

Coupling Natural Language Processing and Animation Synthesis in Portuguese Sign Language Translation

Inês Almeida and Luísa Coheur

INESC-ID

Instituto Superior Técnico, Universidade de Lisboa

name.surname@tecnico.ulisboa.pt

Sara Candeias

Microsoft Language Development Center

Lisbon, Portugal

t-sacand@microsoft.com

Abstract

In this paper we present a free, open source platform, that translates in real time (written) European Portuguese into Portuguese Sign Language, being the signs produced by an avatar. We discuss basic needs of such a system in terms of Natural Language Processing and Animation Synthesis, and propose an architecture for it. Moreover, we have selected a set of existing tools that couple with our free, open-source philosophy, and implemented a prototype with them. Several case studies were conducted. A preliminary evaluation was done and, although the translation possibilities are still scarce and some adjustments still need to be done, our platform was already much welcomed by the deaf community.

1 Introduction

Several computational works dealing with the translation of sign languages from and into their spoken counter-parts have been developed in the last years. For instance, (Barberis et al., 2011) describes a study targeting the Italian Sign Language, (Lima et al., 2012) targets LIBRAS, the Brazilian Sign Language, and (Zafrulla et al., 2011) the American Sign Language. Some of the current research focus on sign language recognition (as the latter), some in translating text (or speech) into a sign language (like the previously mentioned work dedicated to Italian). Some works aim at recognising words (again, like the latter), others only letters (such as the work about LIBRAS). Only a few systems perform the two-sided translation, which is the case of the platform implemented by the Microsoft Asia group system (Chai et al., 2013), and the Virtual Sign Translator (Escudeiro et al., 2013).

Unfortunately, sign languages are not universal or a mere mimic of its country’s spoken counterpart. For instance, Brazilian Sign Language is not related with the Portuguese one. Therefore, none or little resources can be re-used when one moves from one (sign) language to another.

There is no official number for deaf persons in Portugal, but the 2011 census (Instituto Nacional de Estatística (INE), 2012) mentions 27,659 deaf persons, making, however, no distinction in the level of deafness, and on the respective level of Portuguese and Portuguese Sign Language (LGP) literacy. The aforementioned Virtual Sign Translator targets LGP, as well as the works described in (Bento, 2103) and (Gameiro et al., 2014). However, to the best of our knowledge, none of these works explored how current Natural Language Processing (NLP) tasks can be applied to help the translation process of written Portuguese into LGP, which is one of the focus of this paper. In addition, we also study the needs of such translator in terms of Animation Synthesis, and propose a free, open-source platform, integrating state of the art technology from NLP and 3D animation/modelling . Our study was based on LGP videos from different sources, such as the Spread the Sign initiative¹, and static images of hand configurations presented in an LGP dictionary (Baltazar, 2010). The (only) LGP grammar (Amaral et al., 1994) was also widely consulted. Nevertheless, we often had to recur to the help of an interpreter.

Based on this study we have implemented a prototype, and examined several case studies. Finally, we performed a preliminary evaluation of our prototype. Although much work still needs to be done, the feedback from deaf associations was very positive. The whole system is freely available.

This paper is organised as follows: Section 2

¹<http://www.spreadthesign.com>

describes the proposed architecture, and Section 3 its implementation. In Section 4 we present our prototype and, in Section 5, a preliminary evaluation. Section 6 surveys related work and Section 7 concludes, pointing directions for future work.

2 Proposed architecture

Figure 1 presents the envisaged general architecture.

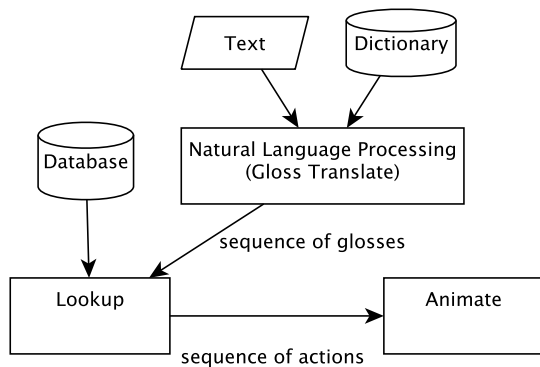


Figure 1: Proposed architecture

We followed a gloss-based approach, where words are associated to their ‘meaning’ through a dictionary. The order of the glosses is calculated according with the LGP grammar (structure transfer). Then, glosses are converted into gestures by retrieving the individual actions that compose it. In the last and final stage, the animation is synthesised by placing each action in time and space in a non-linear combination. The current platform is based on hand-crafted entries/rules, as there is no large-scale parallel corpus available that would allow us to follow recent tendencies in Machine Translation.

