotations.

Figure 7. ELAN user interface with default tier. The video displayed is taken from the website SpreadTheSign.

4.3.1. ELAN annotations and their application on multimodal corpora

ELAN has been used for the annotation of sign language corpora both in the international and (partially) the Italian context (Santoro & Poletti, 2011). With regards to publically available annotated European sign language corpora, it is found that data collection with correlated annotations on ELAN were developed starting from the second half of the 2000s, with the creation of the *Corpus NGT*, the *DGS Corpus*, the *SSL Corpus* and the *BSL Corpus*. Firstly, the *Corpus NGT*[62] (Crasborn & Zwitserlood, 2008) was developed and partially annotated[63] in the Netherlands. Simultaneously, the

[62] https://www.corpusngt.nl/ (Accessed on September 3rd, 2023).
[63] The NGT Corpus developers state that about 15% of the sessions are glossed and translated (Crasborn & Zwitserlood, 2008).

DGS Corpus (Deutschen Gebärdensprache) was compiled as a result of a 15-year-long project developed, with the aim of creating an extensible annotated corpus (Prillwitz *et al.,* 2008) which is, as of 2023, at its third version (Hanke *et al.,* 2020). The *SSL Corpus*, also known as *Svensk teckenspråkskorpus,* was compiled and annotated on ELAN between 2009 and 2011 (Mesch, 2012; Mesch & Wallin, 2015). Lastly, the *BSL Corpus*[64] was compiled and annotated on ELAN between 2008 and 2014 (Schembri *et al.,* 2014; Schembri *et al.,* 2017). Within the context of the *ECHO project*[65] (Emmerik *et al.,* 2003-2005) a collection of the mentioned European corpora, including conventionally annotated and accessible video files of BSL, SSL, NGT and DGS (Crasborn *et al.,* 2007; Nonhebel *et al.,* 2004a; 2004b; 2004c) was gathered. Lastly, outside of Europe, ELAN has been used to annotate public and partially-public corpora such as the *AUSLAN Corpus* (Johnston, 2019) and the *ASL Corpus from Boston University* (Neidle & Vogler, 2012).

The developers of the BSL and NGT corpora, through their extensive collection and annotation efforts, along with the collaborative efforts on the ECHO project, established annotation standards and guidelines for sign language annotation using ELAN. Crasborn *et al.* (2015) and Cormier *et al.* (2016) outline these standards, including: (1) Multilingual annotation in both the spoken language known to the signers of a national sign language and English[66]. (2) The presence of a tier containing

[64] https://bslcorpusproject.org/ (Accessed on September 3rd, 2023).
[65] https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_0000_0000_0001_494E_3 (Accessed on September 3rd, 2023).

[66] The corpora belonging to the ECHO project collect continuous signing. However, they also include a 'lexicon' section where isolated signs for each included corpus (BSL, NGT and SSL) are collected (https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_0000_0000_0001_4ADF_1 (Accessed on October 17th, 2023) . These isolated signs have been collectively included in three videos, as part of the corpora's data. In these videos, signs taken from each one of the three corpora collections are signed sequentially in alphabetical order, following the sign language State vocal language. Each video was annotated on ELAN, where each Sign was assigned a number and a vocal language label in the three different vocal languages of the States in which data collections took place (i.e., English, Dutch and Swedish). Despite the similar type of data included in these videos, their annotations are not as detailed as the official corpora annotation and their annotation is not documented in the corpora's reports. For

vocal language labels (referred to as ID-gloss tier[67]). (3) The use of two tiers to annotate two-handed Signs. (4) A tier aimed at describing meaning to provide context for variations in formational parameters such as hand shape. (5) The indication of lexical variants, repetition, buoys, plurality, number incorporation, Sign names, fingerspelling, deixis, and transfer of person. In the context of this thesis and the specific features of the MultiMedaLIS Datasets, it is important to focus on the first three standards: multilingual annotation (using vocal language labels) and the development of one tier per hand. Since the third version of the presented Dataset consists of controlled, isolated Signs, the guidelines related to two-handed Signs, lexical variants, repetition, buoys, plurality, number incorporation, Sign names, fingerspelling, deixis, and transfer of person are not directly applicable.

Expanding beyond the realm of sign language annotation, ELAN finds application in the broader European context for multimodal analysis of spoken language and gestuality. Luca Lo Re's CORMIP (2022) is relevant in this regard as his research emphasizes multimodality and involves the codification of various body components, including facial expressions, which aligns with the focus of this work. Furthermore, the works of Emanuela Campisi (Campisi & Özyürek, 2013; Slonimska *et al.,* 2016; Cutugno & Campisi, 2022), on gestuality annotation through ELAN have contributed to shaping the annotation process of the MultiMedaLIS_3 Dataset presented in section 6.3. of this work.

4.3.2 Other tools for multimodal data annotation

While ELAN remains the most widely used software for annotating multimodal sign language data, other noteworthy tools exist such as iLex and ANVIL.

---

this reason, aside from the inclusion of multilingual annotations, these videos were not considered in identifying ECHO corpora annotation guidelines.

[67] «An ID-gloss is the (English) word that is consistently used to label a sign within the corpus, regardless of the meaning of that sign in a particular context or whether it has been systematically modified in some way.» (Johnston, 2010, p. 119)

iLex (Integrated Lexicon) is one such software that stands out for its flexibility in visualizing video data for transcription, accommodating diverse user preferences and layered transcriptions. iLex offers a transcription database for sign languages, integrating it with a lexical database to facilitate a swift lemmatization process. This software is designed to support multi-person projects, enabling collaboration among multiple researchers on a single project. In terms of layout, iLex provides two main views for displaying transcript data: a horizontal view and a vertical view. In the horizontal view, time flows from left to right, with tags represented horizontally, and their length corresponds to their duration. This layout resembles ELAN, as discussed in the previous section (see Figure 7). On the other hand, the vertical view features time flowing from top to bottom, with each interval occupying one row, regardless of its length. Users have the flexibility to choose between a vertical or horizontal view based on their task. Notably, iLex focuses on transcription, offering unique perspectives through its vertical and horizontal visualizations, thereby facilitating ease in switching between different viewpoints and aiding in error detection (Hanke & Storz, 2008).

Another software designed for annotating multimodal material is ANVIL (Kipp, 2001; 2012), specifically developed for annotating audiovisual material with a focus on multimodal analysis. Similar to previously discussed software, ANVIL was created to center around video annotation, employing time-bound elements organized on multiple layers that annotators can customize based on their specific needs.

One notable feature of ANVIL is its capacity to establish links across cross-level annotations, enabling annotators to highlight relationships between different annotated segments. Moreover, ANVIL allows for the annotation of elements that don’t necessarily occur at a specific moment in time or persist throughout the data. In ANVIL, tracks assume that all encodings within a track share similar properties. As a result, users are required to predefine the set of corresponding attributes for each layer in the coding scheme. This characteristic adds a level of customization and precision to the annotation process, enhancing the software's versatility for various research needs.

Another tool worth noting, developed for annotating spoken corpora for pragmatic research, is EXMARaLDA (EXtensible MARkup Language for Discourse Annotation) (Schmidt & Wörner, 2009). This tool was specifically designed to meet the requirements of annotating multi-party unpredictable interactions with a focus on pragmatic cues. What makes EXMARaLDA interesting is the developer's intuition and recognition of the necessity to annotate audiovisual data in pragmatic studies, with a specific focus on facial expressions, gestures, and body posture. For this reason, EXMARaLDA adopts an annotation plan that accommodates the diverse aspects of pragmatic studies. Annotators can distribute annotations on different levels, depending on the speaker and the description of the level. Levels can be created and added freely at any time of the process, and their annotations follow a sequential arrangement from left to right. This flexible approach allows annotators to capture the richness of multi-party interactions, considering various modalities beyond speech alone.

In conclusion, although ELAN remains the most used tool for multimodal annotation, tasks of this type can be carried out by other tools, including iLex, ANVIL, and EXMARaLDA which, despite of the reason behind their development, share certain characteristics that contribute to their effectiveness in annotating multimodal sign language data. Each of these tools was crafted with specific purposes in mind, yet they all showcase a shared commitment to multimodality, flexibility, customization, shareability, durability, portability, and effective presentation of the data. Simultaneously, they aim at supporting collaborative research efforts. In essence, the common thread among ELAN, iLex, ANVIL, and EXMARaLDA lies in their provision of valuable assets for annotating multimodal data, addressing a spectrum of needs within this specialized field.

4.4. Spontaneous strategies of sign language representation

Despite the development of transcription systems for sign languages, and the adoption of SW for the transcription of LIS by some Deaf researchers, there has not been a widespread adoption of any of these systems within signing communities. As a result, signers often develop their idiosyncratic methods for annotating Signs or signed discourse. These unofficial systems that have spontaneously emerged within signing communities will be explored, ranging from small groups to individuals deeply embedded within the signing community, or at its periphery. The present analysis will focus on the defining characteristics of this systems as well as their purposes, with the goal of uncovering potential common characteristics or shared similarities.

This section draws from three research studies on spontaneous notation strategies for sign languages by Ardita & Caligiore (2022), Raniolo *et al.* (2023), and Raniolo (2021). In this section, two systems developed by Deaf performers will be taken into account and compared to spontaneous notation strategies employed by LIS students at various levels.

Oliver Schetrit, a Deaf actor and researchers from France, developed *chorésignes*: a system for memorizing the choreographies of his artistic performances. The iconic system aligns with the visual memorization processes of Deaf individuals (Raniolo, 2021). Figure 8 illustrates that the drawings in *chorésignes* are highly iconic in representing the entire human body, including occasional objects. Interestingly, *chorésignes* not only encompasses information about bodily movements but also places importance on facial expressions within the context of the performance. Manual components play a peripheral role, as evidenced by the only two depictions of hands, located at the top and bottom center of the image. Lastly, images are represented onto a pentagram, following a sort of rithmic flow where movements performed higher or lower are rendered on the pentagramed as if they followed a musical scale.

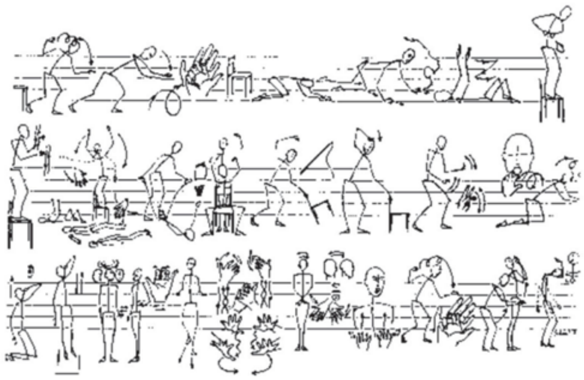


Figure 8. Representation of chorésignes. Schetrit (2016) in Raniolo (2021, p. 213).

Turning to notation systems developed for sign languages, the work of Victor Abbou[68] can be found. Abbou, a Deaf actor from the International Visual Theatre in Paris, used a notation strategy shred with his Deaf colleagues to visually represent LSF Signs that corresponded to the sung passages of a partially musical theatre performance. In this case, Abbou (2017) also employs a pentagram (see Figure 9), but the primary focus is on the hands, their movements, and occasionally on facial expressions as his goal is to represent both manual and body components of LSF.

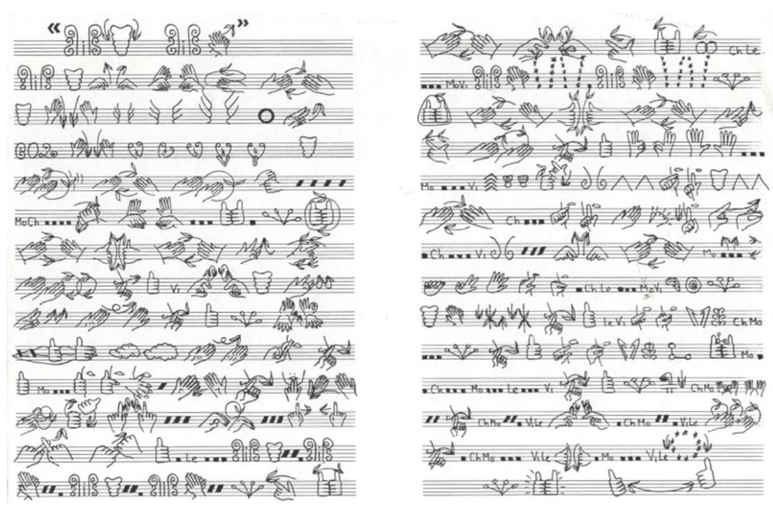



Figure 9. Representation of LSF Signs for the 1990 performance 'Les Pierres'. Abbou (2017) in Raniolo (2021).

[68] It is worth noting that this method, as described by Raniolo (2021), was actually developed in the 1990s. However, for the sake of the logical flow in this passage, it is introduced after discussing *chorésignes*.

While both methods share an iconic nature and make use of the pentagram, it is evident that the components depicted can vary significantly depending on the purpose of the representation. In fact, *chorésignes* provides a comprehensive depiction of the entire body, including facial expressions, whereas Abbou's notation strategy focuses more specifically on the hands and occasionally includes facial expressions. This divergence in representation highlights how these systems adapt to their respective goals and contexts.

Both of these examples stem from artists who primarily use LSF, especially within artistic contexts involving other Deaf performers. Consequently, these systems have been developed by individuals that are well-integrated into the signing communities, the French one in this case. However, for the purpose of a past research (Raniolo *et al.,* 2023), it was considered intriguing to compare these two methodologies, with notation systems developed by individuals that are on the periphery of the signing community or even completely outside it, depending on whether their sign language studies will progress or not. Specifically, focus was placed on university students in this context, aiming to identify both commonalities and significant differences in the notation strategies that spontaneously emerge. To this regard, it is important to emphasize that the primary goal behind the development of notation strategies by students is, generally, the study and memorization of Signs (either during or after class). However, a significant variation in the strategies used by students was discovered, ranging from those using written Italian labels to completely iconic methods. Our data collection, as documented in the qualitative study by Ardita and Caligiore (2023), involved students from the University of Catania and the University of Venice Ca' Foscari. These students were enrolled in LIS classes at the first, second, and third levels during the academic year 2021/2022. Data from students pursuing LIS interpreter training was also included.

In this section, three examples of Sign notations by students are presented. It is important to note that all these students are hearing individuals and tend to incorporate written Italian in their notations, whether in the form of labels or more detailed explanations of movements. This characteristic may reflect their peripheral position within the signing community. In fact, they are not only recent signers, but also hearing individuals who may default to using the traditional strategies they have employed in their previous learning experiences, which involve written Italian. Figure 10 demonstrates that some students rely on non-iconic strategies to memorize Signs, trusting their memory and their ability to recall the Sign during review. As shown, students can indicate multiple translations of a Sign.

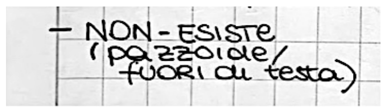


Figure 10. First example of notation strategy by student. The text translates as: 'DOESN'T-EXIST (crazy / out of their mind)' (Ardita & Caligiore, 2023) (CC BY 4.0).

Moving to Figure 11, an evolution in the notation strategy can be observed. Here, the student writes down the vocal language label *MAMMA* (in capital letters, following traditional spoken language labelling standards) and adds an explanation of the manual parameters of the Sign. Additionally, they include a stylized drawing of the Sign being performed, accompanied by arrows that clarify the direction of the movement.

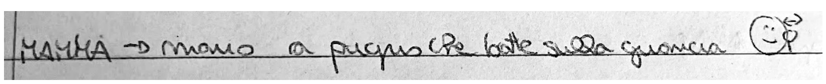


Figure 11. Second example of notation strategy by student. The text translates as: 'MOM → hand closed in a fist taps on the check' (Ardita & Caligiore, 2023) (CC BY 4.0).

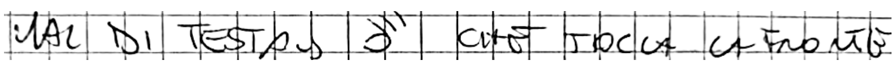


Figure 12. Third example of notation strategy by a student. The text translates ad: HEADACHE [drawing of hand shape] that touches the forehead (CC BY 4.0).

In Figure 12 a similar strategy to the previous one is noted, but with an iconic representation that is limited to the hand shape. In Figure 13, it is possible to see another evolution in the notation strategy, once again shifting towards iconicity. The student includes a spoken language label for the Sign and incorporates symbols. In this case, the iconic representation of the Sign takes a central role and becomes the focal point of the notation. It is also worth noting that the student draws body components, highlighting their importance in this Sign. Furthermore, the student uses written Italian to describe a mouth gesture (puffing cheeks and blowing air), rather than focusing on manual components.

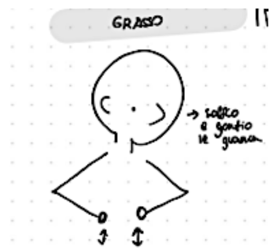


Figure 13. Third example of notation strategy by student. The text translates as: 'FAT / I blow air and puff up my cheeks' (Ardita & Caligiore, 2023) (CC BY 4.0).

After examining various notation strategies for describing movements or sign language in several contexts and from individuals with varying experiences and objectives their similarities and differences can be considered. In all iconic instances, there is some representation of the human body, typically in the form of stick figures. Symbols, such as arrows, are present in both student and Deaf performer notations. Notable differences emerge depending on the purpose of the notation. For example, Abbou and the students are representing a sign language, and do not need to depict anything beyond the signing space, so they typically omit the lower part of the body. One common element overlooked by both Abbou and the students is the movement of the shoulders, possibly because the Signs they are representing do not involve any significant shoulder movement (although this cannot be confirmed for Abbou's performance without access to it). Furthermore, Abbou and Schetrit represent body movements and Signs within a flow or context, while students tend to represent Signs in isolation, essentially creating a vocabulary of isolated Signs. These differences in representation stem from the diverse goals and needs of the notators (Raniolo *et al.,* 2023).

The exploration of these notation strategies reveals their diversity and adaptability. Whether developed by Deaf performers deeply embedded in signing communities or by hearing students, these notation systems prove to be invaluable tools. They showcase how, despite the absence of a standardized writing system, the emergent notation strategies share common traits with the transcription systems artificially created for research or educational purposes. For instance, when comparing the drawings in Figure 11 and 13 to the symbols of SW or the representation of hand shape in Figure 12 to the Typannot glyph presented in section 4.2.2., noticeable similarities arise. These similarities suggest that these methodologies, whether they originate spontaneously or artificially, whether they are developed by Deaf individuals for shared use or by hearing students for individual use, often converge due to a shared understanding of the need for iconic representation. Furthermore, they serve as a natural confirmation of the necessity to move beyond vocal language labels, which are frequently integrated to varying degrees into the iconic representation.

4.5. Practical examples of LIS representation in the Italian context

Electronic media surely provide valuable tools for researchers studying sign languages, facilitating the description of these languages through video recordings that depict Signs. However, the feasibility of this approach relies on the distribution of these publications in digital formats. On the other hand, printed publications on sign languages have historically employed various strategies to represent Signs, either in isolation or within signed utterances. These strategies often include the use of vocal language labels, visual depictions of Signs, or a combination of both. The choice of representation method for sign languages is frequently influenced by the publication's objectives and the approach to sign language studies adopted by the authors. In this section, examples both in print and digital formats that illustrate the representation strategies of LIS Signs and utterances[69] will be presented. For example, in the context of printed LIS-Italian dictionaries, at least two instances make use of drawings, images

[69] While it is true that the examples provided in this section are exclusively from LIS, these strategies are not unique to LIS. In fact, they are general approaches to representing sign languages.

and vocal language labels (written in all caps) to depict Signs in isolation, occasionally is association with symbols. Among these dictionaries, one noteworthy example is the *Dizionario Bilingue Elementare della Lingua Italiana dei Segni* by Elena Radutzky (1992) which is accompanied by the DVD-ROM *I Segni in Movimento*, stands out for its cultural significance. Radutzky organized the Signs in her dictionary based on hand shape. In this dictionary, Sign representation employs a combination of drawings, including arrows, to illustrate each Sign. Each illustration is accompanied by a clear identification label, indicating the page and Sign number within that page. For example, the Sign 576.3, depicted in Figure 14, is the third Sign located of page 576[70]. Furthermore, the author provided transcriptions using symbols adapted from Stokoe notation symbols[71] (see previous section) the corresponding Italian vocal language label, a sentence in Italian where the word translating the Sign is used in context, the grammatical category, synonyms (in Stokoe Notation), and potential Sign variants.

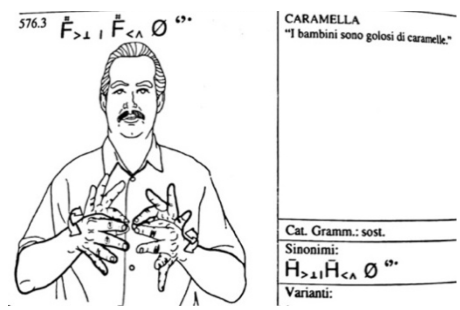


Figure 14. LIS Sign from Elena Radutzky's Dictionary (1992).

Another example of image use in LIS-Italian dictionaries can be found in Orazio Romeo's dictionaries (1991; 1997; 2004; 2021), which feature drawings accompanied by Italian translation labels. Romeo employs different strategies for organizing Signs in his dictionaries. In his 1991 Dictionary, he listed Signs in alphabetical order according to the Roman alphabet. In contrast, his 2004 thematic dictionary grouped

[70] Interestingly, the Sign labels were also included on the accompanying DVD, further enhancing the clarity and unambiguous labeling of both the drawn and video representations of the signs.
[71] Radutzky (1992) proposes a system of symbols specifically developed for LIS. The system included 56 configurations, 16 locations, 48 movements, and 20 manual positions (Celo, 2015).

Signs based on their semantic categories, such as psychology, society, and sports. Below are two examples from the 'A' section of the 1991 Dictionary. As illustrated in Figure 15, Romeo includes labels and possible alternative translations beneath the drawings, which also incorporate arrows to depict movement. As can be seen, Romeo occasionally employs two drawings to represent the movement in certain Signs.

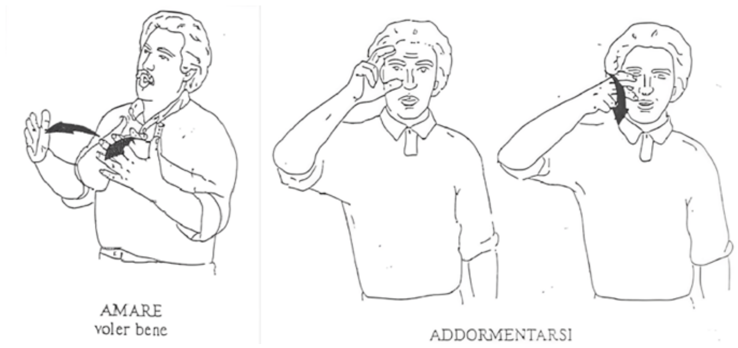


Figure 15. Two instances taken from Romeo (1991, pp. 2- 4).

In the 1990s, when the first two LIS dictionaries were published, digital versions were absent or limited, think of Radutzky's inclusion of a DVD-ROM format which, however, lacks extensibility. Thanks to technological advancements, contemporary access to online resources has become a reality. These resources are continually updated and situated within a multilingual context. One important example of such an online multilingual dictionary is SpreadTheSign, a website that has been mentioned, and will continue to be mentioned, several times throughout this work. The website comprises videos featuring Deaf individuals demonstrating Signs both in isolation and within utterances.

Additionally, it provides Italian translations presented in lowercase letters and offers multiple variations of the Signs. On occasion, basic Signs like colors also include drawings, as shown in Figure 16.

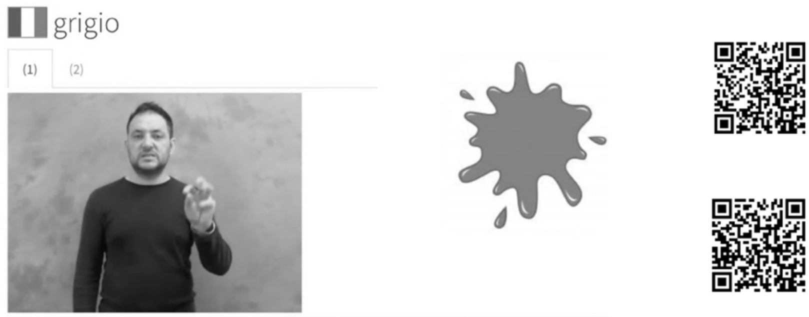


Figure 16. LIS Sign 'grey' ('grigio') from SpreadTheSign.

When it comes to other forms of sign language-related publications, such as articles and printed books, sign language linguists pursuing formalist research efforts frequently make use of vocal language labels. In recent years, these labels have also been complemented by video recordings, made possible by digital platforms, such as QR codes or hyperlinks, a methodology adopted in this work. In contrast, linguists who adopt a functionalist COS-S approach tend to employ language-specific methodologies, where vocal language labels (and more recently, translations) play a more peripheral role, primarily included to enhance accessibility for non-signers.

Moving beyond the context of paper and online dictionaries within the Italian context, the book titled *Scrivere la LIS con il SignWriting* (Writing LIS with SignWriting), authored by researchers from the Institute of Cognitive Science and Technologies (ISTC - *Istituto di Scienze e Tecnologie della Cognizione*) at the CNR, stands as a significant contribution. This initiative emerged from a series of endeavors undertaken by CNR researchers, dating back to 1998 when prompted by Elena Antinoro Pizzuto, to employ SW for transcribing LIS (see section 4.1.1). In addition to the works of the CNR research group, another notable text within the Italian context that uses SW is the 2022 contribution by Volterra *et al.* (as discussed section 4.6).

Lastly, vocal language labels have been a strategy employed in sign language research since Stokoe's description of ASL in the 1960s. Vocal language labels, typically written in uppercase, serve as translations of the meaning of a Sign and are used to 'write' sign languages, alongside their possible corresponding translations in the target language. In some cases, vocal language labels are joned in sequence to construct signed utterances, often accompanied by a free translation of the vocal language label sequence into one or more vocal languages. For example, Geraci and Bayley (2011, p. 127) provide an illustration of this strategy, accompanied by symbols that highlight the specific phenomenon they are analyzing: the distribution of WH- Signs in LIS (see Figure 17).

______wh
a. GIANNI COMPRARE FATTO COSA
b. Cosa ha comprato Gianni?
c. What did Gianni buy?

Figure 17. Combination of vocal language labels with Italian and English translation, adapted from Geraci & Bayley (2011).

As can be seen, only vocal language label representations present several challenges, as a vocal language label alone do not convey information about aspects such as lexical variations used by the signer or the features of a LIS verb and vocal language labels can only be associated within specific contexts, such as a text-containing files. Despite researchers continuing to use vocal language labels, thanks to computer programs like ELAN, it is possible to create time-bound relations between videorecorded Signs and label.

4.6. Italian Sign Language: different approaches, similar strategies.

In the last years, two main attempts at describing LIS have been carried out, from different perspectives. On the one hand, Branchini and Mantovan's recent publication of *A Grammar of Italian Sign Language (LIS)* (2021) offers an in-depth exploration of various aspects of LIS. The digital *A Grammar of Italian Sign Language (LIS)*, initially released in English in 2021 (Branchini & Mantovan, 2021) and later in Italian in 2023, is structured into six parts. The first section discusses the sociocultural context that gave rise to LIS, while the subsequent five sections provide detailed descriptions of phonology, lexicon, morphology, syntax, and pragmatics. On the other hand, the book *Italian Sign Language from a Cognitive and Socio-Semiotic Perspective: Implications for a General Language Theory* (Volterra *et al.,* 2022), holds substantial importance in this context. This volume introduces a new approach to the description of LIS, referred to in this work as the COS-S approach, rooted in an embodied view of language.

Following functionalist paradigms, the authors emphasize the inseparability of linguistic and extralinguistic knowledge, asserting that Signs should not always be segmented into discrete units. They advocate for considering Signs as holistic entities, acknowledging their face-to-face modality, oral nature, and contextual variations. In this paradigm, both manual and body components accomplish equal significance.

These two works, despite representing different approaches to sign language studies, share a common quality: an effort to transcend traditional written vocal language labels of LIS by employing innovative strategies such as images, drawings, and videos of Deaf signers, often accessible through hyperlinks or QR codes (in printed versions). Significantly, *A Grammar of Italian Sign Language (LIS)* (2021) uses vocal language labels written in all caps to describe Signs, complemented by videos of Deaf signers performing the corresponding Signs or utterances. Similarly, the volume authored by Volterra *et al.* (2022) incorporates drawings of LIS Signs or utterances alongside SW notations, written translations in lowercase Italian, and QR codes that link to videos of Deaf signers demonstrating the Signs or utterances on mobile devices, as portrayed in Figure 18.

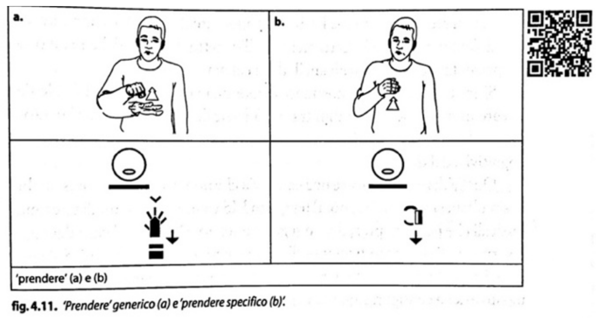


Figure 18. Instance of combination of representation strategies in Volterra et al. (2022). The QR code is part of the original image.

The significance of moving beyond the traditional practice of representing sign languages solely through vocal language labels, is thus recognized transversally. By incorporating visual representations such as images, drawings and videos, researchers embrace a more inclusive and accurate approach to describing sign languages, as visual media allow for the preservation of the visual elements that are essential to sign languages. Videos, in particular, offer the advantage of showing the full range of bodily movements, facial expressions, and temporal aspects that contribute to the richness of signed communication.

Lastly, the use of visual representations enhances comprehension and accessibility for both sign language users and those learning about sign languages. Deaf signers, who are the experts in the language, can more readily engage with materials that accurately reflect their language and communication mode. For learners and researchers who may not be fluent in the sign language being studied, visual representations provide valuable context and insight into the structure and meaning of Signs.

Chapter 5
Ethnographic Tools, Computational Annotation, and Ethical Considerations in Multimodal Data Collection

In this chapter, the integration of ethnographic tools and methodologies with computational annotation for the study of multimodal data will be discussed, particularly with reference to Italian and LIS. A key point of discussion within this chapter are the data collection techniques employed by linguists, sociologists, and anthropologists, emphasizing participant observation and the cooperation with native speakers. The relevance of multimodal electronic recordings in these processes will be highlighted. Following that, an overview of a multimodal corpus of spoken Italian will be provided. This will lead to a comparison of this collection with the *Corpus LIS*, opening a discussion on the delicate balance between collecting data in semi-controlled environments with the need for authentic data, as well as the potential benefits and challenges of formal linguistic analysis. The strategies employed for the collection and processing of the two mentioned multimodal resources share certain characteristics. These can be extrapolated to serve as a reference for the current state of the art for multimodal data collection in the Italian context, as well as provide best practices for future data collections.

5.1. Conjugating ethnographic tools and methodologies with computational annotation

Duranti (see section 2.2.1.) describes data-collection techniques practiced by linguistic anthropologists, widely taken from ethnographic methodologies such as participant observation and work with native speakers to obtain interpretive information on the recorded material, interviews, writing and note-taking (1997; 2007). In the present section, the focus is on the importance of electronic recordings that, since their introduction, have allowed for a deeper analysis of language in context and, additionally, are the tools that allow in the first place for computational annotation.

Particularly relevant for sign language collection and multimodal analysis, electronic recordings allow to stop, restart, rewatch portions of video materials, thus noticing aspects that may have gone previously unnoticed. In fact, the introduction of automatic and electronic means to record human interactions has had a great impact on this field, an aspect that had already been predicted by Stokoe *et al.* (1995), when they wrote:

> Looking ahead, it appears that a future science of language and communication, both visible and acoustic, will be made possible, in all probability, not by refinements in notational systems but by increasing sophistication in methods of recording, analyzing, and manipulating visible and auditory events electronically.
> (Stokoe *et al.,* 1995, p. 354)

Ideally, the mashup considered in this section between ethnolinguistics tools and methodologies and computational annotation should involve joining ethnolinguistic qualitative tools for language collection together with language-specific annotation models. Studies on ethnographic research habits and procedures, have been mainly carried out on vocal languages. However, the principles for data collection can be applied to sign language as they do not refer specifically to vocal language. Moreover, the oral nature of LIS and the unique situation of bimodal bilingualism experienced by Deaf signers, together with the experience of Deafhood, are aspects that cannot be ignored in the collection of LIS data. Likewise**,** these peculiarities call for an in-depth reflection also on the annotation of said data, involving issues such as the assignation of vocal language labels to each Sign for computational processing and the description of the features (both manual and bodily) of annotated Signs.

