MSc

2.º CICLO

FCUP
2017

U.PORTO

3D Character Animation Using Sign Language

Anabel Cristina Alves Ferreira

FC

U.PORTO
FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

# 3D Character Animation Using Sign Language

Anabel Cristina Alves Ferreira

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Área Científica
2017

U.PORTO
FC FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

MSc

# 3D Character Animation Using Sign Language

Anabel Cristina Alves Ferreira
Mestrado em Ciência de Computadores
Departamento de Ciência de Computadores
2017

**Orientador**
Prof. Verónica Costa Teixeira Pinto Orvalho, Professora Auxiliar DCC-FCUP
Faculdade de Ciências da Universidade do Porto

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ______/______/_________

# Acknowledgements

First of all, I would like to thank the University of Porto's Department of Computer Science for all their help and support through all my Master's. It was a great adventure to start a course in a foreign country, but the professors, personnel, and student body I had the honor to met made it worth the risk.

I would also like to thank my thesis advisor Verónica for giving me the opportunity to work on this wonderful project, always giving me feedback and guidance throughout the development of this thesis. In the same way, I would like to thank every person who directly, or indirectly, made possible this thesis.

To my aunt, uncle, and cousin, for making me feel like at home since the first day I arrived in Portugal, I thank you from the bottom of my heart. It is one of my deepest desires to be able to repay you someday for all your kindness and love.

To all my friends, because they were always there for me when I needed it. Every little chat to cheer me up helped me to get where I am.

To those who are like my family, Mercedes and Mariana, for their unconditional support over all these years....thanks. You two have been with me on the good and bad times, there are not enough words in my vocabulary to express how much that means to me. I hope we can keep being silly fangirls together for many, many, many more years in the future.

While you are not with us anymore, I also want to thank you Cuchi. You were the most awesome little angel I ever came across. My best wishes to you wherever you are.

Last but not least important, thanks to my mother. You are the most important person in my life, the one who I know is always there for me and will never stop believing in me. I miss you greatly every single day and I hope we can soon meet again so you can see all the things I have learned, the people I have met and the places I have been since I left home.

# Resumo

Perda de audição é a incapacidade parcial ou total de ouvir que afeta aproximadamente 10% da população mundial. Uma das principais consequências desta condição é a perda da capacidade do indivíduo para se comunicar com outras pessoas através da língua falada. Esta incapacidade limita o seu acesso à informação, e pode ter um impacto negativo significante na sua vida diária, causando solidão, isolamento e frustração.

Hoje em dia, associações especializadas fornecem educação em língua gestual e intérprete para jornais, eventos e conferências, como um meio alternativo de comunicação para pessoas com deficiência auditiva. Apesar disso, já que os intérpretes são escassos, não tem sido possível fornece-lhes o mesmo nível de acesso à informação que as pessoas ouvintes têm. O desafio é como outorgar às pessoas com deficiência auditiva a mesma capacidade de sempre se poder comunicar com a sociedade de modo a que não estejam limitados pela disponibilidade de intérpretes.

Esta tese resolve as limitações na comunicação que as pessoas com deficiência auditiva sofrem atualmente. Isto é alcançado com uma ampla investigação nas disciplinas de processamento de linguagens naturais, computação gráfica, interação humano-computador e línguas gestuais. Os resultados dessa investigação foi um sistema que faz uso de técnicas de processamento de linguagens naturais para decompor frases em português nos sinais equivalentes em Língua Gestual Portuguesa. Estes sinais são então decompostos em unidades mais pequenas que produzem animações 3D desde um conjunto limitado de clipes de animação.

Após esta investigação, um sistema chamado 3DSL for desenvolvido para atestar a viabilidade desta abordagem, tendo a usabilidade como uma parte importante do processo de desenvolvimento. Para além disso, realizou-se uma avaliação qualitativa preliminar da saída do sistema, com resultados globalmente positivos.

Embora a investigação da Língua Gestual Portuguesa (LGP) ainda se encontre nas primeiras fases, a tradução automática utilizando o conhecimento disponível pode ter um impacto positivo na vida das pessoas com deficiência auditiva. 3DSL contribui para estabelecer as bases de sistemas que preencham a lacuna entre as pessoas ouvintes e as pessoas com deficiência auditiva.

# Abstract

Hearing loss is a partial or total inability to hear that affects approximately 10% of the population worldwide. One of the principal consequences of this condition is the loss of the individual's ability to communicate with others through spoken language. This inability limits their access to information, and can have a significant negative impact on everyday life, causing feelings of loneliness, isolation, and frustration.

Nowadays, specialized associations provide sign language education and interpreters for news, events, conferences and meetings as an alternative mean of communication for hard of hearing people. In spite of this, since interpreters are scarce, it has not been possible to provide them with the same level of access to information that hearing people have. The challenge is how to grant hard of hearing people the ability to always be able to communicate with society so that they are not limited by the availability of interpreters.

This thesis solves the limitations in communication that hard of hearing people suffer nowadays. This was achieved by an extended research in natural language processing, computer graphics, human-computer interaction and sign language. The outcome of this research was a system that uses natural language processing techniques to decompose Portuguese sentences into a set of equivalent signs in Portuguese Sign Language. Each sign is then decomposed into smaller units which produce 3D sign animations from a limited set of animation clips.

Following the research, a system called 3DSL was developed to attest the feasibility of this approach, with usability evaluation being an important part of the development process. Additionally, preliminary qualitative evaluation of the system output was carried out, with generally positive results.

Although research on Portuguese Sing Language (PSL) is still in early phases, automatic translation using available knowledge can positively impact the lives of hard of hearing. 3DSL contributes to set the foundation for systems that close the communication gap between hearing and hard of hearing people.

# Contents

# List of Tables

# List of Figures

# Acronyms

**3DSL** Three Dimensional Sign Language.

**API** Application Programming Interface.

**ASL** American Sign Language.

**BSL** British Sign Language.

**CSV** Comma Separated Values.

**GUI** Graphical User Interface.

**KSL** Korean Sign Language.

**LGP** *Língua Gestual Portuguesa*/Portuguese Sign Language.

**MPEG-4** Method of defining compression of audio and visual (AV) digital data by the ISO/IEC Moving Picture Experts Group.

**OSV** Object-Subject-Verb word order.

**PoS** Part of Speech.

**SAPI** Microsoft Speech API.

**SiGML** Signing Gesture Markup Language.

**SSL** Spanish Sign Language.

**UX** User Experience.

**XML** eXtensible Markup Language.

# Chapter 1

# Introduction

Communication has been important since even before the start of humanity. It grants the ability to convey and share ideas and feelings, even more, allows to impart and exchange any kind of information. Given its importance, there have been multiple studies throughout history focused in to explain how communication works. One of the main focus has been how humans have developed languages as an evolved form of communication with a set of rules that transmits information in a structured and conventional way [84].

However, communication is only effective if all participants involved shared a language. This common communication tool allows the sender to encode and transmit a message (using whatever medium is available) knowing that its receiver(s) will be able to decode it accordingly. Therefore, there is an inherent relationship between the capacity of encoding and decoding a language and the ability of participants to extract information from it; when that capability is compromised it is said that there is a communication barrier.

Communication barriers have a wide range of effects that negatively influence many life aspects of affected individuals. As those barriers become bigger, so do the feelings of doubtfulness, mistrust, and isolation of a person who sees that even simple tasks can become difficult because their interlocutors are not able to truly perceive the intended meaning of their words. And the consequences are not only psychological, it is estimated that millions of dollars are lost due problems related to language barriers [37].

Sadly, with more than 6000 languages [6] existing in the world, communication barriers are not an uncommon problem. Even countries with a relatively homogeneous population might use more than one language. In Portugal's case, while most of the population uses Portuguese as its main language, there are two other native languages used in the country: Mirandese and Portuguese Sign Language (LGP).

## 1.1 Motivation

Hearing loss, also known as hard of hearing, is a partial or total inability to hear that affects approximately 10% of the population worldwide [60]. One of the principal consequences of this condition is the loss of the individual's ability to communicate with others through spoken language. To overcome communication barriers, people with hearing loss can learn to communicate using alternative means, such as lip-reading or sign language.

However, people around the world share copious amounts (more than 300 hours [14]) of audio and video every minute, and most of them are not available in any other form. People who became hard of hearing after language acquisition can close this information gap thanks to closed captions or transcripts, and those can be produced with any hearing person since spoken and written language follow the same framework of rules and are interchangeable with each other. Meanwhile, the rest of hard of hearing population are not as lucky. Those who did not achieve communication proficiency in spoken language use sign language as their native way of communication. Furthermore, spoken and sign languages are independent of each other. Their grammar rules and even the way concepts are expressed in each of them (vocal-auditory in contrast to visual-gestural), are not directly related. Likewise, different sign languages are also independent of each other, there is no such thing as a universal sign language. In other words, sign languages evolved as independent languages and are not simple gestural codes representing the surrounding spoken language. Therefore, only trained interpreters (which might be as few as a hundred like on LGP case [66]) can provide a translation between a spoken language and a sign language.

Nowadays, several initiatives around the world are promoting equal rights for deaf and hard of hearing people. The creation of associations to provide sign language education and interpreters for news, events, conferences, or meetings, has been an important step forward along with 2010 Brussels declaration of sign languages in the European Union [33]. Although these projects improve the access to information, there are several instances where this community is still not able to access information because it is not available in a way understandable by them (e.g. Internet videos or conversations with people without sign language knowledge). To make up for those situations, and to help the full integration of deaf and hard of hearing people, automatized alternatives may be considered.

## 1.2 Problem Statement

Automated systems should aim to provide hard of hearing people with means to easily understand information expressed in spoken language; also, translations should take into consideration cultural and linguistic differences while preserving the intended meaning of the original message. Similarly, systems should be available anytime and work reliably with no human intervention (besides the user input) and should be straightforward enough that they can be easily used by any user.

However, machine translation is a complex process surrounded by issues that impact its performance. Those problems arise from the nature of languages and their inherent linguistic characteristics (e.g., bilingual lexical differences and structural ambiguity). Moreover, translation between spoken and signed languages have an additional layer of complexity: how to express visual-gestural features of signed languages in a virtual environment.

Logic problems are not the only ones that affect a system. Even if all features are properly implemented, and for every input, a correct non-ambiguous output is provided; the work is not done. If users are not able to comfortably interact with the system, then the goal has not been achieved. The fact is that there is not a single approach that will work for all types of users because people differ in so many ways.

In consequence, different technologies [35] [85] [32] have been developed through the years. However, not all hard of hearing people have the same quantity of tools at their disposal since only in recent years has the study of LGP started to rise. The first published grammar is barely two decades old [5], and while there are automated translation systems in development [70] [41], none of them are ready for user usage. Need therefore arises to expand the current research about automated systems focused on LGP translation.

Accordingly, this thesis aims to study the state of art of automated sign language translation, specifically based on 3D avatars, and propose a user-centered system for automated LGP translation. This approach seeks to provide usability levels, according to the special accessibility needs of hard of hearing people, while also testing how such translation systems can improve their communication. This thesis will focus on translation from European Portuguese (simply referred as Portuguese through the rest of the document) language to LGP. Thus, while equally important, translation from LGP to Portuguese is out of the scope of this thesis.

To be more specific, this thesis deploys a system called 3DSL that combines natural language processing with SignWriting to provide a way of building animations from a limited set of gestures. This system translates Portuguese into a set of glosses that meet LGP grammar rules and then, divides each sign into smaller sections which later will be used to display a 3D animation through a humanoid avatar.

## 1.3 Objectives

The overall goal of this thesis is to close the communication gap between hearing and hard of hearing people, in order to make hard or hearing less depend on the availability of interpreters. Even more, this thesis expects to reduce the gap between the information available for listeners and the information available to hard of hearing people.

While it is not the intention of this work to provide a definitive solution for all encountered problems; it aspires to present an exploratory overview that analyzes current projects, and provide an architecture that expands them in order to propose solutions for some open problems they present. In particular, using SignWriting as a helper in the animation process.

In order to achieve the overall goal following objectives were proposed:

1. Describe the state of the art of Sign Language translation systems based on 3D animated avatars in order to identify areas of improvement.
2. Using the principles of user-centered design, identify the needs, wants and limitations of end users of Sign Language translation systems.
3. Design the architecture for an automated translation of Portuguese to LGP.
4. Develop a system that automatically translates Portuguese into LGP.

## 1.4 Contribution

The major contributions of this thesis are:

1. An automated translation system that improves the communication of hard of hearing people.
2. The definition of an animation system that can produce signs from a limited number of animations, in contrast of having to create a new animation every time a word is added to the system's dictionary.

3. The development of a system that addresses the needs of automated translation of LGP.

## 1.5 Summary of Contents

The present document is composed of 5 chapters, each one describing a different aspect of the developed work:

- **Chapter one** gives a short introduction to the thesis. A small description of the current state of the background, motivation, and goals of this work; at the same time, a brief overview of how that problem was approached.
- **Chapter two** explains basic concepts about LGP: how it can be represented, compares its grammar with that of Portuguese and explains its importance and applications in the daily life of hard of hearing people.
- **Chapter three** describes the evolution of digital automated translation methods for Signed Languages, focusing on the difficulties found along the way and how current systems are trying to overcome those obstacles.
- **Chapter four** portrays how the developed system is structured, explaining how each of its modules works and which technologies were used in the creating of each.
- **Chapter five** gives the all the information related to the proof of concept developed as part of this thesis.
- **Chapter six** shows and analyses the results obtained after testing the prototype based on the architecture described in the previous chapter.
- **Chapter seven** concludes the document with a discussion about the outcome of this work and how it could be expanded in the future.

# Chapter 2

# Background

Hearing loss is an auditory problem where the ability to hear becomes compromised. In order to measure hearing loss, the individual is tested to find the quietest sounds they can hear by using tones with different frequencies, which are heard as different pitches. This level is called the threshold and defines whatever the person presents mild, moderate, severe or profound hearing loss. People with severe and profound hearing loss might have problems following speech, even with hearing aids [1].

Around the world, 360 million people live with disabling levels of hearing loss[1]. This is approximately 5.3% of the world population. Even more, 32 million of these are children of less than 15 years of age [95]. In Portugal, estimations say that around 120 thousand people suffer some level of hearing loss [9].

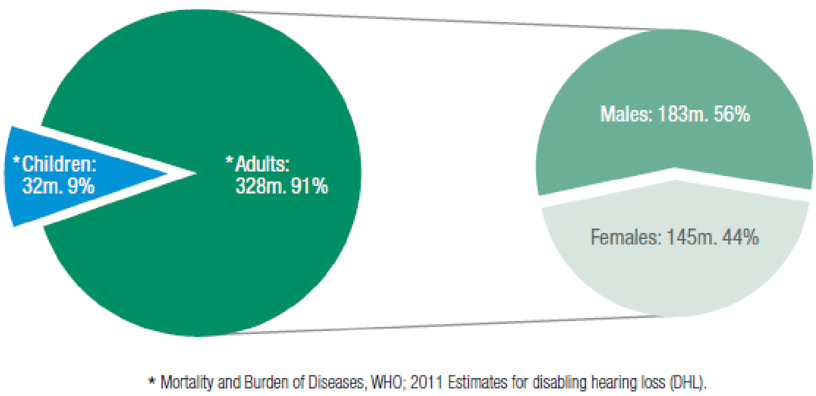



Figure 2.1: Demographics of global hard of hearing population [95]

However, inability of processing speech does not equal to a total loss of the individual's ability to communicate with others. As a medium to overcome communication barriers, people with hearing loss can learn to communicate using alternative means, such as lip-reading or sign language.

[1] Hearing loss greater than 40 dB in the better hearing ear.

This chapter presents an overview of the main characteristics of sign languages, specifically focused on Portuguese Sign Language. Likewise, talks about ways that allow the representation of signed languages and their applications. It also offers a brief description of other concepts that will be useful for understating of future chapters.

## 2.1 Sign Language

In contrast to spoken languages, sign languages are visual-gestural based. They make use of hand, arm and body movements along with facial expressions in order to express thoughts and ideas. Previously, they were considered part of the same category as other systems like body language, home signs, and other manual codes; they are now recognized as proper languages with the same linguistic properties and faculties as spoken languages [77]. Likewise, 2010 Brussels Declaration on Sign Languages [33] in the European Union also recognizes the importance of national sign languages for the communication, and general quality of life, of Hard of hearing. With regard of Portuguese Sign Language (know in Portuguese as língua gestual portuguesa or LGP), it is legally recognized in the Constitution of Portugal since 1997 (article 74) [23].

However, while the language is actively used and officially recognized, it is still classified as developing [66]. Although there is some literature available, it is not yet widespread. To change this, different initiatives [5] [10] [32] have been working to document, standardize and modernize LGP. Similarly to other ways of communication, LGP has evolved through the times. Nowadays two main dialects can be found in Lisbon and Porto regions, both of them present historical influence from Swedish Sign Language. Lexical comparison of non-iconic signs revealed no apparent relationship to Spanish or Catalan Sign Languages[28].

The Ethnologue [66] estimates the number of LGP users 60.000, with at least half of them having it as their native language. Other users include people that became hard of hearing after language acquisition, relatives of hard of hearing people, and professionals that have to interact with hard of hearing people on a regular basis. The degree of fluency on the non-native groups vary, however, most of them are not truly fluent signers [9].

## 2.2 Properties of Portuguese Sign Language

Linguistic studies of LGP started on 1985 [55] on a seminar by the Faculty of Letters of the University of Lisbon. Following it, the first dictionary [13] and grammar [5] were released. As consequence, it was confirmed that LGP was a proper language with its own set of rules and properties. More specifically, it was confirmed as a language with contrastive aspects, creativity, recursivity and whose signs could be not only iconic but also arbitrary.

Those works are the basis of the current knowledge about the structural aspects of LGP that will be described in the sections below.

### Phonetic Aspects

Each sign can be represented as a sequence of segments, each of them composed of two major elements related to the hand's behavior (posture and activity) plus non-manual features. Moreover, in parallel with what happens with writing, signers have a dominant hand in charge of doing most of the needed movements, and a non-dominant one that is limited to some secondary movements.

Postures are described by a set of articulatory features. All taken together, describe how the hand is placed at a particular point in the production of a sign, they do not describe its movement (or lack of). To be more specific, the articulatory features refer to:

- Hand configuration: marks the state of the fingers in a given moment. Figure 2.2 gives an example of them. All configurations for LGP are shown on Appendix B.


Figure 2.2: Examples of hand configurations [5]

- Articulatory space: specifies the location where the hand is situated, along with the spatial relationship between the hand and that location. Those locations include forehead, nose, lips, shoulder, arms chest and thighs; each with even smaller distinguishable areas called contact points. All areas can be seen on Figure 2.3.

  Likewise, the articulatory space can be divided into two main levels: horizontal and vertical. The combination of different horizontal and vertical spaces represents the three-dimensional space where most signs take place and is shown on Figure 2.4.

- Orientation: indicates the direction which the signer's palm faces. Mostly can be defined by general orientations as up, down, forward, side and back.

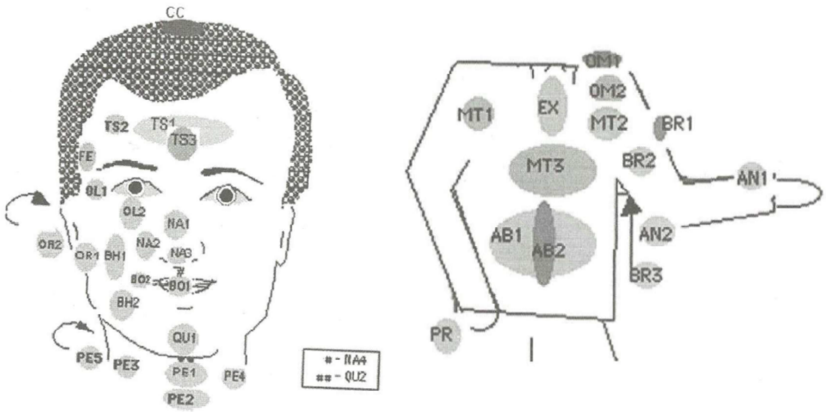



Figure 2.3: Localization of contact points [5]

In the other hand activity defines whether or not the hand is moving, and, if so, in what manner. Those features can be classified on Movements which define when the posture elements are in a transition, and on Holds which characterizes the periods of time during which the position is in a steady state.

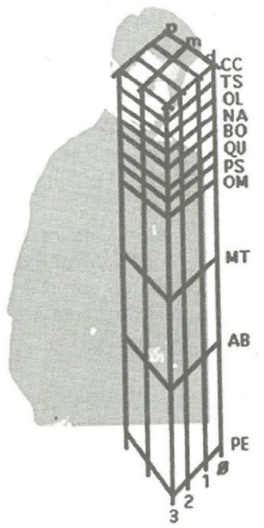



Figure 2.4: 3D representation of the articulatory space [5]

Finally, non-manual features involve all variations on the body position (i.e., trunk rotation). It also involves alterations on facial expressions and head position.

## Morphological Aspects

Similarly to spoken languages, LGP presents different morphological phenomena on parts of speech described bellow:

### Noun

Gender marking is used only when it is strictly necessary. When it is used, feminine is usually indicated by prefixation while male gender is characterized by the lack of any gender mark. Regarding proper nouns, LGP does not present any kind of gender marking (unlike Portuguese).