In the next sections we detail the three main components of this platform, namely the NLP, the Lookup and the Animate components, by focusing on the needs of the translation system and how these components contribute to it.

2.1 The Natural Language Processing component

As usual, the first step consists in splitting the input text into sentences. These are *tokenised* into words and punctuation. Then, possible orthographic errors are corrected. After this step, a basic approach could directly consult the dictionaries, find the words that are translated into sign lan-

guage, and return the correspondent actions, without further processing. However, other NLP tools can still contribute to the translation process.

Some words in European Portuguese are signed in LGP as a sequence of signs, related with the stem and affixes of the word. Therefore, a *stemmer* can be used to identify the stem and relevant suffixes (and prefixes), which allows to infer, for instance, the gender and number of a given word. Thus, we still might be able to properly translate a word that was not previously translated into LGP (or, at least, produce something understandable), if we are able to find its stem and affixes. To illustrate this, take as example the word ‘coelhinha’ (‘little female rabbit’). If we are able to identify its stem, ‘coelho’ (rabbit), and the suffix ‘inha’ (meaning, roughly, female (the ‘a’) and small (the ‘inho’)), we can translate that word into LGP by signing the words ‘female’ + ‘rabbit’ + ‘small’, in this order (which, in fact, is how it should be signed).

A *Part-of-Speech (POS) tagger* can also contribute to the translation process:

- It can couple with the stemmer in the identification of the different types of affixes (for instance, in Portuguese, a common noun that ends in ‘ões’ is probably a plural).
- As there are some morphosyntactic categories that have a special treatment in LGP, it is important to find the correspondent words. For instance, according with (Bento, 2103), articles are omitted in LGP ((Amaral et al., 1994) reports doubts in the respect of their existence), and thus could be ignored when identified. Also, the Portuguese grammar (Amaral et al., 1994) refers a temporal line in the gesturing space with which verbs should concord with in past, present and future tenses. Thus, to be able to identify the tense of a verb can be very important.
- A POS tagger usual feeds further processing, as for instance named entity recognisers and syntactic analysers.

A *Named Entity Recognizer* allows to identify names of persons. It is usual, among the deaf, to name a person with a sign (his/her *gestural name*), often with a meaning in accordance to his/her characteristics. For instance, names of public personalities, such as the current Portuguese prime

minister, usually have a gestural name. However, if this name is unknown, fingerspelling the letters of his/her name is what should be done.

A *Syntactic Analyser* is fundamental to identify the syntactic components of the sentence, such as subject, and object, as LGP is usually Object–Subject–Verb (OSV), while spoken Portuguese is predominantly Subject–Verb–Object (SVO). It does not matter if it is a dependency parser or a constituents-based one. The only requirement is that, at the end, it allows structure transfer rules to be applied to the glosses. Finally, a sentiment analyser would allow to infer subjective information towards entities and the generality of the sentence, so that emotional animation layers and facial expression reinforcement can be added to the result.

After all this processing, a bilingual dictionary (glosses) is consulted, so that meaningful sequences of words (glosses) are identified (lexical transfer), and a set of syntactic rules applied, so that the final order of the set of glosses is identified.

2.2 Lookup stage

Being given a sequence of glosses, the goal of the Lookup stage is to obtain a set of actions' identifiers for the animation.

The difficulty in designing this step is derived from the fact that many Portuguese words and concepts do not have a one-to-one matching in LGP. Also, gestures may be composed of several actions, which in turn, may be compound of several actions (the gestures subunits). Finally, some contexts need to be added to the database in order to help this step.

2.3 Animate

This stage receives a sequence of actions to be composed into a fluid animation, along with a set of hints on how best to do so, for example, if the gestures are to be fingerspelled or not. The animation stage is responsible for the procedural synthesis of the animation by blending gestures and gesture subunits together.

We propose an approach where gestures are procedurally built and defined from an high-level description, based on the following parameters identified in other works (Liddell and Johnson, 1989; Liddell, 2003) as gesture subunits: a) hand configuration, orientation, placement, and movement, and; b) non manual (facial expressions and

body posture).

The base hand configurations are Sign Language (SL) dependent. The parameter definition for orientation, placement and movement is often of relative nature. For example, gestures can be signed 'fast', 'near', 'at chest level', 'touching the cheek' and so on. The definition of speed is dependent on the overall speed of the animation, and the definition of locations is dependent on the avatar and its proportions.

2.3.1 Rig

To setup the character, an humanoid mesh with appropriate topology for animation and real-time playback is needed. Then, we need to associate it with the mechanism to make it move, the *rig*. We suggest a regular approach with a skinned mesh to a skeleton and bones.