Keeping these principles into consideration, a case study within the Italian context is described. The case in question called for the collection natural (or semi-structured) interactions in Italian. The process of data collection and partial annotation of the *Corpus LIS*, which has already been discussed in section 3.2.3.1., will be also taken into consideration, as it is believed that it is its best fit within this work. In the following sections the process of data collection described by the developers will be discussed. After that, a conclusive section will be dedicated to issue of collecting spontaneously elicited language, ethical aspects involved in data collection and the need of using specific tools for the data to be collected and analyzed within an interdisciplinary computational framework.

5.1.1. CORMIP: a Multimodal Corpus of Spoken Italian

In his Ph.D. thesis, Luca Lo Re (2022), describes the development of a prototype corpus of Italian: The Multimodal Corpus of Spoken Italian (CORMIP— *Corpus Multimodale dell'Italiano Parlato)*[72]. The *CORMIP* is based on a pragmatic and perceptive method that frames language as a phenomenon that occurs though the use of the phonic, acoustic, body, and auditory channel (Lo Re, 2022). In this section, the focus is put on the data collection process detailed by the developer.

The primary objective of data collection was to gather naturally elicited data using electronic tools capable of capturing the multimodal aspects of spoken communication, with a specific emphasis on gestures. The goal was to collect a substantial amount of data that not only reflects language production, but also its relationship with the surrounding context and the dynamics of interaction. This is because omissions and nuances in natural communication can be challenging to retrieve. As the methodology chosen for data collection significantly influences the degree of naturalness in the recorded interactions, when aiming to collect spontaneous data, it is important to minimize the influence of external factors on the speakers or signers. This calls for the use of non-invasive recording tools that allow for data collection to be replicated across different contexts. To achieve this, the *CORMIP* was recorded using a GoPro Hero 6 Camera and a panoramic microphone. These tools were placed in familiar settings for the participants, most of whom were acquainted with each other. With regards to the contents of the exchanges included within the corpus, participants were given the

[72] I want to express my gratitude to Dr. Luca Lo Re for providing the access to his corpus, allowing me to directly consult both the videos and associated ELAN annotations.

liberty to discuss any topics of their choice. Therefore, the corpus includes conversations about various subjects such as trip planning, university lectures, or casual gatherings among friends. The *CORMIP* comprises a total of six recordings, divided into three conversational genres (conversation, dialogue, monologue), and it was recorded in two different cities (Catania and Florence), for a total of around 35 to 40 minutes of recorded material.

The collected data was transcribed and annotated referring to the *Language into Act* theory (L-Act) (Cresti *et al.,* 2018). This approach, while distinct from its application in sign languages, may be described as embodied, as it creates a framework where the pragmatic consequences of an utterance are influenced significantly by prosody. Lo Re's annotation methodology primarily focused on the transcription of spoken utterances, with a specific emphasis on annotating the speech flow. This involved segmenting and labeling the identified units following a perception-based methodology, allowing for the description of both illocutionary acts (following Austin, 1962) and semantic acts (Lo Re, 2022, p. 153). However, as mentioned, what sets the *CORMIP* apart from other collections in the Italian context is its focus on multimodality. Lo Re developed an annotation methodology for gestures based on Kendon's framework (2004), through which Kendon demonstrates that gesture and speech co-occur and mutually influence each other in daily interactions in order to convey intended meaning. A particularly interesting aspect of Lo Re's annotation is the specific development of ELAN tiers. In fact, he created a specific segment for the description of gesture with a hierarchically dominant tier, the *Gesture Unit*, that includes all dependent tiers. Beyond the overall annotation framework, relevant features include the identification of distinct tiers to describe the movement of each hand[73] and the presence of facial expression. Most importantly, in fact, the inclusion of

[73] This aspect is relevant as it reflects methodologies developed for sign language corpora annotation, where two hands are described within separate tiers (see Nonhebel *et al.,* 2004a; Mesch & Wallin, 2015).

information on facial expressions is a key aspect of Lo Re's annotation. Lo Re recognizes the importance of facial expression in conveying and modifying meaning by including a level that, although on a superficial level, signals the presence of a non-neutral facial expression.

Lo Re's work sets a theoretical framework and offers valuable insights for those embarking on multimodal annotations of Italian. Within the context of LIS studies, despite being detached from the specific topic of sign language annotation, it still holds relevance as it contributed, together with other annotation methodologies, to shape future annotation frameworks, such as the one presented for the MultiMedaLIS_3 Dataset (see section 6.3.1)

5.1.2. A discussion on spontaneous data collection and annotation strategies for multimodal resources in Italian and LIS

When working with datasets collected in everyday situations, it is crucial to consider the sociolinguistic implications of the data collection process itself. During the second half of the 1920s, a study was conducted at the Hawthorne Works factory in Illinois, to study if workers were more or less efficient depending on lighting conditions. However, the study's results turned out to be mostly irrelevant to its original research question. Instead, the study revealed a different phenomenon: an increase in worker's efficiency was noted when they were aware of being scrutinized and received special social treatment (Jones, 1991). In the context of sociolinguistics, a similar phenomenon, known as the *Observer's Paradox*, was theorized by Labov (1972) during his studies on linguistic methodologies. In fact, while collecting African American Vernacular English in 1968, he noticed that the effort to observe how speakers talked when they were not being observed created the Observer's Paradox, i.e., the authomatic conditioning felt by participants when asked to speak as they are accoustomed to (Labov, 1997; Duranti, 2007). In fact, similarly to the workers at Hawthorne Works, Labov noticed that when people are aware they are being observed,

they tend to alter their speech to match what they think is expected of them, making it challenging to collect authentic casual speech styles (Cukor-Avila, 2000).
In order to address this Paradox, several solutions have been discussed, including the role of the interviewer, their social class compared to the interviewee, and their inclusion within the community being studied.These observations are applicable to the realm of sign language studies, where the interviewee's perception of the interviewer may very well (if not more) influence the signing flow. Within this context, the different approaches that were undertaken by researchers in the collection of multimodal corpora, both for Italian and LIS will be explored.

As expected, reference is made in particular to the collection of the *CORMIP* (Lo Re, 2022) and the *Corpus LIS* (Geraci *et al.,* 2011). Lo Re (2022) states that the adopted strategy was the following: of course, for legal and ethical reasons it was necessary to have the participants approval for recording and, additionally, hiding the recording tools may be unethical as well as compromise the quality of the recorded data. Consequently, speakers were informed that the recordings were made for research purposes, without specifying the field of studies or elements of interest. As for the discretion of the tools of data collection, non-invasive instruments were selected and used within familiar locations to the speakers.

Similar methodologies were applied in the colleciton of the *Corpus LIS* (Geraci *et al.,* 2011). As mentioned in section 3.2.3.1., the authors minimized the observer's paradox by allowing signers to sign within a 'safe space' setting up the cameras within a familiar settings (mostly ENS locations) and without the influence of an external interviewer who was present only when necessary, and was a Deaf person. However, in contrast to Lo Re, who opted for less invasive recording tools, the *Corpus LIS* collectors used three videocameras. In fact, the concept of non-invasiveness of tools becomes more problematic in relation to the collection data sign language corpora. In fact, for the *Corpus LIS*, the choice of using three cameras was led by the need for high quality video data, which was among the main focuses of the collection itself (Geraci *et al.,* 2011; Branchini *et al.,* 2013).

To summarize, the similarities between the two data collections can be highlighted. In fact, as can be seen, the two data collections, despite their different aims, follow similar ethnolinguistic methodologies with the aim of obtaining spontaneous data. Key principles include: disclosing the recording to participants, providing a comfortable and familiar environment for interactions, and using non-invasive recording tools, within certain limits dictated by the study's goals and the language modality.

5.2. The issue of formalization of semi-authentic and authentic multimodal data

Both corpora collections taken into consideration in this chapter were developed with the prospect of being annotated, even though, as mentioned in section 3.2.3.1., the task has not been extensively carried out on the *Corpus LIS*. Without delving into the annotation strategies developed or prospected, which are detailed respectively in Lo Re (2022) and Santoro & Poletti (2011), the objective of this section is to discuss the implication of said annotations: the formalization of (semi)authentic data.

The two corpora show that familiar settings (both in terms of participants and location) are a fertile ground for data collection aimed at gestural and Sign analysis and that, using the available tools such as ELAN, it is inseed possible to annotate the data. A starting point is the integration of available technologies, from recording tools to editing softwares to the storage of the collected and annotated data. In fact, technology plays a central role on different levels of the data collection and annotation. It is taken for granted, nowadays, that in order to collect multimodal data, multimodal tools that allow for an audiovisual rendering of the data to analyse should be employed. This possibility, made feasible by technological advancements, was complex or impossible decades ago.

With regards to the annotation, a key aspect, in this case, is the development of a standardized annotation strategy that is shared, at least, among the annotators of the same corpus. In fact, establishing and documenting clear guidelines and conventions for annotators promotes consistency across different annotators and timelines. This task, as mentioned, was undertaken in both corpora, but the nature and purpose of annotation have led to differences in their approaches. In the case of the *Corpus LIS*, its development by Deaf researchers adds significant value to the annotation. This dual perspective offers options for structured descriptions while, at the same time, incorporating Deaf researchers' insights.

Another crucial aspect to consider is the organized storage of the data, preferably on cloud-based platforms that are available to anyone who needs access and are not at risk of being lost or damaged[74] thus granting the long-term use of the corpus itself, which makes the resource relevant for additional purposes, such as diachronic studies. Moreover, for the annotations to be accessible and replicable, it is important to share them with the community. A key aspect in this case is the possibility of associating annotation and audiovisual data. Relying solely on a text file containing annotations would oversimplify the richness of multimodal data and make the annotators' efforts inaccessible. The interplay of text, audio, and visual elements is crucial for preserving the complexity and depth of the collected sign language data, ensuring that it remains a valuable resource for research and analysis in its multimodal context.

Additionally, the controlled nature of the collected data is a topic that asks for careful consideration. Controlled data collection offers a range of advantages, particularly within multimodal contexts, where variables can be carefully managed by researchers.

[74] Regarding the storage of the two corpora discussed in this chapter, there is a distinction in how they are managed. The CORMIP is stored on a Google Drive, while the *Corpus LIS*, to the best of the author's knowledge, is distributed across multiple hard disks located throughout Europe. This difference in storage methods may reflect the specific infrastructure and management approaches of the organizations responsible for these datasets, as well as the years in which they were collected.

Consider technical aspects such as adjusting lighting conditions or positioning recording equipment for optimal results. In terms of participants, researchers have the flexibility to select individuals who align with the research objectives and to set specific discussion topics. This level of control provides valuable opportunities, such as the ability to test hypotheses effectively. For example, a portion of the *Corpus LIS* is dedicated to semi-structured dialogues, a deliberate choice made by its developers to explore WH- utterances in LIS (Cecchetto *et al.,* 2011). Nevertheless, the application of control over data collection and formalization processes can sometimes be driven by the desire to align multimodal data, especially in the context of sign languages, with predetermined models of language structure and usage. In these cases, controlled data collections and well-structured computational annotations are seen as tools to conform to these formalizations, resulting in data that can be adapted to formal linguistic analysis and modeling. This pursuit of formalization may clash with the authenticity of the data. In fact, overly controlled environments for data collection might restrict the naturalness and variability of the sign language being studied, leading to formal models that do not accurately reflect the language day-to-day usage.

Certainly, reaching a conclusion on this issue is not an easy task. Within this context, what researchers can do is attempt at striking a balance between the need to capture authentic and diverse data and the advantages of controlled data collections. Depending on the purpose of a study or, more broadly, or the research stage of a particular topic, a controlled data collection approach can contribute to the identification of foundational linguistic principles (consider the early attempts at describing LIS in the 1980s). In contrast, naturalistic data collections will provide concrete insight to the studied language and its community.

Regarding annotation, it is undeniable that computational processing of data, even for sign languages, requires some form of written annotation (Antinoro Pizzuto *et al.,* 2008). Whereas this annotation is essential for linguisitc data processing, it should be developed as a result of extensive analyses.

5.3. What can be learnt and adapted from past experiences in multimodal data collection

The data collection efforts mentioned in the previous section undoubtedly serve as a basis for the development of further multimodal resources. After careful consideration and discussions with the developers of the two studied corpora (see section 5.1.), specific characteristics have been identified for incorporation into the data collections presented in this work, encompassing both methodology and content.

Incorporating elements shared by the *Corpus LIS* (Geraci *et al.,* 2011) and the *CORMIP* (Lo Re, 2022), the following key elements will be replicated[75]:

i. The portability and non-invasiveness of the recording tools: priority will be given to the use of portable and non-invasive recording tools to capture data in familiar, if not even domestic, contexts with the aim of ensuring the least possible amount of interference to the participant's environment.
ii. The high resolution of the recording: data should be recorded with the highest possible resolution with while maintaining the principle of portability and non-invasiveness, which are always given priority, as video data below HD quality «[…] may be unlikely to provide recorded content of sufficient quality for detailed analysis in the future» (Hanke & Fenlon, 2022, p. 36).
iii. Selection of tools able to capture of multimodal data: tools capable of capturing multimodal data that is suitable for post-processing will be chosen.
iv. Uniform, logical and motivated processing: all collected data will be labelled with univocal, clear and logically organized labels following the principles of transparency and comparability (Hodge & Crasborn, 2022).

[75] Refer to sections 6.2. and 6.3. for insights into their practical application.

v. Cloud-based storage: recordings will be securely kept on cloud-based platforms that are shared among developers to grant accessibility and long-term usage.

vi. Tools for annotation and storage: an annotation methodology using ELAN for an effective and accessible computational processing of the data will be developed and described. Annotations will be stored, when complete, within the same platform used to store the audiovisual data.

5.4. An ethical statement on working with sign language data

When discussing the collection and annotation of sign languages for linguistic purposes, the ethical and sociological implication of the work carried out should be considered. In fact, «[l]inguists should not always take for granted that their work is by definition beneficial for the communities of signers and should try to explain in a clear fashion what the possible benefit can be of linguistic research on the target sign language»[76] (Finnish Association of the Deaf*,* 2015, p. 39). In essence, conducting research on sign languages without actively involving the community can be criticized for resembling linguistic 'colonialism'. For instance, the development of a LIS treebank (Caligiore, 2020) will be discussed in section 6.1.1. While the project undoubtedly provided valuable insights and intuitions, especially regarding the unambiguous annotation of LIS data, it was executed without consultation with LIS experts. This approach could perpetuate the imposition of tools and instruments for sign language processing, which may not align with the needs signing community and, therefore, may not be accepted by it.

In this broader context, the creation of resources for sign languages should take into consideration the community that employs the language in their daily lives, be it at home, in the workplace or other settings. The motto of oppressed and minority

[76] https://slls.eu/slls-ethics-statement/ (Accessed on September 5th, 2023).

communities in this regard is 'Nothing about us, without us' (Charlton, 1998). In the context of sign language research, and from the point of a hearing researcher, this could translate into 'Working *with* Deaf people, and not *on* Deaf people' by including a group of individuals who fall under the category of 'invisible disability', a classification often disputed by some members of the Deaf community who regard Deafness as a different way of experiencing life.

Consequently, in the data collection efforts presented in this thesis, attention was taken to involve individuals who play various roles within the local signing community. The research team includes LIS experts with diverse backgrounds, ranging from Deaf signers to Children of Deaf Adults (CODA), to interpreters. Additionally, participation in various conferences during the 2020/2023 period facilitated interactions with different signers, both from Italy and abroad, some with experience in sign language research and others without, all of whom provided valuable perspectives and insights into the work presented here.

5.5. Positioning this work within an Open Science framework

While the specific outcomes of this work are yet to be made public (see section 6.3.), this project has been consistently grounded in an Open Science framework from its beginning. For this reason, the adopted approach follows the guidelines established by the European Commission, promoting interdisciplinary collaboration and the replicability of research efforts. In particular, the framework Open Data, which includes the FAIR principles—ensuring data is Findable, Accessible, Interoperable, and Re-usable. This commitment has been integral throughout the development of the presented data collections, as well as in the selection of data collection tools (see section 6), and it will continue to guide potential future developments. In line with the project's nature and intended outputs, adherence to the principles outlined by Wilkinson *et al.* (2016) has been and will be maintained.

Starting from the concept of Findability, the data has been defined through globally unique and persistent identifiers. This applies to both file naming, where each video possesses a distinctive identifier and annotation (see sections 6.3.4.2. and 6.5.3). Since a single individual was responsible for data collection, no user-specific information is disclosed. However, in upcoming developments, additional metadata will be incorporated, including details like age, origin, identified gender, and featuring the identifier corresponding to the described data (i.e., the video). Furthermore, to enhance Findability, the data will be systematically indexed within a searchable resource.

Moving onto the concept of Accessibility, the data is readily accessible through its unique identifier, adhering to protocols explicitly designed for the third data collection (see section 6.3.4.2.) and annotation (see section 6.5.3). These protocols, outlined in Chapter 6 of this work, are characterized by their openness, free availability, and universal implementability, ensuring broad accessibility for all interested parties.

In terms of Interoperability, the data has been collected and annotated using widely applicable language for knowledge representation, a crucial consideration, especially in the context of sign language data. The specifics of the annotation procedure for the third version of the dataset are clarified in section 6.5. The same annotation methodology will be consistently applied to future data collections, ensuring continuity and interoperability.

Concerning Reusability, the data will be released with a clearly defined and easily accessible data usage license. Additionally, it will be linked to comprehensive provenance information, further enhancing its reusability for interested parties.

5.5.1. Publishing data collections: repositories for multimodal corpora

When adhering to the Open Science framework for a data collection output, careful consideration should be given to selecting an appropriate venue for publication. While the FAIR principles (Wilkinson *et al.*, 2016) when applied to a private resource, can facilitate categorization and the subsequent retrieval of items, it is important to recognize that they function as a *means* to an end, emphasizing the broader goal of enhancing accessibility, collaboration, and transparency rather than being an end in themselves. For this reason, choosing a suitable platform for dissemination is important to ensure that the outputs of one's research can reach a wider audience.

When selecting repositories for a sign language data collection, a central requirement is the ability to host video data. As mentioned by Crasborn (2022) the two primary archives that host sign language data collections are the Endangered Languages Archive[77] (ELA) and The Language Archive (TLA)[78]. ELA is a repository dedicated to preserving collections of audio and video recordings of endangered languages. Its primary focus appears to be on preservation rather than being strictly research-oriented. This platform serves as a resource for safeguarding linguistic diversity and cultural heritage by housing recordings of languages facing endangerment. TLA is a platform developed in the Netherlands by the Max Planck Institute for Psycholinguistics (see section 6.5.6). This repository encompasses a diverse range of language corpus data, including audio and video recordings, photographs, notes, and experimental data considered essential for one of the hosted languages' documentation and analysis. Given the topic of the present work, a relevant section within this repository is TLA's Sign Language Corpora, sign language corpora are housed. With nearly 600 openly accessible folders, this section features recordings of Dutch, British, and Swedish sign languages.

In the French context, a prominent choice appears to be Ortolang[79], a platform designed for building a repository of language data along with well-documented tools for processing. Specifically, Ortolang adheres to the recommended data format published by CLARIN, a digital infrastructure offering access to language data and tools.

[77] https://www.elararchive.org/ (Accessed on February 1st, 2024).
[78] https://archive.mpi.nl/tla/ (Accessed on February 1st, 2024).
[79] https://www.ortolang.fr/en/home/ (Accessed on February 1st, 2024).

Ortolang is structured to host raw data in various forms, including text (CSV, TXT), audio (MP3, WAV), and crucially, video (MPEG).

To the author’s knowledge, in the Italian context, no dedicated repositories seem to exist for the collection and preservation of multiple linguistic resources in video or audio format, either LIS or spoken Italian. Currently, most corpora, especially those of spoken Italian, are hosted on private platforms developed for specific purposes. The absence of dedicated repositories is significant, given the potential advantages they could provide in consolidating, disseminating, and preserving the languages present in Italy, whether they be signed or spoken. Moreover, establishing a repository within the Open Science framework, for instance following the Ortolang model, would not only contribute to the preservation of linguistic diversity, but also foster collaboration and enhance accessibility thanks to the establishment of shared characteristics.

Chapter 6
Glosses, Guidelines, and Dataset Development: An Interdisciplinary Approach to LIS in the Medical Domain

Building upon the insights gained from past experiences in relation to the collection and annotation of multimodal and sign language data, as exemplified in Chapter 4 and 5, this chapter introduces the MultiMedaLIS_1, 2 and 3 Datasets, three preliminary LIS Datasets in the medical domain. These datasets represent the concretization of three years of research, which have led to carrying out three different efforts at gathering LIS Data, resulting in the collection of the most extensive RADAR/camera dataset for a sign language: the MultiMedaLIS_3 Dataset.

Before presenting the datasets, as well as the methodologies employed for the data collections and processing, a discussion on the issue of using vocal language labels (see section 4.5.) for sign language annotation will be established, supported by a contextual background, closely linked to a prior work of the author as master's student (described in Caligiore, 2020; Caligiore *et al.,* 2021). This contextual background will provide tangible evidence of the evolution in the author's approach to LIS research, as it is expected, while still preserving certain foundational principles that have remained consistent over the years.

6.1. The issue with glosses

The oral nature of sign languages presents a dual-sided surface. It allows for a multifaceted, multimodal observation free from written representation, yet it also poses challenges in framing these languages and their intricacies (Fontana, 2014). In sign language research, a historical solution to these challenges, regardless of whether representation is digital or analogic (such as manual writing), has been the use of glosses[80] to transcribe Signs (Antinoro Pizzuto *et al.,* 2010). Glosses have historically served a practical purpose, primarily to record signed content for memorization, especially among LIS L2 learners (see section 4.4.), for conducting studies on syntactic structure (see section 3.2.3.1.) or to simply provide examples in printed publishing (see section 4.6.). However, the pervasive use of glossing in sign language research goes beyond mere practicality as it aligns with a deeper theoretical assumption. As explored in Chapter 2 of this thesis, formalist approaches in sign language studies aim to uncover the underlying structure of sign languages and position Signs within specific constraints. Consequently, the adoption of glosses emerges as a direct solution to formal inquiries, allowing for the establishment of parallels among structures of signed and vocal languages while, at the same time, aiding in the identification of the deep structures of sign languages.

Despite their extensive use, glosses inadequately capture the multimodality, multilinearity, and tridimensionality of a Sign or a signed utterance. In fact, annotating sign language solely through glosses, by essentially using another vocal language to describe a Sign, fails to establish an exact Sign-form correspondence and does not allow readers to retrieve the original form of the Sign when it is not visually perceivable. Additionally, glosses inadvertently suggest linguistic categories and theoretical models that can compromise the distinct characteristics of sign languages (Volterra *et al.,* 2022). In fact, sign languages, being visual-spatial languages, use not only manual Signs but also facial expressions, body movements and spatial relationships to convey meaning (see section 2.2.5). These intricate layers of communication are integrated into the signing process, creating a holistic and multisensory experience that cannot be adequately captured through text-based vocal language labels.

[80] As mentioned in the introductory section of this work, throughout the present thesis, the term ‘gloss’ has been replaced by ‘vocal language label’.

Upon closer examination of glosses and considering these factors, it becomes evident that transcribing a sign language solely through glosses results, at the very least, in information loss. Relying solely on glosses for sign language transcription then proves to be an unsustainable practice as glosses falls short in capturing the multifaceted nature of signed communication as vocal language labels inherently flatten the rich nuances of sign languages by reducing them to isolated lexical items devoid of their inherent visual and manual qualities.

Taking LIS as an illustrative example pertinent to this study, it has been noted that the Sign for 'grey' ('*grigio'*) is presented on SpreadTheSign (Cardinaletti, 2016) through three distinct realizations of the sign, as can be seen in Figure 16[81] (see p. 128). At the same time, if a labelling on the Sign shown in Figure 19 were to be carried out using traditional LIS labeling standards, one could write *DARE*, which is the Italian translation of the verb 'to give' in reference to an object.

Figure 19. Depiction of the LIS Sign 'DARE' ('to give') (Romeo, 1991, p. 36).

However, if the iconic representation of the Sign were to become unavailable at a later stage, it would be virtually impossible to reconstruct the original form of the sign, even within context. As highlighted by Volterra *et al.* (2022), formational parameters can vary and convey different information about the agent and the object involved. For instance, the signer could represent giving a pen or a ball (see Figure 20). The agent

[81] In relation to this example, it is essential to note that signs denoting colors exhibit variability across Italy due to sociolinguistic variations (Geraci *et al.,* 2011).

could be the actant or the receiver of this action, and the signer could take the role of a child giving something to an adult, or an adult giving something to a child. In all these instances, the formational parameters will consistently change and convey crucial information that would be lost, if one relied solely on the label *DARE.*

Figure 20. Possible variations in the formational parameters of the LIS Sign labelled as DARE ('to give') depicting the action of giving a pen (on the left) and a ball (on the right) (CC BY 4.0).

Having noted the shortcomings of using labels (or glosses) as the only means of annotation in the annotation process carried out in this work, a decision was made to develop an annotation system that would still rely on vocal language labels— as they appear to be unavoidable at this stage (Antinoro Pizzuto *et al.,* 2010)— but at the same time rethink this practice through a different, language-specific, approach.

6.1.1. Annotating LIS: description and insights from a past experience

The Universal Dependencies treebank for Italian Sign Language (Caligiore, 2020)[82] was developed in 2020, becoming the first LIS treebank comprised of approximately 230 utterances extracted from LIS videos in the storytelling domain, including a fairytale (Little Red Riding Hood, *Cappuccetto Rosso*) and a short story (The Three Brothers, *I Tre Fratelli*). What is relevant in this regard and, therefore, will be discussed in the present section, are the steps that were undertaken prior to the creation of the treebank itself. In fact, it was crucial to conduct an analysis and comparison of existing treebanks, with a specific emphasis on annotation methodologies. Various Universal

[82] Available on GitHub at https://github.com/alexmazzei/LIS-UD (Accessed on October 14th, 2023).

Dependencies treebanks for Italian were examined to gain insights into the analysis of spoken Italian (see section 3.1.), the most used dependency relations, and how words are interconnected. This reliance on spoken language resources was the consequence of a lack of resources of this kind for sign languages, with the only Universal Dependencies treebank available for a sign language being the Swedish Sign Language (SSL) treebank (Ostling *et al.,* 2017), derived from the *Swedish Sign Language Corpus* (SSLC) (Mesch & Wallin, 2015).

The analysis of the SSL treebank served multiple purposes. It not only helped in identifying a starting point for processing LIS videos, but also raised awareness of the ethical considerations that might arise when adapting annotation strategies developed for spoken languages to sign languages. During this phase of research, the objective was twofold: to create an original segmentation and annotation framework for LIS that aligns with the existing sign language resource and, concurrently, introduce new and unique perspectives on annotation strategies. These strategies were intended to be language-specific and unambiguous for LIS.

The initial phases in building a treebank involved tokenization and POS tagging, which has been discussed in section 3.1. The processes of tokenization and POS tagging for sign languages are not as straightforward as they can be for spoken languages. It was essential, at that moment, to devise an innovative annotation strategy that would align with prior research, while at the same time addressing the distinctive characteristics of LIS. Nevertheless, an immediate challenge was encountered. Tokenization, which involves identifying the smallest units within a corpus (Sharma, 2020), posed a fundamental question: how could tokens be described without relying on traditional language labels? Transcribing Signs using all capital letters in written Italian did not fulfill the requirements of unambiguous and language-specific annotation, principles that were considered unavoidable. Consequently, the annotation developed for the LIS treebank offers insights into a potential approach to sign language labeling, with necessary adaptations depending on the annotation's purpose. By using the ELAN function of tier development (see section 4.3.), different tiers were created with the aim of including information that was pertinent in relation to the aim of the annotation itself. The first tier that will be discussed is the *Segno* (Sign) tier.

It has been stated that vocal language labels are essential at this stage (Antinoro Pizzuto *et al.*, 2008), for this reason different methodologies for Sign labelling were used to fill the segments of the *Segno* tier. The primary source for Sign transcription in this work was the *Dizionario Bilingue Elementare della Lingua dei Segni Italiana (LIS)* by Elena Radutzky (1992) in its DVD-ROM version. Similar to the printed version of the dictionary, the DVD-ROM assigns a specific code to each Sign performed in isolation, which includes the Italian translation and a sequence of letters and numbers corresponding to the page and position of the Sign in the printed *Dizionario* (see section 4.5.).

To provide a practical example, in the annotation methodology developed for the LIS treebank, if the signer is narrating the story of *Cappuccetto Rosso* (Little Red Riding Hood), the name of the main character would not be labeled as *'CAPPUCCIO ROSSO'* (which translated literally from written Italian as HOOD and RED). Instead, it would be labeled as *'cappuccio664.2'* and *'rosso202.1'*[83]. While this annotation may not be as intuitive as conventional vocal language labels, it provides a more unambiguous transcription for Signs as it follows similar principles to the ID-gloss system (Johnston, 2010). Anyone with access to the DVD-ROM or printed *Dizionario* can identify the Sign in question by referring to its unique code. While the *Dizionario* served as a great source for the identification of unique vocal language labels, it had limitations due to its completeness, and the fact that it was published in 1992. Consequently, the

[83] These labels consist of various elements. *'Cappuccio'* and *'rosso'* represent the corresponding spoken language labels for the two LIS signs. The numbers before the dot ('664' and '202') refer to the page numbers in the printed version of the *Dizionario*, while the numbers after the dot ('2' and '1') indicate the position of the sign within the page of the printed dictionary, typically accommodating three signs per page, as discussed in section 4.5.

Dizionario could not be the only source for the LIS treebank annotation. For this reason, another crucial resource used was the website SpreadTheSign (again, refer to section 4.5). As an online resource, Signs from this website were labeled with the vocal language label of the Sign as provided by the website itself, followed by the acronym -STS. For example, *'GATTO-STS'* (CAT-STS). As with the Signs identified in the *Dizionario*, a similar strategy was employed for Signs found on the SpreadTheSign website. To ensure access to the labeled Sign even in the absence of video material, direct links to these Signs were included as appendices to the treebank. If the website contained different variations of the Signs, the variation number would be included in the acronym following the Italian label.

However, it became apparent that it was impossible to label all Signs using only these two sources. Since the videos in the treebank primarily featured Transfer Units (Cuxac, 2000; Volterra *et al.,* 2022) especially Transfers of Person, there were variations in manual and body components. To address this, a language-specific system was needed. This led to the use of SW (section 4.1.1.) which was chosen to provide an unambiguous representation of Transfers. Each Sign developed with SW was transcribed as 'SW-Italian_Label'[84]. This methodology allowed for a precise description of Units that are not present within Dictionaries as they are 'non-standard'.

As previously mentioned, Signs often exhibited variations in formational parameters. To mark these variations, specific methodologies were developed. When a variation involved the manual parameter of hand shape, codes such as 'CL5)' were added after the Sign vocal language label. Here, 'CL' stood for CLassifier[85], and '5)' represented

[84] For an in-depth description of this process, see Caligiore (2020).

[85] The term 'classifier' will not appear elsewhere in this thesis. In the COS-S framework adopted in the present work, what Mazzoni (2008) refers to as a 'classifier' is defined as a 'Transfer Unit'. These are structures that signers use to convey morphological information, employing one or more representational strategies to indicate or specify the subject of their signing (Volterra *et al.,* 2022).

the code developed by Mazzoni (2008, p. 120) for a specific hand shape characterized by a semi-bent position for all five fingers. Additionally, tags were created to indicate the start and end of a Transfer of Person (referred to as Role Shift). These were labeled as ‘RS=’ followed by ‘<’ or ‘>’ to mark the beginning and end, respectively.

Now, an additional tier called the *Traduzione* (Translation) tier will be discussed. The tier includes a translation of the LIS utterance into written Italian and was developed to provide a linear translation of the LIS utterance, making it easier to read and understand for non-signers. In summary, the information included in the *Traduzione* tier, and the three different annotation strategies employed in the LIS treebank project for the *Segno,* tier provide insights into what characteristics could be maintained, and potential improvements could be made. Using labels from Radutzky’s *Dizionario* (1992) offered unambiguous information about the original shape of the Signs, mitigating the issue of being unable to retrieve a Sign when its video is not available (Antirono Pizzuto *et al.,* 2008). However, it cannot be taken for granted that the *Dizionario* can be accessed easily by anyone, as it is available behind purchase. In contrast, SpreadTheSign is an updated and easily accessible collection of Signs, making it a valuable resource for translations into written Italian and other sign languages. Considering all these elements, the key takeaway from this labelling methodology is the intuition of the importance of using unambiguous labels that allow for the direct retrieval of Sign forms from the label itself.