Meanwhile, number designation on nouns follows spoken language division on plural and singular. For pluralization LGP uses the following grammar process:

- Incorporation: for small quantities, the noun is followed by the corresponding number (i.e. QUATRO + CASA). Whereas, for hard to count quantities the corresponding determinant (many, few, etc.) is added to the main sign.
- Movement repetition: consists in a regular movement where there are no appreciable alterations to the sign's position structure.
- Redouble: the non-dominant hand repeats the sign made by the dominate one.

### Article

Definite articles are theorized to be represented by a sign with the index finger positioned in front of the signer's body. The order of signs vary according to how they are performed as indicated on Table 2.1.

Table 2.1: Article usage

| Hand(s) used | Sign order |
|---|---|
| Only<br>dominant hand | article ->noun ->verb<br>OR<br>noun ->article ->verb |
| Two<br>hands simultaneously | non-dominant hand: article ->verb<br>dominant hand: noun |

### Personal Pronouns

As in Portuguese, there is a distinction between first, second and third persons for both singular and plural. Each of them has its own sign clearly distinguishable from the others. However, in contrast with Portuguese, there is no distinction on the gender of the third person (neither in singular nor on plural).

Another important aspect is that LGP does not use a pronoun for each verb in a sequence unless it changes. In other words, only when the subject changes is a new sign introduced to mark it. Even more, the subject can appear at the end or at the beginning of the sentence, or in both to put more emphasis in it.

#### Verb

On Portuguese, verbs are inflected according to the person and number of the subject. However, on LGP the verb does not change its form according to the subject, both number and person have to be separately indicated by their respective signs according to whether they are needed or not.

Verb tenses are divided on past, present, and future. However, verbs are not inflected to express those tenses; verbs are always used in their infinitive form, a tense mark will be added to them according to the tense that needs to be expressed. Table 2.2 indicated one way of using tense marks to indicate each of the tenses used on LGP.

Table 2.2: Verb formation

| Tense | Formation |
|---|---|
| Past | near: infinitive + PASSADO sign<br>distant: infinitive + one repetition of PASSADO sign<br>very remote past: infinitive + several repetitions of PASSADO sign |
| Present | infinitive |
| Future | infinitive + FUTURO sign |

Another way of expressing time tenses is by the usage of the infinitive in conjunction with an adverb or temporal expressions like “yesterday” or “the day after tomorrow”.

Additionally, verb formation implies the usage of imaginary temporal lines represented as lines perpendicular to the signer. All of them are indicated on Figure 2.4.

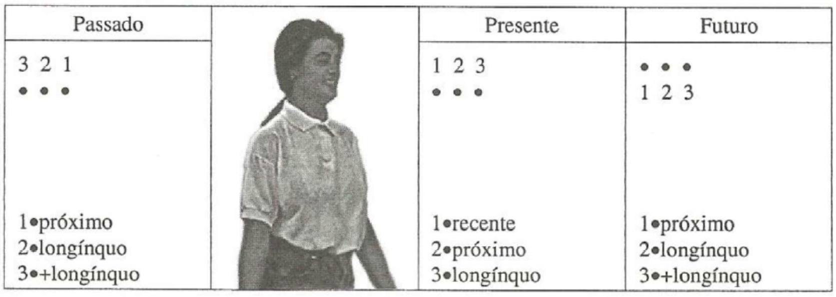



Figure 2.5: Temporal lines on LGP [5]

Besides tenses, LGP verbs can present one of the following verbal aspects: durative, iterative , punctual and habitual. To express them LGP uses different processes:

- Repeating the verb’s sign
- Repeating the sign of the temporal expressions associated with the verb

- Changing the execution time of the verb's sign or of the temporal expressions associated with it
- Changing the amplitude, intensity or tension on the sign execution
- Adding non-manual elements like head movement or face expression

### Lexical Aspects

According to the studies realized, there are three main lexical aspects present on LGP: derived signs, compound signs, and dactylology.

Both derived and Compound signs allow the creation of new signs from smaller units. The difference between the two methods resides in their formation: derived signs are usually formed by sub-fixation of primitive words. Figure 2.7 shows how the sign *PORTA* (door) is formed by the addition of an extra movement (M) the sequence of movements and poses (S) that form the sign *ABRIR-PORTA* (open the door).

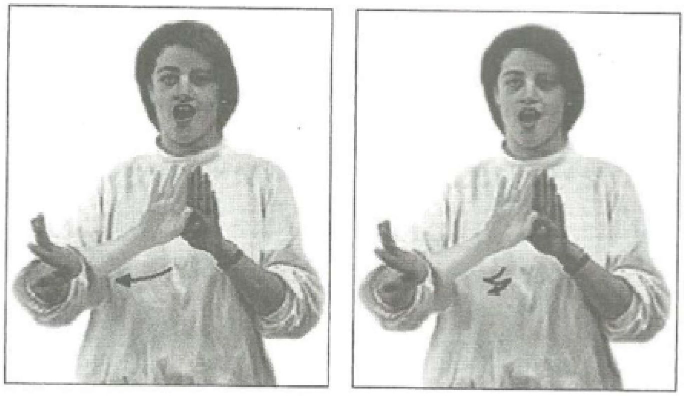



Figure 2.6: Derived sign example [5]

Meanwhile compound signs are through juxtaposition or agglutination as can be found on Figure 2.7. Here two different signs *BOA* (good) and *TARDE* (afternoon) that are formed by different sequences of poses and movements create a new sign by the juxtaposition of the structure of each element. The words created by this process usually have a meaning related to the elements that formed it, however, that is not necessarily the case.

Meanwhile, dactylology is used for words that do not have a sign yet, usually names of people, cities, brands or foreign expressions. In punctual cases, it can also be used to aid the creation of new signs.

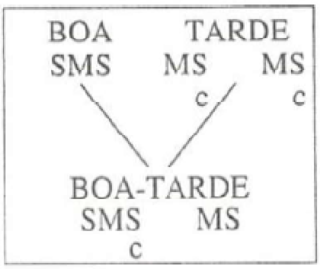



Figure 2.7: Compound sign examples [5]

## Syntactic Aspects

In contrast with Portuguese, LGP presents a different word order Object-Subject-Verb (OSV) according to [5] and Subject-Object-Verb according to [10].

For complex sentences, there is an organizer space that can be seen in detail on Figure 2.8. To be specific, this space defines how verb agreement and nominal establishment are used during the signing process.

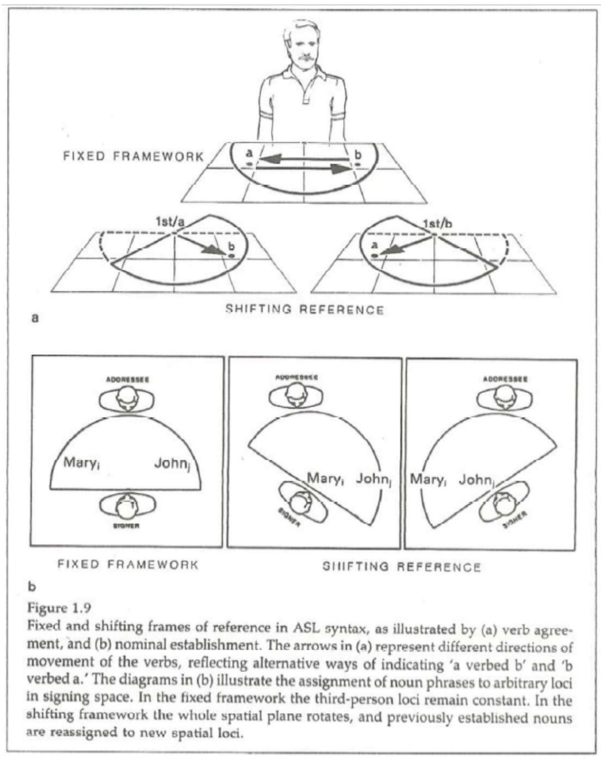



Figure 2.8: Organizer space [5]

Interrogative and exclamatory sentences are expressed mainly by non-manual gestures like body position and facial expressions. Negative sentences are built by adding the NÃO sign or by shaking the head while signing the negative sentence. In regard to reported speech and direct speech, the first is done following the rules described in the previous sections, while the latter implies the movement of the body as if the signer takes each of the different roles needed.

## 2.3 Sign Writing

Analogous to what happens when a learner tries to write in a foreigner language, a Hard of Hearing person might have to simplify their sentences to make their ideas understood while writing using spoken language rules. Moreover, many times they might use words that do not have the same meaning or will misuse the grammar rules. In order to overcome this, Hard of Hearing community started to look for a way to actually represent sign language in writing instead of having to resort to spoken language [78].

The first writing system was created by William C. Stokoe [77]. This notation used Latin letters and numerals to represent the handshapes and iconic glyphs to transcribe the position, movement, and orientation of the hands. Although, hand configuration elements are clearly defined, this system is not able to represent all possible signs. Such fact is derived from its lack of plane of movement representation.

Time after, other systems were developed such as François Neve notation [59], HamNoSys [43], SignWriting [79] [82] and D'Sign System [40]. From those, SignWriting has the most regular use and it is the only that makes part of Portuguese Academic Curricula for Hard of Hearing education (elementary [17] and secondary [56]).

Unlike Stokoe's, SignWriting is a featural script, that is, the shapes of the symbols are not arbitrary, but encode phonological features of the phonemes that they represent. SignWriting symbols are divided into seven categories according to their function [64] [81]:

- Hands: hand shapes.
- Movements: contact symbols, small finger movements, straight arrows, curved arrows, and circles are placed into 10 groups based on spatial planes.
- Dynamics: provide emphasis on a movement or expression.
- Head and Faces: facial expressions (head, eyebrows, cheeks, mouth and tongue positions)
- Body: torso, shoulders, hips, and limbs movement and positions.
- Detailed Location: for detailed analysis of location needed in linguistic research. Not used for everyday writing.
- Punctuation: used when writing complete sentences or documents.

Those symbols are arranged in a two-dimensional layout (as can be seen in Figure 2.9) within a sign box that represents the localization of the hands and other body parts involved in the sign. Therefore, and unlike most writing languages, SignWriting is written in columns following a set of conventions for how symbols are to be arranged relative to each other within a sign.

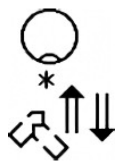


Figure 2.9: SignWriting example representing the word soup in LGP

Those characteristics have their advantages and disadvantages. Iconicity makes reading easier to learn, however its great symbol set size creates a challenge in learning how to write all the fine details. SignWriting's two-dimensional spatial layout is also a double-edged sword, although more iconic than a linear layout, it requires a special software [80] to be easily written and exported into a word processor and other programs. In order to overcome this, Unicode Standard has introduced 672 characters to represent SignWriting's standardized symbol set [34], although it does not address the two-dimension problem.

## 2.4 Corpora for Portuguese Sign Language

In the current context, corpus is a collection of natural language elements ideally designed to be a representative of one (or more) characteristics(s) of a language. As the time of writing, no annotated corpora for LGP or bilingual translations were found.

The biggest publicly available collection of LGP elements is the project SpreadTheSign [32] administered by the Non-Governmental and Non-Profit Organization European Sign Language Centre. It has 15000 video entries, which the administrators mark as a $100\%$ completion.

Other, more limited, corpus is the terminology dictionary provided by the Universidade Católica Portuguesa [88]. It provides LGP translations for handful number of linguistic concepts (102) that are not available on SpreadTheSign project.

## 2.5 Other Concepts

### Gloss Notation

Gloss notation is a brief description of the meaning of a word or wording in a text. In the context of sign languages, they are typically used to transcribe word-by-word the meaning of gestures. To differentiate glosses from spoken language they are written in capital letters, optionally followed by the original word or sentence. An example of this notation is presented in Figure 2.10.

$$NATURE\ SELF-1\ WOW[emph]\ AWESOME\ POWER[focus]$$

Figure 2.10: Example of gloss notation for "Nature can be unbelievably powerful" [3]

# Chapter 3

# State of the Art

Since the recognition of sign languages as proper languages, there has been a clear disparity between the information access available for hearing people versus that available for hard of hearing people. This chapter will give a brief overview of the approaches used to overcome this problem, first through non-computational means which naturally migrated to automated systems as the technology started to make it possible. While not an extensive overview, this chapter's aim is to expose some historically-relevant projects and connect them with current approaches. Afterwards, there is a summary that focuses on LGP projects and the results achieved by them.

Finally, this overview is concluded by a small discussion about the general current state of automated translation systems for sign languages in contrast with those available for LGP.

## 3.1 Telecommunications Relay Services

Formerly, technology was not mature enough to improve the access to information for deaf and hard of hearing people; access was limited to the availability of interpreters in the zone where the information was contained. As technology evolved, relay systems were used to ease the situation; this allowed to overcome distances between the information and the receiver: the data, the interpreter and the person with hearing loss did not need to be in the same place anymore. Those relay systems are known as telecommunications relay services (TRS). The first service of this kind was established in 1974 by Converse Communications of Connecticut and are still active nowadays. The specifics about how TRS works depend on the particular needs of the user and the equipment available, but, in general, works as described below [35]:

1. A person with a hearing or speech disability initiates a TRS call, the person uses a text input device to call the TRS relay center.

2. An operator places an outbound traditional voice call to that person, then serves as a link for the call, relaying the text of the calling party in voice to the called party, and converting to text what the called party voices back to the calling party.

However, this was the definitive solution. As 1997, thanks to the advent of the World Wide Web, relay services evolved into video relay services (VRS). Being first available in Sweden, and expanding to the rest of the world afterwards, [69]. This new system works in a similar fashion as TRS but frees the user from having to know a written language in order to use it. A drawback is that some governments forbid the usage of VRS in cases where the party with hearing loss are in the same place because the service is designated only for remote calls [36]; for those cases, video remote interpreting (VRI) was created. VRI provides off-site interpreters through the use of devices such as web cameras or videophones.

As can be seen, relay services and remote interpreting have been improving the communication skills of people with hearing loss, but they still have two important flaws surrounding them: the dependency of interpreters' availability and the size of the target information. In other words:

- These systems only cover specific situations, mainly communication between hearing people and people with hearing loss.

- And they are highly dependent on the availability of interpreters, which cannot be guaranteed 24/7.

## 3.2 Automated Translation

To undertake the flaws from previous approaches and to take advantage of the newly available technologies, research teams started to develop automated solutions. Some of the most relevant are described bellow.

#### Signing Avatars

As early as 1998, there is registered research about automated methods of Sign Language translators. Seamless Solutions [91] proved that it was feasible to perform interactive communication in American Sign Language over the Internet without needing advanced hardware to achieve acceptable results.

This system consisted of an articulated humanoid model and a library of ”basic” animations (letters, words, and facial expressions). The animation engine was developed in Java using Virtual Reality Modelling Language to represent the current state of the three-dimensional scene in a format capable of being sent across the Internet in real-time.

Different avatars, using different commercial software available at development's time, were designed; each of them had approximately 5000 polygons facets groups into 50 segments controlled by 100 degrees of freedom in motion. However, no facial features were available. This setup was renderable at 15 frames per second on a Pentium II, 300 MHz desktop, and 6 frames per second on a 166 MHz Pentium MMX laptop (state of art personal computers at the time). Animation was performed by a combination of direct manipulation of joint angles and inverse kinematics, which proved satisfactory for a limited vocabulary but optical motion capture was indicated as desirable for bigger scenarios.

There is not any significant information about how the project took an input and transformed it into a correct (both in vocabulary and grammar) sign language output. However, it was indicated that the system used a conjunction of voice processing with scripted behaviors to simulate signing in response to voice, actions, and text.

As a conclusion of this project, sign language communication via the Internet was proved achievable yet not practical outside of limited application areas because the lack of proper coordination of facial expressions and arm/hand movements.

### ViSiCAST

With the start of the new millennium, the European Union funded the Virtual Signing, Animation, Capture, Storage, and Transmission (ViSiCAST [1]) project as part of the Fifth Framework Programme [21]. This project aimed to improve access of information and services, building from the experience acquired on earlier projects which also used 3D avatars for sign language reproduction.

As summary, ViSiCAST workflow consisted of a series of transformations:

- From text to semantic representation.
- From semantic representation to a sign-language specific morphological representation.

[1] http://www.visicast.co.uk/

- From morphology to a signing gesture notation.
- From signing gesture notation to avatar animation

Each of those transformations spawned several smaller projects in order to iteratively increase the performance of the system. One of them is the TESSA System which was developed in cooperation with the UK Post Office to increase access to customer services. TESSA used commercially available speech-to-text software customized in order to capture over 450 of the most frequent transactions. The system workflow is summarized in Figure 3.1.

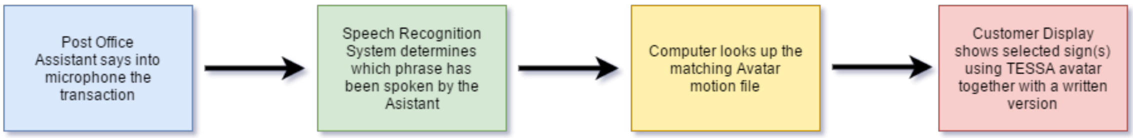


Figure 3.1: TESSA Workflow [25]

Each of the sentences recognized by the system is either complete on its own or has some placeholder that is filled accordingly to the input using a rule-based system. Feedback was favorable, but the lack of facial expressions and the need of more clear handshape animation was noted.

Another subproject called Visia [90] was developed around the same time with the aim to provide weather forecast through a browser plug-in. Given the limited field of usage, it was possible to record all possible signs using motion capture to record posture, motions, handshapes and facial expressions as seen on Figure 3.2. This procedure was done for three different sign languages: Dutch, German and British sign languages.


Figure 3.2: Motion Capture of Visia project

In order to make the data accessible from the Internet, the project developed a markup-language (based on HamNoSys representation) that is interpreted by a visualization software running over ActiveX with a simple interface. With regard to natural language processing, it uses the same approach as TESSA: a basic rule-based translation of previously stored sentences. The user interface for this version can be seen on Figure 3.3.

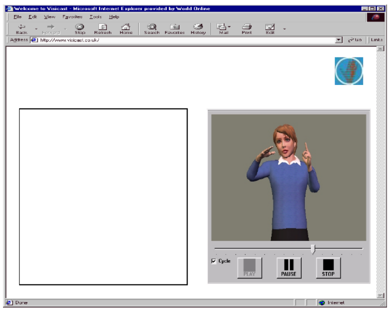


Figure 3.3: ViSiCAST User Interface for weather forecast

To overcome the problems related to natural language processing, further studies were carried out [30]. Those studies explored the linguistic aspects of Sign Languages (mainly British Sign Language) at a morphological and syntactic levels and their difference from those of English. Their approach was based on Discourse Representation Structures which represents each sentence in terms of logical propositions. For example, "*I give X to you*" might be represented as: $[X, Y : i(X)\ \&\ you(Y)\ \&\ give(X, Y)]$. Those structures are just intermediate elements that, subsequently, will be used to generate the signed language representation.

Their work using this representation focuses on representing specific features of sign languages like classifiers, temporal adverbs and grammatical distinction between a single instance of a one-to-many relationship and many instances of the corresponding one-to-one relationship. However, it is acknowledged that this is insufficient to develop a completely automated translation system because the inability of handling ambiguity and anaphoric relationships.

Further development of the project focused on synthetic animations to replace the usage of motion capture [50]. While motion capture provided motion authenticity, it also had its drawbacks. The more prevalent disadvantages were the amount of work involved to capture a large number of signs required for a complete lexicon and how difficult was to modify the motions after their capture. For the animation generation, the system reused the previously established markup language (SiGML) notation.

From the specifications given in SiGML, ViSiCAST system modeled an Inverse Kinematics problem to determine the movement that places the avatar limbs in the designated position. In order to achieve a more natural motion synthesis, the system places a controller variable (usually its angle) in each joint that afterward will be used to model the mass, force, and damping to create a simplified physics model. For a sequence of signs, each of them will be blended with the next, without requiring the avatar to go to the neutral position in-between.

After ViSiCAST three-year research period, the European Union founded another project to continue over its steps. This project was called eSign[2] and tried to improve over ViSiCAST performance, which was reported to produce barely 60% sign understandable by the nine deaf subjects in a test in the Netherlands [97]. However, besides a brief explanation about how it took over ViSiCAST HamNoSys-based notation and added support for non-manual features like movement of body parts shoulders, back, chest, head, and face (eyebrows, eyes, nose, lips, tongue and lower jaw) no further information about testing results or new development was found.

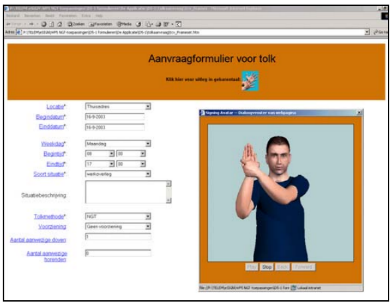



Figure 3.4: eSign User interface

ESign also addresses the need of allowing new content to be created, which is achieved by an editing program by the University of Hamburg. However, as Visia and TESSA were, this translation project is still limited to very specific scenarios and the creation tool only supports the creation of new sentences from previously stored signs. Animation-wise, the project also uses Inverse Kinematics instead of motion capture with the aim to reduce costs. An user interface example is presented in Figure 3.4.

#### PAULA

PAULA is the name of DePaul University avatar capable of translating English into American Sign Language. This project started in 1999 with a system proposal for hand configuration creation [92]. Since its beginnings, the team acknowledged that motion capture, while provides high amount of detail, is not the best solution for sign language animation given the inability of extracting specific phonetic aspects of each sign. 3DS Max[3] was customized in order to create this system which allowed the specification of hand shapes along with their localization and orientation. Figure 3.5 show the user interface for this version.