Bones should be named according to a convention for symmetry and easy identification in the code. For the arms and hands, the skeleton can approximately follow the structure of a human skeleton. The rig ideally should have an Inverse Kinematics (IK) tree chain defined for both arms, rooting in the spine and ending in the hands. All fingers should also be separate IK chains, allowing for precise posing of contacts. Ideally, the IK chains should consider weight influence so that bones closer to the end-effector (hands and fingertips) are more affected, and the bones in the spine and shoulder nearly not so. The rig should also provide a hook to control the placing of the shoulder, and should make use of angle and other constraints for the joints, so as to be easier to pose and harder to place in an inconsistent position.

Finally, the rig should have markers for placement of the hands in the signing space and in common contact areas in the mesh. These markers ensure that gestures can be defined with avatar dependent terms (eg. 'near', 'touching the nose').

The markers in the signing space can be inferred automatically using the character's skeleton measures (Kennaway, 2002; Hanke, 2004), forming a virtual 3D grid in front of the character (Figure 2).

The markers in the mesh need to be defined manually and skinned to the skeleton in a consistent manner with the nearby vertices. Figure 3 shows a sample rig, with key areas in the face and body identified by a bone with a position and orientation in space.

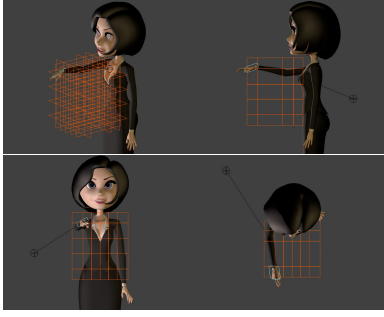


Figure 2: Virtual 3D marker grid defining the signing space in front of the character

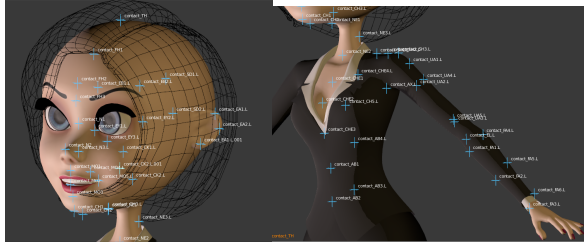


Figure 3: Definition of key contact areas in the rig

2.3.2 Building the gestures

It is now necessary to record (*key*) the poses in a good timing to build a gesture. Whichever the keying methodology, all basic hand poses and facial expressions should be recorded and can then be combined given the high level description of the gesture. The description should specify the gesture using the mentioned parameters: keyed hand configurations, placement and orientation using the spatial marks, and movement also using the marks and the overall speed of the animation.

The intersections defined by the grid from Figure 2, in conjunction with marks from Figure 3 define the set of avatar relative locations where the hands can be placed. Knowing the location where the hand should be, it can be procedurally placed with IK, guarantying physiologically possible animation with the help of other constraints.

Figure 4 shows the result of hand placement in a key area using two distinct avatars with significantly different proportions.

While this approach works well for static gestures, several problems appear when introducing movement. Gestures can change any of its parameters during the realisation, requiring a blending from the first definition (of location, orientation, configuration...) to the second. The type of blending is very important for the realism of the animation. Linear blending between two keys

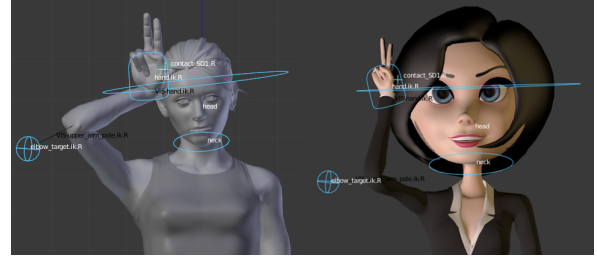


Figure 4: Avatars using the key areas

would result in robotic movements. Linear movement in space from one key location to another will also result in non realistic motions and even in serious collision problems (Elliott et al., 2007). For example, making a movement from an ear to the other. This is a problem of arcs. Additionally, more movements need to be defined in order to accommodate other phenomena, such as finger wiggling and several types of hand waving.

2.3.3 Blending the gestures

Moving to the sentence level, synthesising the final fluid animation is now a matter of agreeing the individual gestures in space, of realistic interpolation of keys in time, and of blending actions with each other in a non-linear way.

A reasoning module, capable of placing gestures grammatically in the signing space, and making use of the temporal line, entity allocation in space and other phenomena typically observed in SLs (Liddell, 2003) is needed.