Regarding the use of SW, this intuition was fundamental, as it employed a language-specific system designed to represent the manual and body components of signed discourse, including important information like eye gaze. However, the drawback of SW— as presented in Table 5 in comparison to other language-specific systems— is that it cannot be directly coded on ELAN, the annotation software used, thus calling for an intermediate labeling step. Additionally, regarding the use of Italian translations of LIS utterances, this characteristic reflects the approach of working with an interdisciplinary mindset that considers that all the individuals who may want to access an annotation may not be familiar with LIS, thus aiming to render the annotation accessible to a broader audience.

In conclusion, this methodology demonstrated promising initial insights into annotating LIS. However, there is certainly room for adaptation and improvement through the implementation of innovative language-specific notation strategies. Always giving priority to the purpose behind the development of an annotation methodology, these strategies should aim to eliminate the need for intermediate labeling processes, thus maximizing the annotating capabilities of ELAN.

6.2. Guidelines for Effective Collection and Annotation of LIS within a COS-S framework

In the previous chapters, the complexities of collecting and annotating sign language data were discussed, and separate guidelines related to collecting and annotating LIS data were drawn, based on multimodal data collection experiences both in the Italian and international context. Drawing from these discussions, a set of comprehensive guidelines that guided the approach to identifying data collection tools and annotation methodologies is presented in this section.

When it comes to collecting data, Geraci *et al.* (2011) and Lo Re (2022) consider the importance of seeking tools that align with the principles of portability and non-invasiveness (see section 5.3). This choice is motivated by the desire to capture data in environments that minimize interference with the participants' natural settings. Additionally, high-resolution recordings are prioritized to maintain data quality while still respecting the principles of portability and non-invasiveness. It is important, in fact, that the selected tools can capture multimodal data suitable for subsequent post-processing and analysis. Moving on to the annotation of LIS data, the choice of the program to use easily falls onto ELAN. As discussed, this decision is influenced by the extensive use of ELAN in sign language research (Crasborn *et al.,* 2007; Crasborn & Zwitserlood, 2008; Santoro & Poletti, 2011; Mesch & Wallin, 2015; Cormier *et al.,* 2016; Schembri *et al.,* 2017; Hanke *et al.,* 2020) as well as its user-friendly interface, making it a practical and effective tool (see section 4.3. and 4.5.). Clear and uniform data labeling is considered as another fundamental aspect of the annotation guidelines as all collected data must be labeled with clear and logically organized labels to allow for efficient and long-lasting data retrieval (Antinoro Pizzuto *et al.,* 2010; Crasborn *et al.,* 2015). Of course, consistency in labeling practices is a key factor in this regard. Multilingual annotation is another guideline that should followed. By including multilingual vocal language labels, the annotated MultiMedaLIS_3 Dataset will be accessible to both signers and non-signers (Crasborn & Zwitserlood, 2008; Mesch & Wallin, 2015; Crasborn *et al.,* 2015). This approach aims at promoting interdisciplinary research, as well as amplifying the potential applications of the Dataset itself.

Furthermore, in the annotation process, a combination of vocal language labels and a language-specific annotation system should be adopted (Antinoro Pizzuto *et al.,* 2010; Bianchini, 2012). This dual approach provides comprehensive information while maintaining clarity and precision. Additionally, multiple language-specific tiers for the annotation should be developed. These tiers are not limited to each hand (as suggested by Crasborn & Zwitserlood, 2008) but can follow language-specific transcription systems such as Typannot (Doan *et al.,* 2019; Danet *et al.,* 2021).

Lastly, regarding data storage, precautions to ensure accessibility for future research and collaborations should be taken. The production of different copies of the data collection can be a strategic measure to store the data and make it readily available for future endeavors, minimizing the risk of losing it.

In summary, these guidelines serve as the backbone of the approach to collecting and annotating the MultiMedaLIS Datasets, as presented in this work. These guidelines reflect not only practical considerations based on past experiences, but also a commitment to producing a dataset that is accessible, comprehensive, and primed for diverse research applications. In the following paragraphs, the guidelines are directly applied to the different steps of development of the Dataset.

6.3. The MultiMedaLIS Datasets

In the upcoming sections, the development of the MultiMedaLIS Datasets will be presented. In particular the three preliminary versions of the Datasets are the outputs of collection processing efforts at the University of Catania, carried out thanks to ongoing collaborations that began in April 2021. These collaborations are the result of a cooperative partnership between two departments at the University of Catania: the Department of Humanities (DISUM— *Dipartimento di Scienze Umanistiche*) and the Department of Electrical, Electronic and Computer Engineering (DIEEI— *Dipartimento di Ingegneria Elettrica, Elettronica e Informatica*). Additionally, international ties with the Max Planck Institute for Intelligent Systems (MPI-IS) in Tübingen were established.

Although the MultiMedaLIS_1 and MultiMedaLIS_2 Datasets will be discussed comprehensively, throughout the following sections particular emphasis will be placed on the MultiMedaLIS_3 Dataset, as it is to be considered the central output of these data collection and annotation processes. The three preliminary datasets can be located within the medical semantic domain, hence the inclusion of 'medical' in their name.

6.3.1. Three data collections

As introduced in section 1.3.1., three distinct data collection efforts, referred to as the MultiMedaLIS Datasets, were carried out, leading to the development of the biggest RADAR/camera dataset for a sign language: the MultiMedaLIS_3 Dataset. In particular, the MultiMedaLIS_3 Dataset is comprised of a total of 25,830 Sign

instances, encompassing 205 repetitions of 100 Signs and the 26 Signs of the LIS alphabet.
The Signs included in the MultiMedaLIS Datasets can be broadly categorized into two groups: semantically marked Signs related to health and health issues, and non-semantically marked Signs. It is important to highlight that, while the first group of Signs is categorized as semantically marked, this classification does not imply that these Signs belong to a jargon lexicon. At the same time, while the semantically marked Signs are of primary importance due to their significance, a decision was made to include other Signs that facilitate constructing meaningful utterances in patient-doctor interactions. In the MultiMedaLIS Datasets' creation process, the selection of Signs aimed to allow them to be combined to form meaningful and coherent utterances. Consequently, with this goal in mind, finding balance between specific medical terminology and ordinary Signs was essential to the creation of a comprehensive resource for Deaf patients and healthcare professionals alike.

With relation to the specific form of the Signs, the MultiMedaLIS_3 Dataset includes a lexicon of standard, isolated Signs not combined within utterances, whose forms are present on online dictionaries and educational materials designed for beginner and intermediate LIS learners. In fact, to ensure data accuracy, Sign variants performed by a professional LIS interpreter during the collection of the MultiMedaLIS_1 Dataset were compared to the same Sign variants on the online dictionary SpreadTheSign and were then selected, with the aim of gathering documented versions of each Sign in the Dataset. This was done to improve the precision, reliability and real-world applicability of the third version of the Dataset.

The MultiMedaLIS_3 Dataset is the result of an extension and re-elaboration of two data collections conducted at the DISUM and the MPI-IS. The first version of the Dataset (MultiMedaLIS_1) was collected using a standard RGB camera, with a professional LIS interpreter simulating patient-doctor interactions based on a script.

Certainly, using a scripted approach for data collection does not facilitate the capture of spontaneous LIS interactions. However, this initial data collection phase played a crucial role by providing essential information on the Sign variants to use, which were then compared to SpreadTheSign and educational resources variants. Additionally, it served as an introduction to LIS for the non-signing members of the group. In order to make the videos accessible to these members, basic annotations were applied using an ELAN layout consisting of two tiers: one for spoken English labels, where segments corresponded to individual Signs, and another for the translation of LIS utterances into English, as presented in Figure 21.

Figure 21. Annotation sample for the first version of the Dataset (CC BY 4.0).

The MultiMedaLIS_1 Dataset was shared with the members of the Optics and Sensing Lab of the Max Planck Institute for Intelligent Systems (MPI-IS) in Tübingen for the second data collection, which marked a significant improvement in terms of precision, accuracy, and quality of LIS data. This improvement was made possible thanks to the integration of new portable tools which made the process more refined, compared to the first attempt. During this phase, a trial-and-error process was followed, in order to select the most adequate tools for the upcoming data collection(s). For instance, challenges with the employment of computer vision tools designed to detect key points on the signer's body were immediately encountered. Figure 22 illustrates an example where OpenPose[86] struggled to detect the key points of the fingers.

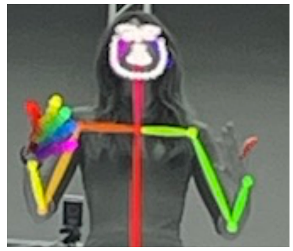


Figure 22. Failure of the use of the OpenPose system in capturing finger key points (CC BY 4.0).

The second data collection process (Figure 23) led to the creation of a second preliminary Dataset (MultiMedaLIS_2) containing 46 individual Signs, 10 utterances composed of Signs selected from the initial 46, and the 26 Signs from the LIS alphabet, repeated three times each. This resulted in a total of 216 isolated Signs and 30 utterances. This second version marked a considerable improvement compared to the previous one, as it served as a superset of Signs to be used in the third data collection.

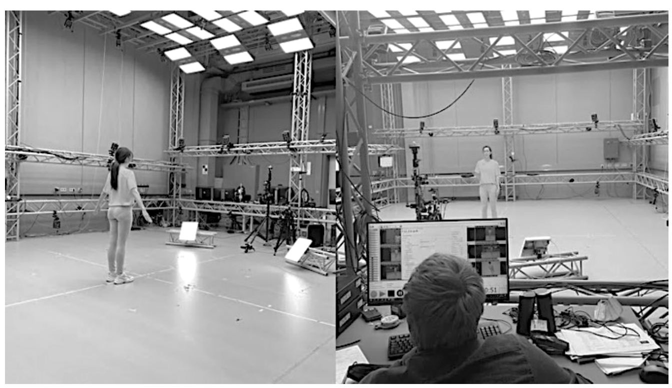


Figure 23. Photos taken during the collection of the second version of the Dataset at the MPI-IS (CC BY 4.0).

The tools employed to collect the second version of the Dataset were provided by the Perceiving Systems Department at the MPI-IS and included:

i. 8 RGB cameras focused on the signer' body from various angles (frontal view of the full body, top view, top-down view, frontal view of the right part of the body,

[86] OpenPose is a real-time system designed for the detection of 2D pose, encompassing body, foot, hand, and facial key points (Cao *et al.,* 2021).

frontal view of the left part of the body, frontal view from the bottom-up, back view of the left part of the body and back view of the right part of the body).

ii. 4 RGB cameras specifically focused on the hands (top view, frontal view of the right hand, frontal view of the left hand, frontal close-up of the hands).

iii. 60Ghz RADAR Infineon Technologies XENSIVTM (without synchronization).

iv. Sony depth sensor paired with an RGB camera, synchronized with the IOI cameras.

v. LED panels to improve the front view.

The standard capture speed for this Dataset was set at 30 frames per second (fps). In the second version of the Dataset, the principle of capturing high-quality data was respected, as it was recorded from various angles by using 12 RGB cameras. However, this setup involving multiple cameras, did not adhere to the principles of portability and non-invasiveness. For this reason, the experience from the second data collection was taken into consideration and applied it to the third data collection, by using a Realsense D455 depth camera (see section 6.3.4). This camera not only provides RGB data but also depth measurements, capturing detailed sign language gestures and movements while maintaining portability and non-invasiveness. Regarding the LED panels, they were initially used to ensure optimal lighting conditions because the department was located underground. However, in the third data collection, natural lighting was used.

6.3.2. Defining the tools for data collection

As outlined in section 6.3.1, The MultiMedaLIS_3 Dataset comprises a total of 25,830 sign instances. In this regard, it is essential to acknowledge that this outcome is the result of an extensive process involving discussions and selections conducted around the recording tools.

In identifying the most adequate tools for collecting LIS data, the aim was that of developing a methodology that would strictly adhere to the principles of portability, non-invasiveness, multimodality, replicability and high quality of captured data. To adhere to these principles, non-invasive and portable instruments were identified, thanks to their ability to capture manual and body components in high resolution and in different contexts. This approach was designed to facilitate the processing of signing from a holistic perspective. In fact, it was considered paramount to comprehensively capture the manual and body components that comprise signed discourse, in alignment with the framework proposed by Volterra *et al.* (2022).

Portability offers distinct advantages. First, it allows signers to express themselves in a more natural way, which is crucial for capturing the full range of semantic and pragmatic aspects of signed discourse. Second, it promotes user-friendliness, making it possible to collect data in different settings and with different signers, including contexts in which access to specialized recording equipment may not be available.

To ensure high-quality data acquisition, guidance from experts in computer vision was sought. Meetings were organized among linguists and engineers from the University of Catania and the Optics & Sensing Lab of the MPI-IS. During these meetings, which first took place in late 2021, a consideration on the best tools to use for data collection began[87]. Following a trial-and-error path, the recording technologies available at the two institutions were looked into. At the Perceiving Systems division of MPI-IS, a great use of Vicon Motion Capture System (54 Vicon V-16 Vantage) was made[88]. This motion capture system uses reflective markers placed on an individual's body to track movement. However, through experimentation during a research visit in early 2022, it was decided that this tool did not align with the principles of portability and non-

[87] I extend my gratitude to Dr. Raffaele Mineo, Dr. Simone Palazzo, Prof. Concetto Spampinato and most importantly, Prof. Egidio Ragonese of DIEEI-UNICT for their efforts in aiding the selection of recording tools and facilitating the collection of the MultiMedaLIS_2 and 3 Datasets. In fact, the development of the second and third multimodal collections owes much to their dedication, especially Prof. Ragonese's suggestion to employ RADAR in capturing sign language data. In particular, these sections would have not come to life without the valuable contribution, patience and advice of Dr. Raffaele Mineo throughout these past months.

[88] https://ps.is.mpg.de/pages/motion-capture (Accessed on October 9th, 2023).

invasiveness. In fact, its reliance on markers placed on specific facial and bodily points could be considered invasive if compared to the established parameters.

The importance of video quality and the ability to capture multimodality has been previously emphasized. To address these requirements, it was ensured that third collection of the Dataset featured high video resolution, allowing for a clear depiction of body components, namely facial expressions, eye gaze, and the production of oral components (Volterra *et al.,* 2022). To guarantee the comprehensive collection of various LIS features, for the third collection of the Dataset, it was decided to integrate a combination of multimodal recording tools, including the 60GHz RADAR (already employed in the third data collection), a Realsense D455 depth camera, a Kinect v1, a Zed v1 and Zed v2[89].

The table below illustrates the types of recording tools employed and the type of captured data.

Recording Tool	Captured Data
60Ghz RADAR Infineon Technologies XENSIV	Employed to capture various types of manual data at a rate of 13 fps. This included: a) Time-Domain data. b) Frequency-Domain data. c) Range Doppler Map data. d) Range Doppler Map data with Moving Target Indication (MTI) collected at the same frame rate.
Realsense D455 Depth Camera	Employed to capture a various type of body data, including: a) Two infrared (IR) cameras in stereoscopic mode at a resolution of 848x480 pixels and a frame rate of 30 fps. b) An RGB camera at 1280x720 pixels and 30 fps.

[89] These tools were ultimately selected by the engineers of the group, following the inputs on data collection requirements from linguists, based on their ability to meet our criteria and provide high-quality data that could be effectively analyzed. To ensure the accuracy of the insights regarding the recording tools used in the third collection of the Dataset, Dr. Raffaele Mineo has provided and reviewed the information concerning the tools employed in this context.

	c) A 16-bit depth map at 848x480 pixels and 30 fps. d) A filtered depth map at 424x240 pixels and 30 fps. e) A filtered depth map aligned with the RGB image at the same resolution and frame rate. f) A 3D point cloud at 848x480 pixels and 30 fps. g) 68 face tracking points at 30 fps.
Kinect v1[90]	Employed to capture various data types specifically related to facial expression: a) An RGB image at a resolution of 640x480 pixels and a frame rate of 30 fps. b) A 16-bit depth map at the same resolution and frame rate. c) 87 face tracking points at 30 fps.
Zed v1 Camera	Employed to capture the following data types: a) Two RGB images in stereoscopic mode at a resolution of 1920x1080 pixels and a frame rate of 25 fps. b) Two 32-bit depth maps at the same resolution and frame rate in stereoscopic mode. c) Two 3D point clouds at the same resolution and frame rate in stereoscopic mode.
Zed v2 Camera	Employed to capture the following data types: a) One RGB image at a resolution of 1920x1080 pixels and a frame rate of 25 fps. b) 18 2D and 3D body tracking points at 25 fps.

Table 6. List of recording tools and types of captured data.

The 60Ghz RADAR Infineon Technologies XENSIV, in particular, was chosen due to its capability to capture the motion and position of the signer's hands and fingers, as well as details of the surrounding environment. RADAR technology appears as a promising solution for the collection of sign language data, mainly due to three reasons: user privacy protection, the ability to detect subtle movements and the absence of physical contact. To protect the privacy of the person being recorded, as reported by

[90] To ensure the capture of facial expressions, the RADAR signal was synchronized with the Kinect v1 camera.

Gurbuz *et al*. (2020), RADAR technology can be employed to capture sign language without capturing visual images of the user, thus ensuring their privacy. Furthermore, RADAR sensors are not affected by ambient light, they have an extended detection range and, therefore, do not require a direct line of sight, and can detect even subtle movements with high sensitivity (Li *et al.*, 2021).

The output of this third multimodal data collection efforts, hence the MultiMedaLIS_3 Dataset, which combines data from various sources including RADAR technology, depth cameras, and facial tracking (see Figure 24), aligns with points i, ii and iii of section 5.3: portability, replicability and high-quality of data capture. In fact, the selected tools can collect high precision data, ensuring the capture of manual and body components.
Moreover, their ability to collect spatial information allows for a detailed examination the spatial aspects of LIS, including hand and body movements as well as their positions in 3D space. A particular focus is put on capturing facial expression, eye gaze and oral components thanks to the use of the facial tracking data from the Kinect camera.

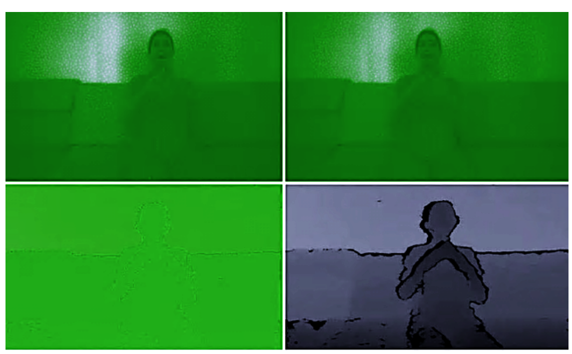


Figure 24. Infrared and depth data from the third version of the MultiMedaLIS Dataset (CC BY 4.0).

Although each tool has a distinct purpose, the synchronization of RADAR data with other camera sources, such as the Kinect, facilitates the visualization and analysis of all facial components together. This synchronization aligns with the holistic approach to studying sign language components, as proposed by Volterra *et al.* (2022), and enhances the MultiMedaLIS_3 Dataset's utility in exploring the interactions between the formational parameters of LIS. Moreover, this third collection offers data at various resolutions to accommodate different research needs. Annotation efforts have and will continue to primarily rely on data captured by the RGB camera of the Realsense D455 depth camera. However, annotations can be potentially extended to all other recorded data, provided it can be uploaded to ELAN which allows for the simultaneous viewing of multiple synchronized video data through split-view. In terms of practical interdisciplinary applications, the various data provided by the multimodal tools are currently undergoing post-processing for training an algorithm designed to automatically recognize and translate isolated LIS Signs.

More specifically, the data from the RADAR will be used by the AI algorithm developed by the DIEEI-UNICT Engineers, to achieve the set goal of automatic translation from LIS to Italian. The RADAR data serves as an integral part of the project's multimodal approach for LIS-Italian automatic translation as it provides information about the position and movements of the signer's body, processed by the developed algorithm to recognize and interpret the Signs. This processing involves converting coordinate points into meaningful information about body positions and movements, essential for understanding and facilitating Sign translation. In this context, the expertise of engineers is vital in transforming raw RADAR data into useful information for the algorithm. Moreover, the choice of using RADAR also aims to preserve patients' privacy in real-life medical contexts. In fact, in contrast to cameras, the RADAR does not capture visual images of signers but rather anonymous data regarding body movements. This approach aims to ensure confidentiality and respect for signers' rights, particularly in sensitive hospital environments (see section 6.3.4).

In terms of operation, the RADAR antennas are arranged in an 'L' shape to obtain a three-dimensional view of the surrounding environment and signers within it. During data collection, RADAR antennas emit electromagnetic pulses in the desired direction. When these pulses encounter objects in their path, some of the energy is reflected back to the antennas. These reflected signals constitute the 'raw' data acquired by the RADAR and contain crucial information about the detected objects, such as distance, speed, and dimensions. The acquired raw data is then processed to generate a map called the 'Range Doppler Map' (RDM), providing information on the position of objects along the distance axis and their speed along the Doppler axis. Through this spatial and temporal representation, different objects in the RADAR's environment can be recognized and distinguished, including the movements of patients, the focus of this work's interest.

The processing technique can be implemented with or without Moving Target Indication (MTI). With MTI, the RADAR can distinguish movements of moving objects from the static background, enhancing the detection and tracking capability of subjects of interest, such as patients. Without MTI, a more general view of the surrounding environment is obtained without specifically emphasizing movements.
Regarding the management and processing of data from cameras, especially from Pose Estimators like AlphaPose[91] and OpenPose[92], the extensive numerical data representing the coordinates of captured points for skeleton construction may be complex and seemingly incomprehensible for non-specialists. In fact, the RADAR does not provide data directly interpretable by individuals. However, the RADAR data is used as input for artificial intelligence algorithms developed by the computer engineers involved in the project. These algorithms leverage RADAR data to identify and understand body movements and positions without compromising patient privacy. Engineers will then process said data to make it compatible with linguistic analyses

[91] https://github.com/MVIG-SJTU/AlphaPose (Accessed on February 14th, 2024).
[92] https://github.com/CMU-Perceptual-Computing-Lab/openpose (Accessed on February 14th, 2024).

and facilitate automatic LIS-Italian translation, without the need for cameras that could increase computational load and compromise privacy.

6.3.2.1. Discussing the Signs included in the third collection of the Dataset

In identifying the Signs to include in the MultiMedaLIS_3 Dataset, the goal was to establish continuity with existing resources in the Italian context. Therefore, it was verified that most of the translations of the captured Signs into written Italian were aligned with the Il Nuovo *Vocabolario di Base della Lingua Italiana* (The New Basic Vocabulary of Italian) (De Mauro, 2016). Out of the 100 Signs (refer to Table 7), 93% of the Italian translations are found in the *Vocabolario*. The remaining seven Signs, which are related to healthcare, health, and Covid-19, are not included in the *Vocabolario* and are highlighted in light grey below.

Sign	Italian Label	English Label
1.	Aiutare	To help
2.	Altro	Other
3.	Ambulanza	Ambulance
4.	Ansia	Anxiety
5.	Battito cardiaco	Heartbeat
6.	Bocca	Mouth
7.	Braccio	Arm
8.	Brivido	Shivers
9.	Calore	Heat
10.	Casa	House
11.	Chiamare	To call
12.	Chiedere	To ask
13.	Collo	Neck
14.	Come	How
15.	Comprare	To buy
16.	Comunicare	To communicate
17.	Contagioso	Infective
18.	Coronavirus	Coronavirus
19.	Cuore	Heart
20.	Curare	To cure
21.	Difficoltà	Difficulty
22.	Disinfettare	Disinfect
23.	Dolore	Pain
24.	Domani	Tomorrow
25.	Dottore	Doctor
26.	Dove	Where
27.	Dovere	To have to
28.	Elicottero	Helicopter
29.	Esserci	To be
30.	Farmacia	Pharmacy
31.	Fegato	Liver
32.	Figlio/a	Son/Daughter
33.	Fronte	Forehead
34.	Fumare	To smoke
35.	Gambe	Legs
36.	Giorno	Day
37.	Gola	Throat
38.	Grave	Bad
39.	Ieri	Yesterday
40.	Inguine	Groin
41.	Iniezione	Injection
42.	Io	Me
43.	Leggero	Light
44.	Letto	Bed
45.	Malattia	Illness
46.	Mamma	Mother
47.	Mangiare	To eat

48.	Mano	Hand
49.	Mascherina	Mask
50.	Mattina	Morning
51.	Medicine	Medicine
52.	Mezzogiorno	Noon
53.	Molto	A lot
54.	Muscolo	Muscle
55.	Naso	Nose
56.	Nausea	Nausea
57.	No	No
58.	Nome	Name
59.	Occhio	Eye
60.	Orecchio	Ear
61.	Ospedale	Hospital
62.	Pallore	Pale
63.	Papà	Father
64.	Perché	Why
65.	Persona	Person
66.	Petto	Chest
67.	Piedi	Feet
68.	Pillola	Pill
69.	Positivo	Positive
70.	Potere	To be able to
71.	Prima	Before
72.	Problema	Problem
73.	Pronto soccorso	Emergency room
74.	Quando	When
75.	Raffreddore	To have a cold
76.	Respirare	To breathe
77.	Restare	To stay
78.	Rumore	Noise
79.	Schiena	Back
80.	Sciroppo	Syrup
81.	Se	If
82.	Sera	Night
83.	Sì	Yes
84.	Sintomi	Symptoms
85.	Smettere	To stop
86.	Sordo	Deaf
87.	Spalle	Shoulders
88.	Stanchezza	Tiredness/ Fatigue
89.	Stomaco	Stomach
90.	Subito	Immediately
91.	Sudore	Sweat
92.	Tempo	Time
93.	Termometro	Thermometer
94.	Testa	Head
95.	Tosse	Cough
96.	Tu	You
97.	Vaccino	Vaccine
98.	Vertigini	Vertigo
99.	Vista	Eyesight
100.	Vomito	Vomit

Table 7. List of the 100 Isolated Signs included in the third collection of the MultiMedaLIS Dataset.

Since, at the present time, no digital collections of LIS with a specific focus on the medical domain are available, the semantically-marked terms included in the MultiMedaLIS_3 Dataset were cross-referenced with a review of Italian and international studies on Covid-19-related symptoms and associated terminology published after 2020. This approach ensured that the Dataset included relevant terminology. To do so, systematic reviews of common Covid-19 symptoms in both anglophone (Alumohamadi *et al.,* 2020) and Italian contexts (Riccardo *et al.,* 2020) were examied, as well as diagnosis and treatment solutions (Cascella *et al.,* 2023[93]). Additionally, an analysis of the *Covid-19 Open Research Dataset was* conducted through the corpus management function of the SketchEngine tool[94], which is based

[93] The cited book was initially accessed in its 2022 release but has been updated in August 2023.
[94] https://www.sketchengine.eu/ (Accessed on October 10th, 2023).

on English. Unfortunately, there are no comparable resources available for Italian or LIS, preventing the conduction of an analysis for these two languages. Nevertheless, this comprehensive comparison helped us ensure that the MultiMedaLIS_3 Dataset included pertinent terminology.

For reference, an analysis of the numbers of occurrences of each word was performed, as shown in Table 8, where words are listed alphabetically.

Sign	**Italian Label**	**English Label**	**Occurrences in the *Covid-19 Open Research Dataset***
1.	Ansia	Anxiety	660,267
2.	Battito cardiaco	Heartbeat	4,387
3.	Brivido	Shiver/s	18,590
4.	Contagioso	Contagious	53,339
5.	Coronavirus	Coronavirus/ Covid-19	7,506,986
6.	Curare	Treatment	2,530,197
7.	Difficoltà	Difficulty	130,475
8.	Disinfettare	Disinfect	7,392
9.	Dolore	Pain	410,064
10.	Dottore	Doctor	197,565
11.	Farmacia	Pharmacy	71,639
12.	Iniezione	Injection	181,244
13.	Malattia	Illness/ Sickness	514,911,124
14.	Mascherina	Mask	275,776
15.	Medicine	Medicine	1,450,451
16.	Nausea	Nausea	49,761
17.	Ospedale	Hospital	1,648,344
18.	Pallore	Pale	8,094
19.	Pillola	Pill	6,829
20.	Positivo	Positive	1,878,263
21.	Pronto soccorso	Emergency Room/ E.R.	165,228
22.	Raffreddore	(to have a) cold	91,219
23.	Sciroppo	Syrup	2,577
24.	Sintomi	Symptoms	1,821,966
25.	Stanchezza	Tiredness/ Fatigue	181,099
26.	Termometro	Thermometer	808,814
27.	Tosse	Cough	221,204
28.	Vaccino	Vaccine	2,190,913
29.	Vertigini	Vertigo	4,891
30.	Vomito	Vomit	64,751

Table 8. Occurrences of English words used to label the Signs of the MultiMedaLIS Dataset in the Covid-19 Open Research Dataset.

To visualize the most frequently occurring terms in the *Covid-19 Open Research Dataset* included in the third version of the MultiMedaLIS Dataset, presented in this work, see to the word cloud displayed in Figure 25.

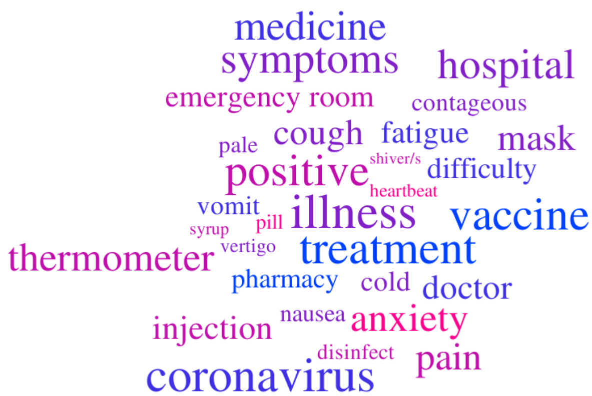


Figure 25. Frequency word cloud of the Covid-19 Open Research Dataset included in the third collection of the LIS Dataset presented in this thesis (CC BY 4.0).

Lastly, it is important to mention that the translations of the Signs in English may have one or more synonyms in Italian. Consequently, in the frequency distribution values presented in this study, unique values resulting from the combination of frequency distribution values for two translations of a Sign may be encountered. For example, the term *stanchezza* was searched within the Covid-19 Dataset both as *tiredness* and *fatigue*. This choice is motivated by the fact that *stanchezza* is itself a translation of the equivalent LIS Sign in the Dataset. This shows that, when selecting synonyms, priority was given to the original meaning of the LIS Sign itself. To further clarify the differences between Italian and LIS, consider the LIS Sign for Coronavirus/ Covid-19. There is no distinction between the two in LIS, so the two translations in Italian (and their equivalent translations in English) can be used interchangeably when translating LIS.

6.3.2.2. Recording protocol for the third data collection

In preparation for the third data collection, a recording protocol designed to ensure the collection of high-quality data was developed. As mentioned, the MultiMedaLIS_3

Dataset comprises 126 Signs, all performed by the user during each recording session. The 100 selected Signs were organized alphabetically, followed by the 26 Signs of the LIS alphabet. To ensure the robustness and reliability of the data, a total of 205 recording sessions were conducted. Choosing an uneven number of sessions, specifically 205, provided 5 extra sets to use in case of any failed or inadequate recordings, while still maintaining a total of 200 successfully recorded sets.

Dr. Raffaele Mineo, a Ph.D. student at the Campus Bio Medico University of Rome and the University of Catania, designed a customized recording software. In addition to the data capturing tools, the DIEEI-UNICT supplied a desktop computer and a keyboard, which Dr. Mineo modified into a pedal board. The pedal board was equipped with two pedals, both connected to the keyboard, allowing the user to move forward (right pedal) and backward (left pedal) within the data collection software without needing to move the upper part of their body and, therefore, not altering the neutral recording position[95]. During the recording sessions, the user was presented with one of the 126 Italian labels or letters of the Italian alphabet on a screen. The duration of display was determined by the user in terms of seconds or milliseconds, giving them time to prepare for signing and transition between adjacent Signs. Each Sign recording started from the user in neutral position, and the user returned to this pose after completing the execution of each Sign. The right pedal was employed by the user to indicate the completion of a Sign. However, if the user made a mistake (for instance, they produced an unclear Sign or accidentally pressed the right pedal while still signing), they had the option to press the left pedal to repeat the recording for that specific Sign, allowing them to re-record it as many times as needed. As can be seen in Figure 26, the recording software features a user-friendly and intuitive interface. Additionally, the software uses variation in background color to separate the

[95] The neutral recording position for the third collection of the Dataset consists in a seated position with arms resting on thighs, left palm on left thigh, right palm on right thigh (see section 4.2.4.1.).

preparation phase (yellow background) and the recording phase (green background). This feature ensures that the user is always aware of the stage of the recording process and, in doing so, it aims at reducing confusion and errors.