[2] http://www.visicast.cmp.uea.ac.uk/eSIGN/

[3] `http://www.autodesk.com/products/3ds-max`

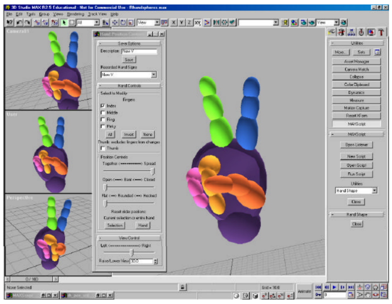


Figure 3.5: Handshape customizer

Following versions focused on turning this sign creator into a complete avatar able to reproduce all needed features for sign language translation. Development process went iteratively, first adding a small set of verbs (27) that only uses one hand for its reproduction (achieving recognition marks above 70%) [86]. Afterward, a version with more variety of signs (170) was developed as a sign language tutor whose interface is represented in Figure 3.6. While the author relates that this version obtained good reviews, not specific were given about the evaluation results. The next step was finger-spelling [87]

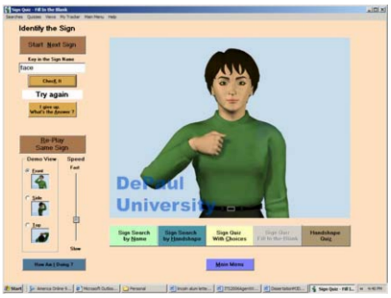



Figure 3.6: PAULA sign language tutor

New versions worked over the linguistics aspects of American Sign Language and how to translate them using a 3D avatar. Special attention is given to facial expressions, which in turn, cause the development of a framework destined to separate and represent syntax, lexical modifiers, and affect of each sign. The studies focus on animations for WH-questions (that is, questions containing the interrogative words what/who/where/which/how) [47]. Tests reveal that at least 60% of subjects were able to perceive clearly which expressions the system tried to convey.

PAULA project also recognizes the importance of addressing facial expressions as a set of simultaneous processes instead of just a set of poses for separated regions [71]. For example, emotions and non-manual features can affect the same area as presented in Figure 3.8.

In order to achieve such expressions, the system uses a weight system where each transformation has a set of tracks and blocks affected by it. The representation used by

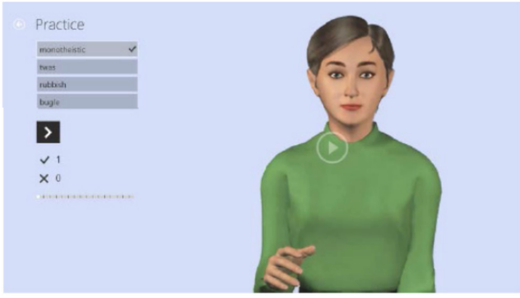


Figure 3.7: PAULA finger spelling software

such systems allows a clear separation between linguistic and animation aspects making possible the creation of an expression builder usable by linguistics with no animation knowledge.

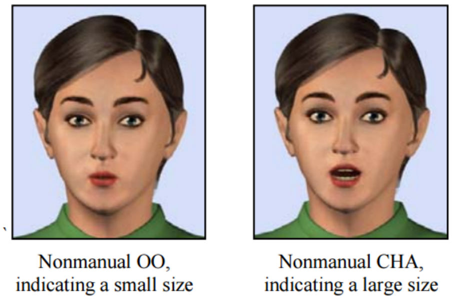


Figure 3.8: Comparison of non-manual features

#### Vsigns

HamNoSys is not the only writing system for sign languages that has inspired automated translators, SignWriting has also been used because its ability to represent movements as they are visually perceived. Even more, it allows the representation of all characteristics of most sign languages without needing any extension or additional features unlike HamNoSys [79]. Vsigns is a project that tried to take advantage from those characteristics.

Vsigns was a web page (not available anymore, but whose interface is shown in Figure 3.9) that using an XML-based format stored signs using their SignWriting representation. For animation synthesis, each sign was separated in a series of ”sign boxes”, which were just the representation of basic parameters like shape, rotation, and orientation. Each of those boxes was used to calculate MPEG-4 body animation parameters which generate a series of keyframes subsequently animated using linear interpolation [63]. No further information is available about which type of translation was used for transforming text into sign language or about evaluation results of the system.

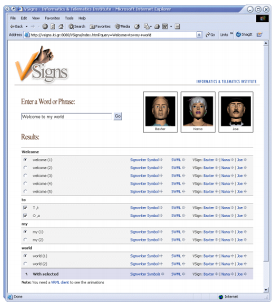



Figure 3.9: Vsigns user interface

### Current works

Nowadays projects are not exclusive of ”major” sign languages. Research has been done on a vast amount of sign languages like Arabic [42], South African [7], Greek [72] and many more. The degree of progress and success varies among them, many affected by the lack of proper corpora or linguistic research. Even more, there are public-available programs for specific and general purposes. Some of them are listed in Appendix chapter A with a brief resume of their main characteristics along with which language(s) they are able to translate.

Also, improvements in machine translation and animation methods have also been advantageous for sign language translations systems.

Research exploring statistical translation has analyzed how to overcome the scarcity of resources to improve results over rule-based approaches. More specifically, phrase-based and hierarchical phrase-based translation have been tested on American and Czech Sign Language [76]. While still not a proper solution, it opens the doors for deeper analysis of state-of-art machine translation techniques into the sign language translation field.

Other approaches included sequence classification[29] and transfer-based [96] machine translation. Where the former approach uses pattern recognition algorithms to associate to each gloss one or more categorical labels that will represent the non-manual components of each sign; and the latter uses linguistic knowledge of both sign and spoken language to produce an intermediate representation that, in conjunction with a rule-based or statistical algorithm, will be capable of transferring syntax (superficial transfer) and semantic (deep transfer) to generate an output on sign language.

There are also studies to discover if the effect of rendering style [49] affects perception of sign language animations, where it was concluded that rendering style indeed does have effect over perception, with non-photorealistic rendering improving the legibility of signs. In the same way, there are studies about the impact of avatar style (stylized vs realistic) [2]. This study did not find out any significant impact over clarity of sign animation that was derived from the style of the avatar, which could suggest that clarity might be influenced by characteristics that go outside the style per se: the textures used in the model, its skeleton structure or how much the user empathizes with the avatar.

Regarding animation, most works make use of some variation of MPEG-4 Controls [46] or Inverse Kinematics [58]. The main focus points of current research are the reduction of robotic movement (that is it, make sign animation and the transitions between them as smooth as possible) and generation in real-time of facial expressions.

## 3.3 Systems for Portuguese Sign Language

First, registered work about automated translation of LGP dates from 2002 by Domingues[26]. This work has as main focus avatar animation of sign languages, therefore addressing none of the natural language processing problems that inherent when translating one language into the other. The animation is achieved with morph target animation using 3DS Max. While the project had interesting ideas including an editor of facial expressions using techniques to currently available grease pencil tools on animation software, this proved insufficient to achieve good feedback on tests. Feedback from evaluation subjects was negative, most of them noting that the system was difficult to use while animations had a score of 3 out of 4. Nonetheless, putting aside the technological aspects, this work noted the importance of more projects for automated translation of LGP, with education being one of the most benefited areas according to the Hard of Hearing surveyed participants. However, even if a need for automated translator systems for LGP was found, there are no registers of any projects for at least 11 years after Domingues work.

Since 2013 new projects have worked over the translation of Portuguese to LGP. The first one them was called *”os meus primeiros gestos”* (my first signs) [54]. This project used motion capture techniques with the help of MotionBuilder[4] along Microsoft Kinect®. It also compares the results of manual animation by keyframes against Kinect-assisted animation, with results similar to previous research: motion capture provides more realistic movements but implies a higher amount of work which might not be feasible for bigger lexicons. As the previous LPG work, *os meus primeiros gestos* does not explore the linguistic characteristics of the sign language, instead, it focuses on the animation processes need to produce signs. Mainly, it studies hand and finger tracking as auxiliary methods for sign creation. Evaluation by Hard of Hearing users noted that speed, bad animation quality and lack of facial expressions had a negative impact on the understating

[4] `http://www.autodesk.com/products/motionbuilder`

of testing sentences. Also, lack of proper grammar structure on the examples (mostly incorrect sentence order) was noted.

2014 was especially active for this research area, with 3 different projects. Two of those are master thesis from Barata [11] and Almeida [70]. They had as a main focus of study how to solve the natural language processing tasks that are involved in the translation of two languages with different syntax and lexical structures as are Portuguese and LGP. Both projects use a rule-based approach coupled with a previous morphological analysis, however Almeida proposed system performs a deeper analysis of the text. To be specific, Almeida performs a complete natural language processing pipeline as can seen in Figure 3.10. One of the most notable characteristics, while not present in the proof-of-concept that was developed, is how her rule-based translation is proposed to be improved using transformation translation by lexical and structure transfer rules. Meanwhile, Barata's system analysis is more limited to an in-house part-of-speech tagging and basic rules to meet the sentence structure required by LGP.

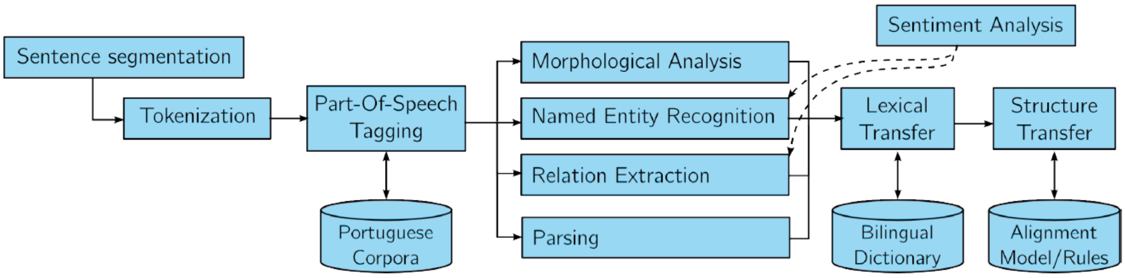



Figure 3.10: Translation pipeline of Almeida's work [70]

Animation-wise Barata's work defines a framework that makes possible to store signs as the set of smaller animations according to the areas they affect, yet it does not define how the animation process was performed. In the other hand, Almeida's work clearly defines that future development would involve an avatar whose rig should have an Inverse Kinematics tree chain so it is possible to place hands and arms into their desired position both manually and procedurally. At the same time, it also defines that individual sign should be blend using linear blending and that interpolation curves should concatenate individual actions for each gesture. For facial expressions proposes the usage of shape keys instead of poses. Neither project published quantifiable results that allow comparing the quality of the output produced.

The final registered project revealed in 2014 was VirtualSign [31]. Contrary to the previous works exposed in this section, VirtualSign is a bidirectional translator (that is it, translates Portuguese into LGP and vice versa). Designed as a learning game that aims to reproduce, as accurately as possible, the movements of real life interpreters in a virtual environment. Figure 3.11 gives a brief overview of the system architecture.

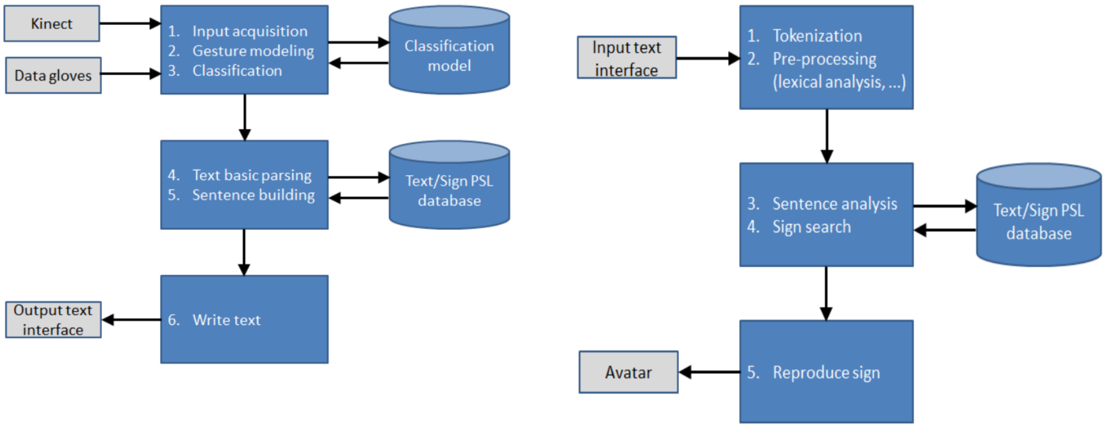


Figure 3.11: VirtualSign [89] architecture

This project, like *os meus primeiros gestos*, uses Microsoft Kinect® for motion capture which is afterward edited on Blender[5]. Blender's Non-Linear Animation (NLA) editor is used to manipulate actions without having to handle keyframes; to be specific, each gesture has its own NLA track to ease the usage of each animation. While natural language processes are specified to be part of the translation pipeline, no more specification of which techniques or what kind of translation method is provided besides to what is shown in Figure 3.11.

## 3.4 Discussion

Automated translation of sign languages is a complex problem that relies on both computational linguistics and animation. However, many projects are focused in only one field instead of seeing the problem as a whole.

Systems that neglect linguistic analysis have had limited success, specially in limited scenarios or used as dictionaries. The lack of a correct grammatical structure hurts the clarity of the system, however, is not always the main concern of the test subjects. One of the main obstacles to improving this situation is the lack of linguistic research on sign languages. And even on those that have deeper linguistic research, as American Sign Language, are limited to rule-based translation methods because their corpora are not big enough to provide acceptable results with statistical approaches. Therefore, improving available corpora would allow deeper research about how sign languages perform while

[5] http://www.blender.org

applying statistical or hybrid approaches that are more aligned with the current state-of-art translation systems for spoken languages.

Animation-wise, manual behaviors have been considered adequate according to the performed research. Nonetheless, sign language translation is not limited to manual features and only in recent years linguistic research has explored the way that non-manual features (specially facial expressions) can be synthesized.

In general, many aspects of translation from spoken languages to sign languages are considered open problems where there is no conclusive research about what are the best methods to perform the different tasks needed to carry about the process. This is specially true for LGP where both linguistic and computational research are in early phases.

# Chapter 4

# System Description

This chapter describes the main characteristics of the proposed system for automatic translations of European Portuguese into Portuguese Sign Language using a 3D avatar. All elements detailed in the following sections are results of the research analyzed in previous chapters. From this point onwards, the system is referred by the name 3DSL (Three-dimensional Sign Language).

First, there is a brief overview about which problem 3DSL is trying to solve. Derived from that aim, there is a series of design challenges attached to it that are detailed along the system features that are considered to be needed in order to provide a solution.

3DSL was conceived using a user-centered design approach. Therefore, besides technical research, there were a number of design tools used for the analysis and design. Those are detailed in section 4.5.

Finally, the rest of this chapter is dedicated to present and describe the architecture of 3DSL. This explanation is done in a technology-agnostic way, that is, exposing the logic behind the work that each of 3DSL modules should execute and the interactions between them, independently of which platform was chosen for 3DSL development.

## 4.1 Problem Statement

While recognized as an official language used in the country, information and education material available in LGP is scarce. Although associations [9] provide education and interpreters for news, events, conferences, or meetings, those only cover punctual scenarios. That is why, in order to cover other daily-life settings, it is proposed an automated translation system to cover the absence of a human interpreter.

After this idea, it becomes natural to wonder which is the equivalent representation of spoken language that can allow hard of hearing people understand it using computational means. Giving the graphical nature of sign language, only something able to reproduce all of its elements would achieve success. Following this train of thought, while there are writing representations for sign languages, animated avatars come as a more intuitive answer to emulate the spatial and motion oriented nature of sign languages. This also derives from the fact that writing systems for LGP are part of the current official academic curriculum for Deaf only since 2007, therefore most Hard of Hearing people are unfamiliar with them [20]; which goes against the goal of making information more accessible to this community.

## 4.2 Design Challenges

However, creating an automated translation system of this nature is not a straightforward task. Any solution needs to couple (in a way transparent to the user) a series of independent elements that can transform any Portuguese input into a clear LGP output. This has to be achieved in real-time without needing specifications above those found on current average computer systems. Moreover, all system characteristics have to be presented in such way that the user does not need any special training to understand how the system works, in other words, the system needs to be usable.

The development of an automated translation system implies a series of design challenges. Input treatment challenges greatly vary according to the specifications of the input. Even more, the information provided by each input type is not the same. For example, audio data has explicit characteristics like voice tone that can completely change the meaning of a sentence without changing its contents. While text can preserve some of those characteristics by the usage of punctuation symbols (e.g., exclamation symbols to put emphasis in sentences), other elements (like emotions of happiness or sadness) are totally lost unless the contents of the sentence are altered to fit the new medium.

On the other hand, input transformation has to deal with the linguistic characteristics of each of the languages involved in the process. At the same time, all levels of language processing have to deal with different degrees of ambiguity [57], this represents another layer of difficulty that should be solved in order to convey the same meaning as the original input.

Finally, animation processes are not as simple as just reproducing random movements. First of all, they have to feel natural enough for the user to understand them, yet cannot be too complex that renders the system unable to reproduce them in real time under a system with average specifications.

## 4.3 Difficulties And Open Problems

Before given the details about 3DSL system architecture, it is important to point out that there are problems that directly affect any system that tries to produce translations between Portuguese and LGP. While the system aims to overcome some of them, others are part of the nature of the tasks needed in order to produce the desired output. Therefore, even if they have negative impacts over how the system performs, it is not possible to totally overcome them until further research is carried out.

**Linguistic studies are still in early phases:** Any translation process deeply relies on linguistic knowledge about the languages involved. While different approaches require different information in order to obtain their results, all of them need to know some characteristics of the language in order to apply the best algorithms for the task.

For example, rule-based approaches require a set of rules in order to transform one sentence in language A into its equivalent in language B. However, such rules can only be created after a lexical and syntactical study of both languages.

While Portuguese is a well-established language, with a vast amount of linguistic studies available, this is not the case of LGP. As exposed on chapter 2, the only formal grammar publicly available is more than two decades old [5] and does not characterize all the linguistic elements in LGP. This limits the amount of syntactical structures that can be correctly translated into those languages.

**Lack of comparative studies between sign languages:** Not all languages have been given the same amount of research, some have many materials available while others barely have any. Comparative studies between languages would be helpful to discover linguistic phenomena shared between them, specially in the case of sign languages that are related or were derived from a common ancestor. In that way, if common characteristics were found it would be possible to take advantage of existent research in languages whose linguistic studies are not as developed because the lack of language stabilization or because the number of speakers is reduced.

In the case of LGP, there is some relationship with Swedish Sign Language, which at the same time derives from British Sign Language. Therefore, comparative studies between those languages could spot common characteristics that would expand the linguistic knowledge of LGP; this, at the same time, would improve the number of characteristics that could be translated between LGP and other languages.

**Scarce corpora:** As was previously exposed in section 2.4, monolingual and annotated of LGP is scarce and mostly only available in a dictionary-like video form. This impacts natural language processing algorithms that could take advantage of the relationship between words in sentences and other knowledge in order to improve their accuracy. Likewise, without corpora, there is no data to properly train and test learning algorithms.

Building corpora is not a trivial task as simple as just gathering random information. In order to assure quality, there are guidelines that need to be followed to have a clear, balanced and useful corpus that can be used by natural language processing techniques. Even more, with sign languages, it needs to careful select storage structures that preserve its linguistic characteristics, yet allow an easy query over its contents.

**Regionalisms:** A common aspect on languages is the existence of differences on language usage between different regions. Those regional differences might be small, therefore not interfering with the understating of what is being said, or they can imply a change in the meaning of a term (or use a different term altogether).

This phenomenon is also seen in LGP, with two main dialects: Porto and Lisbon [5]. In spite of being recognized as two different dialects, there is no clear documentation about their differences nor about which signs belong to which dialect. Such lack of information prevents the development of a ”neutral” translation system.

**Language ambiguity:** Natural languages are ambiguous by nature. This fact has a direct impact in any technique used for processing any language. Furthermore, ambiguity is not limited to the ambiguity of a single word (that is it, lexical ambiguity), it can also entail in syntactic, semantic, anaphoric or even pragmatic types [57].

While some statistical learning techniques improve the results while ambiguity is present [62], they need bilingual parallel data to be applied. As exposed in the previous section, such data is not available for LGP.

**Acquiring animation data:** Animation data has to be generated in some way, either by motion capture, keyframing or any other technique.

Each approach brings about different advantages and disadvantages yet, all need to balance quality and costs. It does not matter the technique, the more natural and accurate animations are, more expensive to generate they are going to be. This is not limited only to a monetary sense, but also to how much time will be needed for their creation. For that reason, the number of terms that an avatar will be able to produce will be tightly connected to the number of animations the system development can afford to generate.

**Facial expressions:** Although earlier chapters pointed out the importance of facial expressions in sign languages, their production in real-time is not an easy task. First, to allow facial expressions more advanced modeling of the avatar also implies an overhead over the general animation process. This means that facial expressions might require a more powerful hardware to run than a system without them.

Likewise, research about the specifics of how facial expressions affect sign language production are far less abundant than those focused on manual gestures.

3DSL design aims to provide a way to make animation data acquisition and facial expression reproduction easier by the usage of SignWriting to build each sign from a limited set of animations instead of having one animation per each sign. At the same time, aspires to use the available linguistic studies to create a rule-based system that carries out structure transfer. This will be done such that translation is not limited to just reproducing signs, but actually reproducing them in the order a native LGP signer would expect.