The interpolation between animation keys is given by a curve that can be modeled to express different types of motion. The individual actions for each gesture should be concatenated with each other and with a 'rest pose' at the beginning and end of the utterance. The animation curves should then be tweaked, following the principles of animation.

Counter animation and secondary movement is also very important for believability and perceptibility. For example, when one hand contacts the other or some part of the head, it is natural to react to that contact, by tilting the head (or hand) against the contact and physically receiving the impact. Besides the acceleration of the dominant hand, the contact is mainly perceived in how it is received, being very different in a case of gentle brushing, slapping or grasping. This may be the only detail that allows distinguishing of gestures that otherwise may convey the same meaning.

Finally, actions need to be layered for expressing parallel and overlapping actions. This is the case for facial animation at the same time as manual signing and of secondary animation, such as blinking or breathing, to convey believability. The channels used by some action may be affected by another action at the same time. Thus, actions need to be prioritised, taking precedence in the blending with less important, or ending actions.

3 Implementation

We have chosen to use the Natural Language ToolKit (NLTK)² for NLP tasks and Blender³ as the 3D package for animation.

The NLTK is widely used by the NLP community and offers taggers, parsers, and other tools in several languages, including Portuguese. Thus, it was chosen for all the tasks concerning NLP.

Blender is an open-source project, which allows accessing and operating on all the data (such as animation and mesh) via scripting. It offers a Python API for scripts to interact with the internal data structures, operators on said data, and with the interface. Moreover, Blender also offers the infrastructure to easily share and install addons. Therefore, the prototype was implemented as an addon, with all the logic, NLP and access to the rig and animation data done in Python. The interface is a part of Blender using the pre-existing widgets, and the avatar is rendered in real-time using the viewport renderer.

3.1 The Natural Language Processing step

The modules implemented in our system can be seen in Figure 5.

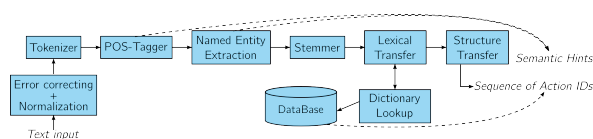


Figure 5: NLP pipeline

We also use the concept of “hint”, that is, a tag that suggests if a word should be signed or spelled. Three different types of hints are possible: GLOSS (words that are not numeric quantities and have a specific gesture associated), FGSPELL (for words that should be fingerspelled), and NUMERAL (for numeric quantities). The NLP module tries to

attribute a label to each word (or sequences of words), which are then used when consulting the dictionary (‘Lexical Transfer’).

In what concerns the NLP pipeline, we start with an ‘Error correcting and normalization’ step, which enforces lowercase and the use of latin characters. Common spelling mistakes should be corrected at this step. Then, the input string is split into sentences and then into words (tokenization). As an example, the sentence ‘o João come a sopa’ (‘João eats a soup’), becomes [‘o’, ‘João’, ‘come’, ‘a’, ‘sopa’]. A stemmer identifies suffixes and prefixes. Thus, the word ‘coelhinha’ (as previously said, ‘little female rabbit’), is understood, by its suffix (‘inha’), to be a female and small derivation of the root *coelh(o)*. Therefore, ‘coelhinha’ is converted into [MULHER, COELHO, PEQUENO], hinted to be all part of the same gloss.

We have used the treebank ‘floresta sintática’ (Afonso et al., 2002) for training our ‘POS-tagger’. The output of the POS-tagger for the sentence ‘o João come a sopa’ is now [(‘o’, ‘art’), (‘João’, ‘prop’), (‘come’, ‘v-fin’), (‘a’, ‘prp’), (‘sopa’, ‘n’)].

We have used a Named Entity Recognizer to find proper names of persons. Our system further supports a list of portuguese names and public personalities names with their matching gestural name. For these specific entities, the system uses the known gesture instead of fingerspelling the name.

The POS-tags and recognised entities also contribute with hints. These hints are then confirmed (or not) in the next step, the ‘Lexical Transfer’, where we converted all the words to their corresponding gloss, using the dictionary, where the word conversions are stored. As an example, the word ‘sopa’ would lead to [‘GLOSS’, [‘SOPA’]], ‘João’ to [‘FGSPELL’, [‘J’, ‘O’, ‘A’, ‘O’]] and ‘two’ to [‘NUMERAL’, [‘2’]] (notice that articles were discarded). Also, we provide the option of fingerspelling all the unrecognised words.