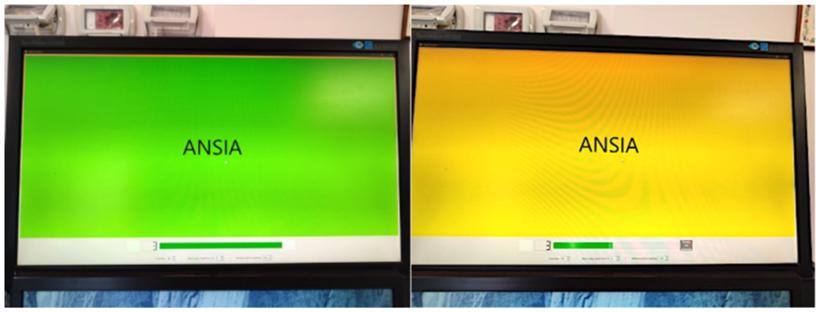


Figure 26. Display of the user interface of the software developed by Dr. Raffaele Mineo for the third data collection (CC BY 4.0).

The software was designed to allow for the expansion of the set of collected Signs (with the possible integration of utterances), as it can accept word lists from text files, offering flexibility and customization options for additional data collections.

6.3.3. Testing the acceptability of the third Dataset collection with a group of Deaf signers

The acceptability of isolated Signs included in the third version of the MultiMedaLIS Dataset was tested via a questionnaire completed by a group of 10 Deaf LIS signers. This questionnaire was entirely anonymous and comprised two sections: one gathering demographic information (such as deafness status, age, and birth location) and another dedicated to assessing the acceptability of Signs within the MultiMedaLIS_3 Dataset. Participants were presented with a GIF depicting each isolated Sign and were asked to evaluate whether, in their opinion, the execution of the Sign and the corresponding Italian label provided were accurate or not. Their responses were collected via a 'yes' or 'no' multiple-choice format.

Regarding geographical distribution, participants that completed the questionnaire came from different regions of Italy. Specifically, one participant each from Lombardy, Veneto, Tuscany, Marche, Lazio, Campania, Sicilia, and two participants from Puglia. While most participants were born in Italy, it is important to note that place of birth was not a selective criterion, as one participant from Tunisia also contributed. In relation to age demographics, groups were divided as follows: 40% of participants fell within the 35 to 44 age range, 30% in the 25 to 34 range, 20% in the 45 to 54 range, and 10% in the 18 to 24 range.

It appears that, among the 100 Signs assessed by 10 participants 47 out of 100 signs were considered acceptable by all participants. Additionally, 17 Signs garnered a 90% approval rate, while 10 Signs received an 80% approval, collectively representing 74% of the Signs assessed. None of the Signs faced unanimous rejection by all participants. However, two Signs denoting body parts, particularly ‘arm’ and ‘leg’, proved to be the most controversial, receiving the lowest approval rate of 30%. Furthermore, five Signs caused a 50% split in opinions, receiving both approvals and disapprovals. Specifically, the divisive Signs belonged to the semantically marked group (refer to Table 7), including sanitary locations like ‘hospital’ and ‘emergency room’ as well as terms Signs broadly associated with the medical domain, such as ‘Coronavirus/ Covid-19’, ‘medicine’ and ‘thermometer’.

Several factors might contribute to the different perceptions of certain Signs. For example, variations in regional Sign usage among participants could explain the divergence in opinions regarding Signs related to the sanitary context. Alternatively, limitations inherent in the data collection process might have influenced how Signs were evaluated. For instance, presenting Signs in isolation rather than within a complete utterance and constraints on movements during data collection due to testing tools could have impacted the signers’ assessments. However, these are all speculative explanations. A comprehensive and expanded statistical analysis involving a larger participant pool could provide deeper insights into the underlying factors influencing these disparities in Sign acceptability.

6.3.4. Limiting the semantic domain of the Dataset

The decision to semantically limit the MultiMedaLIS Datasets to the medical domain was influenced by several factors. First, practical constraints such as limited resources, including time and the number of participants available for data collection led to focus on a specific semantic area that had real-world relevance and applicability. Additionally, the development of the Dataset coincided with an immediately post-pandemic period, as the initial steps were taken in the latter half of 2021. During this time, the challenges faced by D/deaf individuals in medical contexts, particularly related to communication barriers caused by the use of medical masks, became highly evident and pressing issues. This situation prompted a response to these challenges and initiated a discussion on potential solutions to adopt within the Dataset.

The use of masks, especially in healthcare settings such as hospitals, emergency rooms and medical offices, created significant challenges for D/deaf individuals. The masks obscured the lower part of the speaker's face, making it difficult for D/deaf people to access direct communication without the presence of an interpreter. Research on the effects of face coverings on speech transmission efficiency dates back to the 1960s, as discussed by Giovanelli *et al.* (2021), with findings suggesting that even within environments in which mask use was required, people would remove respirators to talk to their peers (Fawcett, 1961).

More recently, studies on the impact of face masks on the transmission of speech signal have been carried out. For instance, a review by Badh & Knowles (2023) was published on this matter. Within the Italian context Giovanelli *et al.* (2021) conducted a study to investigate if, at the same level of acoustic information, the absence of visual information of the speaker's face was an issue in speech comprehension. The study found that hiding lip information indeed limits understanding what is being said while, at the same time, increasing listening effort and decreasing confidence in what has been heard. A similar study conducted by members of the same research group in collaboration with both Deaf and hearing researchers of the CNR, explored the impact of masks on the comprehension of signed information, specifically in LIS. The study, by Giovanelli *et al.* (2023), observed that due to its multilinearity, some information in LIS discourse conveyed through facial expressions and body language may be redundant, at times. Despite the increased effort required when watching a signer with a mask compared to an unmasked signer, it was noted that signers can adapt by extracting information from other visual cues, such as the eyes and eyebrows, even when the lower part of the face is obscured. Giovanelli *et al.* (2023) therefore suggest that signers may compensate for the perceived loss of information by investing more effort and relying on contextual cues.

These two studies represent two instances of the raise of interest towards the possible barriers posed by face coverings respectively in spoken and signed communication. However, if the auditory feedback that is recevied by hearing people is removed, it can be appreciated how medical masks can hinder interpersonal communication with non-signing people for those who rely on lip reading and visual cues (Poon & Jenstad, 2022). On this topic, Gutierrez-Sigut *et al.* (2022) write that for D/deaf and hard-of-hearing people interviewed in the context of their study, communication with people who wore masks was in all cases difficult, leading to a loss of information in day-to-day interactions and a consequent heightened feeling of disconnect from society.

This communication barrier emerged as one of the most significant challenges during the Covid-19 pandemic, particularly in healthcare settings. It highlighted the urgent need to address communication access for D/deaf individuals in healthcare contexts. Initiatives like the *Come ti Senti*[96] campaign were launched to raise awareness and engage with the Deaf community to discuss and address this critical issue.

[96] https://cometisenti.ens.it/ (Accessed on August 24th, 2023).

Beyond these issues, the Covid-19 pandemic had a universal impact on language, introducing new concepts and technical terminology into everyday communication across all languages and cultures (Tomasuolo *et al.,* 2021). As a result, there was a surge in the use of medical and healthcare-related terms during the pandemic. Some terms, like those related to sanitary equipment such as surgical masks, saw increased usage. Moreover, new terms emerged to describe the illness that was affecting the world: Coronavirus. The creation of a Sign for Coronavirus in LIS was a collaborative process facilitated by the active online presence of the Italian Deaf community. This process culminated in the development of the most used Sign for Coronavirus nowadays, represented as 'Variant F' in Gianfreda *et al.* (2021), where one hand is closed in a fist (depending on the signer's dominant hand) and the other hand rotates externally with extended fingers.

These linguistic developments lead back to the inclusion of Lexical Units within the MultiMedaLIS Datasets and the LIS resources employed to retrieve it. In fact, whereas online dictionaries like SpreadTheSign can be updated and can, therefore, reflect the evolution of the language and introduction of new terminology (such as Coronavirus), the situation is different for printed dictionaries published before 2020/2021, as they are obviously not able to account for Signs whose use intensified or that altogether developed after their publication. In fact, except for Orazio Romeo's 2021 self-published printed dictionary *Conoscere i Segni al Tempo del Coronavirus. 600 Segni Selezionati* (Knowing Signs during the Coronavirus Time. 600 selected Signs) all other printed dictionaries of LIS were published well before the pandemic (Radutzky, 1992; Romeo, 1991; 1997; 2004).

As mentioned in the introduction to this section, the limitedness of available resources led to the decision to semantically limit the dataset. This decision was influenced by the broader discourse surrounding the impact of the pandemic on daily lives, which resulted in an ingrease in the use of Covid-related terminology. Additionally, the challenges faced by D/deaf individuals in medical contexts during the pandemic presented an opportunity to align the Dataset with real-world applications.

6.3.5. Contextualizing the MultiMedaLIS Datasets on a sociological level: discussing the issues encountered by Deaf people and their interlocutors in medical contexts

To provide a sociological context for the MultiMedaLIS Datasets, a series of interviews and discussions were conducted with individuals representing the three primary groups involved in communication within medical contexts, following the discussion established in section 5.4. Collaborating with Dr. Erika Raniolo, a triad of participants was identified and parallel interviews were conducted with each of these representative figures. The triad included: a Deaf person, who has experienced medical contexts with and without interpreters, as well as with and without the use of masks, an interpreter with experience working in medical settings and a hearing, non-signing medical doctor with experience in treating D/deaf patients. Figure 27 visually represents this triad, showing the connection of these three figures in the communication flow, as denoted by arrows. The means of communication used within this dynamic are also specified.

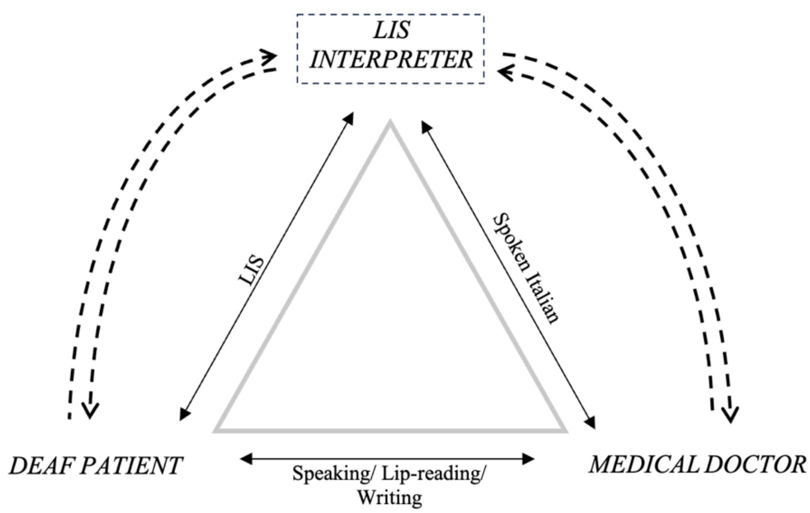


Figure 27. Representation of the communication flow between the triad involved in interactions in the medical context (CC BY 4.0).

In the following sections, insights gathered from the interviews conducted with the three participants in the triad as discussed in Raniolo & Caligiore (2023) are presented, along with discussions that occurred during the conference where the poster was presented: the *5° Convegno Nazionale LIS*. These interviews were conducted with the aim of examining the communication strategies employed, the predominant mode of communication (auditory-oral or visual-gestural), and the role of interpreters in the Italian context, a state in which LIS has been formally recognized, but where the permanent presence of an LIS interpreter within public facilities is not required (with some exceptions).

6.3.5.1. The point of view of the Deaf patient

An interview with a Deaf individual who grew up in a Deaf family and has a Deaf partner and children was conducted. They[97] had limited personal experience with interpreters but mentioned situations where close relatives, like parents of children, had to visit a medical doctor. In these situations, an interpreter was only called when it was considered necessary.

In their personal experience, they usually did not consider an interpreter essential when visiting a medical doctor. Instead, they relied on a combination of lip-reading, gestures, and writing to communicate. In their experience, they found that most medical doctors, but particularly family doctors, understood their situation as Deaf individuals and made efforts towards efficient communication. This included actions like removing masks or placing themselves well-lit environments. However, they noted that some doctors, especially specialized ones, did not always adjust their communication methods, speaking too quickly or not considering the Deaf individual's needs in that context. It is their impression that longer-term relationship with a family doctor allow for a negotiation of effective communication practices that may not be as achievable in brief encounters with specialists.

[97] To ensure anonymity, gender-neutral pronouns will be used when referring to the interviewees.

With relation to resorting to professional LIS interpreters, they mentioned using interpreters in hospitals (although not for personal visits) thanks to the ULSS service[98], a remote interpreting service that was initiated about 4 years ago in the Veneto region. However, they highlighted challenges with remote interpreting, as hospital Wi-Fi may not always be able to grant a smooth experience. Additionally, they discussed the difficulties that older Deaf individuals may face with long-distance interpreting. In fact, older generations may not be familiar with using the technology required for long-distance interpreting, such as videoconferencing platforms. This lack of familiarity can lead to frustration in communication. Furthermore, they mentioned the challenge of compensating interpreters when hospital waiting times exceed expectations.
Lastly, they note that, in response to the pandemic, there has been a shift in favor of email communication for tasks such as sending prescriptions or booking appointments, both in hospitals and at family doctors' offices. This shift has improved the ease of scheduling visits and accessing medications, without the need for hearing relatives or friends to make phone calls on a D/deaf person's behalf.

Following a participation in a conference organized by the ENS, the *5° Convegno Nazionale LIS. La Lingua dei Segni Italiana: patrimonio culturale o capitale umano?*[99], these interviews were discussed with Deaf participants at the conference. What emerged from these discussions was a consistent pattern, where professional LIS interpreters are rarely called upon. Instead, individuals tend to rely on lip-reading or writing when communicating with medical doctors.

6.3.5.2. The point of view of the medical doctor

A family doctor— who had accumulated years of experience with D/deaf patients, having cared for approximately 15 D/deaf patients in their career— was interviewed.

[98] https://regioneveneto.veasyt.com/aulss4 (Accessed on October 9th, 2023).
[99] 5th National LIS Conference. Italian Sign Language (LIS): Cultural Heritage or Human Capital?

The doctor had experience with an equal number of Deaf sign language users and deaf individuals who rely on oral communication. As a non-signer, when asked about communication strategies employed with D/deaf patients, the doctor replied that primary methods involved lip-reading or writing, indicating that the preferred means of communication of the patient (LIS or spoken Italian) did not significantly influence the communication strategy adopted by the doctor. Following each consultation, they ensured that the information had been effectively conveyed by reaching out to a hearing family member to re-explain the diagnosis and treatment shared with the D/deaf patient. During the pandemic, they chose to speak by lowering their mask and speaking more slowly to aid in lip-reading, despite the potential risk of infection.

Much like the perspective shared by our Deaf interviewee, the doctor places great emphasis on the importance of developing meaningful relationships with their patients. In fact, in their opinion, building trust and negotiating the most efficient mode of communication with each individual D/deaf patient are collaborative processes that support an approach to efficient healthcare. During the interview, the doctor also mentioned an interesting development prompted by the pandemic.

They noted an increase in the use of technologies such as e-mails and WhatsApp messages as important tools for staying connected with their D/deaf patients, without having to refer to mediators. As for experience with interpreters, this doctor had the opportunity to work alongside one interpreter who happened to be a CODA (Child of Deaf Adults) and later decided to become a professional interpreter.

As the interview concluded, the doctor emphasized the significance of medical professionals having access to LIS courses, recognizing the importance of facilitating direct communication, in ensuring that D/deaf patients receive the best possible care and attention.

6.3.5.3. The point of view of the LIS interpreter

The third interview was conducted with a LIS interpreter who is also a CODA with 25 years of experience in the field[100]. When discussing the role of interpreters in the medical field in Italy, they expressed that the presence of interpreters in medical contexts is, to some extent, lacking. They attributed this issue to a dual scenario: Italian institutions' limited sensitivity to the Deaf community's needs and the Deaf community itself not advocating strongly enough for accessible communication in healthcare settings. They also mentioned that, in their experience, Deaf individuals often rely on non-professional interpreters (friends or family members), often referred to as *sign language brokers* who, however, can provide inaccurate translations (Napier, 2021). Choosing non-professional interpreters to interpret during medical appointments may trace back to economic reasons (as mentioned by the Deaf interviewee in section 6.3.5.1.), as well as the desire for decisional assistance, rather than neutral interpretation, which is what a professional LIS interpreter provides. However, they added that in more sensitive situations, such as conveying information about a tumor, Deaf individuals may hire an interpreter to ensure accurate and unmediated translation of information.

Regarding the impact of the pandemic on communication in medical settings, the interpreter emphasized that the use of masks intensified the challenges faced by D/deaf individuals in healthcare contexts. They highlighted that medical professionals often lack awareness and preparedness to provide effective information transmission with D/deaf patients.

[100] As can be seen, the presence of CODAs in the medical setting in ubiquitous. The unique perspective of the interpreter, who incidentally is also a CODA, provided a dual insight into the role of interpretation in medical contexts. In fact, CODAs frequently find themselves interpreting for their Deaf parents in various situations, taking on the role of language brokers (Singleton & Tittle, 2000; Moroe & de Andrade, 2018) either earlier or later in life. This experience allowed to gain valuable insight from someone who was at some point in their life a non-professional interpreter and then became one.

6.3.5.4. Discussing the interviews

By carrying out these interviews, the aim was to establish a socioecological discussion, exploring the dynamics established among the triad. The object of the establishment of these discussions was, in particular, to delve into the influence of sociocultural factors on interpersonal relationships, specifically emphasizing the themes of 'understanding' and 'being understood' (Mazzocca & Ottolini, 2022). Additionally, considerations on communication extend to the rights of D/deaf citizens in Italy. The Italian National Health System (*Sistema Sanitario Nazionale*) established with Law No. 833 (1978)[101], views health as a community resource. Within this framework, D/deaf individuals have rights that represent non-negotiable duties for all healthcare professionals, including the right to information about their illness, the right to consent or refuse treatment, and the right to privacy. These rights are based on principles of equality, universality, and equity outlined in the Constitutional Law of December 22, 1947, No. 32[102].

Based on a review of the interviews, some findings have emerged regarding the communication practices of Deaf patients in Italy within healthcare settings, as reported in Raniolo & Caligiore (2023):

i. D/deaf patients in Italy frequently rely on informal communication methods when interacting with medical professionals. This often includes lip-reading and writing.
ii. Professional LIS interpreters are not commonly contacted by Deaf patients during medical interactions. Instead, patients tend to use informal methods or rely on family members for interpretation.
iii. Establishing trust and negotiating communication methods are crucial need recognized both by medical doctors and Deaf patients. Doctors often adjust their communication by speaking more slowly or removing masks to aid lip-reading.

[101] https://www.gazzettaufficiale.it/eli/id/1978/12/28/078U0833/sg (Accessed on October 9th, 2023).
[102] https://www.senato.it/node/129/printable/print (Accessed on October 10th, 2023).

iv. There is a recognized need for increased awareness and training among medical professionals regarding the communication needs of D/deaf patients. Many healthcare providers may lack preparedness to enable efficient communication.

v. The use of masks during the pandemic has intensified communication challenges in healthcare settings, especially for D/deaf patients.

vi. Remote interpreting services, such as the ULSS service, have been introduced in Italy in recent years. However, these services face challenges related to technology availability and older Deaf individuals' familiarity with remote interpreting.

These findings, although deriving from a limited sample of interviews, align with a scoping review by Jin Yet *et al.* (2022) conducted in anglophone states (USA, Australia, Ireland, and South Africa), which also found that Deaf individuals' preferred communication methods include lip-reading and writing. The presence of either a professional or non-professional interpreter was less common, with long-distance interpreting being the least preferred option across all reviewed papers.

In conclusion, understanding the dynamics and challenges faced by D/deaf patients in healthcare settings is not only important for opening a discussion on improving the quality of medical care, but is also important in relation to the methodologies employed in LIS data collections. As the selection of tools for collecting LIS data is discussed, it becomes clear that understanding the experiences and preferences of Deaf individuals calls for a nuanced approach.

6.4. On the importance of working with an interdisciplinary group

As mentioned, the three preliminary MultiMedaLIS Datasets presented were collected as a result of interdisciplinary collaborations between linguists and engineers respectively from the DISUM-UNICT and DIEEI-UNICT. Additionally, a significant contribution to the selection of the multimodal tools for the data collection was given by the researchers of the Optics and Sensing Lab of the MPI-IS, who specialize in multi-sensor data capturing and analysis methods for generating large scale datasets.

During the exploration of best practices for collecting sign language data, the importance of adopting an interdisciplinary approach was recognized. Drawing from the insights of linguists and engineers alike (Crasborn & Zwitserlood, 2008; Geraci *et al.*, 2014; Bragg *et al.,* 2019; 2021), questions concerning the status of LIS documentation in the Italian context, the challenges linked to sign language processing, and how these challenges that are to be faced from an interdisciplinary standpoint, were addressed (Bragg *et al.,* 2019).

Beyond the development of guidelines for sign language collection and processing, the limited availability of LIS documentation in Italy was acknowledged, with resources being developed mainly for personal use and, therefore, not accessible. It was also noted that the required quality of the video data is high, in order for it to be more easily processed for automatic recognition and machine translation tasks and, lastly, that the collection of signed data can occur in a continuous or isolated way. In relation to the key parameters for sign language datasets, especially those designed for automatic translation, Bragg *et al.* (2021) introduced a table outlining the parameters related to the collection and processing of sign language data within an interdisciplinary perspective. This table highlights potential characteristics associated with these parameters, as well as their implications in terms of Fairness, Accountability, Transparency, and Ethics (FATE). FATE is a framework has gained notoriety in the context of artificial intelligence, following the increasing awareness of ethical considerations in its research and applications. In essence, fairness refers to ensuring that processes, decisions, and outcomes are impartial, non-discriminatory, and free from bias. Accountability involves transparent decision-making processes, while Transparency highlights clarity and openness in the processes at play. Lastly, Ethics encompasses the moral principles that guide decision-making throughout the course of a project (Lepri *et al.,* 2017; Memarian & Doleck, 2023), Table 9 provides an adapted version of a table presented by Bragg *et al.* (2021, p. 9), outlining various interdisciplinary parameters related to the collection and processing of sign language datasets. This adapted table includes parameters considered relevant for the goals of this work. The second-to-last column on the right describes the solutions associated with the 'Impact related to FATE' adopted by the interdisciplinary group involved in the Datasets' development. An additional inclusion is the last column on the right, which indicates the research area of the researchers contributing to decision-making on the topic.

Parameter	Possibilities	Impact related to FATE	Proposed solution	Researchers' area of expertise
Signing Data	1. RGB/Depth video (2D/2.5D/3D) 2. Motion Capture 3. Gloves 4. Other sensors	Hardware requirements, recording setup, who can participate, dataset size, quality of resulting models, types of end applications that can be created, privacy concerns	• Setup with portable depth camera with two infrared cameras and an RGB camera • Domestic setting • High definition (HD) • Sole participant (Dataset developer) • Privacy concerns (Dataset not yet public)	Sign language linguistics/ electronic engineering
Label Format	1. Gloss systems 2. Spoken language translation 3. Linguistic notation systems 4. Computer notation systems 5. Sign language writing systems	Annotation granularity, difficulty of temporal alignment, amount of data required for training, labeling process (who can label, with what software), inter-labeller agreement, dataset size (due to cost), model quality	• **Labels in Italian and English**, alongside a language-specific system (Typannot) • **Labeling Process c**onducted by a hearing Ph.D. student • **Inter-labeller Agreement** not required • **Annotation Model** based on documented experiences in multimodal corpora annotation	Sign language linguistics

Grammatical structure	1. Single isolated Signs 2. Continuous signing	What grammatical structures can be modeled, which end applications are possible	• Single isolated Signs: Test multimodal tools operability	Sign language linguistics/ electronic engineering
Vocabulary	1. Limited 2. Unrestricted	How much data is needed to train accurate models, which end applications can be created	• Semantically limited data (medical domain). • 205 repetitions of 126 Signs • Approximately 26.0000 individual Signs.	Sign language linguistics/ electronic engineering
Prompt	1. Prompted/ scripted 2. Unprompted/ unscripted	Data quantity to train accurate models, whether models will work for real-world use cases	• Prompted/scripted data, as the Dataset is preliminary, and the accuracy of the multimodal tools needed testing	Sign language linguistics/ electronic engineering
Recording setup	1. In-lab 2. Real world	Data quality, data quantity to train accurate models, model real-world viability	• HD quality • Approximately 26.000 Signs • Recording tools can be moved as desired	Sign language linguistics/ electronic engineering

Table 9. Adaptation of table presented by Bragg et al. (2021, p.9) on the parameters along which sign language datasets may vary.

In conclusion, the table provided shows that by adopting an interdisciplinary approach it is possible to propose solutions to the challenges faced in developing sign language datasets. This approach ensures that resources are created and processed while taking into account various factors that may influence the collection of signed data.

6.5. Annotating the MultiMedaLIS_3 Dataset: practical solutions from annotation efforts

Gabrielle Hodge and Onno Crasborn are two authors with experience in developing and processing sign language corpora (Hodge, 2014; Hodge *et al.,* 2019; Hodge & Goico, 2022; Crasborn *et al.,* 2007; Crasborn & Zwitserlood, 2008; Crasborn & Sloetjes, 2008; Crasborn *et al*., 2015; Crawford & Eniko, 2016). They recently joined to review good practices in sign language corpora annotation, making this discussion particularly relevant and valuable for those involved in sign language data annotation. Following the principles outlined by Hodge and Crasborn (2022), this section provides an overview of the ongoing annotation efforts for the third version of the MultiMedaLIS Dataset.

In the development of time-aligned annotations for the Dataset, the theoretical framework primarily draws from the COS-S approach (Volterra *et al.*, 2022). This work represents the concretization of an approach to sign language research that has evolved over the years among some members of the international sign language research community. As discussed in Chapter 1, 2 and 3, this approach has its roots in the theoretical works sign languages by Cuxac (2000), Russo Cardona (2004b), Cuxac & Antinoro Pizzuto (2010), and the broader cognitive and embodied language theories of researchers like Lakoff & Johnson (1980) and Langacker (1987).

However, the theoretical approach adopted for the annotation of the MultiMedaLIS_3 Dataset strives for the identification of a balance among language-specific approaches and the need to provide sign language data collection with a concise description of their elements (Chesi & Geraci, 2009) that is also machine-readable. This is undoubtedly a challenging task, however, it is through these attempts and the discussions that derive from them that it may be possible to reach the development of annotation standards for sign languages that are not solely based on formalist approaches, but are able to equip LIS Datasets (intended as the raw video data) with an easily readable transcription that «[…] allows anyone who knows the object language to reconstruct its forms, and its form-meaning correspondences in their contexts, even in the absence of 'raw data'» (Antinoro Pizzuto *et al.,* 2010). To this regard, one could take a step further and aim to make the data accessible even to those who *do not know* the sign language in object. In fact, drawing from the experience gathered from working in an interdisciplinary context it has ben noted that it is not reasonable to expect specialists coming from different backrounds from that of sign language linguistics to become proficient in a sign language, just as all linguists cannot be expected to become experienced programmers. Therefore, the sought-after balance, achieved by combining formalist and functionalist approaches to annotation, could have immediate and practical applications with the potential to offer real-world and social benefits.

Due to the size of the MultiMedaLIS_3 Dataset (approximately seven hours) and the fact that a single person is responsible for the annotation, it will not be possible to entirely annotate it by the publication date of this work. However, in line with Hodge and Crasborn's guidelines (2022), this limitation will be addressed by providing a strategic sample of the Dataset with the associated annotations. Additionally, drawing from annotation procedures used in various sign language corpora, such as the *Corpus NGT* (Crasborn, 2015), the *BSL Corpus* (Cormier & Fenlon, 2014) and the *AUSLAN Corpus* (Johnston, 2019), it is central to emphasize that the annotations currently available may not be considered final. In fact, depending on potential future applications of the Dataset, new annotations might be added following specific research interests and requirements.

In terms of annotation software, the MultiMedaLIS_ 3 Dataset was annotated using ELAN, as discussed in section 6.3.1., and following a process initiated with the annotation of the first data collection (see Figure 21 on p. 159), in adherence to the standards established by the developers of corpora included in the ECHO project, such as the *Corpus NGT*, the *BSL Corpus*, and the *SSL Corpus* (Crasborn *et al.,* 2007), as well as the *AUSLAN Corpus* (Johnston, 2019).

The following sections will be dedicated to the description the effort involved in the creation of different levels of annotation, and the theoretical motivation the decision to include specific information within these segments. Following ELAN naming conventions, annotation levels will be new referred to as tiers.

6.5.1. Types

As introduced in section 4.3., the first step that is usually taken when developing an annotation on ELAN is the creation of types. Types specify the stereotypes (or constraints) applied to a tier, (Wittenburg *et al.,* 2006) that will determine the type of relation that a tier established with other tiers, as well as the information that can be included in the segments of a tier by using controlled vocabularies (refer to section 6.5.2.). All types developed for the annotation of the MultiMedaLIS_3 Dataset, except for the tier Sign, who is the parent tier, are marked by the stereotype *Included in*, which means that: «[a]ll annotations on a tier of this type are linked to the time axis and are enclosed within the boundaries of an annotation on the parent tier» (Wittenburg *et al.,* 2006, 2.1.). Additionally, types marked by this stereotype are time alignable. As can be seen by looking at Figure 28. Current types developed ELAN most types make use of a Controlled Vocabulary.

Current Types

Type Name	Stereotype	Use Controlled Vocabulary	DC ID	Time-alignable
Sign	-	-	-	✔
Articulatory Sub-System	Included In	Articulatory sub system	-	✔
Parts	Included In	Parts	-	✔
DoF FlxExt	Included In	Flexion/Extension	-	✔
DoF AbdAdd	Included In	Abduction/Adduction	-	✔
DoF RinRex	Included In	ProSup/RinRex	-	✔
DoF Fingers	Included In	-	-	✔
DoF Nostrils	Included In	Nostrils	-	✔
DoF Pupils	Included In	Pupils (1 Value)	-	✔
DoF Eyelashes	Included In	Eyelashes	-	✔
DoF Eyebrows	Included In	Eyebrows	-	✔
DoF Jaw	Included In	Jaw (1 Value)	-	✔
DoF Lips	Included In	-	-	✔
DoF Corners of the Mouth	Included In	Corners of the Mouth	-	✔
DoF Tongue	Included In	Tongue (1 Value)	-	✔
DoF Air	Included In	Air (1 Value)	-	✔
Movement Type	Included In	-	-	✔
Hands Contact	Included In	-	-	✔
Meaning Unit Type	Included In	-	-	✔
Vocal Language Label Type	Included In	-	-	✔
Lexical, Deictic, Transfer U...	Included In	-	-	✔

Figure 28. Current types developed ELAN for the annotation of the MultiMedaLIS_3 Dataset.

6.5.2. Controlled Vocabularies

A Controlled Vocabulary (CV) in ELAN is a customizable vocabulary created or imported by the annotator. These vocabularies enable annotators to choose from a defined set of values when selecting information to include within a segment on a tier.

The use of CVs streamlines the annotation process, making it faster and reducing the likelihood of typos and errors (Wittenburg *et al.,* 2006).

Table 10 provides a list of the used CVs. Most of these CVs use the Typannot font with the only exception being the ‘Y’ letter, which represents the eye component, as a symbol for this element has not yet been developed by Typannot designers (see section 4.2.4.4.). The CVs used in this annotation process are an adaptation and integration of those were originally developed by Chevrefils (2022)[103].