## 4.4 3DSL

3DSL is a system designed to translate Portuguese into LGP using 3D humanoid avatars. Its design process was focused in a user point of view, in other words, it was conceived with a user-centered approach. There was a study of the requirements and characteristics of users and those findings influenced the specifications of the system. Also, while not directly embed in the system, 3DSL specifications also acknowledge the need for an interface that allows the insertion of new terms by non-technical experts.

The general specifications of 3DSL architecture should be as technology agnostic as possible, such that it is possible to adapt the system to different platforms with the minimal amount of rework. Even more, the changes of the internal working in a module should not compromise the system as a whole as long as the module meets the architecture conventions.

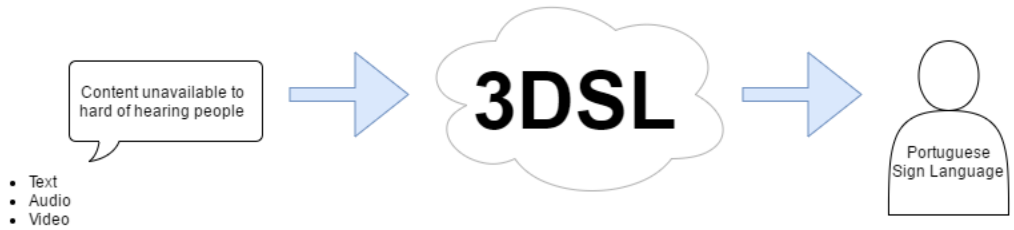



Figure 4.1: 3DSL Overview: User's Perspective

Figure 4.1 characterizes 3DSL from a normal user's point while the following sections will describe the specifications of 3DSL logic and design. This is done from both a user experience (UX) and technical point of view. Everything described in this chapter is from a general point of view, specifics about 3DSL practical implementation are described in the next chapter.

## 4.5 User Experience

It does not matter how accurate or fast a system is if users do not feel comfortable while interacting with it. While closed related to visual design, the way user reacts when interacting with the system also involves other aspects like usability, accessibility, and performance.

Therefore, in order to improve 3DSL's UX, a series of analysis tools were used to study the characteristics and needs of the target users. This section is dedicated to describe how those tools were used to guide the design process.

### Target Users

As described before, 3DSL main goal is to help hard of hearing people improve their communication skills in an easy and simple way in hopes of providing an equal access to information.

However, as studies have shown [67], this community is very diverse. Such diversity has many levels including (but not limited to) age, level of hearing, educational background and communication methods. Under such diversified group of people, it becomes of vital importance to clearly identify which subset is going to be addressed as the target audience. Moreover, this will help the system to focus better on the user's wants and needs, therefore shaping which aspects of the system are going to be the main focus at design and development time. Because all of this, 3DSL target audience was defined in several steps according to the most representative characteristics of the community.

The first step was to identify which demographic profile would provide the most general system, it is of special importance that 3DSL reaches a broad audience. Yet it should preserve an ability to be extended for future adaptation to meet the need of a more specific demographic. As results of this, it was chosen that young and middle-aged adults adapt to this purpose.

Young and middle-aged adults do not only represent more than half of the current Portuguese population [19], but also the deaf and hard of hearing people in this age range make up around 2.72% and 4.41% of Portugal's population according to the World Health Organization [95]. Similar numbers are seen on other countries from inside [1] and outside [53] European Union. Furthermore, according to Central Intelligence Agency numbers, more than 95% of the population are literate (can read and write). Likewise, school life expectancy (primary to tertiary education) is 16 years; numbers also indicate that this group, who at least completed high school, have a computer usage of 96% and Internet usage of 94% [48].

While there are not specific numbers available about hard of hearing people, it is estimated that educational attainment levels are lower than those seen on hearing people [45]. However, the approval of Decree Law 3/2008 guarantees bilingual education for hard of hearing people in Portugal; this fact should mark an improvement of education levels of this community.

After the study detailed above, target users can be described as follows: young and middle-aged adults deaf or with any degree of hearing loss that have at least basic computer skills, are able to follow short and simple instruction and are capable of understanding LGP.

Those users are characterized for using LGP as their main form of communication, therefore, it is desirable for them that any output is provided in this language. They might or might not have written Portuguese knowledge thus graphic metaphors that help them to identify the main elements of the system are essential for using the system successfully. A detailed persona is given by Figure 4.2. This persona was created to help guide decisions about system features, navigation, interactions and visual design.

While there are other groups of hard of hearing people that also require systems that allow them an equal access to information (mainly, hard of hearing people whose primary language is Portuguese instead of LGP), their needs are different from those that would be solved by animated avatars that use sign language. This group would benefit from approaches oriented to closed captions which are outside the scope of 3DSL designed

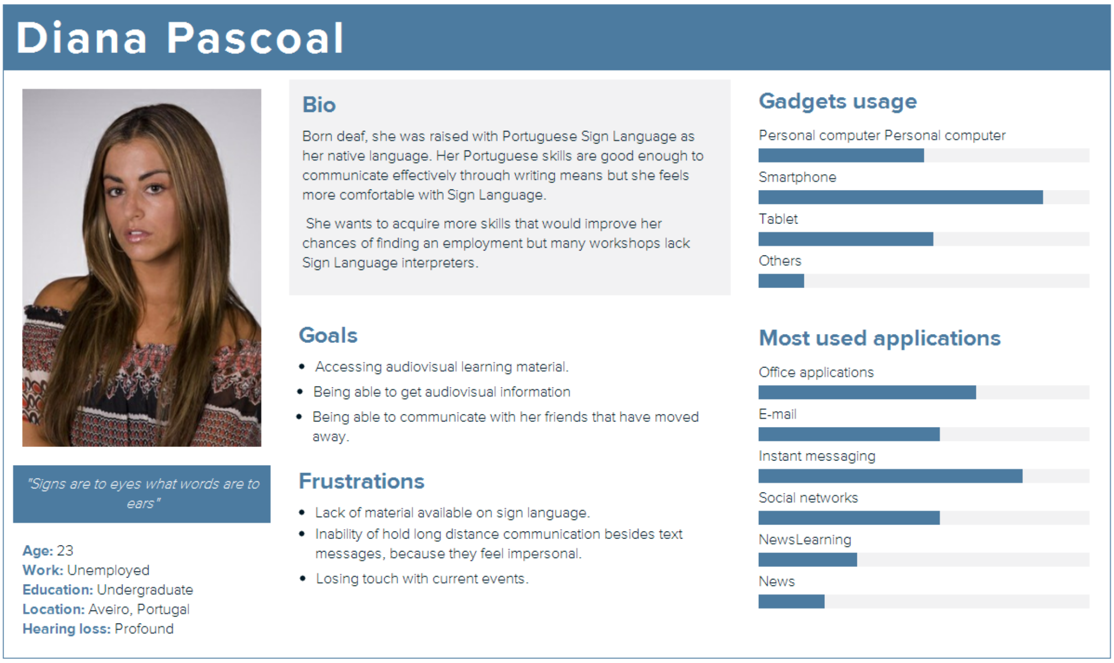



Figure 4.2: Persona designed to represent the user archetype of 3DSL

features.

### User Goals, Tasks, and Actions

In general, every user has a goal in mind when using a system, an end result that wants to achieve. However, those goals are usually complex and require a set of different activities to be executed. Each of those activities, or tasks, are composed of a series of steps that are needed to be undertaken to successfully complete them [8]. The user is called by the name of its representing persona, Diana.

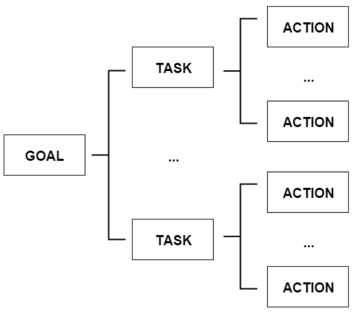



Figure 4.3: Graphic definition of the Goal-Task-Action dependency

3DSL main goal is broad and needs to be decomposed in smaller user-related goals in order to identify which tasks will be needed to achieve them. In concordance with the designed persona, the goals identified are:

- Accessing audio-visual learning material.

- Being able to get access audio-visual information in general (specially news).
- Being able to communicate with friends with hearing loss that live far.

Given those goals, the next step is to develop task scenarios for the sake of identifying the tasks needed to achieve each of goal. Those scenarios will describe the stories and context behind why a specific user or user group uses the system to achieve a given goal. As with the goals, scenarios are written using the persona that was developed previously.

**Accessing audio-visual learning material:** Diana has been interested in learning about how to edit photos. She found several tutorials on the Internet, however, the most useful ones were video-tutorials (usually without any kind of captions).

**Being able to get access audio-visual news:** Rodrigo enjoy being updated about what is going on around the globe. Reading news does not impose any problem for him, however, there are times when he would like to watch live streaming instead of waiting that someone else writes the news. Also, it would make him able to directly know what some said in an interview, instead of waiting that someone else comments about it.

**Being able to communicate with her friends:** Diana used to communicate every day with her friends Luis and Cristina, and even after they moved out they still kept in touch. Technology has helped them to keep in contact, nonetheless, they are not able to express themselves as fluently as before. They can write in Portuguese without any issue, but using Sign Language feels more natural for them; this does not seem like a problem since nowadays there are webcams, but the quality of video sometimes makes almost impossible to understand what the other party is trying to convey.

Scenarios 1 and 2 are closely related to 3DSL goal of enabling information access to people with any degree of hearing loss, and there is a clear underlying process that can be extracted from them:

1. Search topic of interest.
2. Check if the results are available in a medium that the user can understand (sign language).
3. If there is an available medium, then proceed to access the information accordingly.

4. If not repeat steps 1 and 2 until there is an adequate result, or all results are checked...

Nowadays, task 1 is easily covered by the usage of search engines and pose almost no issues to any person with basic computer skills. On the other hand, task 2 can become in the biggest wall between users and the information that they want to access.

3DSL proposal is to provide an additional task on this task that will eliminate step 4, replacing it with a new one: use 3DSL to translate the information.

3DSL tasks will be composed of a set of simple, and short actions:

- Open 3DSL.
- According to the kind of input, select the appropriate option.
- If necessary tweak any options of the system that will improve the experience (speed, change avatar).

When those actions are completed, task 3 can be executed and users will be able to access the information they require.

### Usability Principles

An important feature for 3DSL is usability. Not only is wanted to bring information to users, it is also wanted to do in an effective way that will make the system useful to them. With this in mind, 3DSL has been developed around the following usability requirements:

**Effectiveness:** the system should be accurate, minimizing the opportunities for errors and protecting the users against them. This is achieved in 3DSL with a consistent and simple design. All graphical elements are designed to be clear and separated in groups according to their relationships.

**Efficiency:** Minimize the number of actions required to the user, the less intrusive 3DSL is, the better.

**Easy to learn:** Make the usage of the system as intuitive as it can be, the user should feel comfortable using the system as soon as possible. This is achieved by using familiar metaphors such that the system is as intuitive as possible. Associating each action like ”playing”, ”stopping” or ”pausing” an animation with well-know graphical representations.

**Engaging:** It is not enough to make 3DSL functional, but also try to make the layout as simple as possible and using easy-to-read fonts coupled with eye-catching colors. Even more, the possibility of a mobile version that expands the situations where 3DSL can be useful.

## 4.6 System Architecture

In concordance to what was exposed earlier, the activities needed to be combined in 3DSL execution flow can be divided into three clearly differentiable groups. Each of those composes an independent module whose internal workings are unknown and irrelevant for the work of the others. The modules are named according to the task they are expected to perform: information reception, transformation, and animation. The internal processes and how the information flows from one module to another is described in more detail in the following sections. A general overview of this architecture is presented in Figure 4.4.

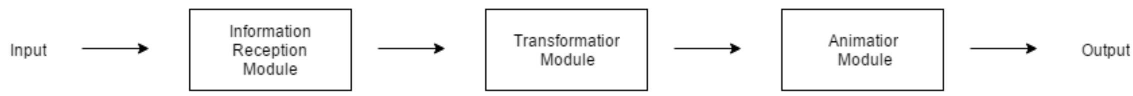



Figure 4.4: 3DSL architecture overview

### System pipeline

Before exploring in detail how each module works internally, it is important to understand how the information flows through the system. The reason behind this is that after understanding why each module is needed for the overall function of the system, it will become more simple to explain how those modules need to work in order to meet the system requirements.

First of all, any translation system needs to capture the data that it is going to be translated. However, 3DSL allows different data types with different characteristics that might need further processing before being usable in future steps. As a result, all data should be unified into a general structure that preserves the most relevant characteristics but allows the rest of the system to be oblivious of how the data was obtained. It was chosen to call this structure *StoredData*, and its internal structure can be seen in Figure 4.5.

In previous chapters, it was discussed how automated translation between sign and spoken languages was not as simple as just animating a sign for each word in sentences.

| **StoredData** |
| --- |
| + Host: identifier of 3DSL host system |
| + Data: raw input data |
| + Text: text representation of Data |
| + InputMode: describes the type of Data |
| + Language: language of Data |

Figure 4.5: StoredData structure

Consequently, it has been acknowledged that natural language processing methods are needed to obtain a sentence whose meaning is as close as the original one. The end result is represented in gloss notation (as described in section 2.5), noting that there might not be a one to one relationship between the glosses and the words from the original sentence.

At last, the only thing left is to take all glosses and translate them into a 3D gestural space. In this step is when the system takes advantage of SignWriting to represent each gloss as phonetic aspects of sign language (like postures and activities). Meanwhile, those aspects will be used to generate the corresponding animation. This step is needed because a gloss just gives written representation of the set of Portuguese concepts that a sign symbolizes; however, it does not give any phonetic information about how those concepts are represented on a 3D space like LGP's.

Everything that just exposed is graphically represented in Figure 4.6.

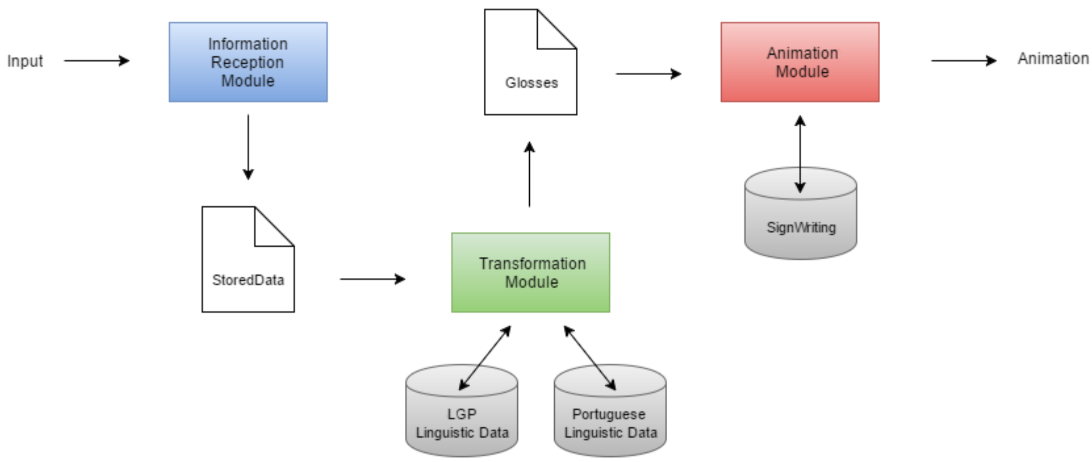



Figure 4.6: General pipeline of 3DSL

The following sections will take a close look at the three main modules that have been presented. Each module is described in function of its main goal, input, output and the internal procedures involved to make it work.

### Information Reception Module

This module, as it name suggests, is in charge to capture the data needed for the system to work. However, any process needs to produce that data is outside of the scope of this module. The reason behind this is that the specifications about how to obtain the data are tightly related to the host system that is running the system (e.g. the hardware and software resources available for audio capture are completely different between desktop and mobile devices). Following this idea, the Information Reception Module is expected to receive any kind of data type that is recognized by the module, but is not directly connected to the user interface that introduces the system input. There is a platform-specific middleware in charge of the connection between the GUI and 3DSL.

The main goal of this phase is to standardize all the different input types into a unique data type that will abstract the following phases about what kind of input the system received, yet it will allow them to access the specifics of each data type if needed. This data type is named *StoredData*. In general, all data types will be transformed in text as part of the conversion process.

How the standardization process will be carried out varies according to the input type:

**Text:** no further treatment is required for text data, it is directly stored in a *StoredData* instance as it is.

**Audio:** The process of transforming audio into text is known as Speech Recognition [12]. It is a field of study which involves several disciplines, mainly linguistics and computer science. Any system using Speech Recognition will need to achieve two tasks: process audio signals and associate them with a set of words that conform the language represented by those audio signals [94].

This association is done using a speech engine. The specifications of how the engine works vary between implementations, however, the general Speech Recognition process can be summarized as follows:

- The engine will be feed with a set of words that should be recognized by the system, this set of words is usually known as grammar.
- Additionally, other elements related to the physical properties related to those words are also fed to the engine. Those properties represent the acoustic model, and systems might need to be trained to accurately recognize words according to how the speech engine is designed. This is a consequence of the

high variety found on the human voice, every person has a different voice; even more, the human voice can vary from the same person because a wide range of causes, like the person mood [65].

- After previous elements are set, the speech engine is ready to start processing audio using techniques like statistical analysis, pattern matching or neural networks.

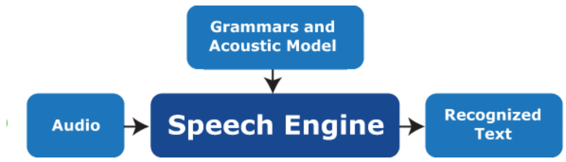


Figure 4.7: Speech Recognition process [52]

Currently, Speech Recognition is still a maturating field, which many research efforts trying to improve its accuracy and efficiency. Nowadays there are several tools that can be used for SR, ranging from command languages focused on it like Vocola[1], to specialized software like Google Now[2], and several API like SAPI [3].

**Video:** For video treatment, the system will transcode the video in order to extract its audio track which will be converted into text using the process previously described.

**Handwriting:** For handwriting recognition, the process is divided into three general steps: preprocessing, feature extraction and classification [68].

Preprocessing usually consists of the process of discarding irrelevant information in order to increase the performance and accuracy of the process. Common tasks in this step are binarization, normalization, sampling, smoothing, and denoising.

In contrast, feature extraction obtains all the important information needed for the classification process. Common highlighter properties include aspect ratio, number of strokes and average distance from the image center. If the case of automatic conversion of the text as it is written instead of image to text conversion (online versus offline handwriting recognition), the system can also take in account pen pressure, velocity or the changes of writing direction.

Finally, the classification process uses a learning model to map the instance into one of the valid characters recognized by the system.

[1] http://www.lumenvox.com/products/speech_engine
[2] https://www.google.com/chrome/demos/speech.html
[3] https://msdn.microsoft.com/en-us/library/ee125077

Information Reception Module sends the standardized input to the next module, which is where the real translation process will be executed.

### Transformation Module

Taking any instance of *StoredData*, the translation module will output a sequence of glosses that are expected to express the same idea represented by the original input. Machine translation research for sign languages has explored statistical and rule-based approaches. Each of those approaches has different advantages and disadvantages [24] which are summarized in Table 4.1.

Table 4.1: Comparison between rule-based and statistical machine translation approaches

| Rule-based | Statistical |
|---|---|
| Based on linguistic theories, therefore needs well-defined linguistic rules and dictionaries | Minimal or no linguistic knowledge required, needs parallel corpora instead |
| Translations are language-dependent | Trained with human translations independent from the language pair of languages |
| Expensive to maintain and extend given all the linguistic knowledge needed and natural language open problems like disambiguation | Easier to maintain if (enough) data is available |
| Acceptable performance even with limited resources if rules are correctly defined | Low performance if not enough data is available. Language pairs with different morphology are specially problematic |

While statistical approaches have been proven to be better for machine translation between several different spoken language pair [16], the results have not been as successful for sign language studies [75] [15] [62] [76]. The main reason behind this is the small size of available corpora, even for languages like American Sign Language, which have access to a significantly higher amount of resources. In contrast, as could be seen on projects described in chapter 2, rule-based approaches have had acceptable results even with limited linguistic knowledge. Based on this fact, 3DSL proposes a rule-based approach.

Generally, rule-based methods can be divided according to the process used to generate the output. This can be appreciated in Figure 4.8.

Direct translation is not suggested because the morphological differences between LGP and Portuguese. Meanwhile, transfer-based techniques take advantage of the limited linguistic knowledge available without the added complexity of interlingual representations that require deeper (and sometimes domain-specific) knowledge [27].

Following this idea, 3DSL proposes an architecture for machine translation that uses transfer rule based principles. Figure 4.9 shows an overview of the whole process. The module tasks can be divided into two phases: analysis and restructuring.