Finally, in what respects *Structure Transfer*, the current implementation only supports basic re-ordering of sequences of ‘noun - verb - noun’, in an attempt to convert the SVO ordering used in Portuguese to the more common structure of OSV used in LGP. We have also implemented another type of re-ordering, which regards the switching of adjectives and quantities to the end of the affected noun. Following

²<http://www.nltk.org>

³<http://www.blender.org>

this process, $[[\text{'GLOSS'}, \text{'SOPA'}]], [\text{'FGSPELL'}, \text{'J'}, \text{'O'}, \text{'A'}, \text{'O'}]], [\text{'GLOSS'}, \text{'COMER-SOPA'}]]$ is the final output for the sentence *O João come a sopa*, and the input *dois coelhos* (‘two rabbits’) results in $[[\text{'GLOSS'}, \text{'COELHO'}]], [\text{'NUMERAL'}, \text{'2'}]]$.

3.2 The Lookup step

The Lookup step, given a gloss, is done via a JSON file mimicking a database constituted of a set of *glosses* and a set of *actions*. Action ids are mapped to blender actions, that are, in turn, referenced by the glosses. One gloss may link to more than one action, which are assumed to be played sequentially.

Figure 6 shows that *coelho* (‘rabbit’) has a one-to-one mapping, that *casa* (‘house’) corresponds to one action and that *cidade* (‘city’) is a composed word, formed by *casa* and a morpheme with no isolated meaning.

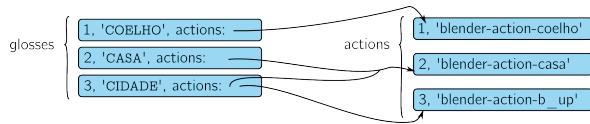


Figure 6: Database design

Knowing that gestures in LGP can be heavily contextualised, we added to the gloss structure an array of contexts with associated actions. Figure 7 shows the case of the verb *comer* (‘to eat’) that is classified with what is being eaten. When no context is given by the NLP module, the default is considered to be the sequence in ‘actions’.

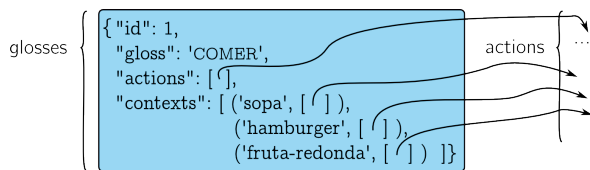


Figure 7: Supporting gloss contextualisation

3.3 The animation step

We start by setting the avatar by rigging and skinning. We chose *rigify* as a base for the rig, that needs to be extended with the spatial marks, to be used when synthesising the animation. The animation is synthesised by directly accessing and modifying the action and f-curve data. We always start and end a sentence with the rest pose, and, for concatenating the actions, we *blend* from one

to the other in a given amount of frames by using Blender’s Non Linear Action (NLA) tools that allow action layering. Channels that are not used in the next gesture, are blended with the rest pose instead. Figure 8 illustrates the result for the gloss sentence ‘SOPA J-O-A-O COME’.

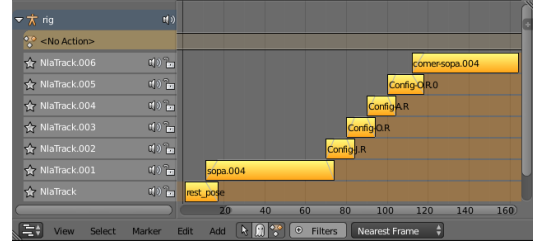


Figure 8: Action layering resulting of a translation

We adjust the number of frames for blending according to the hints received. For fingerspelling mode, we expand the duration of the hand configuration (that is originally just one frame) and blend it with the next fingerspelling in less frames than when blending between normal gloss actions. We also expand this duration when entering and leaving the fingerspell.

3.4 The interface

The interface consists of an input text box, a button to translate, and a 3D view with the signing avatar, which can be rotated and zoomed, allowing to see the avatar from different perspectives. Figure 9 shows the main translation interface (blue). Additionally, we provide an interface for exporting video of the signing (orange) and a short description of the project (green).

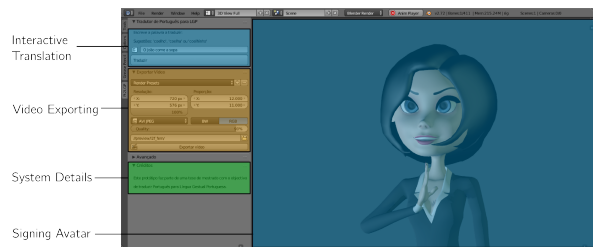


Figure 9: User Interface for the prototype

In what concerns the choice of the avatar, the character only needs to be imported into Blender and skinned to a *rigify* armature. Several characters were tested with success, with examples in Figure 10.