CV name	Description	Values and value name
ProSup/RinRex	General pronation (+) and supination (-) values	= pro/rin +4/4 = pro/rin +3/4 = pro/rin +2/4 = pro/rin +1/4 = pro/sup 0 = sup/rex −1/4 = sup/rex −2/4 = sup/rex −3/4 = sup/rex −4/4 = hidden
Nostrils	Nostrils values	= nostrils up = nostrils neutral = nostrils down
Eyelashes	Eyelashes values	= wide = neutral = semi-closed +1 = almost closed +2 = closed
Jaw (1 Value)	Jaw value	= jaw neutral
Flexion/Extension	General flexion (+) and extension (-) values	= flex +4/4 = flex +3/4 = flex +2/4 = flex +1/4 = flex/ext 0 = ext −1/4 = ext −2/4 = ext −3/4

[103] I would like to thank Léa Chevrefils for kindly sharing her CVs with me, and Prof. Claudia S. Bianchini for putting us in contact.

		= ext −4/4 = hidden
Abduction/Adduction	General abduction (+) and adduction (-) values	= abd +4/4 = abd +3/4 = abd +2/4 = abd +1/4 = abd/add 0 = add −1/4 = add −2/4 = add −3/4 = add −4/4 = Hidden
Tongue (1 Value)	Tongue Value	= tongue neutral
Articulatory sub system	Articulatory sub system values	= upper limb right = upper limb left = fingers right = fingers left Y = eye = mouth
Pupils (1 Value)	Pupils value	= pupil towards interlocutor
Parts	Value for body parts: arm, forearm, and hand.	= hand = forearm = arm
	Rotation of the arm values. Internal rotation (+) or exterior rotation (-).	= internal forearm rotation interior +4/4 = internal forearm rotation +3/4 = internal forearm rotation +2/4 = internal forearm rotation interior +1/4 = internal/external forearm rotation 0 = external forearm rotation −1/4 = external forearm rotation −2/4 = external forearm rotation −3/4 = external forearm rotation −4/4 =forearm hidden
Corners of the Mouth	Corners of the mouth values	= outwards = inwards = neutral
Lips (1 Value)	Lips value	= lips
Air (1 Value)	Air value	= air neutral

Table 10. List of Controlled Vocabularies developed for the annotation of the MultiMedaLIS_3 Dataset, together with their values.

6.5.3. Tiers

The ELAN annotation template for the Dataset has been designed to meet various requirements. Following existing corpora annotation standards (Cormier *et al.,* 2016; Hodge & Crasborn, 2022) it includes Vocal Language Label tiers (named 'Italian VLL' and 'English VLL') to ensure interdisciplinary accessibility, allowing non-signers and English speakers to understand the form/meaning relationship of the LIS Signs, as well as information on Unit of Meaning type (Cuxac, 2000), see Figure 29. Annotation sample showing the tiers dedicated to the description of Vocal Language Labels and Units of Meaning (Volterra et al., 2022). In addition to describing Units of Meaning, the dependent tiers— including Lexical Unit, Deictic Unit and Transfer Unit (see section 2.2.5.)— provide information about the type of Unit represented by the sign. In fact, while all Signs in the Dataset fall into the category of Signs that 'tell without showing' and are, therefore, Lexical Units, the annotation model is designed with expandability in mind. It will allow for future annotations that might encompass other types of data, including Deictic or Transfer Units, which can indicate the presence of a Transfer of Person, Transfer of Form and Size, Transfer of Situation.

The structure of this annotation methodology enables the marking of Double Transfer, as described by Volterra *et al.* (2022).

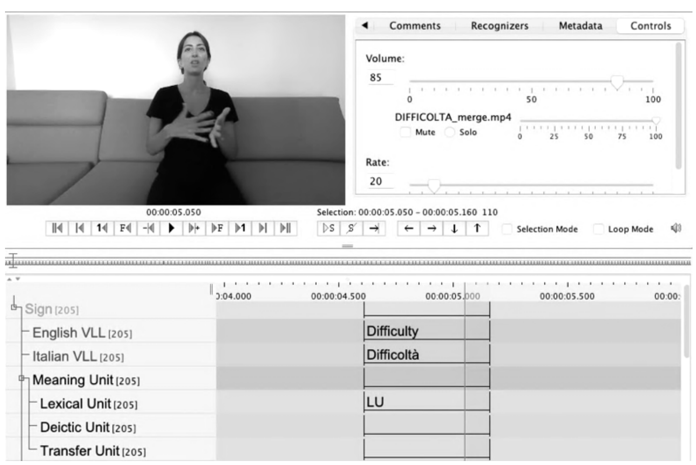


Figure 29. Annotation sample showing the tiers dedicated to the description of Vocal Language Labels and Units of Meaning (Volterra et al., 2022) developed for the annotation of the MultiMedaLIS_3 Dataset (CC BY 4.0).

As shown in Figure 30 below, the information presented in the Typannot font encodes features related to the movement of the upper limbs (arm, forearm, and hand), the configuration of the hands, and the presence or absence of contact between hands.

Figure 30. List of tiers developed for the annotation of the MultiMedaLIS_3 Dataset.

Although the annotation methodology is based on the ontology presented by Volterra *et al.* (2022), it is important to note that the distinction between manual and body components is not applied in this annotation or, at least, is not applied following the mentioned ontology, as priority is given to the codification standards of the transcription system itself. In fact, Typannot describes the positions of the human body and the relationships established between body components leading to these positions (e.g., arm ➔ forearm ➔ hand ➔ fingers). Using the manual and body element categorization from Volterra *et al.* (2022) would have required adapting the transcription system, which was not preferred. This choice does not stand as a disregard of the COS-S ontology, but rather represents a practical application aligned with the authors' recommendation to increase the use of language-specific transcription systems. For this reason, tiers that describe arm, forearm, hand, and fingers, as well as tiers that encode information related to the eyes and mouth will be found. Following Typannot's subdivision of the face, information on the eyes includes the movement of nostrils, pupils, eyelashes, and eyebrows, while information on the mouth encompasses movements of the jaw, lips, corners of the mouth, tongue, and airflow (see section 4.2.4.).

Standardizing the starting and final positions for each Sign performance became necessary due to constraints posed by the data collection tools. This process led to the selection of a seated position as the 'standard' starting point. This decision was largely influenced by the observation that all corpora collections included in this study featured video recordings of signers in seated positions (Crasborn *et al.*, 2007; Cecchetto *et al.,* 2011; Geraci *et al.,* 2011; Gianfreda, 2011; Mesch, 2012; Schembri *et al.,* 2017; Johnston, 2019). Therefore, in contrast to the MultiMedaLIS_1 and 2 Datasets, where signers were in a standing position, the MultiMedaLIS_3 Dataset adopts a seated starting position for Sign performances. Choosing a seated starting position has implications for the calculation of movement and position values within the Typannot framework, as issue that had already been discussed by Léa Chevrefils in her Ph.D. thesis (2022). In fact, her prior work influenced the decision to consider this seated position as the starting point for Sign performances, which, in turn, affects the selection of movements like flexion, extension, rotation, adduction, and abduction (see section 4.2.4.1).

By looking at Figure 30, it can be seen that the tiers containing the label ‘Mov’ (movement) and ‘Hands Contact’ have not been discussed yet[104]. With regards to hand contact, either YES or NO are included in the tier, depending on its presence or absence. In terms of codifying movement, since it is still a work in progress, Typannot developers suggested the development of a specific methodology to encode it by adapting existing Typannot symbols and combining them with Arabic numerals.

Finally, it is worth noting that the intentional order of the tiers facilitates the conversion from *Formule Générique* to *Glyphes Composés* (see section 4.2.2). In fact, by extracting the symbols that correspond to Typannot *Formule Générique,* it is possible to obtain the *Glyphes Composés* form of the system. This is achievable since the order of tiers follows Typannot’s syntax, a practical example that aligns with the developers' intention to create a system that is both machine-readable and human-readable (Bianchini, 2023).

6.5.3.1. Codifying body components: the eyes and mouth

In describing body components, a methodology that combines Typannot, developed by Boutet *et al.* (2016) and Boutet (2018) and the COS-S framework proposed by Volterra *et al.* (2022) is adopted. Is doing so, Typannot symbols are employed to encode information relating to the eyes, which, in the Typannot framework, have an impact on the movement of nostrils, eyelashes, eyebrows, and pupils. Additionally, data related to the mouth, that, in the Typannot framework, affects the movement of the jaw, lips, corners of the mouth, tongue, and airflow is codified. This hybrid methodology provides a comprehensive depiction of body components.

---

[104] Regarding these matters, a discussion was held with Prof. Claudia S. Bianchini during a research visit at the University of Poitiers in June 2023. The conversation focused on the codification of movement and hand contact for the Dataset.

The inclusion of the ‘Eyes’ tier and its child tiers, namely ‘Nostrils’, ‘Pupils’, ‘Eyelashes’ and ‘Eyebrows’, allows for the codification of essential body information. Following Cuxac and Antinoro Pizzuto’s Framework (2010) and its 2022 adaptation by Volterra and colleagues, tiers like ‘Pupils’ and ‘Eyelashes’ (which is an eye aperture tier) are crucial in that the information codified within them becomes essential in distinguishing Units of Meaning during the processes of conveying information without showing and conveying information through showing. This differentiation is confirmed in all LIS Sign instances included within the Dataset. In fact, they are Lexical Units, a characteristic sustained by the information included in the ‘Pupils’ tier, where the symbols ⊡⊕ indicate that the signer's pupils, representing their eye gaze, are directed towards the interlocutor.

The ‘Mov (Lips)’ tier is specifically designed to contain information related to mouth actions, which are further divided into mouth gestures and mouthings. On this topic, refer to Fontana (2008) for the description of LIS mouth actions that has been adopted in this work, and to Crasborn & Bank (2014) for a cross-linguistic analysis. The formalization of the use of mouth components in LIS is a complex matter, and various studies have explored its significance from both etic (research-oriented) and emic (user-centered) perspectives. Mouth actions have been recognized as fundamental contributors to signed utterances in research (Fontana, 2008), and they also play a role in the day-to-day communication of LIS users, considering their epilinguistic awareness (Fontana & Raniolo, 2015).

To capture these nuances, the annotation methodology developed for the annotation of the MultiMedaLIS_3 Dataset is closely linked to the execution of Signs from a manual perspective. Mouth gestures, which can be configurations of the mouth (Fontana *et al.,* 2022), are codified using Typannot, as the system that accommodates such descriptions. The ‘Mouth’ tier serves as a parent tier encompassing the sub-tiers: ‘Jaw’, ‘Lips’, ‘Corners of the mouth’, ‘Tongue’ and ‘Air’. This hierarchical structure enables a more granular and comprehensive analysis of mouth actions. In contrast, in order to transcribe mouthings, a dedicated tier that will only include Roman alphabet letters has been introduced. This approach acknowledges the close relationship between mouthing and the rhythm of the Sign (Fontana, 2008) and emphasizes that mouthings tend to occur frequently in conjunction with manual components (Fontana *et al.,* 2022).

In relation to the annotation methodology for mouthings, transcription conventions for the analyzed corpora were investigated, with particular reference to conventions developed for corpora taking part in the ECHO project (Crasborn *et al.,* 2007; Johnston, 2019). In reference to mouth actions, it was determined that SSL materials is annotated following a set of codes that describes lexically determined and non-morphemic mouth gestures (Nonhebel *et al.,* 2004c) and that a development of the system was carried out to include more mouth positions. However, given the inclusion of Typannot, these types of annotating conventions were not useful, as the mentioned typographic system is already capable of codifying this information and therefore, at this stage, the focus of the enquiry were annotation conventions for mouthings, in particular. Therefore, for the annotation of mouthings, it was decided to implement an adaptation of the guidelines given by Johnston *et al.* (2015) for the AUSLAN Corpus, which were adapted to align with Fontana's (2008) classification.

Building on these practices, the Roman alphabet is employed to represent the spoken Italian word being articulated by the signer. When the word is fully articulated, the complete word is written. In cases where only portions of the word are articulated whether they are initial, medial, or final, these portions are enclosed within brackets (Johnston *et al.,* 2015). To distinguish this transcription from the Italian vocal language label tier, caps lock is used, as demonstrated in Figure 31.

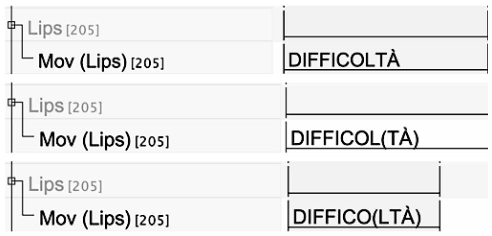


Figure 31. Example of mouthing annotation.

The current annotations in the Dataset offer limited information on the use of mouth actions in spontaneous signing contexts, given the Dataset's preliminary nature. However, the development of this annotation methodology is oriented towards accommodating and codifying mouth actions for *future* LIS data collections. The emphasis is, in fact, on developing a methodology that will be able to host and codify future expansions of the Dataset.

6.5.3.2. Codifying movement

The use of ad adaptation of Typannot symbols to codify movements is a strategy that allows for a detailed description of changes in the position of the formational parameters of a sign. For instance, Figure 32 depicts the strategy used to signal a repetitive movement of the right forearm (refer to Table 10 for information on the values) marking a shift in internal and external rotation and flexion and extension.

Figure 32. Example of movement annotation using Typannot.

Following Typannot syntax, from the symbols included in the 'RinRex Forearm R' tier, the following formula will be obtained . This formula marks the starting position of the Sign before initiating the movement. Right below this tier, as depicted in Figure 33, is the 'Mov (RinRex Forearm R)' tier, where movement is described by indicating the variation from neutral position ( ) to a +2 internal rotation ( ). The repetition of the movement is indicated by ‘x2’, signifying a total of two movement repetitions.

6.5.4. Tokenization

After having identified tier and types and having defined CVs, the concrete annotation of the MultiMedaLIS_3 Dataset began with the most basic and necessary level of corpus annotation: *tokenization*. In relation to sign language data collection, tokenization has been described as «[…] process of segmenting and identifying type-token relationships between the realizations of ‘words’ or ‘Signs’ in a corpus (i.e., standalone forms that are small and substantive that may vary in degree of conventionalization)» (Hodge & Crasborn, 2022, p. 59).

Since the Dataset consists of isolated Signs, there was no need to establish a discussion on the segmentation of signed utterances[105]. To determine the placement of tokens, reference is made to Adam Kendon’s concept of *stroke* in gestural actions considering it as the phase of the Sign (originally gestural excursion) where the pose and movement are most clearly defined in relation to the entire gestural excursion (Kendon, 2004, p. 124). From a practical standpoint, David McNeill’s annotation practices, which can be found in the Appendix to his 2005 book *Gesture and Thought* were adopted and *adapted*. In doing so, the steps to tokenizing the Dataset are presented:

i. Observe the complete elicitation of the individual Sign.
ii. Identify the Sign stroke, which should be considered, similar to a gesture stroke, as the moment of greatest effort for the signer:
    a) The Sign stroke is considered as the reach of the initial form of the Sign.

[105] Despite being beyond the scope of this work at its stage, a discussion on the segmentation of signed utterances will be reserved for the future development of the MultiMedaLIS Datasets, which will ideally include signed utterances, in addition to isolated Signs.

b) The Sign stroke includes the *movement*, which is a fundamental element, and the possible presence of *holds*, which are components that can be combined in LIS.
c) The phases of preparation and retraction, which convey no semantic meaning in this context and are discussed in point B of McNeill's appendix (2005, p. 267), are not included in the token.

6.5.5. Data analysis

Within the dataset encompassing 126 Signs, the outcomes derived from annotations carried out on ELAN using the Typannot system are presented. Specifically, the focus is on two Signs: 'pain' and 'difficulty' (displayed in Figure 33, respectively as Sign n° 21 and n° 23 in Table 6).

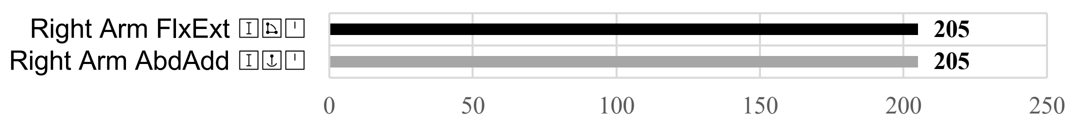
Figure 33. On the left, depiction of the LIS Sign 'pain' ('dolore') on the right, depiction of the Sign 'difficulty' ('difficoltà'). Both images are taken from SpreadTheSign.

These Signs were chosen for analysis due to distinct characteristics: 'difficulty' involves a two-handed execution, while 'pain' employs a one-handed execution. The following discussion will uncover the results derived from Typannot annotations across 205 repetitions of each Sign, developed using ELAN. In fact, ELAN offers the option to export annotations in various formats. For the post-processing intended in this study, the .eaf (Elan Annotation Format) files were exported as tab-delimited text files, enabling visualization of annotations in rows and columns on Microsoft Excel. In the obtained files, each tier corresponded to a column, and each of the 205 annotations represented a row. Therefore, every exported file contained 206 rows, with the first row as the column title and the subsequent 205 rows containing the annotated values. For the focus on Typannot annotations in this section, only the data annotated with this typographic system will be included, excluding labels or other information, and thus excluding some of the details illustrated in Figure 30. As previously mentioned, the positions codified within Typannot were chosen to represent the stroke of the Sign (Kendon, 2004; McNeill, 2005). Therefore, the following sections will delineate and cross the positions adopted by various parts of the upper body throughout the 205 executions of the selected Signs during their stroke phases. The analysis will be carried out as follows. It will begin by reviewing the initial positions for arms, forearms and hands concerning both Signs. After that, the final positions annotated in the dataset, representing movement, alongside the corresponding repetition count will be explored. The movement categorization considers the final position reached by the selected body part before the retraction phase, marking the conclusion of the Sign stroke. Movement repetitions, from the stroke's onset to its conclusion in the pre-retraction final position, are symbolized by 'x' followed by the number of repetitions (e.g., x2, x3, etc.). Lastly, employing pivot graphs, the initial and final positions of the forearms and hands, which present the highest variability in position values, will be cross-referenced.

6.5.5.1. Data analysis of the annotations for 'difficulty'

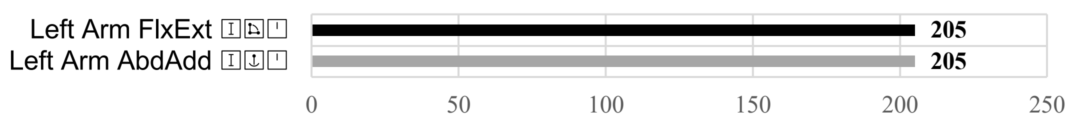
This section will be devoted to the analysis of the annotation of the Sign 'difficulty' specifically focusing on the positions of right and left arm ( ), forearm ( ) and hand ( ). These positions are categorized as follows: flexion and extension (marked by the symbol), abduction and adduction (marked by the symbol), forearm internal and external rotation and, lastly, hand pronation and supination (all marked by the symbol). As shown in Graphs 1 and 2, for both the right and left arms, a singular value for flexion-extension, and another for abduction-adduction is observed, both registering as neutral (□). This indicates that throughout all 205 instances documented in the third version of the MultiMedaLIS Dataset, the arms consistently maintained a neutral position while executing the LIS Sign ‘difficulty’.

Values occurrence for right arm



Graph 1.

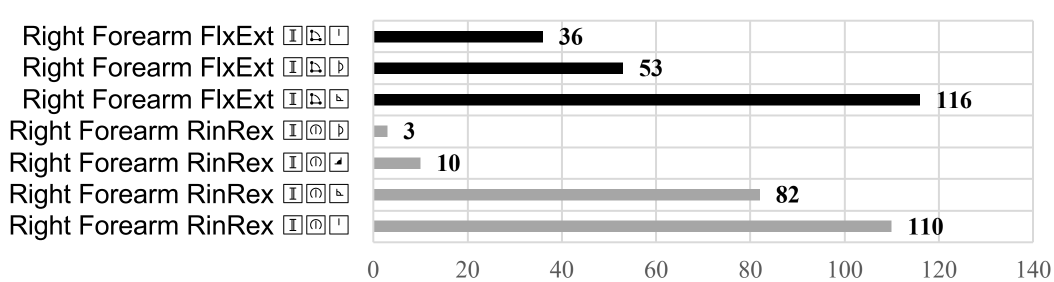
Value occurrence for left arm



Graph 2.

However, concerning the values attributed to the right and left forearm’s flexion-extension and internal-external rotation, a wider spectrum of values is noticed, indicating a greater variability in acquired positions. The right forearm (Graph 3) demonstrates three potential flexion-extension positions, recorded in ascending order: 36 instances of neutral (□), 53 instances of +2/4 (□) and 116 instances of +1/4 (□). Consequently, it can be inferred that in 56.59% of the Sign repetitions, the right forearm stroke position exhibited a slight flexion of approximately +1/4 of the total possible flexion range.

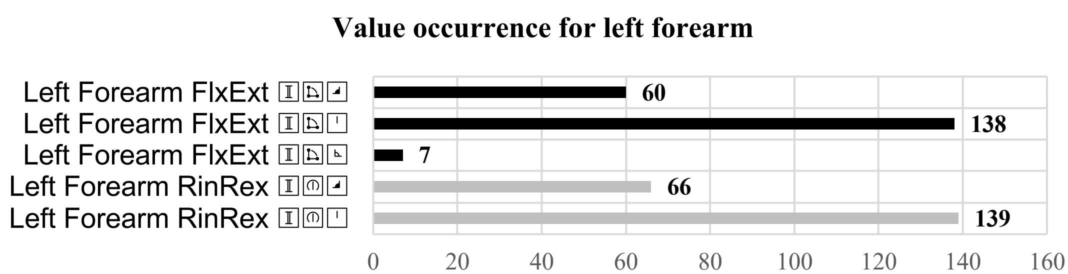
Values occurrence for right forearm



Graph 3.

Conversely, the left forearm (Graph 4) displays three potential initial flexion-extension positions: 7 instances of +1/4 () flexion, 60 instances of –1/4 () extension and 138 instances of neutral () position. This forearm thus showcases a wider extension range in comparison to the right forearm, where only flexion is present, although the neutral value predominates.

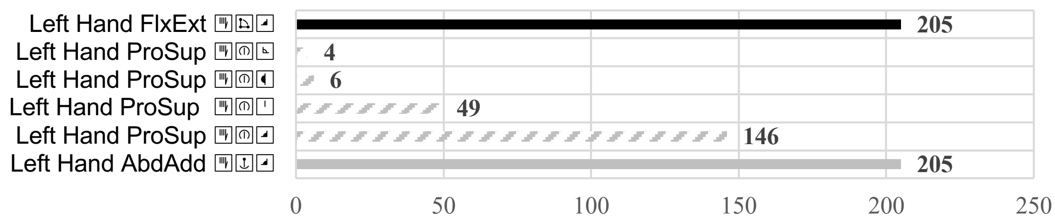
Similar attitudes are also shown by left and right forearm in relation to internal and external rotation initial positions (), as shown in Graphs 3 and 4. Both forearms exhibit a predominant neutral () internal-external rotation position, with the left forearm displaying neutrality in 67,80% of instances and the right forearm in 53,66%. When crossing these internal-external rotation position values, a 32,68% co-occurrence of the neutral position for both the right and left forearms is observed, emerging as the most recurrent combination of left and right forearm flexion-extension positions.



Graph 4.

Turning to the right and left hands, a notable observation is that both hands exhibit only one value for flexion-extension: +3/4 flexion () for the right hand and –1/4 extension () for the left hand. However, a broader range of positions is observed for hand pronation and supination. As shown in Graph 5, the left hand displays 146 instances of –1/4 supination (), 49 instances of neutral position (), along with 6 instances of –2/4 supination () and 4 instances of +1/4 pronation ().

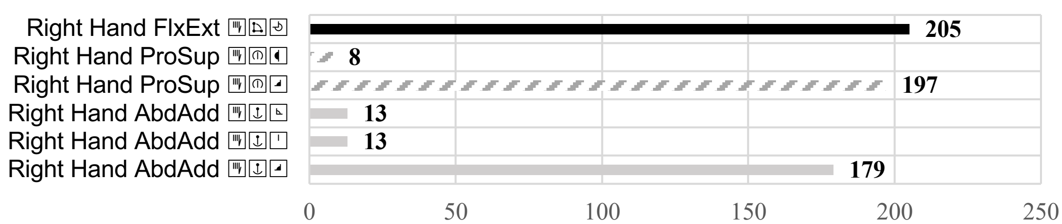
Value occurrence for left hand



Graph 5.

Comparatively, the right hand (Graph 6) exhibits less variability in pronation-supination positions, with 197 instances of –1/4 supination ( ) and 8 instances of –2/4 supination ( ). Lastly, concerning adduction, the left hand consistently shows –1/4 adduction ( ) across all 205 repetitions (Graph 5). The right hand presents more variability (Graph 6), although the most frequent value, similar to the left hand, is –1/4 adduction ( ).

Values occurrence for right hand

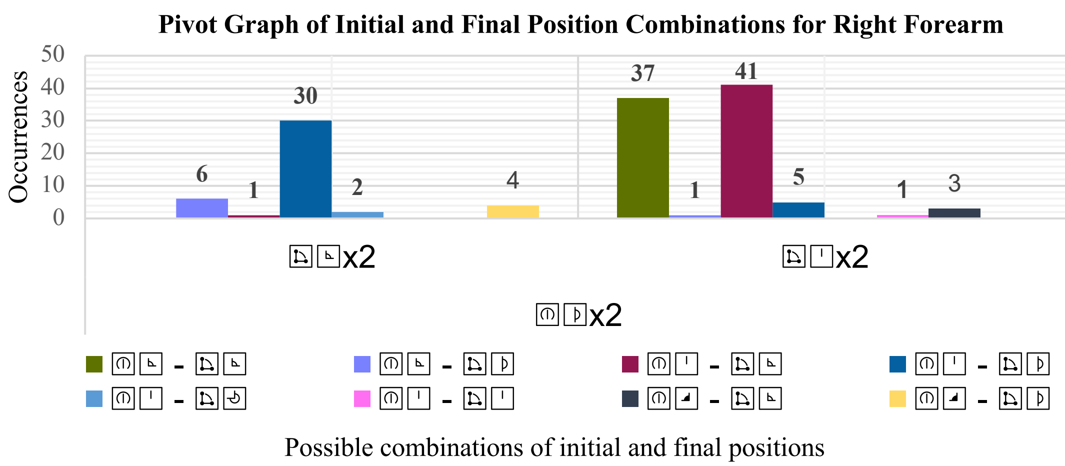


Graph 6.

These analyses feature the consistent nature of certain arm positions while highlighting a greater variability in forearm and hand positions during the execution of the ‘difficulty’ LIS Sign. Overall, the data indicates the presence of slight variations in Sign execution. However, in most instances, these variations remain relatively close to each other. Lastly, the prevalent trend of using neutral positions is observed across the arms and forearms. This consistent adoption of neutral positions within different components of the Sign could imply a fundamental or default orientation maintained during the execution of the ‘difficulty’ LIS Sign, while variations seem to predominantly stem from the positioning of the hands.

Examining the movements executed during the ‘difficulty’ Sign, annotations reveal a prevalent recurring movement present in all 205 instances: a variation in the starting and final positions of the internal-external rotation of both right and left forearms. Specifically, the most frequently occurring combination value among initial positions of left forearm internal-external rotation and flexion-extension is represented as: □□ + □□ + □□x2. Translated, this combination signifies that the most common initial positions of the left forearm (see Graph 4), combined in 113 out of 205 repetitions, result in the final position of +1/4 internal rotation (□□), with the transition occurring twice from initial to final position.

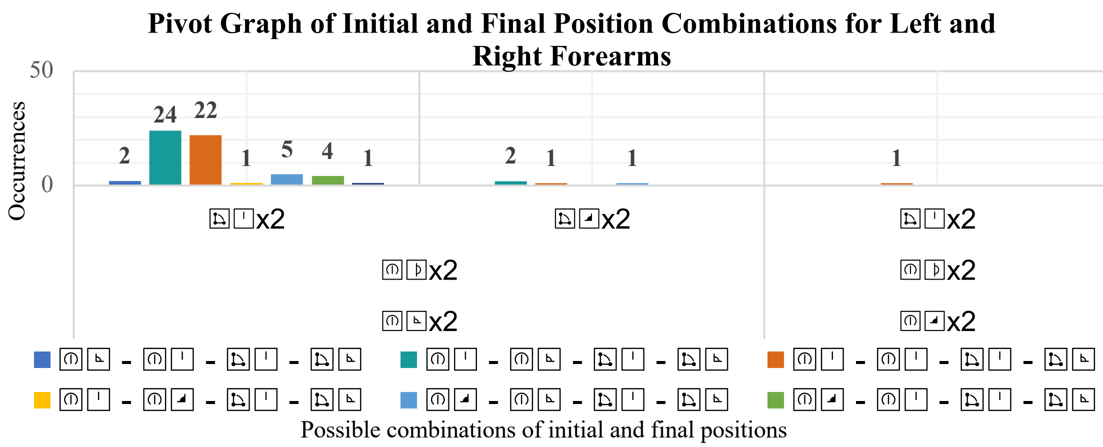
In contrast, the right forearm, displays a more pronounced presence of variations in final positions, concerning both internal-external rotation and flexion-extension, as shown in Graph 7.



Graph 7.

When examining the cross between initial and final positions for the right forearm, the most recurring values manifest as a combination of the initial positions of □□ and □□ with the final positions of □□x2 and □□x2 (41 occurrences). The second most recurring values are a combination of the initial positions of □□ and □□ with the final positions of □□x2 and □□x2 (37 occurrences). This data aligns with the earlier observation that the most common left initial forearm positions are □□ and □□, thus affecting the observed variations in the right forearm positions (see Graph 3).

Upon combining the initial and final positions for both left and right forearm, two predominant groupings emerge, as represented in Graph 8 by the orange and green columns.



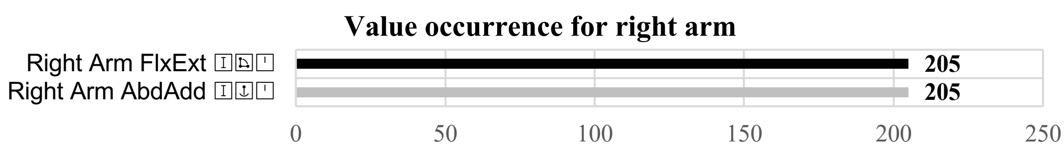
Graph 8,

Firstly, there are 24 instances where the initial positions consist of neutral left forearm rotation (□□), +1/4 right forearm internal rotation (□□), alongside neutral left forearm flexion-extension (□□) and +1/4 right forearm flexion (□□). These initial positions combine with the final positions of +1/4 left forearm internal rotation (□□x2), +2/4 right forearm internal rotation (□□x2) and neutral right forearm flexion-extension (□□x2). Secondly, there are 22 instances where the initial positions consist of neutral left forearm rotation (□□), neutral right forearm rotation (□□), alongside neutral left forearm flexion-extension (□□) and +1/4 right forearm flexion (□□). These combine, once again, with the final positions of +1/4 left forearm internal rotation (□□x2), +2/4 right forearm internal rotation (□□x2), and neutral right forearm flexion-extension (□□x2). While certain combinations occur frequently, the presence of variability within these combinations, as seen in the different initial positions but similar final positions in the second predominant combination, suggests very small flexibility or variation in executing the Sign while maintaining specific components of the sequence.

6.5.5.2. Data analysis of the annotations for 'pain'

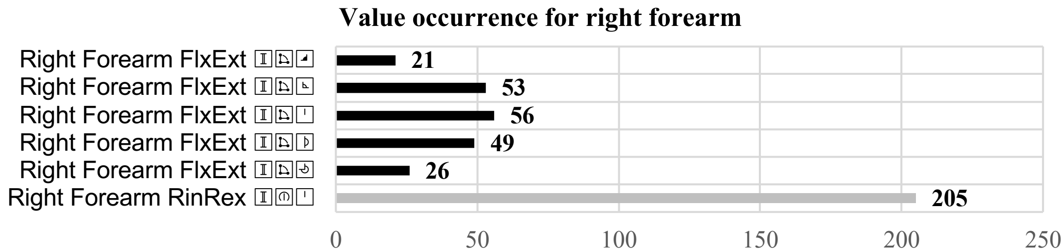
This section will be devoted to the analysis of the annotation of the sign 'pain' specifically focusing on the positions of right arm (□), forearm (□) and hand (□), as the Sign is performed with one hand. These positions are, once again, categorized as follows: flexion and extension (marked by the symbol □), abduction and adduction (marked by the symbol □), internal and external rotation, pronation and supination (all marked by the symbol □).

The analysis will begin by examining initial positions. After that, a cross-reference of the initial and final positions for isolated body parts, specifically the forearm and hand, will be carried out. Following this, the crossed values and visualize the shared initial and final positions of the forearm and hand will be combined, identifying most frequent combinations. Similar to the previously analyzed Sign, as shown in Graph 9, a singular value for flexion-extension (□) and another for abduction-adduction (□) are observed, both registering as neutral (□). This indicates, once again, that throughout all 205 instances documented in the Dataset, the right arm consistently maintained a neutral position while executing the LIS Sign 'pain'.