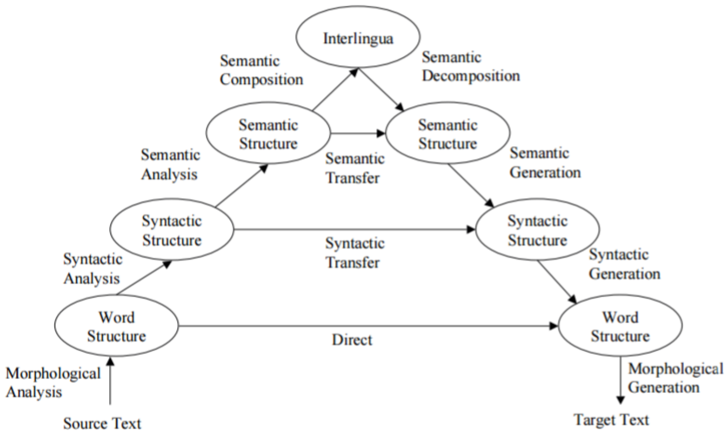



Figure 4.8: Rule-based machine translation [27]

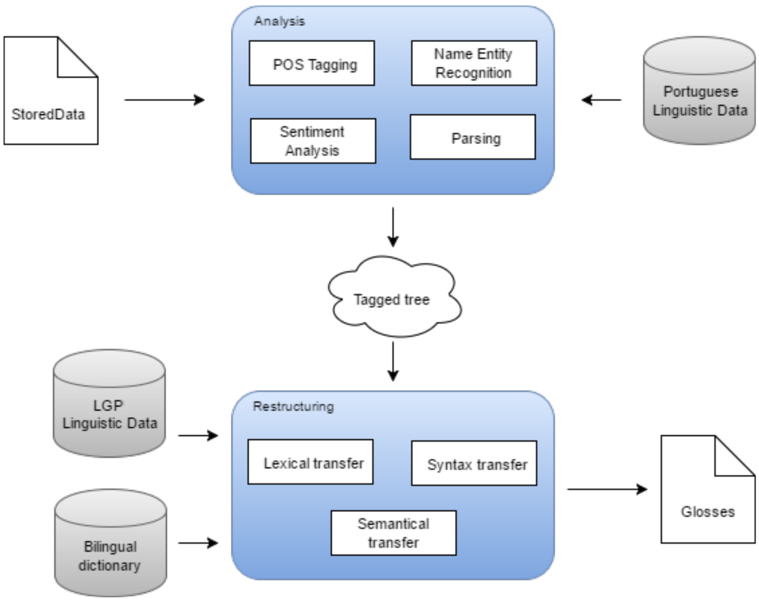



Figure 4.9: Transformation Module diagram

### Analysis Phase

This phase encompasses all morphological and syntax analysis steps that supply the needed information for the structure align that will map a sentence from one language into the other. The performance of this phase will impact directly the quality of the mapping done by the restructuring phase. An important note is that all the process involved in this phase only require monolingual (Portuguese) tools.

First of all, each sentence will be normalized and tokenized. Each token will represent a word or punctuation symbol. As part of the normalization process, orthographic errors will also be corrected.

Following this, every token will be identified according to its word-category. In other words, based on its definition and its context, it will be classified as a noun, verb, adjective or any other category valid in the Portuguese language. Part of Speech (PoS) can be used to identify general groups or to identify sub-categories, e.g. not only identifying that a word is a verb but also its case, tense, gender, and person. Different methods exist that specify how the tags will be identified and how ambiguity will be solved, some include hidden Markov models, dynamic programming, rule-based, stochastic, and neural networks. A comparison between them can be found in studies like [22], in general, most modern approaches achieve accuracy above 95%.

Additional information can be extracted in order to provide more syntactical and semantic information to the restructuring phase.

Name entity recognition helps to identify how tokens would be treated by the system. Some categories, like person and organization, are more likely to be spelled using dactylology while the identification of dates and quantities affect the signing space used within the sentence and the symbol order. It is possible for other processes to also be benefited by name entity recognition, however, more linguistic studies are needed to prove this hypothesis. Name entity recognition is still an area in evolution, whose general-purpose systems achieving moderate success in Portuguese [4].

Another important process to be applied is sentiment analysis. Sign languages are gestural-rich by definition, and emotions are modifiers of how words and sentences are signed. They are mainly expressed by facial but can also imply other non-manual gestures [47]. Unlike the previous process which involved only the text representation of the original sentence, this step can be enriched by additional analysis of other elements present in audio-visual formats [93].

However, while all this information enriches the mapping process it is insufficient to identify the structure of a sentence and transfer it to LGP. Parsing the sentence will convert it into a tree that represents exactly how the words are joined together. More specifically, parse trees can be divided according to how they define the relation between the words, two of the most used in literature are constituency trees and dependency trees. Constituency trees break the sentence into sub-phrases from which leaves will hold their respective tags but the rest of the tree will be unlabeled. Meanwhile, in a dependency tree every node represents a token where its children are the words that are dependent.

While having an Object-Subject-Verb word order in most cases, LGP can present flexible word order structures. Furthermore, sign languages present special structures (absent in spoken languages) called classifiers that help to clarify a message or highlight specific details about it [5]. Classifiers representation and free word order languages could benefit from dependency trees since they take advantage of syntactic functions but use a

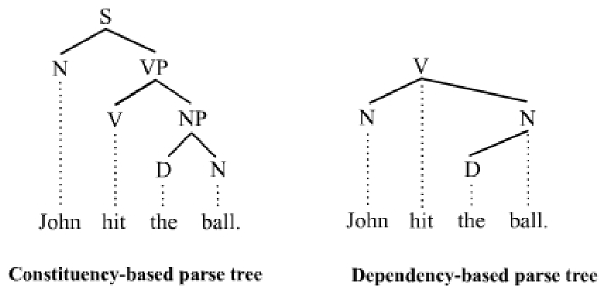



Figure 4.10: Constituency and dependency parse trees comparison

simple representation. Finally, the tree representation of the sentence will be fed to the next phase of the process.

### Restructuring Phase

According to the linguistic information obtained in the previous phase, 3DSL will proceed to transform Portuguese into LGP. This transfer has three levels: lexical, syntax and semantic.

This phase, in contrast with the analysis process, requires bilingual information along with linguistic information about LGP therefore, the performance of this step is limited by the available data.

A first proposed approach is a very basic rule-based reordering according to the current studied linguistic phenomena [5]. This set of rules can be extended as more research is carried out in order to treat more complex sentence structures.

The output will be defined by a set of glosses that directly correspond to the LGP sign that will be animated. This means that elements that do not exist in sign language (e.g. prepositions) were already erased. In the same way, terms that do not have a one-to-one translation are already treated, whether this means that separate words become compounds (e.g. *bom dia* becomes BOM-DIA instead of BOM DIA) or classifiers need to be applied (e.g. *abrir* verb who does not exist per se, but acts as a modifier of the subject). All additional information about the words (name entity recognition, sentiment analysis, and others) is also inherited by the glosses.

## Animation Module

Animation module will be responsible for transforming all the previous work into the final output. It has to transform the list of glosses given by the transformation module into an animation representing the translation into LGP of the original sentence. In order to create the 3D animation there are two separate components: the model and its animation movements.

### Model

A 3D model represents a physical body using a collection of points in a 3D space. There are two primary types of 3D models, the most apparent differences being in the way they are created and manipulated [73].

**Curve models:** A Non-uniform rational B-spline, or NURBS surface is a smooth surface model created through the use of Bezier curves. To form a NURBS surface, the artist draws two or more curves in 3D space, which can be manipulated by moving handles called control vertices along the x, y, or z-axis.

**Polygonal models:** Points in 3D space, called vertices, are connected by line segments to form a Polygon mesh. A mesh is just a collection of vertices, edges, and faces that defines the shape of a polyhedral object.

These models can be created by hand, algorithmically, or scanned. 3DSL proposes the usage of polygonal models for its animation architecture.

When an avatar modeling is finished, it is a static 3D mesh, almost like a marble sculpture. Before a 3D character model can be animated, it must be bound to a system of control handles so that the animators can pose the model. This is achieved by rigging, where a character rig is essentially a digital skeleton bound to the 3D mesh. Like a real skeleton, a rig is made up of joints and bones, each of which acts as a ”handle” that animators can use to bend the character into a desired pose [74].

3DSL is only concerned with the specification of bones concerning the upper part of the body according to the articulatory space defined by Figure 2.4.

### Animations

In order to transform the list of glosses into proper LGP the first step is to translate the glosses into its SignWriting representation making use of a SignWriting dictionary. Each sign is not limited to just one element of each type. Compound signs might be composed of a series of multiple basic configurations. In Figure 4.11 is possible to compare the representation of compound signs against the basic pieces that form it.

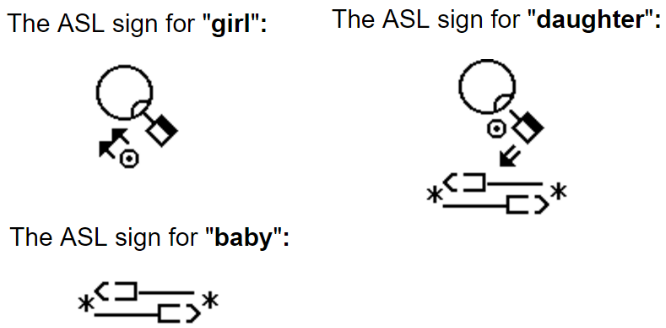



Figure 4.11: Basic and compound signs comparison

Therefore, it is necessary to firstly identify how many clusters make up each sign. Afterward, according to SignWriting definition, each cluster can be described according to its morphological attributes.

3DSL will map those characteristics into a 3D avatar using animation layers. Figure 4.12 represents this distribution. Each layer will be in charge to animate a specific set of attributes:

- Face: head and faces
- Torso: body
- Right and left arms: movements and dynamics according to the arm involved
- Right and left hands: hands according to the hand involved

Each cluster will represent the signer's posture and non-facial elements as a set of key frame poses that at the same time will be characterized according to the activity described by the movement attribute. Those poses will be animated interpolating the values between the points of each keyframe.

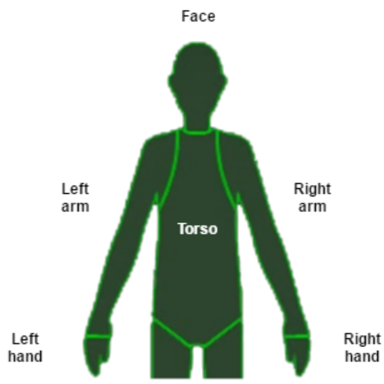



Figure 4.12: Animation layers

However, this process only defines how to create a single sign. In order to transition between them, it is necessary to use blending. In this process, the transition will use the join animations of both animation clips. To be more precise, each joint position will be equivalent to $(signA * wA) + (signB * wB)$, where $wA$ and $wB$ correspond to the weight of each animation clip in a given time. Initial weights are $wA = 1$ and $wB = 0$, and during the transition process $wA$ will go down until it reaches $0$ (and vice-versa).

## 4.7 Dictionary Updater

One of the required elements in any translation system is a bilingual dictionary. However, for a visual language as LGP comes the question: how to make the dictionary easy to update such that there is no need to consult a developer for doing it?

Using SignWriting any LGP speaker may be able to write any word into the system which will be able to store it as the set of cluster elements needed for its eventual animation. The only major obstacle for this process is the lack of a standard usage of SignWriting, which might increase the difficulty of identifying the clusters that compose a sign. Yet, this problem can be minimized through the imposition of a grid layout that makes obligatory the visual separation between each cluster. This will create a updater similar to SignWritting Studio™[4] interface but instead of creating an CSV or XML representation it will store the gloss as a set of parameters described in the previous section.

[4] http://signwriterstudio.com

# Chapter 5

# Implementation

This chapter first describes the complementary background study that was carried out as part of the development process. This is followed by a brief description of the iterations needed in the design process of 3DSL's user interface from both a visual design and usability point of view.

Afterward, the chapter discusses the implementation process of the system, coupled with an exploration of the technologies and standards used for its development. For each module developed it is exposed which features were implemented, any tools and external dependencies that were used, and a general description of how they are integrated into concordance with what was exposed in previous chapters.

## 5.1 Background Study

In addition to the UX research detailed on chapter 4, a short survey was conducted in order to further explore the characteristics of 3DSL prospect users, along with their opinions and expectations about avatar-based translated systems. A sample of this survey can be found on appendix chapter C.A total of 27 people aged between 11 and 56 years answered the survey as shown in Figure 5.1.

Of those, a total of 9 people had some degree of hearing loss ranging from mild to profound. To be precise, 3 people presented mild, 2 moderate and 4 profound hearing loss. Meanwhile, the other participants did not have any kind of hearing loss. The hearing distribution is detailed in Figure 5.2. Out of all participants, only 2 indicated that they use some kind of device to compensate their hearing loss.

While 100% of the participants indicated that their highest completed level of education was at least high school, only 9 of them indicated that they do frequently use of computers

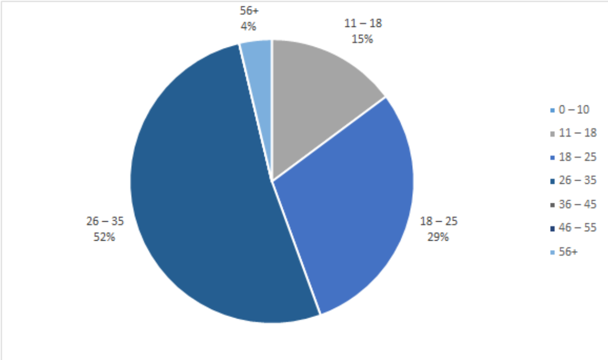



Figure 5.1: Survey Participants Age Distribution

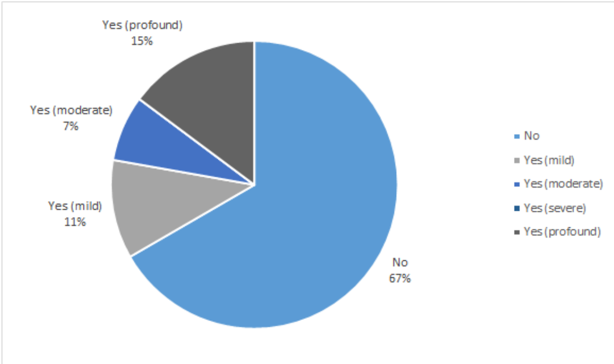



Figure 5.2: Survey Participants Hearing Loss Distribution

or smartphones. The most used applications between those whose use computers are social networks and instant messaging as can be seen in Figure 5.3.

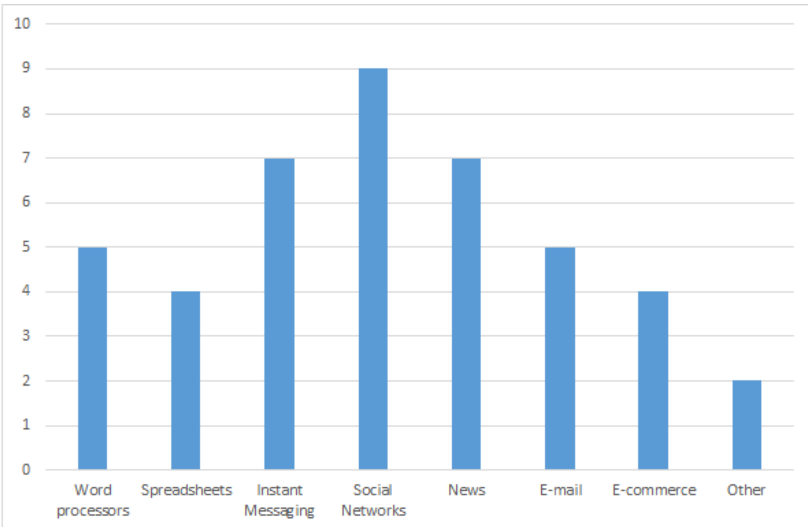



Figure 5.3: Most Used Applications Between Survey Participants

Likewise, all participants showed a positive reaction to the possibility of LGP translation using avatar-based systems. Suggestions and expectations were mainly related to the usability and cost of such system. To be more precise, the majority of people expressed that such systems have to be as simple to use as possible, while also providing high animation quality that allows for easy interpretation of all important elements of LGP (manual and non-manual). Moreover, participants pointed out the need of improving the access to either interpreters or providing close captioning in situations where it is not possible to have an interpreter.

In the end, the conclusions that can be extracted from the results of this survey are aligned with the UX research previously detailed in section 4.5. As a consequence, development was carried out with those guidelines in mind.

## 5.2 User Interface Development

After the user study was completed, 2 survey participants helped in the usability inspection needed for the prototyping and graphical user interface design. Mainly, each user interface iteration was evaluated by the feedback given by those participants as they tried to perform the tasks defined in section 4.5. Mainly using the questions:

- Will the user try to achieve the effect that the task has?
- Will the user notice that the correct action is available?
- Does the user get appropriate feedback?

One of the first questions that arose was how to expose the two different input modes in a simple manner. Analysis done in the prototyping process revealed that each of them warranted a separate set of components in order to work properly. In consequence, the first screen to be designed was the selection screen whose result can be seen in Figure 5.4.

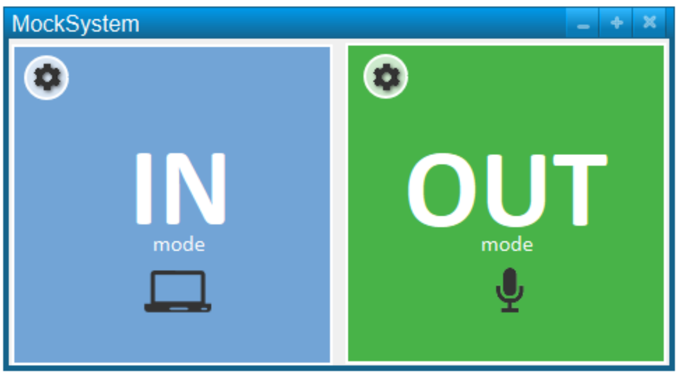



Figure 5.4: First Selection Screen Prototype

Separation was done by where was located the source of the input inside the computer as text (IN mode) or outside the computer as audio (OUT mode). However, this proved to be extremely confusing for users. The idea was scrapped and replaced by directly stating what is the expected input, also the metaphors associated with each option were exchanged for more explicit ones according to user requests. This resulted in two different preliminary designs that are shown in Figure 5.5.


Figure 5.5: Second and Third Selection Screen Prototypes

Afterward, user preferred the simplistic second prototype, pointing out that the third one while informative, was too distracting and overcharged. Screens for each mode were designed following a style according to those suggestions.


Figure 5.6: Text Mode First Screen Prototype

Subsequently, since all input types are processed in a similar way after going through the Information Reception Module (which results in their eventual transformation into text), the next interface designed was the text mode interface. The first prototype can be seen in Figure 5.6.

However, users declared that even if it was simplistic, the interface was not clear enough. Also, following the example of other applications already available for other sign languages, they suggested the addition of exact text input that was going to be translated. All changes were taken into consideration and after some iterations, the interface took a look and feel similar to a video player. This design is shown in Figure 5.7.

Other parts of the system were also subject to evaluation like the loading screen and

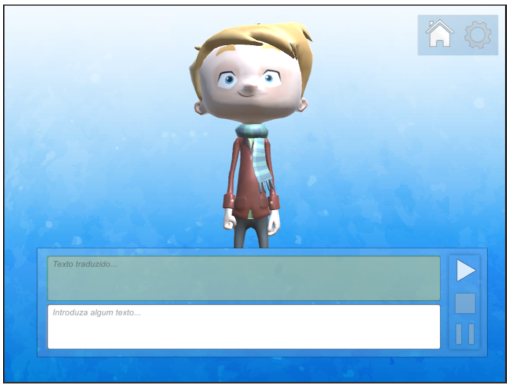


Figure 5.7: Text Mode Screen Final Design

options panel. Additionally, some designs for a mobile version were evaluated, but none of the designs evolved into a functional version in consequence of the development focus on the desktop version.

## 5.3 System

As to prove the practicability of 3DSL architecture, a system was developed. This development was inspired by the previous projects on avatar-based translation that were detailed on chapter 2. In particular, the system development was focused on providing a complete workflow of text transformation into a set of animation triggers capable of reproducing grammatically correct LGP sentences.

According to the user characteristics, it was decided that a desktop application was more acceptable for a first approach. However, mobile capabilities are especially important nowadays, making it desirable to carry out development in such way that it could also be ported to other platforms. For that reason, 3DSL was developed on Unity[1]. This game engine provides an environment where not only desktop development is possible but also supports porting to Android and iOS platforms with relative ease.

Taking this into account, the following sections detail the characteristics and implementation choices of this system. All development was done on Unity $5.4$ using C# as scripting language unless stated otherwise.

### Features

In concordance with the goal of improving the communication skills of deaf and hard of hearing people, the main functionality of 3DSL is the ability to transform an input (text) into

[1] https://unity3d.com/

LGP. Derived from this fact and following the architecture design defined on section 4.6, it is possible to outline the features as:

- Automatic translation and animation.
- Basic rule-based word reordering following LGP grammar rules.
- 3D avatar animation using animation layers.
- Signs can be generated from a relatively small set of animations.
- Can be ported to other platforms thanks to Unity 3D's cross-platform features.

### External Dependencies

The quality of the final translation will be dependent on many factors that influence each step of the process. While some of them are tightly connected to the implementation, others are outside control of the system because they either depend on external components or the quality of the provided data. Among those factors are:

**Spelling and grammar errors:** While some probabilistic error correction can be implemented, no method provides 100% accuracy. As a consequence of an incorrect input, the output can change radically from the expected result. Even something as small, as a case change, can comprise the proper tagging on future steps.

**PoS tagging:** It is necessary for the translation process that each word is properly identified in order to provide a correct translation. This is true for both the reordering and look-up process. It is also important for the differentiation between homonyms like *casa* which can be either a verb (simple present third person singular of the verb to marry -*casar*-) or a noun (home).

**Dependency parsing:** Structure of the resulting parse tree is determined by the relation between a word and its dependents, and from those relationships the system will perform the necessary reordering operations. Consequently, the output quality and system performance are tightly related to the quality of the parse tree used.