Figure 10: Example of some of the supported avatars

4 Case studies

Parallel to the development of the prototype, we devised a series of case studies to test the flexibility of the architecture and technology choices. We started with posing base hand configurations in a limited context case, passing then to full words, their derivations and blending between them. Finally, we tested the prototype with full sentences.

4.1 Basic gestures

All the 57 different hand configurations for LGP were manually posed and keyed from references gathered from (Baltazar, 2010; Amaral et al., 1994; Ferreira, 1997), and also from the Spread the Sign project videos. These hand configuration are composed of 26 hand configurations for letters, 10 for numbers, 13 for named configurations and 8 extra ones matching greek letters. This task posed no major problem.

4.2 Numbers

Numbers can be used as a quantitative qualifier, as the isolated number (cardinal), as an ordinal number, and as a number that is composed of others (eg. 147). Gestures associated with each number also vary their forms if we are expressing a quantity, a repetition or a duration, and if we are using them as an adjective or complement to a noun or verb.

Reducing the test case to ordinal numbers, the main difficulty is to express numbers in the order of the tens and up. Most cases seem to be “fingerspelt”, for example, ‘147’ is signed as ‘1’, followed by ‘4’ and ‘7’ with a slight offset in space as the number grows. Numbers from ‘11’ to ‘19’ can be signed with a blinking movement of the units’ number. Some numbers, in addition to these system, have a totally different gesture as an abbreviation, as is the example of the number ‘11’.

Doing a set of base hand configurations to start, proved to be a good choice as it allowed to test the

hand rig and basic methodology. The ten (0 to 9) hand configurations are shown in Figure 11.



Figure 11: Hand configurations for numbers (0-9)

4.3 Common nouns and adjectives

A couple of words were chosen, such as ‘coelho’ (‘rabbit’), with no serious criteria. Several words deriving from the stem ‘coelho’ were implemented, such as ‘coelha’ (‘female rabbit’) and ‘coelhinho’ (‘little rabbit’). In the former, the gesture for “female” is performed before the gesture for “rabbit”. In the latter, the gesture for the noun is followed with the gesture for the adjective (thus, ‘coelho pequeno’ (‘little rabbit’) and ‘coelhinho’ result in the same translation). Figure 12 illustrates both cases.



Figure 12: Gestures for ‘coelha’ and ‘coelhinho’

4.4 Proper Nouns

As previously said, if the person does not have a gestural name that is known by the system, the letters of his/her name should be fingerspelt. This morpho-syntactic category posed no major problem.

4.5 Verbs

When the use of the verb is plain, with no past or future participles, the infinitive form is used in

LGP. For instance, for the regular use of the verb ‘to eat’, the hand goes twice to the mouth, closing from a relaxed form, with palm up. However, this verb in LGP is highly contextualised with what is being eaten. The verb should be signed recurring to different hand configurations and expressiveness, describing *how* the thing is being eaten.

4.6 Sentences

After testing isolated words, we proceed to the full sentence: ‘O João come a sopa’, an already seen example, often used as a toy example in Portuguese studies. The verb gesture had to be extended, as for eating soup, it is done as if handling a spoon (for instance, for eating apples, the verb is signed as if holding the fruit)⁴. Considering the previous mentioned re-ordering from SVO (spoke Portuguese) to OSV (LGP), Figure 13 shows the resulting alignments.

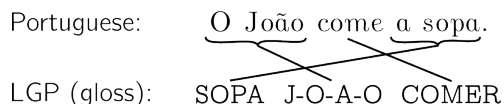


Figure 13: Alignment for European Portuguese and LGP of the sentence ‘John eats the soup’

5 Evaluation

A preliminary evaluation was conducted by collecting informal feedback from the deaf communities of two Portuguese deaf associations.

5.1 Usefulness

Both associations were asked for comments on the whole idea behind this work, and if and how such application would be useful. Both were skeptical towards the possibility of achieving accurate translations, or of animating enough vocabulary for a final product, but the feedback was positive for the idea of an application that would translate to LGP, even if just isolated words were considered.

5.2 Translation Quality

The correctness and perceptibility was evaluated by six adult deaf persons and interpreters. The avatar was set to play the translations for *coelha* (‘female rabbit’), *casa* (‘house’) and *coelhinho* (‘small rabbit’). The viewers were asked, individually, to say or write in Portuguese what was

⁴These contextualisations are not evident in the most recent and complete LGP dictionary (Baltazar, 2010).

being signed, with no previous information about the possibilities. In the second interaction of the system, a full sentence was added with limited variability of the form ‘A eats B’, where the verb ‘to eat’ is signed differently according to ‘B’. All the gestures were recognised as well as the sentence’s meaning, except for the inflection of the verb with a ‘soup’ object, that is signed as if handling a spoon. All of the testers recognised correctly the results, without hesitations, saying that the signs were all very clear and only lacking facial reinforcement to be more realistic.