Graph 9.

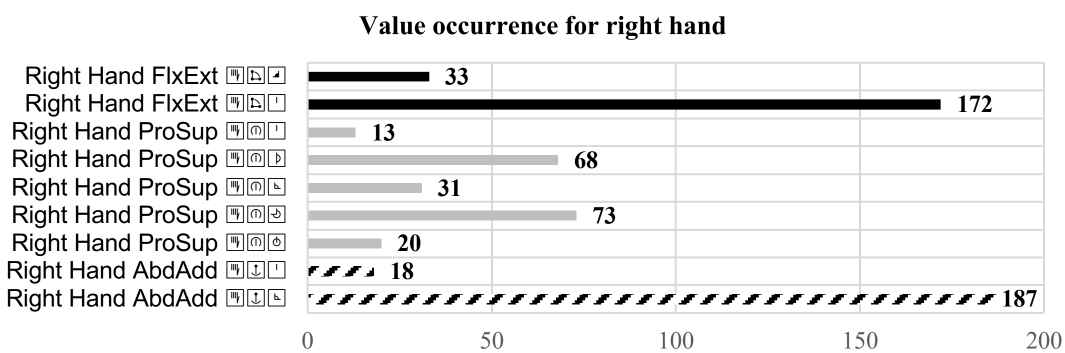
As shown in Graph 10, there is no observed variability in the right forearm's internal-external rotation, as all instances exhibit a neutral rotation position (□□). However, variable values are present with regards to the initial flexion-extension positions of the right forearm.



Graph 10.

Regarding the initial positions of the right hand (Graph 11), movements involve flexion-extension, pronation-supination and abduction-adduction. The predominant initial position for flexion-extension is neutral (□□) occurring 172 times out of 205 (83.90% of instances). A similar prevalence is observed in the occurrence of +1/4 abduction (□□), observed 187 times (91,22% of instances).

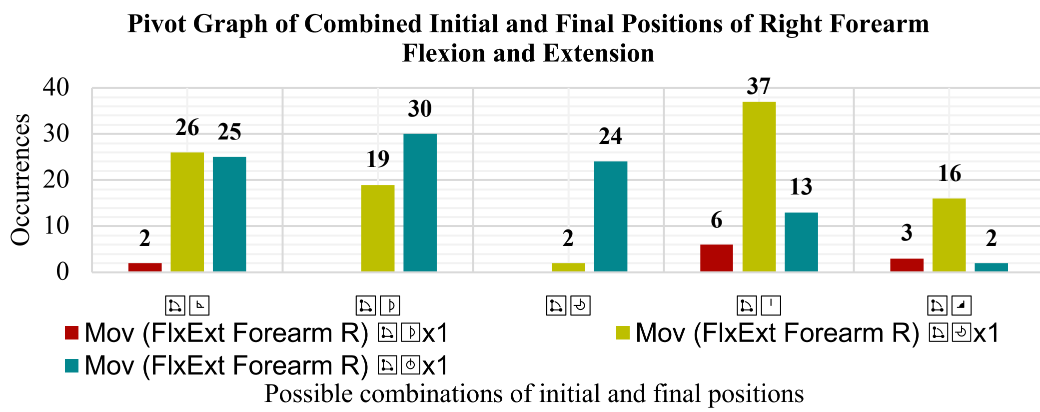
Nonetheless, more variability is present in the initial pronation-supination position. In fact, considering the formational parameters of the LIS Sign ‘pain’, depicted in Figure 33, the signer is expected to start with some degree of hand pronation. As anticipated, the observed variables align with this expectation. The most frequent variables correspond with the expected initial pronation position of the right hand: 73 instances of +3/4 pronation (□□) and 68 instances of +2/4 pronation (□□).



Graph 11.

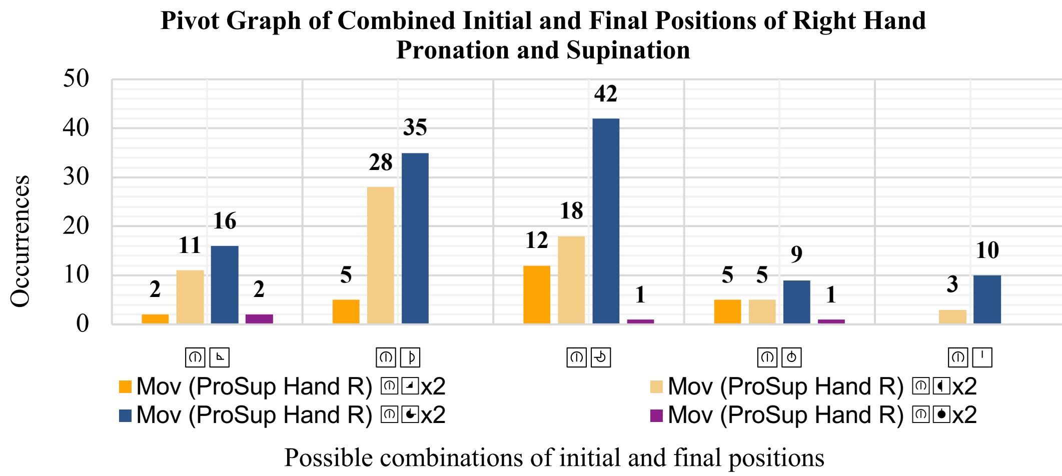
Now, having identified the most common and variable values for the initial positions of the right forearm flexion-extension and hand pronation-supination, let's combine these values with the annotated final positions to identify the most prevalent crossed combinations. In the following graphs, the final position is denoted with the label 'mov' for 'movement'. This transition from the initial to final position indicates the movements of the body part.

The arm is excluded as no shifts from the initial position were detected. Regarding the right forearm, variations are noted in the initial-final flexion and extension positions. Among the recorded final positions, the most prevalent is neutral flexion-extension initial position (□□) combined with +3/4 flexion (□□) final position (37 occurrences). Additionally, the initial position of +2/4 flexion (□□) combines with a final position of +4/4 flexion (□□) of the forearm (observed 30 times), as depicted in Graph 12.



Graph 12.

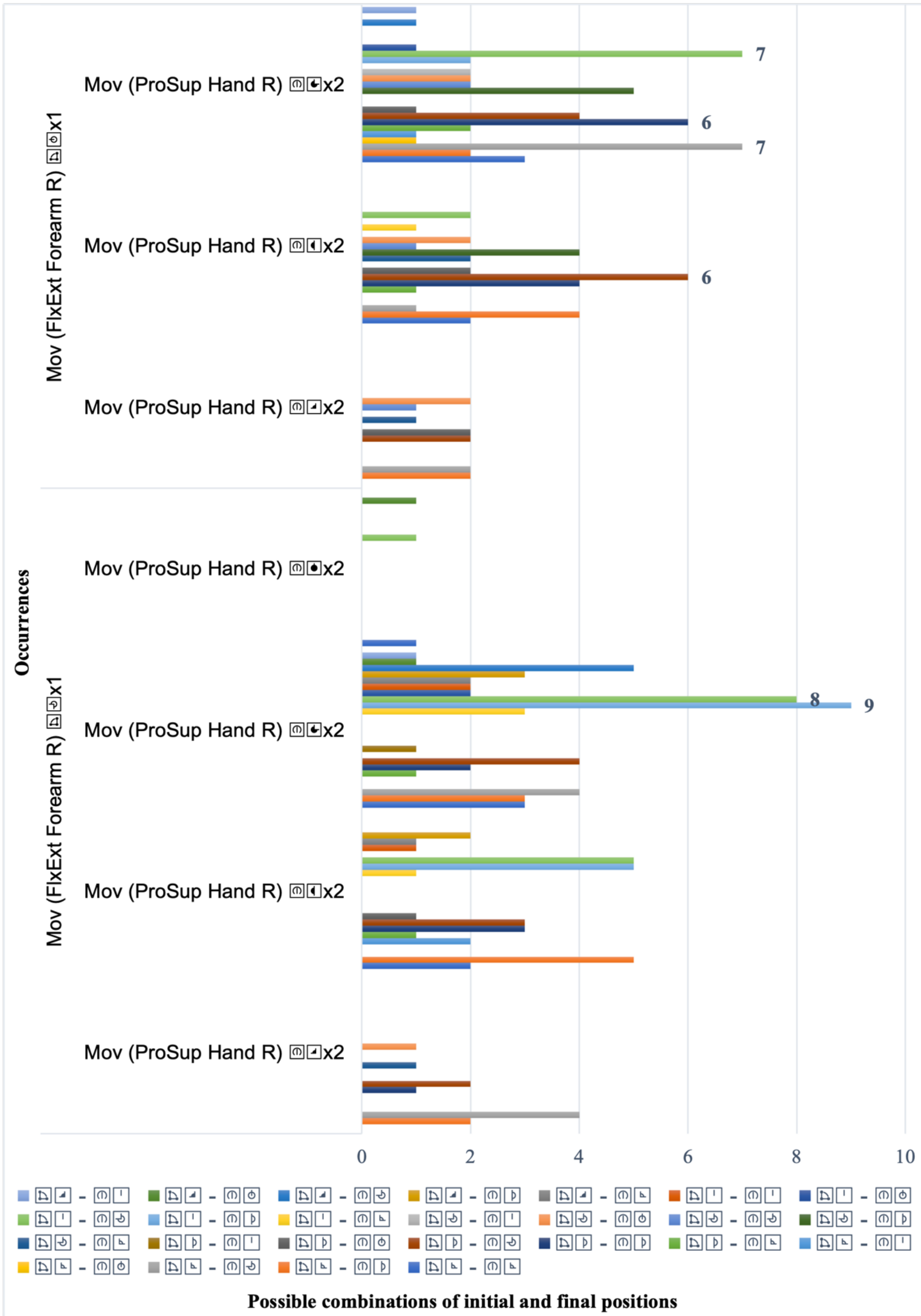
Regarding hand pronation and supination of the right hand (Graph 13), the most recurring final position is –3/4 supination (□□) which combines with the most prevalent initial position values (refer to Graph 11), namely +3/4 pronation (□□) and +2/4 pronation (□□), describing a rotational movement of the hand.



Graph 13.

Upon combining the initial and final positions of the right hand pronation and supination and right forearm flexion-extension throughout the 205 repetitions of the 'pain' LIS Sign, diverse distribution of value combinations is encountered. While all possible combinations account for the total 205 Sign repetitions, for clarity, focus will be put on combinations of initial and final positions with occurrences that are equal or higher that 6.

The data taken into consideration is highlighted in Graph 14 by colored squares. By looking at Graph 14, it can be seen that the highest combination occurrence is the initial position of neutral forearm flexion-extension () and hand +2/4 pronation (), with the final position of +3/4 forearm flexion () and –3/4 hand supination (), with a shift from initial to final pronation-supination position repeated twice. As a matter of fact, it has been in Graph 13 that a combination of the initial +2/4 pronation () position with the final –3/4 supination (), is the second-most common initial and final pronation-supination position (35 occurrences). Therefore, this value aligns with previously calculated data. The cross of this alignment with initial and final flexion-extension movements, gives back a value that corresponds to the most frequent initial neutral flexion-extension forearm position () and the final +3/4 forearm flexion (), coherent with the data presented in Graph 12.


Mov (FlxExt Forearm R) x1
Mov (ProSup Hand R) x2
Mov (ProSup Hand R) x2
Mov (ProSup Hand R) x2
Mov (FlxExt Forearm R) x1
Mov (ProSup Hand R) x2
Mov (ProSup Hand R) x2
Mov (ProSup Hand R) x2
Mov (ProSup Hand R) x2
Occurrences
7
6
7
6
8
9
0
2
4
6
8
10
Possible combinations of initial and final positions

Graph 14.

6.5.6. The advantages of using sign language-specific systems to annotate LIS

Upon examining the visualized Typannot annotations developed on ELAN for the LIS Signs ‘pain’ and ‘difficulty’, a key observation emerges. In both Signs, the arms consistently maintain a neutral arm position for flexion-extension and abduction-adduction, implying a foundational posture during execution. This data may reflect the controlled recording setting for the MultiMedaLIS_3 Dataset, where the signer adhered to consistent Sign execution due to recording protocol constraints. Because the Signs were recorded in isolation, the arms consistently maintained a neutral position. However, in a different context with spontaneous signing, variations in flexion-extension and adduction-abduction might occur. Contextual factors such as Signs being executed within an utterance or interactions with other signers, including turn-taking, could influence these variations. Furthermore, the increased variability observed in the forearms and hands positions is expected due to the Signs being performed as isolated Lexical Units (Volterra *et al.,* 2022). As the signer’s eye gaze is directed towards the interlocutor (in this context, the camera), and with no notable body components accounted for in the annotations beyond labializations, the manual components stand out as the most prominent. Hence, they were chosen as the primary focus of the analysis presented in this study.

These insights highlight Typannot’s utility for post-analyzing annotations developed on ELAN, thus maximizing ELAN’s post-processing applications. Despite Signs being performed under controlled conditions, the recurring annotations from the 205 instances of the MultiMedaLIS_3 Dataset unveil valuable general characteristics. This practical example, although limited, was specifically included to emphasize how annotating sign language data using language-specific annotation systems can aid in defining Signs. These annotations allow to identify commonalities and distinctions among Signs. They aid in characterizing these Signs both in isolation and when compared with each other, providing valuable insights into their nature and nuances.

This analysis thus demonstrates the capability to surpass basic glossing tasks and explore Signs through sign language-specific transcription systems. Although time-consuming in their annotation, these systems enable us to analyze Signs independently, free from external influences or conditioning. Furthermore, it is worth mentioning that ongoing efforts are being made to advance Typannot transcription through the employment of motion capture systems. An example of this is the ANR-LexiKHuM project[106], where the sensor-free motion capture software AlphaPose[107] is employed for this purpose (Bianchini *et al.,* 2022).

6.5.7. Future developments in terms of annotation

The annotation guidelines for the MultiMedaLIS_3 Dataset were designed with the flexibility to accommodate future developments and potential improvements. While they primarily align with the COS-S framework, they also adapt to a computational annotation approach that enables diverse groups with varying backgrounds to work with sign languages. Despite this adaptability, it is important to recognize that as the annotated data expands, new annotation challenges may emerge, requiring careful consideration and the development of solutions.

In the case of the MultiMedaLIS_3 Dataset, where Signs were collected individually, there was no requirement to discuss the annotation of utterances or the codification of Sign transitions and their impact on the tokenization process. However, the possibility of introducing utterances in future collection will require the development of additional tiers to annotate and include a free translation in Italian and English for the entire utterance, a practice observed, for instance, in the *Corpus NGT* (Crasborn & Zwitserlood, 2008). Moreover, the introduction of new participants to future data collection tasks could have several implications. From the perspective of participant identification, it would necessitate: (1) the inclusion of metadata such as information

[106] https://anr.fr/Project-ANR-20-CE33-0012 (Accessed on February 15th, 2024).
[107] https://github.com/MVIG-SJTU/AlphaPose (Accessed on February 17th, 2024).

on the signer's gender, age, origin, age of acquisition of sign language, and deafness status. (2) The implementation of a unique code to identify each participant in the data collection, a practice previously proposed for the *Corpus LIS* by Santoro & Poletti (2011). Incorporating this information will align with FAIR principles and guidelines (Wilkinson *et al.*, 2016), as detailed in section 5.5.

Regarding tier development, the introduction of new participants engaging in spontaneous signing, potentially producing utterances, might require (1) the lemmatization of Signs and (2) the identification of Sign variants and their distinctions using standardized vocal language labels, possibly adopting the ID-Gloss convention employed in the corpora of the ECHO project (Crasborn *et al.,* 2007). In this regard, the issue of vocal language labels, or glosses, will always call for requires extensive discussions. To reach the goal of enabling anyone familiar with the sign language in question to access the original form of the Sign— even in the absence of the raw data, as recommended by Antinoro Pizzuto *et al*. (2010)— and to go a step further by making the information accessible to those who do not know the sign language, it will be crucial to continue using labels that unambiguously identify the Sign, or at, least its lemmatized form.

The method used in a previous project (Caligiore *et al.,* 2021), where Elena Radutzky's *Dizionario* (1992) was referenced, still holds potential, with necessary adaptations. The idea of using unambiguous labels that refer to a dictionary can be updated by relying on digital resources like SpreadTheSign or other similar platforms. This approach allows anyone accessing the annotation, even without the raw data, to trace back the original form of the Sign in its lemmatized form. This type of annotation layer is not entirely new, and it is already being applied through resources like SignBank[108]. SignBank, originally created for AUSLAN, has expanded to host repositories for

[108] https://auslan.org.au/ (Accessed on October 12th, 2023).

different sign languages[109] and can be connected to ELAN, enabling the display of videos for SignBank entries by hovering over a segment (Sloetjes, 2023). However, despite the availability of an online dictionary for LIS, SpreadTheSign is not integrated with SignBank and cannot be directly connected to ELAN, as this functionality is exclusive to SignBank-hosted data.

One potential solution to this limitation is to introduce SpreadTheSign video URLs into the annotation tiers. This possibility was explored in March 2022, by discussing it with ELAN developers through the Language Archive Forum, hosted by the Max Planck Institute for Psycholinguistics in Nijmegen. Developers have responded that preparations for this functionality, including the ability to include URLs in ELAN's .eaf format and partially in the source code, are in progress but not fully implemented yet[110]. Once fully implemented, this feature could potentially open up the possibility of using existing videos from online sign language dictionaries in ELAN, making this service accessible to all sign language dictionaries hosted on SpreadTheSign.

In summary, the annotation of the Dataset was created with flexibility in mind, linking the COS-S framework and a computational approach to sign language studies, and is designed to adapt to users with various backgrounds either within, close to, or outside of sign language research. However, the future may bring new insights, depending on how data collection efforts will evolve[111].

In conclusion, the development of this annotation methodology and the openness to potential Dataset expansions demonstrate its adaptability and a commitment to

[109] https://github.com/Signbank (Accessed on October 12th, 2023).
[110] https://archive.mpi.nl/forums/t/including-url-in-segment/3456 (Accessed on October 12th, 2023).
[111] Currently, the data collection process and annotations have received approval from both Deaf and hearing researchers specializing in sign languages and multimodality. During a research visit at the Max Planck Institute for Psycholinguistics in Nijmegen in October 2023, discussions with Deaf researchers focused on the data collection and processing methodologies employed for this thesis were held. These discussions yielded valuable insights for future annotation and collection efforts.

inclusivity and accessibility in the context of LIS collection and processing, as this approach aims to serve the diverse needs of researchers in the field.

Chapter 7
Insights and trajectories of this work: discussing implications, limits, and conclusions

The primary objective of this thesis, as stated in section 1.3., was to propose a new perspective on the collection and processing of sign language data, operating within a framework that recognizes LIS as a multilinear and multimodal sociocultural entity. In the context of linguistic studies, which have historically diverged into two main paths of formalist and functionalist approaches, the focus of this research predominantly lies within the functionalist COS-S framework and its application to LIS (Volterra *et al.,* 2022). At the same time, this work aimed to adapt the COS-S framework while simultaneously attempting to establish a cohesive balance between this functionalist approach and formal sign language studies. Specifically, it was pointed out that this study aimed to navigate the complex relationship between formalist and functionalist paradigms, seeking to harmonize the COS-S approach to LIS studies while, at the same time, addressing the practical necessity of using established resources and methodologies for collecting and processing LIS data. Thus, in pursuing this balance, the present study would have positioned itself within an intersegmental space, envisioning opposing framework in a perspective of mutual contribution, finding alignment with both formal and functionalist approaches. After reviewing the literature, the hypothesis that research on sign languages and the development of recognition and translation systems are often compartmentalized into autonomous and non-communicating areas of enquiry was confirmed (Bragg *et al.,* 2019; De Coster *et al.,* 2023). In establishing this peculiarity, this thesis aimed to surpass existing literature by not only advancing the hypothesis of a computational approach within the COS-S framework, but also by implementing it within an experience of interdisciplinary collaboration that led to the development of the MultiMedaLIS Datasets. To accomplish this objective, the study outlined four interrelated research questions:

RQ1. Is it possible to balance structure and variation in the collection, description and analysis of LIS?

RQ2. Can sign language resources be developed to facilitate interdisciplinary collaboration among researchers, providing both human-readable and machine-readable data through language-specific annotation methods? How can LIS-specific multimodal linguistic features be effectively represented during the annotation of LIS data without compromising the language's integrity for machine processing?

RQ3. Are there tools available for the collection of LIS data that can effectively capture its unique characteristics and multimodal linguistic features while ensuring accurate, high-quality, and comprehensive data collection for research and analysis purposes?

RQ4. Which guidelines should be established in selecting tools and developing annotation methodologies within a COS-S framework?

To practically address these questions, a theoretical groundwork was established serving as the foundation for collecting and processing three distinct MultiMedaLIS Datasets.

7.1. Discussion

For the purpose of this section, it is perhaps better to start answering the research questions starting from RQ2, and then conclude by reconnecting to RQ1. RQ2 touches on two points: (A) can sign language resources be developed to facilitate interdisciplinary collaboration among researchers, providing both human-readable and machine-readable data through language-specific annotation methods? (B) How can LIS-specific multimodal linguistic features be effectively represented during the annotation of LIS data without compromising the language's integrity for machine processing?

To practically address this inquiry, Chapter 6 was dedicated to describing three experimental data collection processes, while also providing a detailed discussion on the accompanying annotation processes employed during the experiments.

Point A of RQ2 was approached from various angles. Previous experiences in multimodal data collections emphasized the necessity of consistent and coherent labeling of data (Hodge & Crasborn, 2022). This characteristic holds particular relevance when the goal of an annotation task is to provide annotations that are both machine-readable and human-readable. Considering the oral nature of LIS and the requisite of Unicode-compatible or alphanumerical representation in computational sign language processing, a compromise was considered the most viable solution. In fact, the contemporary discussion surrounding computational annotation in any sign language highlights the necessity of some form of Unicode-compatible or alphanumerical representation. This need, as originally highlighted for LIS by Antinoro Pizzuto *et al.* (2008), remains constant. In addressing this significant concern, various specific factors pertinent to the conducted data collection project and its intended outcomes, had to be considered.

In fact, working within interdisciplinary research groups offers the advantage of integrating diverse perspectives. However, it also introduces constraints that stem from the distinct viewpoints of each discipline, necessitating mutual respect and consideration for these boundaries. In terms of annotation, aside from using alphabetic writing, it was also crucial to ensure the resource's accessibility for both signers and non-signers within the group. In this context, a new annotation methodology was developed to offer diverse labels for the annotated resource, ensuring human and machine readability. Simultaneously, this methodology adhered to the COS-S framework principles, accommodating detailed representations of language-specific features, encompassing manual and body components (Volterra *et al.,* 2022).

Given the need to develop multi-layered annotations, ELAN was chosen as the designated annotation software (Crasborn & Sloetjes, 2008). This aligns with previous efforts in the annotation sign language and multimodal resources (Nonhebel *et al.,* 2004a; Crasborn & Zwitserlood, 2008; Crasborn *et al.,* 2007; Santoro & Poletti, 2011; Mesch, 2012; Mesch & Wallin, 2015; Johnston, 2019; Lo Re, 2022). Another prevalent noted characteristic in literature was the use of vocal language labels, a feature that resonates with formalist approaches to describing LIS (Cecchetto *et al.,* 2011; Geraci *et al.,* 2011). The incorporation of vocal language labels was adopted as it served a dual purpose: facilitating accessibility for non-signers (both Italian and English speakers) and establishing a Sign-word equivalence to support the creation of machine-readable data. Additionally, the decision to code Units of Meaning (Volterra *et al.,* 2022) was made, considering potential future expansions of the dataset or adaptations of the annotation methodology encompassing Units beyond Lexical ones.

However, the focus of point A of RQ2 is the employment of language-specific annotation methods that are both machine *and* human readable. This objective was achieved by implementing the Typannot typographic system (Doan *et al.,* 2019; Danet *et al.,* 2021; Bianchini, 2023), specifically developed for this purpose. In employing the Typannot font in its *Formule Générique* format, manual and bodily components were codified by marking their initial and final positions, accounting for movement. Using ELAN's multilevel design and integrating vocal language labels and language-specific annotations significantly enhanced annotation productivity, establishing an annotation framework that fulfills the objective presented in point A of RQ2. Thus, this annotation framework accounts for variations in manual and body components while simultaneously facilitating statistical analysis. In doing so, it directly aligns with point B of RQ2, by enabling an effective representation of the specificity of LIS as a visual-manual language.

The third research question addressed (RQ3) was: are there tools available for the collection of LIS data that can effectively capture its unique characteristics and multimodal linguistic features while ensuring accurate, high-quality, and comprehensive data collection for research and analysis purposes? To answer this question, Chapter 3, 4 and 5 have drawn upon various studies in sign language and, more broadly, multimodal data collections (Geraci *et al.,* 2011; Mesch & Wallin, 2015; Johnston, 2019; Lo Re, 2022; Hanke & Fenlon, 2022). During this process, certain defining qualities of the tools used for multimodal data collection were identified. Primarily, the selection of recording tools appeared to have been based on their capacity to provide high resolution video data. In fact, high resolution ensures the suitability of captured data for subsequent processing and in-depth analyses. To this it can be added that an essential aspect in terms of recording quality also involves the integration of multimodal recording tools capable of capturing the linguistic features of LIS. The employment of these tools would accommodate the tridimensionality of LIS, facilitating clear depictions of body components including facial expressions, eye gaze, and oral component production (Volterra *et al.,* 2022).

Moreover, it was established that recording tools should prioritize portability and non-invasiveness to facilitate data capture across diverse contexts, while minimizing interference with the participant's environment (see sections 5.3. and 6.3.2). These standards not only enable data collection in varied settings but also ensure minimal disruption to the participant's surroundings. Furthermore, considering the goal of ongoing data collection in spontaneous settings, the advantages of portability and non-invasiveness become even more significant. These qualities enable signers to naturally express themselves, a central element in capturing the entire spectrum of semantic and pragmatic aspects in signed discourse.

To address RQ3, three experimental data collections were conducted, culminating in the creation of MultiMedaLIS_3 Dataset. This dataset resulted from the use of depth cameras and facial tracking, ensuring the capture of high-resolution alignable HD videos and depth data (see section 6.3.2). Consequently, it offers a detailed tridimensional representation of both body and manual components. Furthermore, the experimental integration of a RADAR, among the recording tools, enhanced adaptability, enabling the capture of subtle movements and facilitating data collection in low-light environments. The various outputs can be synchronized, allowing for the simultaneous visualization of all manual and body components. This synchronized approach is able to provide a comprehensive multimodal representation of LIS.

The fourth research question (RQ4) was: which guidelines should be established in selecting tools and developing annotation methodologies within a COS-S framework? The answer to this question can be presented thanks to the establishment of a theoretical background for carrying out the experimental LIS data collections presented in this thesis, as described in Chapter 1 and 2. Firstly, as mentioned in the answer to RQ3, the characteristics of portability, non-invasiveness and high resolution of the recorded tools are to be prioritized, ensuring data capture in minimally disruptive environments while maintaining high-resolution quality for subsequent analyses. The recognition of these qualities aligns with the COS-S model, particularly concerning ethnopragmatic theories discussed in Chapter 2. In fact, the portability and non-invasiveness of these data collection tools facilitate the capture of sign language data in natural, everyday contexts (Russo Cardona, 2004b), enabling their examination as cultural practices (Duranti, 1997). While the experimental data collections presented in this work laid the groundwork for identifying these recording tools (hence the preliminary nature of the Datasets), future iterations of data collection in spontaneous settings will allow to record signers and, consequently, study LIS in natural environments. Moreover, while not a direct consequence of the selected tools, the necessity to capture sign language through videos, derived from its visual-manual nature, allows for the documentation of the embodied experience of LIS signing (Lakoff & Johnson, 1980; Cuccio *et al.,* 2022). In relation to the second guideline, concerning the annotation process, currently, the most viable option for software is ELAN due to its widespread usage in sign language and multimodal research across various studies, like Crasborn *et al.* (2007), Santoro &

Poletti (2011), Hanke *et al.* (2020) and Lo Re (2022). ELAN serves as an excellent tool for capturing the multilinear and multimodal characteristics of LIS (Fontana & Roccaforte, 2015) due to the possibility of developing entirely customizable multi-level annotations.

Thirdly, consistency and clarity in data annotation, as emphasized by Antinoro Pizzuto *et al.* (2010) and Crasborn *et al.* (2015), are fundamental to facilitate efficient data retrieval and longevity. Additionally, multilingual vocal language labelling emerges as a central guideline, enabling accessibility for both signers and non-signers and fostering interdisciplinary (Crasborn & Zwitserlood, 2008; Mesch & Wallin, 2015). Adopting a hybrid approach of multilingual vocal language labels and language-specific annotation systems, recommended by Bianchini (2012; 2021) and Danet *et al.* (2021), ensures comprehensive yet precise annotations.

Lastly, the development of tiers where information is coded through language-specific systems is central. These guidelines, adopted for the development of the MultiMedaLIS_3 Dataset, not only address practical considerations, but also support the present project's dedication to producing a versatile, multimodal and accessible LIS resource for interdisciplinary research efforts, further addressing RQ2. In addition to employing language-specific methodologies, incorporating labels referencing the Unit of Meaning contributes to describing how signers use LIS to create meaning through embodied processes. Adhering to the scheme developed by Volterra *et al.* (2022), information related to the annotated Sign as a mechanism of pointing (using Deictic Units), telling (using Lexical Units), and showing (using Transfer Units) can be codified.

The answers presented for RQ2, RQ3 and RQ4 lead back to RQ1: is it possible to balance structure and variation in the collection, description and analysis of LIS? Revisiting RQ1 in light of the methodologies and tools discussed across this work and the answers provided to the other three RQs, offers a general vision the effort made to strike a delicate balance between structure and variation within the collection, description and analysis of LIS. The pursuit of this balance was central in the selection of recording tools, where the identification of high-resolution, portable, and non-invasive devices aimed not only for comprehensive data capture, but also for minimal interference, allowing signers to naturally express themselves across diverse contexts. The consideration given to these tools inherently acknowledges the need for structured *yet* adaptable approaches in capturing the multimodal and multilinear nature of LIS, guaranteeing a balance between the formalist framework that calls for the precise definition of Signs, and their embodied and social nature.

Moreover, the development of annotation methodologies, particularly through the integration of ELAN and the Typannot typographic system, stands as a further demonstration to this balance-seeking effort. The use of ELAN's multilevel design alongside language-specific annotations and vocal language labels characterizes an approach that aims at joining structured annotation frameworks with the inherent variability and complexity of LIS. By encoding manual and bodily components while accommodating for movement and language-specific features, the annotation framework attempts at striking a balance, enabling both machine and human readability without compromising the integrity of LIS.

Furthermore, the establishment of guidelines for tool selection and annotation methodologies, as presented in response to RQ4, highlights an effort to maintain this balance in further data collections. The emphasis on consistency, clarity, and the inclusion of language-specific systems within a COS-S framework not only facilitates efficient data retrieval and interdisciplinary collaboration, but also reinforces the commitment to structurally robust yet adaptive approaches in studying LIS.

Ultimately, the exploration and application of recording tools, annotation methodologies and guiding principles stand as concretizations of the possibility to navigate the delicate balance between structure and variation in the study of LIS. This pursuit of balance throughout the present work aligns with RQ1.

7.2. Implications of the present work

At its core, this thesis aims to reshape how LIS is collected, described, and analyzed by focusing on finding harmony between structure and variation in sign language studies. Four key research questions guide this work practically, seeking real-world solutions rather than purely theoretical reflections. The results obtained from this research carry some implications. Firstly, the study has attempted as describing LIS as a multifaceted cultural entity. By considering the complex interplay between different linguistic approaches, it has laid a proposal based on the COS-S theoretical framework for striking balance between formalist and functionalist approaches to sign languages studies in terms of sign language data collection and processing.

The developed methodologies, and the creation of the MultiMedaLIS_3 Dataset, mark a step forward in the collection of LIS through multimodal recording tools. Moreover, the annotation methodology proposed, designed to be both machine and human-readable, marks a practical attempt of applying the theoretical framework of the COS-S approach to data that is suitable for computational processing tasks. This practical application could improve sign language recognition systems through ongoing post-processing of the Dataset. Moreover, it establishes a model for collecting comprehensive multimodal datasets in different sign languages.