**LGP grammar:** Given the lack of documentation, many LGP grammar rules are undefined or ambiguous. Therefore, some inputs might present reordering errors that might not be detected until extensive testing is performed.

**Animation:** while providing a technically correct output, poor animation quality on the clips used to generate each sign can render a correct result into something unrecognizable by end users. Consequently, it is important to guarantee that each clip properly represents the pose or movement associated with its trigger.

In the same way, all elements defined in the difficulties and open problems section (section 4.3) have a direct impact on the system's performance.

### System Requirements

Along the external dependencies previously listed, it is also important that in order to provide an acceptable performance the host system for 3DSL must meet at least these requirements[2]:

**Operating system:** Windows Vista+ with .NET Framework 4 or later.

**Graphics card:** DX9 (shader model 3.0) or DX11 with feature level 9.3 capabilities.

**Processor:** 1 GHz with SSE2 instruction set support.

**RAM:** 1 GB.

**Others:** Internet connection that allows all needed requests to DepPattern[3] and CitiusTagger[4].

### Input

Immediately after starting 3DSL, the system will load some configuration parameters from a file in order to tell the system which taggers and parsers it will use during running time. At the same time, it initializes every module in the system in order to reduce the waiting time need by the translator.

Afterward, the user can access the different menu options which eventually lead them into the text mode screen. Here, while the system does not limit the length of the input text, the system was designed to translate one sentence per play. This limitation was set as a consequence of the focus on the reordering rules needed to carry out structure transfer on a parse tree.

[2] https://unity3d.com/unity/system-requirements
[3] http://gramatica.usc.es/~gamallo/php/deppattern/
[4] http://gramatica.usc.es/~gamallo/php/ProlnatTagger/index.php

In this first version, the Information Reception Module is only in charge of capturing the text from the user interface and storing it on the StoredData (Figure 4.5) data structure. No further processing is realized in this phase.

### Transformation Module

The first step on the Transformation Module is dedicated to the analysis of the morphological characteristics of the input data. The system uses the online version of DepPattern [61] which is a linguistic package that provides a grammar compiler, PoS taggers, and dependency based parsers for Portuguese and other Romance languages. It was chosen between other tools because its web interface allowed its direct usage on Unity projects, unlike other tools that either required being ported to C# or did not support European Portuguese.

For PoS tagging, DepPattern provides interfaces for Freeling[18] and TreeTagger[44]. Freeling was chosen among the two because of its higher performance on Portuguese tasks [38]. As part of its pipeline, Freeling carries out the tasks listed below:

**Tokenizer:** this step converts a sentence into a set of tokens according to a set of tokenization rules. Those rules are composed of regular expressions that are matched against the input text. The first matching rule is used to extract the token, which deleting the matching substring from the sentence. This process continues until the whole sentence has been matched.

**PoS tagging:** using a Markovian tagger and the Bosque 8.0[5] corpora, this phase automatically identifies each word with one of the tags listed bellow:

- ADJ (adjective)
- ADV (adverb)
- DET (determinant)
- NOUN (noun)
- CARD (cardinal)
- PRO (pronoun)
- PRP (preposition)
- CONJ (conjunction)
- VERB (verb)

**Lemmatisation:** Depending on the identified part of speech and meaning of a word in a sentence, the word will be reduced to its lemma equivalent. That is it: words like running, run, ran, runs can all be reduced to a common lemma run. This is especially useful for LGP processing because words lack inflections.

[5] http://www.linguateca.pt/Floresta/corpus.html#bosque

After those tasks are completed, DepPattern uses a dependency grammar [6] to identify relationships between the different elements of the sentence. Each rule of this grammar is constituted by two elements: a pattern of PoS tags and the name of a head-dependent relation found within the pattern.

One of the characteristics of DepPattern is the usage of the ”Uniqueness principle” to reduce the search space. This principle states that a dependent word only has one head and makes possible to remove dependents from the input when they are not the head of any word. Following this, rules are applied sequentially in an iterative process that stops when no rule can be applied. The final dependency tree is then returned using CoNLL [51] format. Unless the sentence lacks it, the main verb can be found at most on the first level from the root.

While Freeling allows for more characteristics to be extracted in this phase (like gender, number or tense of each word), DepPattern processing discards them on the final output. Therefore, in order to extract more information about each token other tools are required. Since those attributes are used in the reordering process, an extra tool is needed to recover them. In this version CitiusTagger [39] is the tool that allows to access all the characteristics defined in Freeling's PoS tagset for Portuguese [7] and additionally provides rule-based Named Entity Classification.

Subsequently, the dependency tree is created and populated with the extra attributes. This marks the transition from the analysis phase to the restructuring phase. According to the grammar rules detailed in chapter 2, the dependency tree will be reordered before going to a look-up of the gloss equivalents of each lemma contained in the input sentence. The reordering process works as follows:

1. Extract all subjects from the dependency tree. If the sentence has not explicit subjects, create a pronoun subject according to the inflection of the first verb from the root of the dependency tree.

2. For each subject, go to its children and extract all the adjectives and adverbs that describe it.

3. Reorder each set of $DESCRIBERS + SUBJECT$ as $SUBJECT + DESCRIBERS$.

4. Extract all verbs from the dependency tree.

[6] `http://gramatica.usc.es/pln/tools/deppattern/grammars.html`

[7] `https://talp-upc.gitbooks.io/freeling-user-manual/content/tagsets/tagset-pt.html`

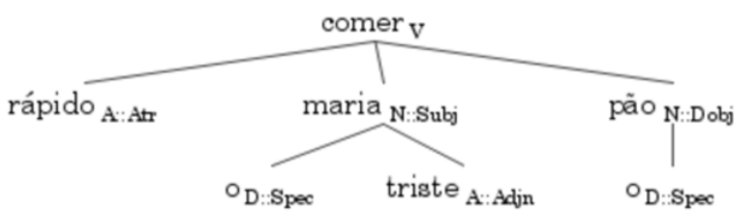
5. For each verb apply the tense rules described in chapter 2. In the same way, extract all the adjectives and adverbs that describe the verb.

6. Reorder each set of $VERB + DESCRIBERS$ as $DESCRIBERS + VERB$.

7. Extract the predicate associated with the verb, discarding prepositions and conjunctions.

8. Reorder the whole sentence following SOV order.

Furthermore, some describers of both subject and verbs can directly modify the word that they are affecting, transforming it in a new word when translating it to LGP. An example of this is phrase *bom dia* (good morning) where the adjective *bom*" (good) affects the word *dia* creating in LGP the word "BOM-DIA" instead of being translated as two separate words. 3DSL checks if any describer could be a modifier (from a list of possible modifiers) and marks it as such. An detailed example of this process can be seen on Figure **??**.




| Step # | Result for sentence "a triste Maria come rápido o pão" |
|---|---|
| 1 | Subjects[ ] = [ "Maria" ] |
| 2 | [ "Maria" ] $_{children}$ = [ "o", "triste" ] |
| 3 | OldSubjects [ "o", "triste", "Maria" ] => NewSubjects [ "Maria", "Triste"] |
| 4 | Verbs[ ] = [ "comer" ] |
| 5 | [ "comer" ] $_{children}$ = [ "rápido" ] |
| 6 | OldVerbs [ "comer", "rápido" ] => NewVerbs [ "rapido", "comer" ] |
| 7 | Predicate["o", "pão" ] = NewPredicate ["pão" ] |
| 8 | NewSentence = NewSubjects + NewPredicate + NewVerbs = [ "Maria", "triste", "rápido", "pão", "comer" ] |
| 9 | NewSentence$_{glosses}$ = [ "MARIA", "TRISTE", "RÁPIDO", "PÃO", "COMER" ] |

Figure 5.8: Example of sentence transformations performed by 3DSL

The final step in this phase is the lookup of glosses in the Portuguese - LGP dictionary. For each word in the reordered sentence, the system will search its gloss translation. If the word has a modifier, then the system will try to find the $WORD + MODIFIER$ combination first. If such combination does not exist, then the modifier is re-added to the sentence and treated as an independent word. Any words that were not encountered in the dictionary are marked for fingerspelling.

While the system could directly lookup for the SignWriting equivalent of each word, the gloss representation allows the system to give feedback about the exact translation that is being played by the avatar. This is helpful in early versions like the described in this thesis where animations are still being evaluated and need more refinement in order to ensure the clarity of each sign.

### Animation Module

When the Transformation Module finishes, 3DSL is ready to start calling the respective triggers that will form each sign. However, for such process to take place there are several elements involved: the avatar, the animation clips, and the animation system.

As noted before, studies have shown that avatar style does not have any significant effect on the clarity of sign language reproduction [2]. Therefore, while 3DSL currently uses Ty avatar provided by Mixamo[8] which presents a cartoon style, any other fbx avatar model compatible with Unity's animation system can be used as long as they are exported using humanoid mode. Even more, as long as bone names match animation clips can be retargeted without major issues. Ty avatar and its skeleton are shown in Figure 5.9.

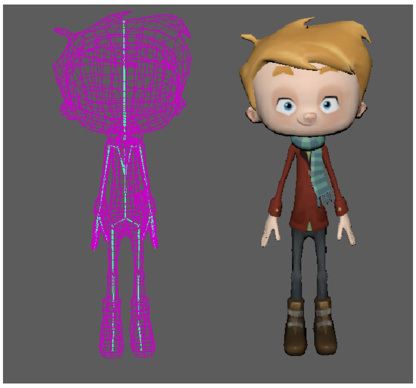


Figure 5.9: Ty avatar used by 3DSL

Animation clips were created using Maya[9] Student version. Each of them is named according to the zone they affect as detailed in Figure 4.12. In Unity, each clip is imported such that only the curves presented in the mask of the respective layer are used. For example, an animation clip for the right hand will be imported using a right hand mask which only imports the curves affecting the bones of that given hand. 3DSL uses all the layers defined in Figure 4.6.

[8] https://www.mixamo.com
[9] http://www.autodesk.com/products/maya/overview

For this development, only fingerspell animations were created. While each hand configuration could be composed of one frame, all clips were expanded such that their duration is 30 frames. Having all clips with the same duration allows reducing the difficulty of layer synchronization.

All clips are imported into Unity's animation state machine system (Mecanim). Each clip will correspond to a state and transitions are going to be triggered according to the results of each gloss lookup. To be precise, each gloss can be associated with one or more different gestures and, at the same time, each gesture is composed by a set of triggers corresponding to each layer in the animation as shown in Figure 5.10.

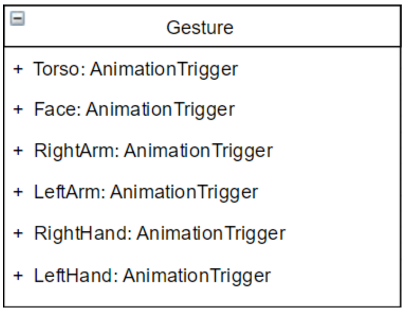


Figure 5.10: Gesture structure

Each of those triggers will activate only one animation per layer, each of those clips is going to be blended using Override Blending type. This means that information from other layers will be ignored. The frames needed to transition from one clip to another is controlled directly by Unity, which allows for an instant transition or for the usage of bleeding between the two animations. This feature was not used in the current version but can be evaluated per animation transition if needed.

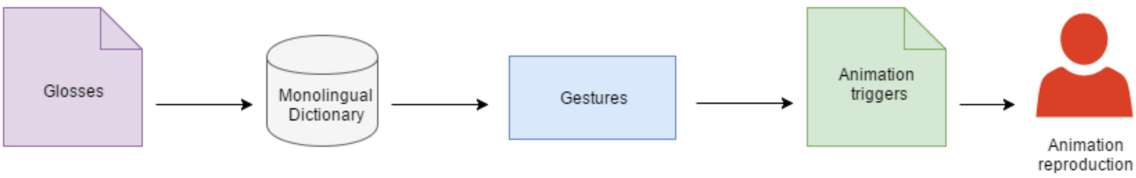


Figure 5.11: Animation Phase Pipeline

In summary, the animation module will go to the monolingual LGP dictionary which will return one or more gestures. Each of those gestures will contain values based on SignWriting notation for the different configurations of each body part, thus triggering different animations in their respective layers. Finally, those triggers will result in the reproduction of the animation clips needed to perform the selected gesture. This process will continue until there are no more gestures to reproduce or the user manually stops the reproduction. This process is graphically represented in Figure 5.11.

# Chapter 6

# Evaluation and Results

As a follow up to the development process, 3DSL was evaluated in concordance with the objectives proposed at the beginning of this thesis. The evaluation was centered around the following points: usability, grammatical correctness of the output and ability to create signs through bleeding of gesture animations.

Usability was evaluated in an iterative process throughout the development process as described in section 5.2. In the other hand, the other points were evaluated only after finalization of the development phase and are the main focus of the present chapter.

Evaluation was carried out through the collection of feedback from volunteer testers and available LGP sentence examples extracted from Amaral's [5] and Baltazar's [10] reference materials. Therefore, any results presented are by no means exhaustive, instead of seeking to provide proof of the system's feasibility. At the same time, this analysis explores the areas where further research is needed.

## 6.1 Grammatical Correctness

Given the differences between LGP and Portuguese grammars, syntax structural transfer is an essential feature in the translation process. The correctness of the resulting output is dependent on both the reordering rules implemented and the dependency parse tree generated by DepPattern.

Simple Portuguese sentences that follow Subject-Verb-Object pattern (e.g. *o gato bebe o leite*) are correctly reordered to LGP's Subject-Object-Verb which matches with all example sentences provided by Baltazar [10]. However, this reordering makes some output sentences differ from the expected results on Amaral [5]. This is true for any test positive and negative sentence on those materials that was in simple past, present, or future verb tenses.

However, some problems were detected in additional sentences with a similar structure. Specifically, those containing specific dates or times. For example, *eram duas horas da manhã* is reordered just as $SER$ and all other elements are discarded. This is a consequence of how 3DSL PoS tagset is smaller than DepPattern's. The object of this sentence (*duas horas da manhã*) gets tagged as DATE, even more, DepPattern replaces this syntax by its simplified equivalent "2:00 AM". While this replacement offers no problem on Portuguese texts, there is not documented equivalent on LGP of such structure. Thus, this kind of structure should be treated as an independent set of words, each one of their own PoS tag.

Regarding the treatment of implicit subjects, 3DSL delivers the correct explicit replacement as long as it can be extracted from the main verb. Thus, sentences like *bebo leite* get translated as $EULEITEBEBER$ instead of just $LEITEBEBER$. For ambiguous cases (like *gostaria de comer um bolo*) it is not possible to extract the correct explicit subject because 3DSL is not context-aware, resulting in a subjectless output $COMERBOLOGOSTAR$.

In a similar way, object pronouns are translated into pronouns according to the person they are referring to. Therefore, in the sentence *abri os braços ao vê-lo* the object pronoun (*lo*) is separated from the verb and translated separately as $VERELE$.

Modifiers work as expected, with examples like *bom dia* being translated as $BOM-DIA$ as long as the word is correctly defined in the bilingual dictionary. In spite of this, other grammatical processes that can affect verbs and nouns, namely, verbal aspects and noun number, are not implemented.

Abnormalities were detected in sentences that made use of the adjective *salgado/salgada* (salty). The online version of DepPattern used by 3DSL wrongly tags it as the verb *salgar* (to salt) which has two consequences: wrong rules are applied and the returned lemma differs from what the users expect.

Grammatical constructions outside of those previously described do not have a defined behavior. More research is needed in order to create rules that can properly describe them.

## 6.2 Animation

The model to be used was chosen by the volunteers between Mixamo catalog [1] and characters generated with Autodesk Character Generator[2] and Adobe Fuse CC[3]. In the end, Ty avatar (Figure 6.1) was selected, resulting in its usage throughout the development and testing process.

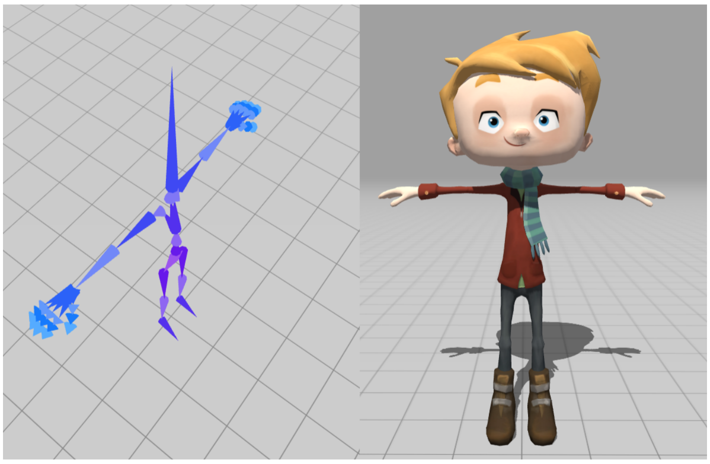


Figure 6.1: Rig and Texture Used of Tests

All layers and their respective masks were created before testing the system, yet due to time constraints, only a limited set of animations were added to them. Each of them was created in Maya by manual joint manipulation. Those animations correspond to some fingerspelling right-hand poses and one right arm pose. Some of them are shown in Figure 6.2.

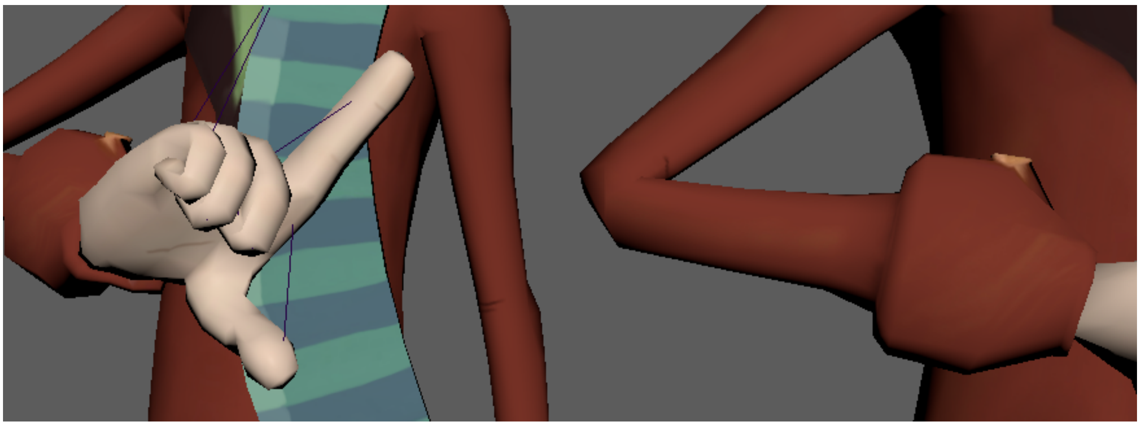


Figure 6.2: Right Hand And Right Arm Pose Examples

Every pose was converted into an animation that lasts one second and plays at 30fps and exported into Unity's animation format (.anim) afterward. All triggers and animation

[1] https://www.mixamo.com/store/#/search?page=1&type=Character
[2] http://charactergenerator.autodesk.com/
[3] http://www.adobe.com/products/fuse.html

transitions were tested and proven to be working correctly. Right hand and right arm animations blended successfully, playing smoothly without any noticeable glitching derived from the fade-in and fade-outs between animation states.

Unfortunately, there were not LGP signers available between volunteer evaluators, therefore it was not possible to test the exactitude of the created poses, nor the adequacy of the current play speed. Meanwhile, non-signer volunteers classified each pose as correct after being exposed to photos of human signers performing the same poses.

Additional notes pointed out the lack of facial expression changes through the different tests. While the trigger system supports animation changes in any layer, the creation, and management of the blend shapes needed to achieve this was outside of the scope of the developed system; however, will be taken in consideration on future versions. In the same way, users complained about the waiting time required before the animation starts playing, which is mostly derived from the usage of non-dedicated web services for the natural language operations needed before the triggers are generated.

## 6.3 Discussion

The developed system successfully translates Portuguese into LGP, proving that it is possible to generate signs from separated gestures by the usage of animation layers. In the same manner, it was proven possible that rule-based syntax transfer can improve the grammatical correctness of automatic translation of sign languages.

However, LGP automatic translation does not come without limitations. The tools used would need to be replaced by faster alternatives, and the size of available gestures greatly expanded, in order to increase the scope of 3DSL real life applications. Moreover, any further development will require more involvement of hard of hearing people and LGP experts.

# Chapter 7

# Conclusion and Future Work

The first part of this chapter summarizes the results and contributions of the work presented throughout this thesis. Meanwhile, the rest of the chapter covers proposals of how to expand that work accordingly to what has been previously exposed.

## 7.1 Conclusions

One of the principal consequences of hearing loss is the reduction of the individual's ability to communicate with others through spoken language. Limiting their access to information, which can have a significant negative impact on everyday life.

Derived from this fact, the main goal of this thesis was to design a system capable of automatic translation of Portuguese into LGP with the aim to improve communication and access to information of hard of hearing people. In the same way, other objectives were defined around this goal as to guide the development process of such system.

After a study of previous and current sign language translation systems, a system architecture capable of translating video, audio or text was designed. Moreover, it also takes account of the impact that the grammar correctness has into the clarity of the final output. Specifically, this work proposes a system that uses natural language processing techniques to decompose Portuguese sentences into a set of equivalent signs in LGP. At the same time, those signs are also decomposed into smaller units derived on the SignWriting system representation of each sign, this allows the reproduction of signs using a limited set of animations.