5.3 Adequacy of the Avatar

The feedback from the deaf testers regarding the avatar looks was also very positive. There were no negative comments besides the observation that there is no facial animation. All hearing testers were also highly engaged with the system, testing multiple words and combinations, frequently mimicking the avatar.

The interest and attention observed, indicates that users had no difficulty in engaging with the avatar and found it either neutral or appealing. When asked about it, the answers were positive and the gesture blending and transitions, when noticed, was commented to be very smooth. However, sometimes the animation was deemed too slow or too fast. The animation generation should take play speed in consideration according to the expertise of the user.

6 Related Work

As ours, several systems also target the mapping of text (or speech) in one language into the correspondent signed language. Some of these systems resulted from local efforts of research groups or from local projects, and are focused in one single pair of languages (the spoken and the correspondent sign language); others aggregate the efforts of researchers and companies from different countries, and, thus, aim at translating different languages pairs (some using an interlingua approach). For instance, Virtual Sign (Escudero et al., 2013) is a Portuguese funded project that focus in the translation between European Portuguese and LGP, while eSIGN⁵ was an EU-funded project built on a previous project, ViSi-

⁵<http://www.sign-lang.uni-hamburg.de/esign/>

CAST⁶, whose aim was to provide information in sign language, using avatar technology, in the German and British sign languages, as well as in Sign Language of the Netherlands.

Our proposal follows in a traditional transfer machine translation paradigm of text-to-gloss/-avatar. Due to the lack of parallel corpora between European Portuguese and LGP, a data-driven method, example- and statistical-based approaches were not an option (see (Morrissey, 2008) for a study on this topic). Approaches such as the one of VISICAST (and eSIGN) (Elliott et al., 2008), which rely on formalisms, such as Discourse Representation Structures (DRS), used as intermediate semantic representations, were also not a solution, as, to the best of our knowledge, there are no free, open-source tools to calculate these structures for the Portuguese language. Thus, we focused in a simpler approach, that could profit from existing open-source tools, which could be easily used for Portuguese (and for many other languages).

We should also refer recent work concerning LGP, namely the works described in (Bento, 2103), (Gameiro et al., 2014) and (Escudeiro et al., 2013). The first focus on the mapping of human gestures into the ones of an avatar. The second targets the teaching of LGP, which the previously mentioned Virtual Sign also does (Escudeiro et al., 2014). The third contributes with a bidirectional sign language translator, between written portuguese and LGP, although it is not clear their approach in what respects text to sign language translation.

7 Conclusions and future work

We have presented a prototype that couples different NLP modules and animation techniques to generate a fluid animation of LGP utterances, given a text input in European Portuguese. We have further conducted a preliminary evaluation with the deaf community, which gave us positive feedback. Although a working product would be highly desirable and would improve the lives of many, there is still much to be done before we can reach that stage.

As future work we intend to perform a formal evaluation of our system, so that we can properly assess its impact. Also, we intend to extend

the existing databases. Particularly inspiring is ProDeaf⁷, a translation software for LIBRAS, the Brazilian Sign Language, that, besides several features, allows the crowd to contribute by adding new word/sign pairs. In our opinion, this is an excellent way of augmenting the system vocabulary, although, obviously, filters are needed in this type of scenarios. In the current version of the system, words that are not in the dictionary are simply ignored. It could be interesting to have the avatar fingerspelling them. Nevertheless, the system will probably have to be extended in other dimensions, as a broader coverage will lead to finer semantic distinctions, and a more sophisticated NLP representation will be necessary. We will also need to explore a way of simplifying the information concerning the contextualisation of a verb. For example, by storing categories of objects rather than the objects themselves. Moreover, we intent to move to the translation from LGP to European Portuguese. Here, we will follow the current approaches that take advantage of Kinect in the gesture recognition step.

Acknowledgements

We would like to thank to Associação Portuguesa de Surdos and Associação Cultural de Surdos da Amadora for all their help. However, the responsibility for any imprecision lies with the authors alone. This work was partially supported by national funds through FCT – Fundação para a Ciência e a Tecnologia, under project PEst-OE/EEI/LA0021/2013. Microsoft Language Development Center is carrying this work out in the scope of a Marie Curie Action IRIS (ref. 610986, FP7-PEOPLE-2013-IAPP).