Lastly, the present study emphasizes the central role of collaborative interdisciplinary efforts. By merging insights from sign language linguistics and computational language processing, it showcases how this type of collaboration can produce original approaches in the domain of sign language studies. The outputs of this research thus stand as a testament to the fertile ground that collaborative interdisciplinary approaches offer in shaping new, impactful methodologies.

7.2. Limits

Regarding the limitations of this study, the main one resides in the authenticity of the captured data. In fact, the constraints imposed by formal recording settings influenced the capture of spontaneous signing. Confined to these controlled settings, the recorded data thus lacks spontaneity. Additionally, the MultiMedaLIS_3 Dataset is comprised of isolated Signs. The absence of exploring LIS utterances within this study's scope does not reflect a disregard for their importance. Rather, it stems from limitations imposed by the recording tools used during the third data collection. These collections were considered preliminary, requiring testing to confirm the reliability and multimodal collection capability of the tools. Consequently, tools were initially tested on isolated signing and within controlled environmental contexts, prioritizing the confirmation of their adequacy.

Recognizing these limitations serves to incentivize future developments which, in particular, could aim at surpassing the mentioned constraints and conduct data collections that encompass signed utterances by Deaf signers within authentic contexts. Carrying out such a project will also entail a consideration of FAIR principles and guidelines (Wilkinson *et al.*, 2016) and their applications to multimodal data collection.

7.3. Conclusion

This thesis aimed to combine diverse research approaches and apply them to LIS collection and processing. In doing so, it demonstrated that achieving a balance between formalist and functionalist theories in this context is achievable, without having to compromise the multimodal representation of LIS in its embodied and multilinear essence. This theoretical formulation finds practical success in the creation of three preliminary multimodal LIS Datasets within the medical domain, the MultiMedaLIS Datasets.

The outcomes of this thesis highlight the practicability of conducting sign language collection tasks using specialized multimodal tools. Furthermore, they highlight the possibility of employing annotation methodologies on collected data, resulting in sets of raw data accompanied by vocal language labels and language-specific annotation systems.

References

Abbou, V. (2017). *Une clé sur le monde*. Riom: Eyes Editions.

Adamou, E. (2019). Corpus linguistic methods. In J. Darquennes, J. C. Salmons, & W. Vandenbussche (Ed.), *Language contact: An International Handbook*, (pp.638-653). Berlin: De Gruyter.

Albano Leoni, F. (2007). *Il corpus CLIPS.* Accessed on August 12, 2023, http://www.clips.unina.it/it/documenti/presentazione_clips.pdf

Alumohamadi, Y., Sepandi, M., Taghdir, M., & Hisamirudsari, H. (2020). Determine the most common clinical symptoms in COVID-19 patients: a systematic review and meta-analysis. *Journal of Preventive Medicine and Hygiene, 61*(3).

Antinoro Pizzuto, E., Chiari, I., & Rossini, P. (2008). The representation Issue and its Multifaceted Aspects in Constructing Sign Language Corpora: Questions, Answers, Further Problems. *Proceedings of the Third Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora* (LREC 2008) (pp. 150- 158). ELRA.

Antinoro Pizzuto, E., Chiari, I., & Rossini, P. (2010). Representing Signed Languages: Theoretica, Methodological and Practical Issues. In M. Pettorino, A. Giannini, I. Chiari, & F. M. Dovetto (Ed.), *Spoken Communication* (pp. 205- 240). Cambridge Scholars Publishing.

Antinoro Pizzuto, E., Rossini, P., Sallandre, M.-A., & Wilkinson, E. (2006). Deixis, Anaphora and Highly Iconic Structures: Cross-linguistic Evidence on American (ASL), French (LSF) and Italian (LIS) Signed Languages. In R. M. De Quadros (Ed.), *Sign Languages: spinning and unraveling the past, present and future. TISLR9, forty five papers and three posters from the ninth Theoretical Issues in Sign Language Research Conference* (pp. 476- 497). Editora Azara Azul.

Arbib, M. (2000). The Mirror System, Imitation, and the Evolution of Language. In C. Nehaniv & K. Dautenhahn (Ed.), *Imitation in Animals and Artifacts*. Cambridge: The MIT Press

Arbib, M. (2003). From monkey-like Action Recognition to Human Language: an Evolutionary Framework for Neurolinguistics. *Behavioral and Brain Sciences, 20*, 105-124.

Ardita, G., & Caligiore, G. (2022) *Annotare la LIS all'Università. Strategie degli apprendenti, criticità didattiche e implicazioni metodologiche* [Conference presentation]. Insegnare la LIS e la LISt all'Università. Venice.

Armstrong, D. F., Stokoe, W. C., & Wilcox, S. E. (1995). *Gesture and the Nature of Language* (1996 Edition). Cambridge- New York: Cambridge University Press.

Assunção, C., & Araújo, C. S. (2019). Entries on the History of Corpus Linguistics. *Linha D Água , 32*(1), 39-57.

Austin, J. L. (1962). *How to do things with words. The William James Lectures delivered at harvard University in 1966.* London: Oxford University Press.

Badh, G., & Knowles, T. (2023). Acoustic and perceptual impact of face masks on speech: A scoping review. *PLoS One, 18*(8).
Barbera, E. F. (2013). *Linguistica dei corpora e linguistica dei corpora italiana. Un'introduzione.* Milan: Qu.A.S.A.R.
Bates, E. (1979). *The Emergence of Symbols. Cognition and Communication in Infancy.* New York- San Francisco- London: Academic Press.
Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merril Palmer Quarterly, 21*, 206-226.
Bates, E., Thal, D., MacWhinney, B. (1991). A Functionalist Approach to Language and its Implications for Assessment and Intervention. In T. M. Gallagher (Ed.) *Pragmatics of Language,* 133-161. Berlin: Springer.
Bavelier, D., & Neville, H. J. (2000). Cross-Modal Plasticity: Where and How? *Nature Reviews Neuroscience, 3*(6), pp. 443- 452.
Bébian, R. A. A. (1825). *Mimographie, ou Essadi d'Écriture Mimique Propre a Régulariser le Langage des Sourds-Mutes.* Paris: Ches Louis Colas.
Benetti, S., van Ackeren, M. J., Rabini, G., Zonca, J., Foa, V., Baruffaldi, F., Rezk, M., Pavani, F., Rossion, B., & Collignon, O. (2017). Functional selectivity for face processing in the temporal voice area of early deaf individuals. *Proceedings of the National Academy of Sciences, 114*(31).
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)* [Ph.D. thesis, University of Paris 8, University of Perugia]. Paris, Perugia.
Bianchini, C. S. (2016). Regard sur la nature de SignWriting (SW), un système pour représenter les Langues des Signes (LS). Dossiers d'HEL [*Histoire Épistémologie Langage], 9*, 404- 421.
Bianchini, C. S. (2021). Représenter les langues des signes sous forme écrite: questions qui ont besoin (encore aujourd'hui) d'être posées. *Corela [En ligne] 19*(2).
Bianchini, C. S. (2023). *(D)écrire les Langues des Signes. Approche grapho linguistique des Langues des Signes.* Brest: Fluxus.
Bianchini, C. S., Boutet, D., Contesse, A., Danet, C., Doan, P., Goguely, T., & Rébulard, M. (2017). Écrire les langues des signes : inscription et transcription. *8ème Colloque "Printemps de la typographie 2017: supporter l'écriture".* Ècole Estienne.
Bianchini, C. S., Danet, C., Chevrefils, L., Catteau, F., Chloé, T., Rébulard, M., Contesse, A., Doan, P., Champion, E. (2022). Transcribing and comparing sign languages and co-speech gestures: the contribution of the Typannot typographic transcription system to the study of epistemicity [Poster presentation] *Sign Café 2 Conference*, Ragusa.
Borgia, F. (2015) *Informatisation de forme graphique des Langues des Signes: application à l'écriture de SignWriting.* [Ph.D. Thesis, Université de Toulouse 3 Paul Sabatier, Sapienza University of Rome]. Toulouse, Rome.

Borgia F., Bianchini, C. S., Dalle, P., De Marsico, M. (2012). *Resource production of written forms of Sign Languages by a user-centered editor, SWift (SignWriting improved fast transcriber),* in Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012) (pp. 3779-3784). ELRA.

Borghi, A. M., Capirci, O., Gianfreda, G., & Volterra, V. (2014). The body and the fading away of abstract concepts and words: a sign language analysis. *Frontiers in Psychology, 5.*

Boutet, D. (2018). *Pour une approche kinésiologique de la gestualité Synthèse.* [Habilitation à Diriger des Recherches, Université de Rouen-Normandie]. *HAL Open Science*.

Boutet, D., Bianchini, C., Doan, P., Goguely, T., Rébulard, M., & Danet, C. (2016). Typannot: a glyphic system for the transcription of handshapes. *Proceedings of the Seventh Conference of the International Society of Gestures Studies "Gesture-Creativity-Multimodality"*, (p. 139). Paris.

Bragg, D., Caselli, N., Hochgesang, J. A., Huenerfauth, M., Katz-Hernandez, L., Koller, O., Kushalnagar, R., Vogler, C., Ladner, R. E. (2021). The FATE Landscape of Sign Language AI Datasets: An Interdisciplinary Perspective. *ACM Transactions on Accessible Computing, 14*(2), 1- 45.

Bragg, D., Koeller, O., Bellard, M., Berke, L., Boudrealt, P., Braffort, A., Caselli, N., Huenerfauth, M., Kacorri, H., Verhoef, T., Vogler, C., Ringel Morris, M. (2019). Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective. *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 1- 16). Pittsburgh: Association for Computing Machinery.

Branchini, C., Cecchetto, C., & Chiari, I. (2013). La lingua dei segni italiana. In G. Iannaccaro (Ed.), *La linguistica italiana all'alba del terzo millennio (1997-2010)* (pp. 369- 404). Rome: Bulzoni Editore.

Branchini, C., & Geraci, C. (2011). L'ordine dei costituenti in LIS. In A. Cardinaletti, C. Cecchetto, & C. Donati (Ed.), *Grammatica, Lessico e Dimensioni di Variazione nella LIS* (pp. 113- 126). Milan: FrancoAngeli.

Branchini, C., & Mantovan, L. (2021). *A Grammar of Italian Sign Language (LIS).* Venice: Edizioni Ca' Foscari.

Brentari, D. (2010). *Sign Languages.* Cambridge: Cambridge University Press.

Buccino, G., & Mezzadri, M. (2013). La teoria dell'embodiment e il processo di apprendimento e insegnamento di una lingua. *Enthymema, VIII*, 5- 20.

Busatta, L. (2022). The legal recognition of sign languages in an intersectional perspective. *Comparative Law and Language, 1*(1), 74- 87. Trento: University of Trento.

Butera, C., & Aziz-Zadeh, L. (2022). Mirror Neurons and Social Implications for the Classroom. In S. L. Macrine & J. M. B. Fugate (Ed.), *Movement Matters: How Embodied Cognition Informs Teaching and Learning,* (pp. 261- 273). Cambridge, Massachusetts: The MIT Press.

Caligiore, G. (2020). *Universal Dependencies for Italian Sign Language: a Treebank From the Storytelling Domain* [Master's thesis: University of Turin]. Turin.

Caligiore, G. (2023*). Italian Sign Language: Searching for Common Ground Between the Sociosemiotic Perspective and Computational Annotation* [Conference presentation]. Sign Café 2. Ragusa.

Caligiore, G., Bosco, C., & Mazzei, A. (2021). Building a Treebank in Universal Dependencies for Italian Sign Language. *Proceedings of the Seventh Italian Conference on Computational Linguistics* (pp. 1- 6). CEUR.

Camgöz, N. C., Hadfield, S., Koeller, O., Ney, H., & Bowden, R. (2018). Neural Sign Language Translation. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 7784- 7793). Salt Lake City.

Camgöz, N. C., Koller, O., Hadfield, S., & Bowden, R. (2020). Sign Language Transformers: Joint End-to-end Sign Language Recognition and Translation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 10023- 10033). Seattle.

Campisi, E., & Özyürek, A. (2013). Iconicity as a communicative strategy: Recipient design in multimodal demonstrations for adults and children. *Journal of Pragmatics, 47*(1), 14- 27.

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2021). OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *EEE Transactions on Pattern Analysis and Machine Intelligence, 43*(1), 172- 186.

Capirci, O., Iverson, J. M., Antinoro Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language, 23*(03), 645- 673.

Cardinaletti, A. (2016). Il progetto Spread The Sign. *Blityri Studi di storia delle idee sui segni e le lingue, 1-2*, 175- 182.

Cardinaletti, A., & Mantovan, L. (2022). Le Lingue dei Segni nel "Volume Complementare" e l'Insegnamento della LIS nelle Università Italiane. *Italiano Lingua Seconda. Rivista internazionale di linguistica italiana e educazione linguistica, 14*(2), 113-128.

Cascella, M., Rajnik, M., Aleem, A., Dulebohn, S. C., & Di Napoli, R. (2023). *Features, Evaluation, and Treatment of Coronavirus (COVID-19).* Treasure Island: StatPearls.

Cecchetto, C., Giudice, S., & Mereghetti, E. (2011). La raccolta del Corpus LIS. In A. Cardinaletti, C. Cecchetto, & C. Donati (Ed.), *Grammatica, Lessico e Dimensioni di Variazione della LIS* (pp. 55- 68). Milan: FrancoAngeli.

Celo, P. (2015). *La scrittura come traduzione; proposta di un metodo per l'apprendimento della lettura e della scrittura in bambini sordi segnanti* [Ph.D. thesis, University of Milano-Bicocca]. Milan.

Charlton, J. I. (1998). *Nothing about us without us: disability oppression and empowerment.* Berkley: University of California Press.

Chesi, C., & Geraci, C. (2009). *Segni al computer. Manuale di documentazione della lingua italiana dei segni e alcune applicazioni computazionali* (2010 Edition). Siena: Cantagalli.

Chevrefils, L. (2022). *Formalisation et modélisation du mouvement en Langue des Signes Française: pour une approche kinésio-linguistique des productions gestuelles* [Ph.D. thesis, University of Rouen-Normandie]. Rouen.

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

Chomsky, N. (1959). A Review of B. F. Skinner's Verbal Behavior. *Language*, *35*(1), 26- 58. Retrieved from https://web.archive.org/web/20150929070654/http://www.chomsky.info/articles/1967----.htm

Chomsky, N. (1965). *Aspects of the Theory of Syntax.* Cambridge, Massachussets: The MIT Press.

Chomsky, N. (1970). Remarks on Nominalization. In R. Jacobs, & P. S. Rosenbaum (Ed.), *Readings in English Transformational Grammar* (pp. 184- 221). Waltham: Ginn and Company.

Chomsky, N. (1975). *Reflections on Language.* New York: Pantheon Books.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin and Use*. New York: Praeger Publishers.

Chomsky, N. (2002a). *On Nature and Language.* (A. Belletti, & L. Rizzi, Ed.) Cambridge: Cambridge University Press.

Chomsky, N. (2002b). *Syntactic Structures* (2nd Edition). Berlin- New York: De Gruyter.

Chomsky, N., & Halle, M. (1968). *The Sound Pattern of English.* New York: Harper & Row.

Church, K., & Liberman, M. (2021). The Future of Computational Linguistics: On Beyond Alchemy. In S. Nirenburg (Ed.), *Frontiers in Artificial Intelligence, 4*.

Cignarella, A. T., Bosco, C., & Rosso, P. (2019). Presenting Twittiro-UD: An Italian Twitter Treebank in Universal Dependencies. *Proceedings of the Fifth International Conference on Dependency Linguistics* (pp. 190- 197).

Conte, G., Santoro, M., Geraci, C., & Cardinaletti, A. (2011). Perché alzi le sopracciglia? Le funzioni linguistiche marcate del sollevamento in LIS. In A. Cardinaletti, C. Cecchetto, & C. Donati (Ed.), *Grammatica, Lessico e Dimensioni di Variazione della LIS* (pp. 161- 170). Milan: FrancoAngeli.

Conte, S. R. (2018). Nomi di persona e di luogo nella comunità sorda in Italia: interviste, analisi e primi risultati [Poster presentation] *IV Convegno Nazionale LIS "La Lingua dei Segni Italiana: una risorsa per il futuro",* Rome.

Cormier, K., & Fenlon, J. (2014). BSL Corpus Annotation Guidelines. London: Deafness Cognition and Language (DCAL) Research Centre of the University College London.

Cormier, Kearsy, Crasborn, O., & Bank, R. (2016). Digging into Signs: Emerging Annotation Standards for Sign Language Corpora. *Proceedings of the Seventh Workshop on Representation and Processing of Sign Languages: Corpus Mining.* ELRA.

Crasborn, O. (2015). Transcription and notation methods. In E. Orfanidou, B. Woll, & G. Morgan (Ed.), *Research methods in sign language studies: A practical guide* (pp. 74- 88). New York: John Wiley & Sons.

Crasborn, O., & Bank, R. (2014). *An annotation scheme for the linguistic study of mouth actions in sign languages.* Nijmegen: Radboud University.

Crasborn, O., & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. *Proceedings of the Third Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora* (LREC 2008). ELRA.

Crasborn, O., & Zwitserlood, I. (2008). The Corpus NGT: an online corpus for professionals and laymen. *Proceedings of the Third Workshop on the Representation and Processing of Signed Languages: Construction and Exploitation of Sign Language Corpora* (pp. 44- 49). Paris: ELDA.

Crasborn, O., Bank, R., & Cormier, K. (2015). *Digging into Signs: Towards a gloss annotation standard for sign language corpora.* Radboud University (CLS) & University College London (DCAL).

Crasborn, O., Mesch, J., Waters, D., Nonhebel, A., van der Kooij, E., Woll, B., & Bergman, B. (2007). Sharing sign language data online: Experiences from the ECHO project. *International Journal of Corpus Linguistics, 12*(4), 535- 562.

Crawford, W. J., & Eniko, C. (2016). *Doing Corpus Linguistics.* New York: Routledge.

Cresti, E., Gregori, L., Moneglia, M., & Panunzi, A. (2018). The Language into Act Theory: A Pragmatic Approach to Speech in Real-Life. *Proceedings of the LREC 2018 Workshop "LB-ILR2018 and MMC2018 Joint Workshop"* (pp. 30- 35). ELRA.

Croft, W., & Cruse, A. (2004). *Cognitive Linguistics.* Cambridge: Cambridge University press.

Cuccio, V. (2015). Embodied Simulation and Metaphors. On the Role of the Body in the Interpretation of Bodily-based Metaphors. *Epistemiologia , XXXVIII*, 97- 112.

Cuccio, V., Di Stasio, G., & Fontana, S. (2022). On the Embodiment of Negation in Italian Sign Language: An Approach Based on Multiple Representation Theories. *Frontiers in Psychology, 13*.

Cukor-Avila, P. (2000). Revisiting the Observer's Paradox. *American Speech, 75*(3), 253- 254.

Cutugno, F., & Campisi, E. (2022). Linguistica applicata alle esperienze di cultura digitale. Analisi multimodale nel progetto CHROME. In J. Saturno, & L. Spreafico (Ed.), *Fare Linguistica Applicata con le Digital Humanities* (pp. 9- 25). Milan: AltLa.

Cuxac, C. (2000). La Langue des Signes Française (LSF) — Les voies de l'iconicité. *Collection Faits de Langues n° 15-16*(98), 56- 57.

Cuxac, C., & Antinoro Pizzuto, E. (2010). Émergence, norme et variation dans les langues des signes: vers une redéfinition notionnelle. *Langage et Société* (131), 37- 53.

Cuxac, C., & Sallandre, M.A. (2002). Iconicity in Sign Language: a theoretical and methodological point of view. *Lecture notes in computer science*, 173- 180.

Cuxac, C., & Sallandre, M.A. (2007). Iconicity and arbitrariness in French Sign Language: Highly iconic structures, degenerated iconicity and diagrammatic iconicity. *Empirical Approaches to Language Typology, 36*(12).

Danet, C., Boutet, D., Doan, P. S., Contesse, A., Chèvrefils, L., Rébulard, M., Thomas, C., Dauphin, J. F. (2021). Transcribing Sign Languages With TYPANNOT: The Typographic System That Retains and Displays Layers of Information. *Proceedings of Grapholinguistics in the 21st Century 2020, 5*, 1009- 1037.

Danet, C., Thomas, C., Contesse, A., Rébulard, M., Bianchini, C., Chevrefils, L., & Doan, P. (2022). Applying the Transcription System Typannot to Mouth Gestures. *Proceedings of the Tenth Workshop on the Representation and Processing of Sign Languages* (sign-lang@LREC 2022) (pp. 42- 47). ELRA.

De Coster, M., Shterionov, D., van Herreweghe, M., & Dambre, J. (2023). Machine translation from signed to spoken languages: state of the art and challenges. *Universal Access in the Information Society, 22*(3).

de Marneffe, M., & Nivre, J. (2019). Dependency Grammar. *Annual Review of Linguistics, 5*, 197- 218.

de Marneffe, M. C., Dozat, T., Silveira, N., Haverinen, K., Ginter, F., Nivre, J., & Manning, C. D. (2014). Universal Stanford Dependencies: A cross-linguistic typology. *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC 2014) (pp. 4585- 4592). ELRA.

De Mauro, T. (1991). Ancora Saussure e la semantica. In L. J. Prieto (Ed.) *Cahiers Ferdinand de Saussure, 45*, 101-109. Genève: Librairie Droz.

De Mauro, T. (2005). *Scritti inediti di linguistica generale.* Rome: Editori Laterza.

De Mauro, T. (2016, december 23). *Il Nuovo vocabolario di base della lingua italiana* Internazionale.https://www.dropbox.com/s/mkcyo53m15ktbnp/nuovovocabolariodibase.pdf?dl=0

De Mauro, T. (2019). *Linguistica Elementare*. Rome- Bari: Editori Laterza.

de Saussure, F. (2005). *Corso di linguistica generale.* (T. De Mauro, Ed.) Rome: Editori Laterza. (Original work published 1916).

de Saussure, F. (2011). *Course in General Linguistics.* (T. B. Baskin, Ed.) New York: Columbia University Press. (Original work published 1916).

Di Renzo, A., & Slonimska, A. (2018). L'uso delle Strutture di Grande Iconicità nei testi narrativi segnati: primi dati su bambini prescolari, scolari e adulti [Poster presentation] *IV Convegno Nazionale LIS "La Lingua dei Segni Italiana: una risorsa per il futuro",* Rome.

Di Renzo, A., Lamano, L., Lucioli, T., Pennacchi, B., & Ponzo, L. (2006). Italian Sign Language (LIS): can we write it and transcribe it with SignWriting? *Proceedings of the Second Workshop on the Representation and Processing of Sign Languages "Lexicografic matters and didactic scenarios"* (LREC 2006) - *Fifth International Conference on Language resources and evaluation* (pp. 1-6). ELRA.

Di Renzo, A., Lamano, L., Lucioli, T., Pennacchi, B., Gianfreda, G., Petitta, G., Bianchini, C., Rossini, P., Antinoro Pizzuto, E. (2011). *Scrivere la LIS con il Sign Writing. Mauale introduttivo. Adattamento alla LIS del manuale di Valerie Sutton "Lessons in Sign Writing".* Rome: ISTC-CNR.

Doan, P., Boutet, D., Contesse, A., Bianchini, C., Danet, C., Rébulard, M., Dauphin, J. F., Chevrefils, L., Thomas, C., Reguer, M. (2019). Handling Sign Language handshapes annotation with the typannot typefont. *CogniTextes, 19, 1- 24*.

Donges, N. (2021). *Introduction to NLP*. Builtin. https://builtin.com/data-science/introduction-nlp

Duranti, A. (1997). *Linguistic Anthropology.* New York: Cambridge University Press.

Duranti, A. (2007). *Etnopragmatica. La forza nel parlare.* Roma: Carocci.

Elmetti, V. (2010). *Insegnare e apprendere la LIS come lingua straniera.* [Bachelor's thesis, University Ca' Foscary of Venice]. Venice.

Emmerik, W., Nonhebel, A., van der Kooij, E., Mesch, J., van Kampen, A., Crasborn, O., Sutton-Spence, R., Waters, D., Hiddinga, A., British Broadcasting Corporation (BBC), Leendert, P. (2003 - 2005). *ECHO- Sign Language* [Data set]. The Language Archive. https://archive.mpi.nl/tla/islandora/object/tla%3A1839_00_0000_0000_0001_494E_3

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429- 492.

Everson, M., Hosken, M., Slevinski, S., & Sutton, V. (2012). *Proposal for encoding Sutton SignWriting in the UCS.* https://www.unicode.org/L2/L2012/12321-n4342-signwriting.pdf

Fawcett, H. H. (1961). Speech transmission through respiratory protective devices. *American Industrial Hygiene Association Journal, 22*, 170- 174.

Fenlon, J., & Hochgesang, J. A. (Ed.). (2022). *Signed Language Corpora.* Washington, D.C.: Gallaudet University Press.

Ferrari, P. F, & Coudé, G. (2018). Mirror Neurons, Embodied Emotions, and Empathy. In K. Z. Meyza & E. Knapska (Ed.), *Neuronal Correlates of Empathy. From Rodent to Human,* (pp. 67- 77). Cambridge, Massachusetts: Academic Press.

Finnish Association of the Deaf. (2015). Working Together Manual for Sign Language Work within Development Cooperation. Finnish Association of the Deaf, World Federation of the Deaf, Disability Partnership Finland, Ministry of Foreign Affairs of Finland.

https://www.slwmanual.info/PDF/Working%20Together%20-%20PDF%20manual.pdf
Fodor, J. A. (1983*). The Modularity of Mind*. Cambridge, Massachusetts: The MIT Press.
Fogassi, L. (2008). Neuroni specchio e funzioni cognitive motorie. Mirror neurons and cognitive and motoric functions. *Giornale di Neuropsichiatria dell'Età Evolutiva, 28,* 329- 350.
Fontana, S. (2008). Mouth Actions as Gesture in Sign Language. *Gesture, 8*(1), 104-123.
Fontana, S. (2009). *Linguaggio e Multimodalità. Gestualità e oralità nelle lingue vocali e nelle lingue dei segni*. Pisa: Edizioni ETS.
Fontana, S. (2013). *Tradurre lingue dei segni. Un'analisi multidimensionale.* Modena: Mucchi.
Fontana, S. (2014). Traduzione e traducibilità tra lingue dei segni e lingue vocali. *Rivista Italiana di Filosofia del Linguaggio, 3*, 102- 111.
Fontana, S. (2022a). Grammar and Experience: The Interplay Between Language Awareness and Attitude in Italian Sign Language (LIS). *International Journal of Linguistics, 14*(5), 1- 18.
Fontana, S., & Caligiore, G. (2021). Italian Sign Language (LIS) and Natural Language Processing: an Overview. In P. Basile, V. Basile, D. Croce & E. Cabrio (Ed.), *Proceedings of the 4th Workshop on Natural Language for Artificial Intelligence (NL4AI 2020)*. CEUR Workshop Proceedings.
Fontana, S., Corazza, S., Boyes-Braem, P., & Volterra, V. (2015). Language research and language community change: Italian Sign Language (LIS) 1981-2013. *International Journal of the Sociology of Language, 236*.
Fontana, S., & Cuccio, V. (2013). Metafora e metonimia tra linguistica cognitiva e teoria della pertinenza: uno studio sulle lingue dei segni primarie. *Rivista di Psicolinguistica Applicata/Journal of Applied Psycholinguistics, XIII*(1), 83-112.
Fontana, S., & Mignosi, E. (2012). *Segnare, Parlare, Intendersi. Modalità e Forme.* Milan: Mimesis.
Fontana, S., & Raniolo, E. (2015). Interazioni tra oralità e unità segniche: uno studio sulle labializzazioni nella Lingua dei Segni Italiana (LIS). In G. Schneider, M. Janner, & B. Élie (Ed.), *Proceedings of the VII Dies Romanicus Turicensis* (pp. 241- 258). Bern: Peter Lang.
Fontana, S., & Raniolo, E. (2020). I toponimi in LIS (Lingua dei Segni Italiana) quando la comunità Sorda denomina la Sicilia. *Rivista Italiana di Onomastica, 26*(1), 230- 253.
Fontana, S., & Roccaforte, M. (2015). Lo strutturarsi e il destrutturarsi dei suoni nell'interazionecon la Lingua dei Segni Italiana (LIS). In M. Vayra, C. Avesani, & F. Tamburini (Ed.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio* (pp. 371-382). Milan: ASIV.

Fontana, S., & Roccaforte, M. (2019). Oltre l'approccio assimilazionista nella descrizione LIS: quando la prassi comunicativa diventa norma. In B. Moretti, A. Kunz, S. Natale, & K. Etna (Ed.), *LII Congresso Internazionale di Studi della Società di Linguistica Italiana* (pp. 273- 286). Milan: Officinaventuno.

Fontana, S., Tomasuolo, E., & Gianfreda, G. (2022). Naming in Two Modalities: Manual and Mouth Gestures in Hearing and Deaf Children. *International Journal of Linguistics, 14*(4), 120- 140.

Fornasiero, E. (2018). La morfologia valutativa in LIS: una descrizione preliminare [Poster presentation] *IV Convegno Nazionale LIS "La Lingua dei Segni Italiana: una risorsa per il futuro"*, Rome.

Fowley, F., & Ventresque, A. (2022). Sign Language Fingerspelling Recognition using Synthetic Data. *Irish Conference on Artificial intelligence and Cognitive Science (AICS2021)*. Dublin.

Francis, W., & Kucera, H. (1964). *Brown Corpus Manual of Information* (1979 Edition). Providence: Brown University. http://icame.uib.no/brown/bcm.html

Friston, K. (2018). *Embodied Cognition* [Video]. Serious Science. http://serious-science.org/embodied-cognition-9027

Gallaudet, E. M. (1881). The Milan Convention. *American Annals of the Deaf and Dumb, 26*(1), 1- 16.

Gallaudet, T. H. (1848). On the Natural language of Signs and its Value and Uses in the instrubtion of the Deaf and Dumb. *American Annals of the Deaf and Dumb, 1*(2), 79- 93.

Gallese, V. (2005). Embodied simulation: from neurons to phenomenal experience. *Phenomenology and the Cognitive Sciences, 4,* 23- 48.

Gallese, V. (2007). Before and below theory of mind: Embodied simulation and the neural correlates of social cognition. *Proceedings of the Royal Society in Biological Sciences, 362,* 659- 669.

Gallese, V. (2008). Mirror neurons and the social nature of language: The neural exploitation hypothesis. *Social Neuroscience, 3,* 317- 333.

Gallese, V., Fadiga, L., Fogassi, L., Rizzolati, G. (1996). Action recognition in the premotor cortex. *Brain, 119*(2), 593- 609.

Gallese, V., Freedberg, D., Umiltà. M. A. (2022). Embodiment and the Aesthetic Experience of Images. In A. Chatterjee & E. R. Cardillo (Ed.), *Brain, Beauty, & Art* (pp. 88- 92). New York: Oxford University Press.

Garcia, B. (2010). *Sourds, surdité, Langue(s) des Signes et épistémiologie des sciences du langage : problématiques de la scripturisation et modélisation des bas niveaux en Langue des Signe Française (LSF).* [Habilitation à Diriger des Recherches, Université Paris 8 Vincennes Saint-Denis]. HAL Open Science.

Geraci, C. (2004). L'ordine delle parole nella LIS. In F. A. Leoni, F. Cutugno, M. Pettorino, & R. Savy (Ed.), *Atti del Convegno nazionale della Società di Linguistica Italiana.* Naples: D'Auria Editore.

Geraci, C., & Bayley, R. (2011). Chi, cosa, dove, perché, quando: la distribuzione dei segni wh- in LIS. In A. Cardinaletti, C. Cecchetto, & C. Donati (Ed.), *Grammatica, Lessico e Dimensioni di Variazione della LIS*, (pp. 127- 144). Milan: FrancoAngeli.

Geraci, C., Battaglia, K., Cardinaletti, A., Cecchetto, C., Donati, C., Giudice, S., & Mereghetti, E. (2011). The LIS Corpus Project. *Sign Language Studies, 11*(4), 528- 571.

Geraci, C., & Mazzei, A. (2014). Last train to "Rebaudengo Fossano": the case of some names in avatar translation. *6th Workshop on the Representation and Processing of Sign Languages: Beyond the Manual Channel* (pp. 63- 66). Reykjavik: ELRA.

Geraci, C., Mazzei, A., & Angster, M. (2014). Some issues on Italian to LIS automatic translation. The case of train announcements. In R. Basili, A. Lenci, & Bernardo Magnini (ED), *The First Italian Conference on Computational Linguistics* (pp. 190- 196). Pisa: Pisa University Press.