Furthermore, a system (3DSL) was developed to prove the feasibility of such architecture. It takes text, reorders it according to a set of pre-established rules extracted from LGP's documented grammar [5], and finally reproduces an animation according to the values associated to each word finger-spelling. As part of 3DSL development process, a user study was carried out in order to find out the best way to expose the system features to its target users.

Finally, 3DSL was qualitatively evaluated by some volunteers, which described the overall experience as positive. All of them showing great interest in contributing with any future development of the project.

## 7.2 Discussion

While many people take the ability to communicate for granted, hard of hearing have many difficulties to express their ideas and needs because of their inability to understand spoken language. To overcome communication barriers, people with hearing loss can learn to communicate using alternative means, such as lip-reading or sign language.

However, many misconceptions surround sign languages, and the initial phases of this thesis were not exempt from them. Particularly, the beliefs that only a unique sign language existed and that it was just a graphic representation of the spoken language. Therefore, only computer graphics techniques were needed to recreate it using 3D avatars. As research was carried out, it was proven that both ideas were wrong.

First of all, as happens with spoken languages, there are several different sign languages that are not mutually intelligible with one another. And, even if two countries share a spoken language, they do not necessarily share a sign language.

In the same way, a sign language is not just a graphic representation of a spoken language. They are independent languages with their own linguistic elements. Therefore, their grammar is independent of the spoken language used in the country where the sign language is used.

Those discoveries implied that other areas had to be involved in the development of the system, mainly natural language processing techniques. Additionally, it was essential to obtain linguistic information about LGP.

While limited by several factors related to both the language and the technologies used (as explained in section 4.3), this thesis produced a system that solves the communication limitations that hard of hearing suffer to understand spoken Portuguese, and sets the foundation for the next generation of automated translation systems of Portuguese and LGP.

## 7.3 Future work

This thesis expects to instill further research on how to perform automatic translation of European Portuguese into LGP. The presented solution was limited by the reasons listed in the ”Open Problems” (section 4.3) and because time constrains that did not allow further research and development of all the planed elements.

According to the results obtained, the following list covers a general list of topics that would serve as follow up to this work.

**Linguistic Studies:** In order to improve translation quality and allow more complex inputs to be processed, it is needed to expand the knowledge of how syntactically complex sentences are created in LGP.

Improvements in this area include the study of exactly how aspect, sentence type and general context affect the translated sentences. At the same time, it also includes how information extraction tasks like named-entity recognition or sentiment analysis could improve the obtained output.

**Corpora creation:** The increasing of available corpora would open the gates for not only better validation of results, but it could also create a new line of research centered around other types of automatic translation like pattern recognition-based statistical approaches.

Given that some of those approaches require a (relatively) small corpus [83], it would be possible to test its performance against rule-based systems even on early phases of corpora creation.

**Expand animation database:** For this work only a small set of gestures were animated in order to prove that it is possible to generate signs from them. However, in order to be full functional, all SignWriting gestures would be needed to be animated and added to their respective animation layers.

**Speech mode:** Using libraries like Microsoft SAPI[1] or Google's Cloud Speech [2] would allow the transformation of speech into text. This could be coupled with the infrastructure presented in thesis thesis in order to provide a complete speech to LGP translation.

[1] `https://msdn.microsoft.com/en-us/library/ee125077`
[2] https://cloud.google.com/speech

**Offline usage:** The current system can only be used if the host is connected to Internet, and also depends of the availability of DepPattern and CitiusTagger. So that to break this dependency, it would be needed to port those tools to C# or implement new ones from scratch.

**Mobile version:** Making possible for a hard of hearing person to have their own personal interpreter anywhere, anytime would greatly improve their access to any kind of information that they normally cannot access. Although a desktop application represents a step forward into that direction, a mobile application would decrease even more the gap between the available information and the information that hard of hearing can access.

As noted, Unity allows projects to be ported into different platforms. However, this does not mean that the process is automatic. Each platform has its own characteristics and as to guarantee an acceptable performance it is likely that some process would need to be optimized or completely replaced. Additionally, the user interface would need to be redesigned to adapt to the different design guidelines expected on mobile software.

**Dictionary Updater:** As to create a complete translation suite, it is needed a simple way to add new words to the system. With this in mind, an external application could be deployed in order to manage all the dictionaries needed by 3DSL. This was briefly referenced on section 4.7.

This application could be based on SignMaker[3] so that any person with SignWriting knowledge could manage 3DSL's database without needing any additional technical background.

*"Much unhappiness has come into the world because of bewilderment and things left unsaid."*

Fyodor Dostoyevsky

[3] http://www.signbank.org/signmaker.html

# Bibliography


[1] Action on Hearing Loss Information. “Facts and Figures on Hearing Loss and Tinnitus”. In: *Action on Hearing Loss Information* (July 2011). URL: http://www.actiononhearingloss.org.uk/your-hearing/about-deafness-and-hearing-loss/statistics/~/media/56697A2C7BE349618D336B41A12B85E3.ashx (visited on 12/09/2015).

[2] Nicoletta Adamo-Villani, Jason Lestina, and Saikiran Anasingaraju. “Does Character’s Visual Style Affect Viewer’s Perception of Signing Avatars?” In: *E-Learning, E-Education, and Online Training*. Springer, 2015, pp. 1–8. URL: http://link.springer.com/chapter/10.1007/978-3-319-28883-3_1 (visited on 03/17/2016).

[3] Nicoletta Adamo-Villani and Ronnie B. Wilbur. “ASL-Pro: American Sign Language Animation with Prosodic Elements”. In: *Universal Access in Human-Computer Interaction. Access to Interaction*. Ed. by Margherita Antona and Constantine Stephanidis. Vol. 9176. Cham: Springer International Publishing, 2015, pp. 307–318. ISBN: 978-3-319-20680-6. URL: http://link.springer.com/10.1007/978-3-319-20681-3_29 (visited on 07/22/2016).

[4] Daniela OF Amaral et al. “Comparative Analysis of Portuguese Named Entities Recognition Tools”. In: (2014). URL: http://www.lrec-conf.org/proceedings/lrec2014/pdf/513_Paper.pdf (visited on 07/30/2016).

[5] Maria Augusta Amaral, Amândio Coutinho, and Maria Raquel Delgado Martins. *Para Uma Gramática da Língua Gestual Portuguesa*. Linguística. Lisbon, Portugal: Caminho, Dec. 1994. 166 pp. ISBN: 972-21-0981-2.

[6] Stephen Anderson. *Languages: A Very Short Introduction*. Oxford University Press, Sept. 2012. 152 pp. ISBN: 0-19-959059-1.

[7] Kgatlhego Aretha Moemedi. “Rendering an Avatar from Signwriting Notation for Sign Language Animation”. Master’s Thesis. Cape Town, South Africa: University of the

Western Cape, Nov. 2010. 98 pp. URL: http://etd.uwc.ac.za/xmlui/bitstream/handle/11394/2608/Moemedi_MSC_2010.pdf.

[8] Lora Aroyo. *Lecture 2: Human-Computer Interaction Course (2015)*. English. University Amsterdam, June 2015. URL: http://es.slideshare.net/laroyo/lecture-2-humancomputer-interaction-course-2015-vu-university-amsterdam (visited on 12/22/2015).

[9] Associação Portuguesa de Surdos. *Informação – Comunidade*. 2011. URL: http://www.apsurdos.org.pt/index.php?option=com_content/&view=article/&id=43/&Itemid=57 (visited on 12/09/2015).

[10] Ana Bela Baltazar. *Dicionário de Língua Gestual Portuguesa*. Dicionários Temáticos. 2010. 1168 pp. ISBN: 978-972-0-05282-7.

[11] João Pedro Barata de Sousa. "Modelo De Língua Natural E Transformação Em Modelo De LGP". Master's Thesis. Lisbon, Portugal: Universidade Nova de Lisboa, Sept. 2014. 85 pp.

[12] BBC. *Voice-Recognition Software – An Introduction*. Mar. 2009. URL: http://www.bbc.co.uk/accessibility/guides/factsheets/factsheet_VR_intro.pdf (visited on 12/10/2015).

[13] José Bettencourt. *Gestuário: Língua Gestual Portuguesa*. 1st ed. 1992.

[14] Statistic Brain. *STATS — YouTube Company Statistics – Statistic Brain*. English. bibtex: statistic_brain_stats_2016. Sept. 2016. URL: http://www.statisticbrain.com/youtube-statistics/ (visited on 04/29/2017).

[15] Jan Bungeroth and Hermann Ney. "Statistical Sign Language Translation". In: *In Workshop on Representation and Processing of Sign Languages, LREC 2004*. 2004, pp. 105–108.

[16] Chris Callison-Burch et al. "(Meta-) Evaluation of Machine Translation". In: *Proceedings of the Second Workshop on Statistical Machine Translation*. StatMT '07. Stroudsburg, PA, USA: Association for Computational Linguistics, 2007, pp. 136–158. URL: http://dl.acm.org/citation.cfm?id=1626355.1626373 (visited on 07/29/2016).

[17] Helena Carmo et al. *Programa Curricular de Língua Gestual Portuguesa – Educação Pré-Escolar e Ensino Básico*. Dec. 18, 2007. URL: http://www.dge.mec.pt/sites/default/files/ficheiros/programacurricularlgportuguesa.pdf.

[18] Xavier Carreras et al. “FreeLing: An Open-Source Suite of Language Analyzers”. In: Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC’04). 2004.

[19] Central Intelligence Agency. *The World Factbook*. Central Intelligence Agency. Nov. 2015. URL: https://www.cia.gov/library/publications/the-world-factbook/geos/po.html (visited on 12/09/2015).

[20] Rafaela Coda Da Silva. “SignWriting: Um Sistema De Escrita Das Línguas Gestuais – Aplicação À Língua Gestual Portuguesa”. Master’s Thesis. Lisbon, Portugal: Universidade Lusófona de Humanidades e Tecnologias, 2012. 208 pp. URL: http://recil.grupolusofona.pt/handle/10437/4066.

[21] European Commission. *Fifth RTD Framework Programme (1998-2002)*. 1998. URL: http://cordis.europa.eu/fp5/ (visited on 07/20/2016).

[22] Association for Computational Linguistics. *POS Tagging (State of the art) - ACL Wiki*. Wiki of the Association for Computational Linguistics. July 2016. URL: http://www.aclweb.org/aclwiki/index.php?title=POS_Tagging_(State_of_the_art)&oldid=11577 (visited on 07/29/2016).

[23] *Constituição da República Portuguesa*. Portuguese. 2005. URL: http://www.parlamento.pt/Legislacao/Paginas/ConstituicaoRepublicaPortuguesa.aspx (visited on 05/06/2017).

[24] Marta R. Costa-Jussa et al. *Study and Comparison of Rule-based and Statistical Catalan-Spanish Machine Translation Systems*. Jan. 1, 2012. URL: https://www.researchgate.net/publication/290610124_Study_and_comparison_of_rule-based_and_statistical_catalan-spanish_machine_translation_systems (visited on 07/29/2016).

[25] Stephen Cox, M. Lincoln, and Judy Tryggvason. *School of Computing Sciences (CMP) - TESSA*. 2001. URL: http://www.visicast.cmp.uea.ac.uk/Tessa.htm (visited on 07/20/2016).

[26] Leonel Domingues Deusdado. “Ensino da Língua Gestual Assistido por Personagens 3D Virtuais”. PhD thesis. Braga, Portugal: Universidade do Minho, Apr. 2002. 106 pp. URL: http://sim.di.uminho.pt/mestrados/deusdado/teseLeonel.pdf (visited on 07/21/2016).

[27] Bonnie J. Dorr, Eduard H. Hovy, and Lori S. Levin. “Machine Translation: Interlingual Methods”. In: *Encyclopedia of Language & Linguistics*. Second Edition. Elsevier, 2006, pp. 383–394. ISBN: 978-0-08-044854-1. URL: http://www.sciencedirect.com/science/article/pii/B0080448542009391.

[28] Daniel Eberle and Sarah Eberle. “Sociolinguistic Survey Report of the Portuguese Deaf Community”. In: *SIL International - Digital Resources*. Electronic Survey Report 2012 (Dec. 2012). URL: http://www-01.sil.org/SILESR/2012/silesr2012-044.pdf (visited on 07/10/2016).

[29] Sarah Ebling and Matt Huenerfauth. “Bridging the gap between sign language machine translation and sign language animation using sequence classification”. In: *Proceedings of the 6th Workshop on Speech and Language Processing for Assistive Technologies (SLPAT)*. 2015. URL: http://www.aclweb.org/website/old_anthology/W/W15/W15-51.pdf#page=9 (visited on 07/21/2016).

[30] R. Elliott et al. “The development of language processing support for the ViSiCAST project”. In: ACM Press, 2000, pp. 101–108. DOI: 10.1145/354324.354349. URL: http://portal.acm.org/citation.cfm?doid=354324.354349 (visited on 07/20/2016).

[31] Paula Escudeiro et al. “Virtual Sign Translator”. In: *ICCNCE : proceedings*. International Conference on Computer, Networks and Communication Engineering (ICCNCE). Beijing, China: Atlantis Press, May 2013, pp. 290–292. ISBN: 978-90-78677-67-3. URL: http://hdl.handle.net/10400.2/2893 (visited on 12/09/2015).

[32] European Sign Language Center. *Sign language dictionary – SPREADTHESIGN*. 2012. URL: https://www.spreadthesign.com/pt/ (visited on 12/22/2015).

[33] European Union of the Deaf. *2010 Brussels Declaration on Sign Languages in the European Union*. Nov. 2010. URL: http://www.eud.eu/uploads/brussels_declaration_English.pdf (visited on 12/09/2015).

[34] Michael Everson et al. *Proposal for encoding Sutton SignWriting in the UCS*. Oct. 2012. URL: http://www.unicode.org/L2/L2012/12321-n4342-signwriting.pdf (visited on 07/15/2016).

[35] Federal Communications Commission. *Telecommunications Relay Service (TRS)*. Federal Communications Commission. Dec. 21, 2015. URL: https://www.fcc.gov/consumers/guides/telecommunications-relay-service-trs (visited on 12/21/2015).

[36] Federal Communications Commission. *Video Relay Services*. Federal Communications Commission. Dec. 21, 2015. URL: https://www.fcc.gov/consumers/guides/video-relay-services (visited on 12/21/2015).

[37] James Foreman-Peck and Yi Wang. *The Costs to the UK of Language Deficiencies as a Barrier to UK Engagement in Exporting: A Report to UK Trade & Investment*. May 2014. URL: https://www.gov.uk/government/publications/the-costs-to-the-uk-of-language-deficiencies-as-a-barrier-to-uk-engagement-in-exporting.

[38] Pablo Gamallo and Marcos Garcia. *FreeLing e TreeTagger: um estudo comparativo no âmbito do Português*. Tech. rep. Centro Singular de Investigação em Tecnologias da Informação (CITIUS), 2013. URL: http://gramatica.usc.es/~gamallo/artigos-web/PROLNAT_Report_01.pdf (visited on 11/03/2016).

[39] Pablo Gamallo et al. "Análisis morfosintáctico y clasificación de entidades nombradas en un entorno Big Data". Spanish. In: *Procesamiento del Lenguaje Natural* 53.0 (Sept. 2014), pp. 17–24. ISSN: 1989-7553. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/5046 (visited on 11/04/2016).

[40] Brigitte Garcia. "Contribution à l'histoire des débuts de la recherche linguistique sur la Langue des Signes Franc¸aise (LSF). Les travaux de Paul Jouison." PhD thesis. Paris, France: Paris University, 2000. URL: http://www.theses.fr/2000PA05H060.

[41] Inês Gomes Sá Neiva. "Desenvolvimento de um tradutor de Língua Gestual Portuguesa". Master's Thesis. Lisbon, Portugal: Universidade Nova de Lisboa, Dec. 2014. 107 pp. URL: http://run.unl.pt/handle/10362/14753.

[42] Sami M Halawani. "Arabic Sign Language Translation System on Mobile Devices". In: *IJCSNS International Journal of Computer Science and Network Security* 8.1 (2008), pp. 251–256. URL: http://paper.ijcsns.org/07_book/200801/20080136.pdf (visited on 12/10/2015).

[43] Thomas Hanke. "HamNoSys-representing sign language data in language resources and language processing contexts". In: *Workshop Proceedings : Representation And Processing Of Sign Languages*. LREC 2004. Vol. 4. Paris, France, 2004, pp. 1–6. URL: http://www.sign-lang.uni-hamburg.de/dgs-korpus/files/inhalt_pdf/HankeLRECSLP2004_05.pdf (visited on 03/15/2016).

[44] Schmid Helmut. “Probabilistic Part-of-Speech Tagging Using Decision Trees”. English. In: Proceedings of International Conference on New Methods in Language Processing. Manchester, UK, 1994. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.28.1139&rep=rep1&type=pdf (visited on 11/02/2016).

[45] Western Interstate Commission for Higher Education. *Information Gaps on the Deaf and Hard of Hearing Population: A Background Paper*. May 2006. URL: http://www.wiche.edu/pub/12452 (visited on 07/26/2016).

[46] Matt Huenerfauth and Hernisa Kacorri. “Augmenting EMBR Virtual Human Animation System with MPEG-4 Controls for Producing ASL Facial Expressions”. In: *ResearchGate*. 5th International Workshop on Sign Language Translation and Avatar Technology (SLTAT). Apr. 10, 2015. URL: https://www.researchgate.net/publication/302898154_Augmenting_EMBR_Virtual_Human_Animation_System_with_MPEG-4_Controls_for_Producing_ASL_Facial_Expressions (visited on 07/21/2016).

[47] Matt Huenerfauth, Pengfei Lu, and Andrew Rosenberg. “Evaluating importance of facial expression in american sign language and pidgin signed english animations”. In: ACM Press, 2011, p. 99. ISBN: 978-1-4503-0920-2. DOI: 10.1145/2049536.2049556. URL: http://dl.acm.org/citation.cfm?doid=2049536.2049556 (visited on 07/21/2016).

[48] INE and PORDATA. *Indivíduos Com 16 E Mais Anos Que Utilizam Computador E Internet Em % Do Total De Indivíduos: Por Nível De Escolaridade Mais Elevado Completo*. PORDATA. Dec. 2015. URL: http://www.pordata.pt/Portugal/Indiv%C3%ADduos+que+utilizam+computador+e+Internet+em+percentagem+do+total+de+indiv%C3%ADduos+por+n%C3%ADvel+de+escolaridade+mais+elevado+completo-1141 (visited on 12/10/2015).

[49] Tiffany Jen and Nicoletta Adamo-Villani. “The Effect of Rendering Style on Perception of Sign Language Animations”. In: *Universal Access in Human-Computer Interaction. Access to Interaction*. Ed. by Margherita Antona and Constantine Stephanidis. Vol. 9176. Cham: Springer International Publishing, 2015, pp. 383–392. ISBN: 978-3-319-20680-6. URL: http://link.springer.com/10.1007/978-3-319-20681-3_36 (visited on 07/21/2016).

[50] Richard Kennaway. “Synthetic Animation of Deaf Signing Gestures”. In: *Gesture and Sign Language in Human-Computer Interaction*. Ed. by Ipke Wachsmuth and Timo Sowa. Red. by G. Goos, J. Hartmanis, and J. van Leeuwen. Vol. 2298. Berlin,

Heidelberg: Springer Berlin Heidelberg, 2002, pp. 146–157. ISBN: 978-3-540-43678-2. URL: http://link.springer.com/10.1007/3-540-47873-6_15 (visited on 07/20/2016).

[51] D. Lin. “Dependency-based Evaluation of MINIPAR”. In: *Workshop on the Evaluation of Parsing Systems*. Granada, Spa in, 1998.

[52] LumenVox, LLC. *LumenVox Speech Recognizer*. Dec. 10, 2015. URL: http://www.lumenvox.com/products/speech_engine/ (visited on 12/10/2015).

[53] Michella Maiorana-Basas and Claudia M. Pagliaro. “Technology Use Among Adults Who Are Deaf and Hard of Hearing: A National Survey”. In: *Journal of Deaf Studies and Deaf Education* 19.3 (Mar. 2014), pp. 400–410. URL: http://jdsde.oxfordjournals.org/content/19/3/400.full.pdf (visited on 12/09/2015).

[54] José Carlos Malhado Bento. “Avatares em Língua Gestual Portuguesa”. Master’s Thesis. Lisbon, Portugal: Universidade de Lisboa, 2013. 157 pp. URL: http://repositorio.ul.pt/handle/10451/15945.

[55] Maria Raquel Delgado Martins. “Integração da criança surda em jardim infantil: Uma experiência de linguagens alternativas”. In: *Análise Psicológica*. Vol. 5. Instituto Superior de Psicologia Aplicada, 1986, pp. 115–120. URL: http://hdl.handle.net/10400.12/2153.

[56] Mariana Martins, Marta Morgado, and Paula Estanqueiro. *Programa Curricular de Língua Gestual Portuguesa – Ensino Secundário*. URL: http://www.dge.mec.pt/sites/default/files/EEspecial/prog_curric_lgp_sec.pdf.

[57] Harish Chandra Maurya, Pooja Gupta, and Nalin Choudhary. “Natural Language Ambiguity and its Effect on Machine Learning”. In: *International Journal Of Modern Engineering Research* 5.4 (Apr. 2015), pp. 25–30. URL: http://www.academia.edu/download/38975976/D0504_01-2530.pdf (visited on 07/26/2016).