References

- S Afonso, E Bick, R Haber, and D Santos. 2002. Floresta Sintáctica: A treebank for Portuguese. *LREC*, pages 1698–1703.
- M.A. Amaral, A. Coutinho, and M.R.D. Martins. 1994. *Para uma gramática da Língua Gestual Portuguesa*. Coleção universitária. Caminho.
- Ana Bela Baltazar. 2010. *Dicionário de Língua Gestual Portuguesa*. Porto Editora.
- Davide Barberis, Nicola Garazzino, Paolo Prinetto, and Gabriele Tiotto. 2011. Improving accessibility for

⁶See, for instance, http://www.visicast.cmp.uea.ac.uk/Visicast_index.html

⁷<http://web.prodeaf.net/>

- deaf people: An editor for computer assisted translation through virtual avatars. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '11, pages 253–254, New York, NY, USA. ACM.
- José Bento. 2103. Avatares em língua gestual portuguesa. Master's thesis, Faculdade de Ciências, Universidade de Lisboa, Lisbon, Portugal.
- Xiujuan Chai, Guang Li, Xilin Chen, Ming Zhou, Guobin Wu, and Hanjing Li. 2013. Visualcomm: A tool to support communication between deaf and hearing persons with the kinect. In *ASSETS 13: Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, New York, NY, USA. ACM.
- Eleni Efthimiou, SE Fontinea, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and François Goudenove. 2010. Dicta-sign?sign language recognition, generation and modelling: a research effort with applications in deaf communication. In *Proceedings of the 4th Workshop on the Representation and Processing of Sign Languages: Corpora and Sign Language Technologies*, pages 80–83.
- R. Elliott, John Glauert, J. R. Kennaway, I. Marshall, and E. Safar. 2007. Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society*, 6(4):375–391, October.
- R. Elliott, J.R.W. Glauert, J.R. Kennaway, I. Marshall, and E. Safar. 2008. Linguistic modelling and language-processing technologies for avatar-based sign language presentation. *Universal Access in the Information Society*, 6(4):375–391.
- Paula Escudeiro, Nuno Escudeiro, Rosa Reis, Maciel Barbosa, José Bidarra, Ana Bela Baltazar, and Bruno Gouveia. 2013. Virtual sign translator. In Atlantis Press, editor, *International Conference on Computer, Networks and Communication Engineering (ICCNC)*, Chine.
- Paula Escudeiro, Nuno Escudeiro, Rosa Reis, Maciel Barbosa, José Bidarra, Ana Bela Baltazar, Pedro Rodrigues, Jorge Lopes, and Marcelo Norberto. 2014. Virtual sign game learning sign language. In *Computers and Technology in Modern Education*, Proceedings of the 5th International Conference on Education and Educational technologies, Malaysia.
- A.V. Ferreira. 1997. *Gestúário: língua gestual portuguesa*. SNR.
- João Gameiro, Tiago Cardoso, and Yves Rybarczyk. 2014. Kinect-sign, teaching sign language to listeners through a game. *Procedia Technology*, 17(0):384 – 391.
- Thomas Hanke. 2004. HamNoSys-representing sign language data in language resources and language processing contexts. *LREC*.
- Instituto Nacional de Estatística (INE). 2012. Cens 2011, xv recenseamento geral da população, v recenseamento geral da habitação, resultados definitivos – portugal. Technical report, INE.
- J. R. Kennaway. 2002. Synthetic animation of deaf signing gestures. In *Gesture and Sign Language in Human-Computer Interaction*, pages 146 – 157. Springer.
- Scott K Liddell and Robert E Johnson. 1989. American Sign Language: The Phonological Base. *Sign Language Studies*, 1064(1):195–277.
- Scott K Liddell. 2003. *Grammar, gesture, and meaning in American Sign Language*. Cambridge University Press, Cambridge.
- M. A. S. Lima, P. F. Ribeiro Neto, R. R. Vidal, G. H. E. L. Lima, and J. F. Santos. 2012. Libras translator via web for mobile devices. In *Proceedings of the 6th Euro American Conference on Telematics and Information Systems*, EATIS '12, pages 399–402, New York, NY, USA. ACM.
- Sara Morrissey. 2008. *Data-driven machine translation for sign languages*. Ph.D. thesis, Dublin City University.
- Zahoor Zafrulla, Helene Brashear, Thad Starner, Harley Hamilton, and Peter Presti. 2011. American sign language recognition with the kinect. In *Proceedings of the 13th International Conference on Multimodal Interfaces*, ICMI '11, pages 279–286, New York, NY, USA. ACM.