Gianfreda, G. (2011). Un corpus di conversazioni in lingua dei segni italiana attraverso videochat: una proposta per la loro trasceizione e analisi. In A. Cardinaletti, C. Cecchetto, & C. Donati (Ed.), *Grammatica, Lessico e Dimensioni di Variazione della LIS* (pp. 95- 109). Milan: FrancoAngeli.

Gianfreda, G., Gulli, T., Lamano, L., & Volterra, V. (2021). Coronavirus, quarantena e positività nella Lingua dei Segni Italiana (LIS). Ripensare i segni in occasione della pandemia. *Rivista di Psicolinguistica Applicata, XXI*(1), 61- 82.

Giovanelli, E., Gianfreda, G., Gessa, E., Valzolgher, C., Lamano, L., Lucioli, T., Tomasuolo, E., Rinaldi, P., Pavani, F. (2023). The effect of face masks on sign language comprehension: performance and metacognitive dimensions. *Consciousness and Cognition, 109*.

Giovanelli, E., Valzolgher, C., Gessa, E., Todeschini, M., & Pavani, F. (2021). Unmasking the Difficulty of Listening to Talkers With Masks: lessons from the COVID-19 pandemic. *Iperception, 12*(2).

Goddard, C. (2006). *Ethnopragmatics: Understanding discourse in cultural context*. Berlin: Mouton de Gruyter.

Gurbuz, S. Z., Gurbuz, A. C., Malaia, E. A., Griffin, D. J., Crawford, C., Rahman, M. M., Aksu, R., Kurtoglu, E., Mdrafi, R., Anbuselvam, A., Macks, T., Ozcelik, E. (2020). A Linguistic Perspective on Radar Micro-Doppler Analysis of American Sign Language. *2020 IEEE International Radar Conference (RADAR)* (pp. 232- 237). Florence: IEEE.

Gutierrez-Sigut, E., Lamarche, V. M., Rowley, K., Ferreiro Lago, E., Pardo-Guijarro, J. M., Saenz, I., Frigola, B., Aliaga, D., Goldberg, L. (2022). How do face masks impact communication amongst deaf/HoH people? *Cognitive Research: Principles and Implications, 7*(81).

Hanke, T. (2004). HamNoSys—Representing sign language data in language resources and language processing contexts. *Proceedings of the Workshop on the Representation and processing of Sign Languages. Fourth International Conference on Language Resources and Evaluation* (LREC 2004). ELRA.

Hanke, T., & Fenlon, J. (2022). Creating Corpora: Data Collection. In J. Fenlon, & J. A. Hochgesang (Ed.), *Signed Language Corpora* (pp. 18- 45). Washington D.C.: Gallaudet University Press.

Hanke, T., Schulder, M., Konrad, R., & Jahn, E. (2020). Extending the Public DGS Corpus in Size and Depth. *Workshop on the Representation and Processing of Sign Languages: Sign Language Resources in the Service of the Language Community, Technological Challenges and Application Perspectives* (pp. 75-82). ELRA.

Hanke, T., Storz, J. (2008). iLex – A Database Tool for Integrating Sign Language Corpus Linguistics and Sign Language Lexicography. *Proceedings of the Third Workshop on the Representation and Processing of Sign Languages: Construction and Exploitation of Sign Language Corpora* (LREC 2008) (pp. 64- 67). ELRA.

Haspelmath, M. (2018). Chomsky now rejects universal grammar (and comments on alien languages). *Hypotheses*. https://dlc.hypotheses.org/1269

Hausser, R. (2001). *Foundations of Computational Liguistics. Human-Computer Communication in Natural Language* (2nd Edition). Berlin: Springer.

Hilzensauer, M., & Krammer, K. (2015). A multilingual dictionary for sign languages: "SpreadTheSign". *ICERI2015, the 8th annual International Conference of Education, Research and Innovation,* Seville.

Hockett, C. F. (1960). The Origin of Speech. *Scientific American, 203*, 89- 96.

Hodge, G. (2014). *Patterns from a signed language corpus: Clause-like units in Auslan (Australian Sign Language)* [Ph.D. thesis, Macquarie University]. Macquarie Park.

Hodge, G., & Crasborn, O. (2022). Good Practices in Annotation. In J. Fenlon, & J. A. Hochgesang (Ed.), *Signed Language Corpora* (pp. 46- 89). Washington D.C.: Gallaudet University Press.

Hodge, G., & Goico, S. A. (2022). Natural and elicited: Sign language corpus linguistics and linguistic ethnography as complementary methodologies. *Journal of Sociolinguistics, 26*(1), 126- 135.

Hodge, G., Sekine, K., Schembri, A., & Johnston, T. (2019). Comparing signers and speakers: Building a directly comparable corpus of Auslan and Australian English. *Corpora, 14*(1), 63- 76.

Hopkins, J. (2008). Choosing how to write sign language: a sociolinguistic perspective. *International Journal of the Sociology of Language, 2008*(192), 75- 89.

Hudecová, E. (2020). *Challenges of Machine Translation*. Vyučovanie prvého (L2) a druhého (L3) cudzieho jazyka z pohľadu aplikovanej lingvistiky. Kooperácie – kontrasty – analógie – interdisciplinarita, Zvolen.

Hymes, D. (1964). The Ethnography of Communication. *American Anthropologist, 66*(6), 1- 34.

Hymes, D. (1974). *Foundations in Sociolinguistics An Ethnographic Approach* (2nd Edition). Abingdon: Routledge.

Inoue, T. (2006). Memory in deaf signers and embodied cognition of sign languages. *Japanese Psychological Association, 48*(3), 223- 232.

Jin Yet, A. X., Hapuhinne, V., Eu, W., Ye-Chuen Chong, E., & Palanisamy, U. D. (2022). Communication methods between physicians and Deaf patients: A scoping review. *Patient Education and Counseling, 105*, 2841- 2849.

Johnson, M. (1992). *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago- London: The University of Chicago Press. (Original work published 1949)

Johnston, T. (2010). From archive to corpus. Transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics*, *15*(1), 106- 131.

Johnston, T. (2019). Auslan Corpus Annotation Guidelines (August 2019 version). Sydney : Macquarie University and University of Syndey.

Johnston, T., van Roekel, J., & Schembri, A. (2015). On the Conventionalization of Mouth Actions in Australian Sign Language. *Language and Speech, 59*(1), 1- 40.

Jones, S. R. (1991). *Was there a Hawthorne Effect? Working Paper no. 91- 01*. Hamilton: McMaster University.

Jurafsky, D., & Martin, J. H. (2020). *Speech and Language Processing. An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition* (3rd Edition). Englewood Cliffs: Prentice-Hall Inc.

Karmiloff-Smith, A. (1995). *Beyond Modularity. A Developmental Perspective on Cognitive Science*. Cambridge, Massachusetts: The MIT Press.

Keller, R. (1995). *On Language Change, The invisible hand in language.* London- New York: Routledge.

Kendon, A. (2004). *Gesture: Visible Action as Utterance.* Cambridge: Cambridge University Press.

Kendon, A. (2008). Some reflections on the relationship between 'gesture' and 'sign'. *Gesture, 8*(3), 348- 366.

Kendon, A. (2012). Language and kinesic complexity, reflection on "Dedicated gestures and the emergence of sign language" by Wendy Sandler. *Gesture, 12*(3), 308- 316.

Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language'. *Philosophical transactions of the Royal Society, 369*(1651).

Kendon, A., Sebeok, T. A., & Umiker-Sebeok, J. (Ed.). (1981). *Nonverbal Communication, Interaction, and Gesture Selections from Semiotica.* The Hague- Paris- New York: Mouton Publishers.

Kipp, M. (2001) ANVIL. A Generic Annotation Tool for Multimodal Dialogue. *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech),* 1367- 1370.

Kipp, M. (2012). Multimedia Annotation, Querying, and Analysis in Anvil. *Multimedia Information Extraction*, 351- 367.

Kusters, A., Spotti, M. S., & Tapio, E. (2017). Beyond languages, Beyond modalities: transforming the study of semiotic repertoires. *International Journal of Multilingualism, 14*(3), 1- 14.

La Grassa, M. (2016). Il Nome in LIS nel Segnato di Adulti Udenti: una Indagine Preliminare sul Corpus LISAU. *Italiano LinguaDue. Rivista internazionale di linguistica italiana e educazione linguistica, 8*(1), 65- 83.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society, 1*, 97- 120.

Labov, W. (1997). Some Further Steps in Narrative Analysis. *The Journal of Narrative and Life History, 7*(1- 4), 395- 415.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago- London: The University of Chicago Press.

Langacker, R. W. (1984). A Dynamic Usage-Based Model. In Barlow, M., & Kemmer, S. (Ed.), *Usage Based Models of Language.* Stanford: Center for the Study of Language and Information Publications.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites.* Stanford: Stanford University Press.

Langacker, R. W. (2008). *Cognitive Grammar. A Basic Introduction.* New York: Oxford University Press.

Laudanna, A. & Volterra, V. (1991). Order of words, signs and gestures: A first comparison. *Applied Psycholinguistics, 12,* 135- 150.

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P. (2017) Fair, transparent and accountable algorithmic decision-making processes. *Philosophy & Technology 31*(4), 611- 627.

Li, B., Yang, J., Yang, Y., Li, C., & Zhang, Y. (2021). Sign Language/Gesture Recognition Based on Cumulative Distribution Density Features Using UWB Radar. *IEEE Transactions on Instrumentation and Measurement, 70*, 1- 13.

Lo Re, L. (2022). Corpus Multimodale dell'Italiano Parlato: basi metodologiche per la creazione di un prototipo [Ph.D. thesis, University of Florence]. Florence.

Marmaridou, S. S. (2000). *Pragmatic Meaning and Cognition.* Amsterdam-Philadelphia: John Benjiamins Publishing Company.

Martin, J. (2003). *A Linguistic Comparison: Two Notation Systems for Signed Languages. Stokoe Notation & Sutton SignWriting.* Western Washington University.

Mazzei, A. (2015). Translating Italian to LIS in the Rail Stations. A. Belz, A. Gatt, F. Portet, M. Purver (Ed.), *Proceedings of the 15th European Workshop on Natural Language Generation (ENLG)* (pp. 76- 80). Associaton for Computational Linguistics.

Mazzei, A., Lesmo, L., Battaglino, C., Vendrame, M., & Bucciarelli, M. (2013). Deep Natural Language Processing for Italian Sign Language Translation. *Advances in Artificial Intelligence. Lecture Notes in Computer Science. 8249*, 193-204. Cham: Springer.

Mazzocca, M., & Ottolini, R. (2022). Curare la sordità o prendersi cura della persona sorda? In C. Galluccio, & A. Zuccalà (Ed.), *Pensare la sordità. Tra dimensione scientifica e promozione sociale. Contributi della Terza Conferenza Nazionale sulla Sordità.* Milan: FrancoAngeli.

Mazzoni, L. (2008). *Classificatori e impersonamento nella Lingua dei Segni Italiana.* Pisa: Pisa University Press.

McElvenny, J. (2019). *Form and formalism in linguistics.* Berlin: Language Science Press.

McEnery, T., & Wilson, A. (1996). *Corpus Linguistics An Introduction*. Edinburgh: Edinburgh University Press.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. Chicago: University of Chicago Press.

McNeill, D. (2005). *Gesture and Thought.* Chicago: The University of Chicago Press.

McNeill, D. (2006). Gesture: a Psycholinguistic Approach. In K. Brown (Ed.), *The Encyclopedia of Language and Linguistics*. Amsterdam: Elsevier.

McNeill, D. (2012). *How Language Began Gesture and Speech in Human Evolution.* Cambridge: Cambridge University Press.

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., & Iacoboni, M. (2007). The Essential Role of Premotor Cortex in Speech Perception. *Current Biology, 17*(19), 1692- 1696.

Memarian, B., & Doleck, T. (2023). Fairness, Accountability, Transparency, and Ethics (FATE) in Artificial Intelligence (AI), and higher education: A systematic review. *Computers and Education Artificial Intelligence, 5*(4), 1-12.

Mesch, J. (2012). Swedish Sign Language Corpus. *Deaf Studies Digital Journal, 3*.

Mesch, J., & Wallin, L. (2015). Gloss annotations in the Swedish Sign Language Corpus. *International Journal of Corpus Linguistics (IJCL), 20*, 102-120.

Mick, D. G. (1986). Consumer Research and Semiotics: Exploring the Morphology of Signs, Symbols, and Significance. *Journal of Consumer Research, 13*(2), 196-213.

Mitkov, R. (2003). *The Oxford Handbook of Computational Linguistics.* Oxford : Oxford University Press.

Montanini, M., Facchini, M., & Fruggeri, L. (1979). *Dal Gesto al Gesto: il bambino sordo tra gesto e parola.* Bologna: Cappelli.

Morgenstern, A., Caët, S. (2021). *Signes and famille corpus*. Ortolang.

Morgenstern, A., Chevrefils, L., Blondel, M., Vincent, C., Thomas, C., Jego, J.F., & Boutet, D. (2021). "Of thee I sing": An opening to Dominique Boutet's kinesiological approach to gesture. *Languages and Modalities, 1*, 3- 16.

Morlicchio, E., Micillo, V., & Dovetto, F. M. (2021). *Dalla 'langue' alla 'parole': verba manent. Scritti di Federico Albano Leoni.* Napoli: UniorPress.

Moroe, N. F., & de Andrade, V. (2018). Hearing children of Deaf parents: Gender and birth order in the delegation of the interpreter role in culturally Deaf families. *African Journal of Disability, 7*(365).

Müller de Quadros, R., & Lillo-Martin, D. (2010). Clause Structure. In D. Brentari (Ed.), *Sign Languages* (pp. 225- 252). Cambridge: Cambridge University press.
Müller, M., Ebling, S., Avramidis, E., Battisti, A., Berger, M., Bowden, R., Braffort, A., Camhöz, N. C., España-Bonet, C., Grundkiewics, R., Jiang, Z., Koller, O., Moryossef, A., Perrollaz, R., Reinhard, S., Rios, A., Shterionov, D., Sidler-Miserez, S., Tissi, K. (2022). Findings of the First WMT Shared Task on Sign Language Translation. *Proceedings of the Seventh Conference on Machine Translation (WMT)* (pp. 744- 772). Association for Computational Linguistics.
Napier, J. (2021). *Sign language brokering in deaf-hearing families.* London: Palgrave Macmillan Cham.
Narayan, J., Kalita, B., & Kumar Dwivedy, S. (2021). Development of Robot-based Upper Limb Devices for Rehabilitation Purposes: A Systematic Review. *Augmented Human Research, (6)*4.
Neidle, C., & Vogler, C. (2012). A New Web Interface to Facilitate Access to Corpora: Development of the ASLLRP Data Access Interface. *Proceedings of the Fifth Workshop on the Representation and Processing of Sign Languages: Interactions between Corpus and Lexicon.* ELRA.
Nivre, J., de Marneffe, M. C., Ginter, F., Goldberg, Y., Hajič, J., Manning, C. D., McDonald, R., Petrov, S., Pyysalo, S., Silveira, N., Tsarfaty, R., & Zeman, D. (2016). Universal Dependencies v1: A Multilingual Treebank Collection. *Proceedings of the Tenth International Conference on Language Resources and Evaluation* (LREC 2016) (pp. 1659- 1666). ELRA.
Nonhebel, A., Crasborn, O., & van Der Kooij, E. (2004a). *Sign language transcription conventions for the ECHO project.* Nijmegen: University of Nijmegen.
Nonhebel, A., Crasborn, O., & van Der Kooij, E. (2004b). *Sign language transcription conventions for the ECHO project. BSL and NGT mouth annotations.* Nijmegen: University of Nijmegen.
Nonhebel, A., Crasborn, O., & van Der Kooij, E. (2004c). *Sign language transcription conventions for the ECHO project. SSL mouth annotations.* Nijmegen: University of Nijmegen.
Oakley, T. (2007). Image Schemas. In D. Geeraerts, & H. Cuyckens (Ed.), *The Oxford Handbook of Cognitive Linguistics* (pp. 214- 236). Oxford: Oxford University Press.
O'Keeffe, A., & Michael, M. (Ed.). (2010). *The Routledge Handbook of Corpus Linguistics.* London- New York: Routledge Taylor & Francis Group.
Onelli, C., Proietti, D., Seidenari, C., & Tamburini, F. (2006). The DiaCORIS project: a diachronic corpus of written Italian. *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (LREC 2006). ELRA.
Ostling, R., Borstell, C., Gardefors, M., & Wiren, M. (2017). Universal Dependencies for Swedish Sign Language. *Proceedings of the 21st Nordic Conference of Computational Linguistics* (pp. 303- 308). Linkoping University Electronic Press.

Peirce, C. S. (2021). Part IV: The Logical Traccts (1903) Logical Tracts. No. 2 (R 492). In A. Pietarinen (Ed.), *Logic of the Future, Writings on Existential Graphs Volume 2/1: the Logical Tracts.* Berlin- New York: De Gruyter. (Original work published 1903).

Peressotti, F., & Navarrete, E. (2019). Iconicity norms for 234 sign in Italian Sign Language (LIS). *Giornale Italiano di Psicologia, 46*(3), 685- 706.

Perniss, P. (2018). Why We Should Study Multimodal Language. *Frontiers in Psychology, 9*.

Peters, C. L. (2000). *Deaf American Literature From Carnival to the Canon* . Washington, D.C.: Gallaudet University Press.

Petrov, S., Dipanjan, D., & McDonald, R. (2012). A Universal Part-of-Speech Tagset. *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (LREC 2012) (pp. 337- 342). ELRA.

Piaget, J. (1976). Piaget's Theory. In B. Inhelder & H. H. Chapman (Ed.), *Piaget and his School. A Reader in Developmental Psychology* (pp. 11- 23). New York, Heidelberg, Berlin: Springer-Verlag.

Piaget, J. (2005). *The Psychology of Intelligence.* London: Routledge. (Original work published 1950)

Piazza, M., & Pavani, F. (2022). *Le neuroscienze cognitive. Come il cervello genera la mente.* Rome: Carocci Editore

Pietrandrea, P. (2012). L'approccio socio-semiotico alle lingue dei segni. Una testimonianza. In A. Thornton, & M. Voghera (Ed.), *Per Tullio De Mauro* (pp. 265- 285). Rome: Aracne.

Plato. (2018). *Cratylus: The Dialogues of Plato* (B. Jowett, Trans). CreateSpace Independent Publishing Platform. (Original work published ca. 360 B.C.E.)

Poon, B. T., & Jenstad, L. M. (2022). Communication with face masks during the COVID-19 pandemic for adults with hearing loss. *Cognitive Research: Principles and Implications, 7*(24).

Pretato, E., Peressotti, F., Bertone, C., & Navarrete, E. (2017). The iconicity advantage in sign production: The case of bimodal bilinguals. *Second Language Research, 3*(4), 1- 14.

Prillwitz, S. (1989). *HamNoSys. Version 2.0. Hamburg Notation System for Sign Languages: an Introductory Guide.* Hamburg: Signum Press.

Prillwitz, S., Hanke, T. K., Konrad, R., Langer, G., & Schwarz, A. (2008). DGS Corpus Project – Development of a Corpus Based Electronic Dictionary German Sign Language / German. *Proceedings of the Third Workshop on the Representation and Processing of Signed Languages: Construction and Exploitation of Sign Language Corpora* (pp. 159- 164). ELDA.

Pugeault, N., & Bowden, R. (2011). Spelling It Out: Real–Time ASL Fingerspelling Recognition. *Proceedings of the First IEEE Workshop on Consumer Depth Cameras for Computer Vision- ICCV'2011* (pp. 1114- 1119). IEEE.

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language, 2*(110), 81- 94.
Radden, G. (2003). The Metaphor Time and Space across languages. *Zeitschrift für Interkulturellen Fremdsprachenunterricht, 8*(2/3), 226-239.
Radutzky, E. (1992). Dizionario Bilingue Elementare della Lingua Italiana dei Segni. Oltre 2500 significati. Con DVD-ROM. Rome: Kappa.
Raheem Jabbar, S., & Bani Madhi, M. (2021). The Role of Translators in Machine Translation. *International Journal of Development in Social Science and Humanities, 11*(January-June 2021), 1- 21.
Raniolo, E. (2021). *Senso, Ritmo, Multimodalità. Uno Studio Comparativo dei Processi Traduttivi nelle Lingue dei Segni (LIS e LSF*) [Ph.D. thesis, University of Palermo]. Palermo.
Raniolo, E., Ardita, G., & Caligiore, G. (2023). Strategie di notazione delle lingue dei segni: processi spontanei all'interno e ai confini della comunità segnante [Poster presentation] *La comunicazione parlata 2023. I 20 anni del GSCP.* Rome.
Raniolo, E., & Caligiore, G. (2023). Comunicazione tra sordi e udenti in ambito medico: sfide e opportunità per un'assistenza sanitaria inclusiva [Poster presentation] *5° Convegno Nazionale LIS. "La Lingua dei Segni Italiana: patrimonio culturale o capitale umano?",* Rende.
Rébulard, M., Danet, C., Contesse, A. S., Dauphin, J. F., Chevrefils, L., Doan, P., & Boutet, D. (2018). TYPANNOT, a new glyphic system to annotate handshapes in any sign languages. 8*th International Conference of the International Society of Gesture Studies "Gesture and diversity" (ISGS8)* (pp. 139- 140). Cape Town: Heather Brooks.
Riccardo, F., Ajelli, M., Andrianou, X. D., Bella, A., Del Manso, M., Fabiani, M., Bellino, S., Boros, S., Urdiales, A. M., Marziano, V., Rota, M. C., Filia, A., D'Ancona, F., Siddu, A., Punzo, O., Trentini, F., Guzzetta, G., Poletti, P., Stefanelli, P., Castrucci, M. R., Ciervo, A., Di Benedetto, C., Tallon, M., Piccioli, A., Brusaferro, S., Rezza, G., Merler, S., & Pezzotti, P., (2020). Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. *Eurosurveillance, 25*(49).
Ricci, A., Bonsignori, C., & Di Renzo, A. (2018). Che giorno è oggi? Prime analisi e riflessioni sull'espressione del tempo in LIS [Poster presentation] *IV Convegno Nazionale LIS "La Lingua dei Segni Italiana: una risorsa per il futuro", Rome.*
Rizzolatti, G., & Arbib, M. A. (1998). Language within our grasp. *Trends in Neuroscience, 21*, 188- 194.
Rizzolati, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental Brain Research, 71*, 491- 507.

Rizzolati, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research, 3*(2), 131- 141.

Rizzolati, G., & Luppino, G. (2001). The Cortical Motor System. *Neuron, 31,* 889-901.

Roccaforte, M. (2015). Componenti orali e riconoscibilità nella Lingua dei Segni Italiana (LIS). In M. Vayra, & F. Tamburini (Ed.), *Il farsi e il disfarsi del linguaggio. Acquisizione, mutamento e destrutturazione della struttura sonora del linguaggio* (pp. 383- 390).

Roccaforte, M. (2016). *Studi Linguistici sulle Componenti Orali della Lingua dei Segni Italiana (LIS)* [Ph.D. thesis, Sapienza University of Rome]. Rome.

Romeo, O. (1991). *Dizionario dei segni. La lingua dei segni in 1400 immagini.* Bologna: Zanichelli.

Romeo, O. (1997). *Grammatica dei segni. La lingua dei segni in 1300 immagini e 150 frasi.* Bologna: Zanichelli.

Romeo, O. (2004). *Il dizionario tematico dei segni.* Bologna: Zanichelli.

Romeo, O. (2021). *Conoscere i segni al tempo del coronavirus. 600 segni selezionati.* StudioLIS.

Russo Cardona, T. (2004a). Iconicity and Productivity in Sign Language Discourse: An Analysis of Three LIS Discourse Registers. *Sign Language Studies, 4*(2), 164- 197.

Russo Cardona, T. (2004b). *La mappa poggiata sull'isola. Iconicità e metafora nelle lingue dei segni e nelle lingue vocali.* Rende: Centro Editoriale e Librario Università degli Studi della Calabria.

Russo Cardona, T. (2005). A Crosslinguistic, Cross-cultural Analysis of Metaphors in Two Italian Sign Language (LIS) Registers. *Sign Language Studies, 5*(3), 333-359.

Russo Cardona, T., Giuranna, R., & Antirono Pizzuto, E. (2001). Italian Sign Language (LIS) poetry: iconic properties and structural regularities. *Sign Language Studies, 2*(1), 84-112.

Sallandre, M. A. (2003). *Les unités du discours en Langue des Signes Française. Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité.* [Ph.D. thesis, Université Paris VIII Vincennes-Saint Denis]. Paris.

Sandler, W., & Lillo-Martin, D. (2006). *Sign Language and Linguistic Universals.* Cambridge: Cambridge University Press.

Santoro, M., & Poletti, F. (2011). L'Annotazione del Corpus. In A. Cardinaletti, C. Cecchetto, & C. Donati (Ed.), *Grammatica, Lessico e Dimensioni di Variazione della LIS* (pp. 69- 78). Milan: FrancoAngeli.

Schembri, A. F., Rentelis, R., & Cormier, K. (2017). *British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008-2017* (3rd Edition). London: University College London.

Schembri, A., & Cormier, K. (2022). Signed Language Corpora: Future Directions. In J. Fenlon, & J. A. Hochgesang (Ed.), *Signed Language Corpora* (pp. 196- 220). Washington D.C.: Gallaudet University Press.

Schembri, A., Fenlon, J., Rentelis, R., & Cormier, K. (2014). *British Sign Language Corpus Project: A corpus of digital video data and annotations of British Sign Language 2008-2014* (2nd Edition). London: University College London.

Schetrit, O. (2016). *La culture sourde. Approche filmique de la création artistique et des enjeux identitaires des Sourds en France et dans les réseaux transnationaux* [Ph.D. thesis, École des Hautes Études en Sciences Sociales (EHESS)]. Paris.

Schubert, L. (2020). Computational Linguistics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2020 Edition). Stanford: Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/entries/computational-linguistics/

Shapiro, L. (2011). *Embodied Cognition.* London, New York: Routledge.

Sharma, A. (2020, July 29). *How Part-of-Speech Tag, Dependency and Constituency Parsing Aid In Understanding Text Data?* Analytics Vidhya. https://www.analyticsvidhya.com/blog/2020/07/part-of-speechpos-tagging-dependency-parsing-and-constituency-parsing-in-nlp/

Shmidt, T., & Wörner, K. (2009). EXMARaLDA- Creating, Analysing and Sharing Spoken Language Corpora for Pragmatic Research. Pragmatics, 19(4)565- 582.

Shormani, M. Q. (2014). The Nature of Language Acquisition: Where L1 and L2 Acquisition Meet? *Journal of Literature, Languages and Linguistics - An Open Access International Journal , 4*, 24- 34.

Singleton, J. L., & Tittle, M. D. (2000). Deaf Parents and Their Hearing Children. *The Journal of Deaf Studies and Deaf Education, 5*(3), 221–2 36.

Skinner, B. F. (1938). *The Behavior of Organisms, An Experimental Analysis.* New York: Appleton Century Crofts Inc.

Skinner, B. F. (1957). *Verbal Behavior.* New Tork: Appleton Century Crofts Inc.

Slevinski, S. (2022). *Formal SignWriting.* Datatracker. https://datatracker.ietf.org/doc/html/draft-slevinski-formal-signwriting#name-symbols

Sloetjes, H. (2023). *ELAN – Linguistic Annotator.* ELAN – Linguistic. Annotator: https://www.mpi.nl/tools/elan/docs/manual/index.html

Slonimska, A., Özyürek, A., & Campisi, E. (2016). Ostensive signals: markers of communicative relevance of gesture during demonstration to adults and children. In G. Ferré & M. Tutton (Ed.), *Gesture & Speech in Interaction 4th Edition* (pp. 217- 222). Universiry of Nantes.

Stefanowitsch, A. (2020). *Corpus linguistics. A guide to the methodology (Textbooks in Language Sciences 7).* Berlin: Language Science Press.

Stokoe, W. (1960). *Sign language structure: an outline of the visual communication systems of the American deaf.* Buffalo, New York: University of Buffalo.

Stokoe, W. C. (2005). Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf. *The Journal of Deaf Studies and Deaf Education, 10*(1), 3- 37.

Stokoe, W. C., Armstrong, D. F., & Wilcox, S. E. (1995). *Gesture and the Nature of Language.* New York: Cambridge University Press.

Stokoe, W. C., Casterline, D. C., & Croneberg, C. G. (1965). *A dictionary of American Sign Language on linguistic principles.* Washington D.C.: Gallaudet College Press.

Tallerman, M., & Gibson, K. R. (2012). *The Oxford Handbook of Language Evolution.* Oxford: Oxford University Press.

Taylor, A., Marcus, M., & Santorini, B. (2003). The Penn Treebank: An Overview. In A. Abeillé (Ed.), *Treebanks. Text, Speech and Language Technology Volume 20* (pp. 5- 22).

Taylor, J. R. (2003) *Linguistic Categorization* (3rd Edition). Oxford: Oxford University Press. (Original work published 1989)

Thanaki, J. (2017). *Python Natural Language Processing. Explore NLP with machine learning and deep learning techniques*. Birmingham, Mumbai: Packt.

Thiessen, S. M. (2011). *A grammar of SignWriting* [Master's thesis, University of North Dakota. Grand Forks.

Tognini Bonelli, E. (2001). *Corpus Linguistics at Work.* Amsterdam, Philadelphia, Philadelphia: John Benjamins Publishing Company.

Tomasuolo, E., Gulli, T., Volterra, V., & Fontana, S. (2021). The Italian Deaf Community at the Time of Coronavirus. *Frontiers in Sociology, 5*.

Unicode. (2017). *About the Unicode Standard. Characters for the World*. Unicode. https://unicode.org/standard/standard.html

van Valin, R. D. (2003). In M. Aronoff & J. Rees-Miller (Ed.), *The Handbook of Linguistics* (pp. 300- 336). Hoboken: Blackwell Publishing.

Vigliocco, G., Perniss, P., & Vinson, D. (2014). Language as a multimodal phenomenon: implications for language learning, processing and evolution. *Philosophical Transactions of the Royal Society, (369).*

Volterra, V. (Ed.). (1981). *I segni come le parole: la comunicazione dei sordi*. Torino: Boringhieri.

Volterra, V. (Ed.). (1987). *La lingua dei segni italiana. La comunicazione visivo-gestuale dei sordi.* Bologna: Il Mulino.

Volterra, V., & Antinoro Pizzuto, E. (2002). La Lingua dei Segni Italiana. In C. Lavinio (Ed.), *Linguistica italiana alle soglie del 2000 (1987-1997 e oltre)* (pp. 1- 20). Rome: Bulzoni

Volterra, V., Capirci, O., Caselli, M. C., Rinaldi, P., & Sparaci, L. (2017). Developmental evidence for continuity from action to gesture to sign/word. Language, *Interaction and Acquisition, 8*(1), 13- 41.

Volterra, V., Laudanna, A., Corazza, S., Radutzky, E. & Natale, F. (1984). Italian Sign Language: The order of elements in the declarative sentence. In F. Loncke, P. Boyes-Braem & Y. Lebrun (Ed.), *Recent Research on European Sign Languages* (pp. 19- 48). Lisse: Swets & Zeitlinger B.V.

Volterra, V., Roccaforte, M., Di Renzo, A., & Fontana, S. (2022). *Italian Sign Language from a Cognitive and Socio-semiotic Perspective. Implications for a general language theory* Amsterdam, Philadelphia: John Benjamins Publishing Company.

Weaver, W. (1955). Translation. In W. N. Locke, & A. D. Booth (Ed.) *Machine translation of languages: fourteen essays.* Cambridge, Massachussets: John Wiley & Sons Inc.

Wig, G. S. (2017). Segregated Systems of Human Brain Networks. *Trends in Cognitive Sciences, 21*(12), 981- 996.

Wilcox, S., & Wilcox, P. (1995). The Gestural Expression of Modality in ASL. In J. Bybee, S. Fleischman, J. L. Bybee, & S. Fleischman (Ed.) *Modality in Grammar and Discourse* (pp. 135- 162). Amsterdam, Philadelphia: John Benjamins Publishing Company.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos. L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*.

Winter, B., Perlman, M. P., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies, 18*(3).

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: a professional framework for multimodality research. In N. Calzolari, K. Choukri, A. Gangemi, B. Maegaard, J. Mariani & D., Tapias (Ed.), *Proceedings of the Fifth International Conference on Language Resources and Evaluation* (LREC 2006). ELRA.