[58] John McDonald et al. “An automated technique for real-time production of lifelike animations of American Sign Language”. In: *Universal Access in the Information Society* (May 14, 2015). ISSN: 1615-5289, 1615-5297. DOI: 10.1007/s10209-015-0407-2. URL: http://link.springer.com/10.1007/s10209-015-0407-2 (visited on 07/21/2016).

[59] François Neve. “Phonologie or gestematique des langue de signes des sourd”. In: *Linguistic and Language Behaviour Abstracts* 26.3 (1996), pp. 1954–1964.

[60] Naoki Oishi and Jochen Schacht. “Emerging treatments for noise-induced hearing loss”. In: *Expert Opinion on Emerging Drugs* 16.2 (June 2011), pp. 235–245. ISSN: 1472-8214, 1744-7623. DOI: 10.1517/14728214.2011.552427. URL: http://www.tandfonline.com/doi/full/10.1517/14728214.2011.552427 (visited on 03/17/2016).

[61] Pablo Gamallo Otero and Isaac González. “DepPattern: a Multilingual Dependency Parser”. English. In: International Conference on Computational Processing of the Portuguese Language (PROPOR 2012). Coimbra, Portugal, Apr. 2012, pp. 659–670. DOI: 10.1.1.295.6249. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.295.6249&rep=rep1&type=pdf (visited on 09/02/2016).

[62] Achraf Othman and Mohamed Jemni. “Statistical Sign Language Machine Translation: From English Written Text to American Sign Language Gloss”. In: *IJCSI International Journal of Computer Science Issues* 8.5 (Sept. 2011), pp. 65–73. ISSN: 1694-0814. URL: http://arxiv.org/abs/1112.0168 (visited on 03/15/2016).

[63] Maria Papadogiorgaki et al. “Synthesis of Virtual Reality Animations from SWML Using MPEG-4 Body Animation Parameters”. In: *Workshop on the Representation and Processing of Sign Languages on the occasion of the Fourth International Conference on Langauge Resources and Evaluation*. Lisbon, Portugal, May 2004. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.97.7465&rep=rep1&type=pdf (visited on 03/15/2016).

[64] Dianne Parkhurst and Stephen Parkhurst. “A Cross-Linguistic Guide to SignWriting®”. In: *A phonetic approach* (2010). URL: http://is.muni.cz/el/1490/jaro2014/SPNZJ_SW/um/Cross-Linguistic_Guide.pdf (visited on 05/06/2017).

[65] Carlos Patrick, Aldebaro Klautau, and Isabel Trancoso. “Free tools and resources for Brazilian Portuguese speech recognition”. English. In: *Journal of the Brazilian Computer Society* 17.1 (Mar. 2011), pp. 53–68. ISSN: 1678-4804. DOI: 10.1007/s13173-010-0023-1. URL: http://link.springer.com/article/10.1007/s13173-010-0023-1 (visited on 12/10/2015).

[66] Lewis Paul M., Gary F. Simons, and Charles D. Fennig. *Portuguese Sign Language*. Ethnologue: Languages of the World. 2015. URL: http://www.ethnologue.com/18/language/psr/ (visited on 07/10/2016).

[67] PeopleGroups.org. *People Name: Deaf Portuguese of Portugal*. Dec. 2015. URL: http://www.peoplegroups.org/Explore/groupdetails.aspx?peid=495 (visited on 12/09/2015).

[68] R. Plamondon and S. N. Srihari. "Online and off-line handwriting recognition: a comprehensive survey". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.1 (Jan. 2000), pp. 63–84. ISSN: 0162-8828. DOI: 10.1109/34.824821.

[69] I. Placencia Porrero and E. Ballabio. *Improving the Quality of Life for the European Citizen: Technology for Inclusive Design and Equality*. Vol. 4. Assistive Technology Research Series. IOS Press, 1998. 544 pp. ISBN: 978-4-274-90230-7. (Visited on 12/19/2015).

[70] Inês Rodrigues Almeida. "Exploring Challenges in Avatar-based Translation from European Portuguese to Portuguese Sign Language". Master's Thesis. Instituto Superior Técnico, Oct. 2014. 101 pp. URL: http://web.ist.utl.pt/~ist163556/pt2lgp/.

[71] Jerry Schnepp et al. "Generating Co-occurring Facial Nonmanual Signals in Synthesized American Sign Language". In: (2013). URL: http://scholarworks.bgsu.edu/vcte_pub/8/ (visited on 07/21/2016).

[72] Merkourios Simos and Nikolaos Nikolaidis. "Greek sign language alphabet recognition using the leap motion device". In: ACM Press, 2016, pp. 1–4. ISBN: 978-1-4503-3734-2. DOI: 10.1145/2903220.2903249. URL: http://dl.acm.org/citation.cfm?doid=2903220.2903249 (visited on 07/21/2016).

[73] Justin Slick. *7 Common Modeling Techniques for Film and Games*. About.com Tech. Dec. 2014. URL: http://3d.about.com/od/3d-101-The-Basics/a/Introduction-To-3d-Modeling-Techniques.htm (visited on 12/22/2015).

[74] Justin Slick. *How Are 3D Models Prepared for Animation?* About.com Tech. Nov. 2014. URL: http://3d.about.com/od/Creating-3D-The-CG-Pipeline/a/What-Is-Rigging.htm (visited on 12/22/2015).

[75] Daniel Stein, Jan Bungeroth, and Hermann Ney. "MorphoSyntax Based Statistical Methods for Automatic Sign Language Translation". In: *In Proceedings of 11 th EAMT Annual Conference*. 2006.

[76] Daniel Stein, Christoph Schmidt, and Hermann Ney. "Analysis, preparation, and optimization of statistical sign language machine translation". In: *Machine Translation* 26.4 (Dec. 2012), pp. 325–357. ISSN: 0922-6567, 1573-0573. DOI: 10.1007/s10590-012-9125-1. URL: http://link.springer.com/10.1007/s10590-012-9125-1 (visited on 07/21/2016).

[77] William Stokoe. *Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf*. 1960. URL: http://saveourdeafschools.org/stokoe_1960.pdf.

[78] Marianne Stumpf. *Escrita das Línguas Gestuais*. Vol. 14. Língua Gestual Portuguesa. Lisbon, Portugal: Universidade Católica Editora, Sept. 2011. 180 pp. ISBN: 978-972-54-0319-8.

[79] Valerie Sutton. *International SignWriting Alphabet*. SignWriting. Jan. 2012. URL: http://www.signbank.org/iswa/ (visited on 07/15/2016).

[80] Valerie Sutton. *SignMaker 2015*. 2015. URL: http://www.signbank.org/signmaker (visited on 07/15/2016).

[81] Valerie Sutton. *SignWriting Symbol Lessons 2010*. URL: http://www.signwriting.org/lessons/iswa/ (visited on 05/06/2017).

[82] Valerie Sutton. *What is SignWriting?* English. Sept. 1996. URL: http://www.signwriting.org/about/what/what02.html (visited on 05/06/2017).

[83] George Tambouratzis, Sokratis Sofianopoulos, and Marina Vassiliou. "Language-independent hybrid MT with PRESEMT". In: *Proceedings of HYTRA-2013 Workshop*. 2013, pp. 123–130. URL: http://www.aclweb.org/anthology/W13-28#page=137 (visited on 11/08/2016).

[84] Michael Tomasello. *Origins of Human Communication*. The Jean Nicod lectures 2008. Cambridge, Mass: MIT Press, 2008. 393 pp. ISBN: 978-0-262-20177-3.

[85] Gary Tonge Freng and Michele Wakefield. "Visicast: Enhanced Broadcast Services for the Deaf Community". In: *Ingenia* 14 (Nov. 2002), pp. 35–40. ISSN: 1472-9768. URL: http://www.ingenia.org.uk/Content/ingenia/issues/issue14/Tonge.pdf (visited on 12/21/2015).

[86] Jorge Toro. "Automated 3D animation system to inflect agreement verbs". In: *Proc. 6th High Desert Linguistics Conf.* 2004. URL: https://pdfs.semanticscholar.org/48c9/63b251bf95774dd2ea37a828b46cf0516d46.pdf (visited on 07/21/2016).

[87] Jorge Andres Toro, John C. McDonald, and Rosalee Wolfe. "Fostering better deaf/-hearing communication through a novel mobile app for fingerspelling". In: *International Conference on Computers for Handicapped Persons*. Springer, 2014, pp. 559–564. URL: http://link.springer.com/chapter/10.1007/978-3-319-08599-9_82 (visited on 07/21/2016).

[88] Universidade Católica Portuguesa. *Dicionário Terminológico em LGP – Ciências da Linguagem*. 2011. URL: http://pro-lgp.com/dicionario/ (visited on 12/22/2015).

[89] Carlos Velasco et al. "Proceedings of the 6th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusionVirtual Sign – A Real Time Bidirectional Translator of Portuguese Sign Language". In: *Procedia Computer Science* 67 (Jan. 1, 2015), pp. 252–262. ISSN: 1877-0509. DOI: 10.1016/j.procs.2015.09.269. URL: http://www.sciencedirect.com/science/article/pii/S1877050915031154 (visited on 07/22/2016).

[90] Margriet Verlinden, Corrie Tijsseling, and Han Frowein. "A Signing Avatar on the WWW". In: *Gesture and Sign Language in Human-Computer Interaction*. Ed. by Ipke Wachsmuth and Timo Sowa. Red. by G. Goos, J. Hartmanis, and J. van Leeuwen. Vol. 2298. Berlin, Heidelberg: Springer Berlin Heidelberg, 2002, pp. 169–172. ISBN: 978-3-540-43678-2. URL: http://link.springer.com/10.1007/3-540-47873-6_17 (visited on 07/20/2016).

[91] Carol J. Wideman, Edward M. Sims, and Inc. Seamless Solutions. "Signing Avatars". In: CSUN 1998 Conference. Los Angeles, United States, Feb. 1998. URL: http://www.dinf.ne.jp/doc/english/Us_Eu/conf/csun_98/csun98_027.html (visited on 07/20/2016).

[92] Rosalee Wolfe et al. "An Interface for Transcribing American Sign Language". In: *SIGGRAPH*. Vol. 99. 1999, p. 229. URL: http://asl.cs.depaul.edu/papers/SIGGRAPH99.pdf (visited on 07/21/2016).

[93] M. Wöllmer et al. "YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context". In: *IEEE Intelligent Systems* 28.3 (May 2013), pp. 46–53. ISSN: 1541-1672. DOI: 10.1109/MIS.2013.34.

[94] Chris Woodford. *Voice Recognition Software*. Explain That Stuff. June 7, 2015. URL: http://www.explainthatstuff.com/voicerecognition.html (visited on 12/10/2015).

[95] World Health Organization. *Millions of People in the World Have Hearing Loss That Can Be Treated or Prevented*. World Health Organization. Apr. 2013. URL: http://www.who.int/pbd/deafness/news/Millionslivewithhearingloss.pdf?ua=1 (visited on 12/09/2015).

[96] Chung-Hsien Wu et al. “Transfer-based statistical translation of Taiwanese sign language using PCFG”. In: *ACM Transactions on Asian Language Information Processing* 6.1 (Apr. 1, 2007), 1–es. ISSN: 15300226. DOI: 10.1145/1227850.1227851. URL: http://portal.acm.org/citation.cfm?doid=1227850.1227851 (visited on 07/21/2016).

[97] Inge Zwitserlood et al. “Synthetic signing for the deaf: ESign”. In: *Proceedings of the conference and workshop on assistive technologies for vision and hearing impairment, CVHI, Granada, Spain*. Citeseer, 2004. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.572.3022&rep=rep1&type=pdf (visited on 07/21/2016).

# Appendix A

# List of Sign Language applications

| Name | Lang. | Type | Plataform | Paid | Description |
|---|---|---|---|---|---|
| Signing Savvy | ASL | Static Videos | Web/iOS | Yes | Learning videos using 3D animated avatars without any further customization. |
| Sign Language With Sammi Signs | ASL | Dictionary | iOS | Yes | Limited number of words but the (non-human) avatar has smooth movements and fully facial expressions |
| Baby Sign and Learn | ASL | Dictionary | iOS/Android | Yes | Only around 300 signs are included but also supports Australian Sign Language, British Sign Language and New Zealand Sign Language. Lacks word search. The avatar has limited movement. |
| Affective Social Computing Laboratory | ASL | Text translator | Desktop | Yes | Part of a research project that seeks to integrate 3D animations in IDRT English-to-ASL and ASL-to-English translation engine, Sign Generator which already has over 30,000 English words, idioms, phrases, numbers, and symbols. |

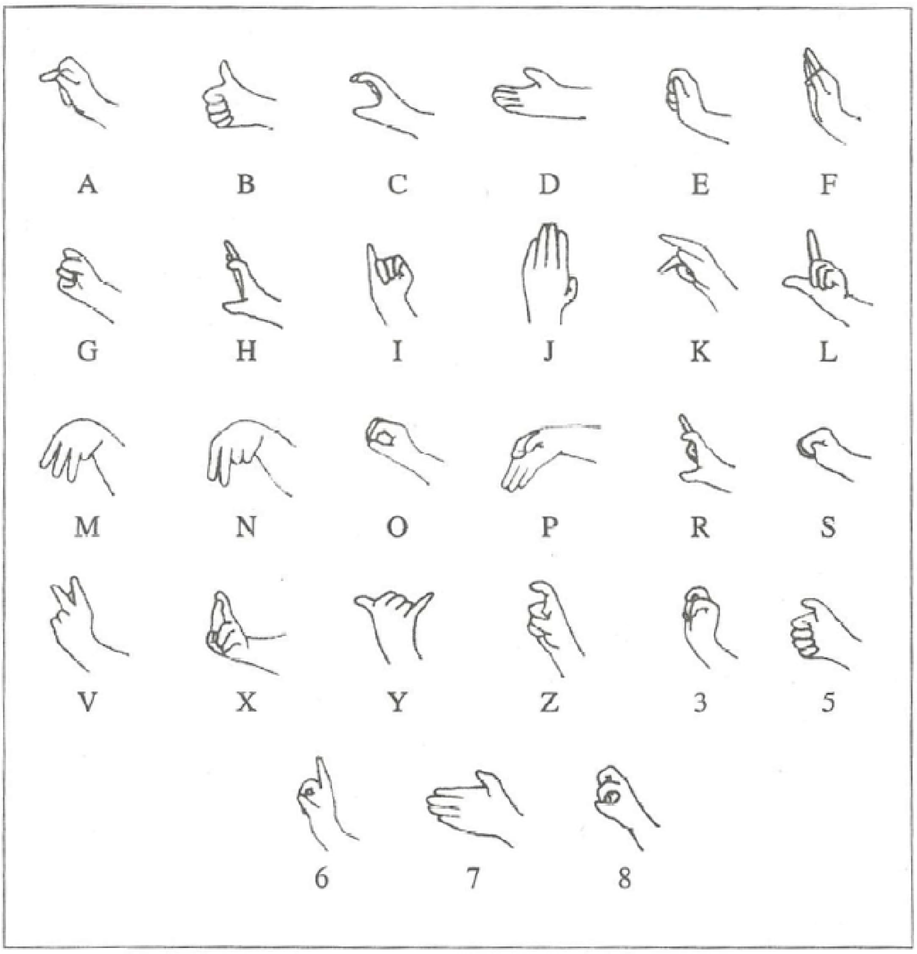
| | | | | | |
|---|---|---|---|---|---|
| Finger spell | ASL | Text translator | Web | No | Limited to only finger spelling but offers some customization (like controlling speed). |
| Sign 4 Me | ASL | Text translator | iOS | Yes | Fully turnaround 3D avatar with speed configuration and possibility of saving previous translations. Fluid movements, but lacks facial expressions. |
| Mimix | ASL | Translator | Android | Yes | Up to 5500 different signs available. Translates speech or text into Sign Language, and also translates text into spoken words. Semi-fluid movements and lack of facial expressions. |
| KBS 수화 날씨 | KSL | Translator | Android | No | Non-configurable 3D avatar that translates weather news (from text captions) to Korean Sign Language (KSL). |
| Surdophone | ASL | Translator | n/a | n/a | Application in testing process that allows both text and speech translation into ASL. The 3D avatar is cosmetically, however current version does not provide speed configuration. Movements are faster than desirable and lacks facial expressions. |

| | | | | | |
|---|---|---|---|---|---|
| ProDeaf | BSL | Translator | iOS/Android | No | Both text and speech can be translated into BSL through a 3D avatars that reproduces all the needed corporal and facial expressions and have speed configuration. Some physical characteristics of the avatar are exaggerated to improve understanding. Can connect to Internet to improve sign's accuracy and also allows request for any missing sign in the system. |
| Baby Signing Animated Words | ASL | Dictionary | iOS/Android | Yes | Limited set of signs expressed with a slow 3D avatar that lacks facial expressions. More suitable for hearing people that want to learn sign language because the audio instructions given along the signs. |
| textoSIGN | SSL | Translator | Desktop and iOS/Android | Yes | Allows text to Spanish Sign Language (SSL), Catalan Sign Language, German Sign Language or International Sign Language. Uses real-time integrated graphics engine that gives extremely fluid corporal and facial movements. Avatar customization is also available. |
| SiGame | ASL | Dictionary | iOS/Android | Yes | Words in this dictionary can be translated to American Sign Language, German Sign Language or International Sign Language. |

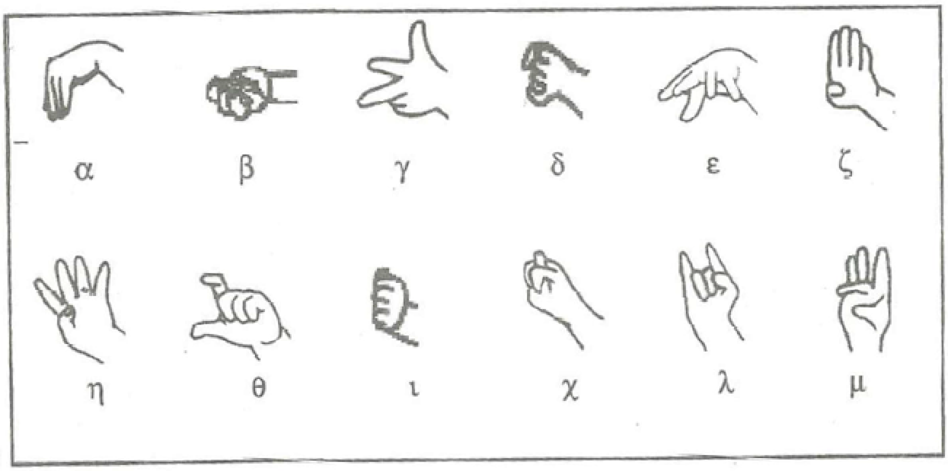
# Appendix B

# Hand configurations for LGP

The 52 hand configurations present on LGP according to Amaral's studies [5].




GANCHO DUPLO
BICO DE PÁSSARO
MÃO ABERTA
PINÇA FECHADA
GARRA ABERTA
BICO DE PATO
PUNAISE ABERTA
GARRA FECHADA
BICO DE PATO
PUNAISE FECHADO
CONCHA
BICO DE ÁGUIA ABERTO
PISTOLA


α
β
γ
δ
ε
ζ
η
θ
ι
χ
λ
μ

# Appendix C

# User Survey

## Survey

* Required

1. **Age:** *

*Mark only one oval.*

- 0 – 10
- 11 – 18
- 18 – 25
- 26 – 35
- 36 – 45
- 46 – 55
- +56

2. **Do you have any degree of hearing loss?** *

*Mark only one oval.*

- No *After the last question in this section, skip to question 11.*
- Yes (mild) *After the last question in this section, skip to question 4.*
- Yes (moderate) *After the last question in this section, skip to question 4.*
- Yes (severe) *After the last question in this section, skip to question 4.*
- Yes (profund) *After the last question in this section, skip to question 4.*

3. **What is the highest level of education you have completed?** *

*Mark only one oval.*

- None
- Elementary school
- High school
- Undergraduate
- Master's
- Doctorate
- Postdoctoral

4. **Do you use any devices for hearing loss?** *

*Mark only one oval.*

- Yes
- No

5. **Indicate if you use any of the following:** *

*Check all that apply.*

- [ ] Personal computer
- [ ] Smartphone
- [ ] Tablet

6. **How often do you use those devices?** *

*Mark only one oval.*

- Every day
- Weekly
- Monthly
- Never

7. **Select your most used applications:** *
*Check all that apply.*

- [ ] Word processor (Word, WordPerfect, Writer, etc.)
- [ ] Spreadsheet (Excel, Numbers, Calc, etc.)
- [ ] Instant messaging (Whatsapp, LINE, Messenger, etc.)
- [ ] Social networks (Facebook, Twitter, Instagram, etc.)
- [ ] News
- [ ] E-mail
- [ ] Buying or selling products and services
- [ ] Other: ..............................

8. **Which activities (that you are not performming nowadays) would you want to perform using those devices?** *
Please describe why you are not currently able to perform those activities.

..............................

..............................

9. **Would increase your usage of those devices if you were able to communicate more effectively through them?** *
*Mark only one oval.*

( ) Yes ( ) No

10. **What elements do you considered important when communicating with other people?** *

..............................

..............................

## Thank you for your participation.

*Stop filling out this form.*

11. **Would you be interested in a program that would improve communication with deaf and hearing impaired people?** *
*Mark only one oval.*

( ) Yes
( ) No

12. **Describe briefly which characteristics you find essential on such program:** *

..............................

..............................

..............................


## Thank you for your participation.

*Stop filling out this form.*