Grammatical Facial Expressions Recognition with Machine Learning

Fernando de Almeida Freitas
Incluir Tecnologia
Itajubá, MG, Brazil
Universidade de São Paulo
São Paulo, SP, Brazil

Sarajane Marques Peres
Clodoaldo Aparecido de Moraes Lima
Felipe Venâncio Barbosa
Universidade de São Paulo
São Paulo, SP, Brazil

Abstract

The automated analysis of facial expressions has been widely used in different research areas, such as biometrics or emotional analysis. Special importance is attached to facial expressions in the area of sign language, since they help to form the grammatical structure of the language and allow for the creation of language disambiguation, and thus are called Grammatical Facial Expressions (GFEs). In this paper we outline the recognition of GFEs used in the Brazilian Sign Language. In order to reach this objective, we have captured nine types of GFEs using a Kinect™ sensor, designed a spatial-temporal data representation, modeled the research question as a set of binary classification problems, and employed a Machine Learning technique.

1 Introduction

Sign languages are the natural means of communication that are used by deaf people all over the world. This language modality emerges spontaneously and evolves naturally within deaf communities. These languages consist of manual (handshapes, position and movements) and non-manual components (facial expressions, head movements, poses and body movements), and hence are intrinsically multimodal languages.

In the sign languages, one of the functions of facial expressions is to convey grammatical information in a signed sentence. When this is the function employed for facial expressions, they are called Grammatical Facial Expressions (GFE). These non-manual signs are important for comprehension in all sign languages, since unless they are used, a sentence might be ungrammatical. For this reason, an analysis of facial expressions has been conducted in an attempt to automate the recognition of sign language. Some studies have shown improvements in their results when the study of facial expressions is included within a multimodal analysis approach, e.g. Nguyen and Ranganath (2012) who work on American Sign Language and von Agris, Knorr and Kraiss (2008) who work on German Sign Language. The automated analysis of GFE can bring objectiveness to several researches in the descriptive linguistic field and to the clinical practice, for example on diagnosis of language impairment expressed in signed languages, currently called atypical sign language.

There are only a limited number of facial movements that can be made by a human being, and thus, the number of possible facial expressions is also restricted. Moreover, within each sign language, there is a set of facial expressions that can be regarded as GFEs. In view of this, it is feasible to use automated techniques to implement GFE recognition. However, the usage of GFEs in signed speech faces two serious difficulties: (a) there is a variations in the GFEs carried out by different people; (b) the co-occurrence of the GFEs and other manual or non-manual features of the signed language, can cause frequent facial occlusions. Both difficulties cause problems for automated recognition methods.

This paper describes a study that involves modeling automated recognition of GFE using Machine Learning. The purpose of this is to analyze the complexity of the problem by taking account of sentences that fall within the scope of Brazilian Sign Language (Libras)[1], when used by a fluent signer. Machine Learning was chosen for this experience because of its ability to generalize, and the well-known neural network Multilayer Perceptron was used to induce the recognition models in the experiments.

This paper is structured as follows: Section 2 outlines theoretical concepts related to sign language and GFEs; related works are examined Section 3; in Section 4 there is a brief description of MLP architecture and the training strategy employed in the experiments; the GFE recognition problem is defined in Section 5, which includes a description of the datasets and data representation used in the experiments; the results of the experiments and analysis are given in Section 6 and, finally, the conclusion of the study and suggestions for future work are discussed Section 7.

2 Sign Language and Grammatical Facial Expressions

One of the earliest studies to formalize the structure of sign language was carried out by William C. Stokoe Jr. (1960), in 1960. Stokoe proposed that, in a sign language, signs have

[1] Our approach is extensible for other sign language, since they also apply GFE. However, due the differences among the languages, specific classifiers must be trained for each case.

three parts or parameters that are combined simultaneously. These parameters are as follows:

- *place of articulation*: the region where the hands are located in front of the body, while making a sign;
- *hand configuration*: the configuration resulting from the position of the fingers;
- *movement*: path or motion, and also speed, of the hands while the sign is being made.

Later, Battison (1974) proposed two more parameters: *palm orientation*, which refers to the direction the palms are facing during a sign execution; and *non-manual signs*, facial expressions, body posture, head tilt. Thus, these parameters are elements that must be combined to form signs, in the same way that phonemes must be combined to form words.

The facial expressions are relatively important in sign language because they communicate specific grammatical information in a sign sentence. In fact, they help to build the morphological and/or syntactic level in sign languages and are called Grammatical Facial Expression[2]. At the morphological level, the signer uses GFEs to qualify or quantify the meaning of a sign. At the syntactic level, the signer uses GFEs to build a particular type of sentence, or to specify the role of a phrase or clause within a sentence.

In this paper, we examine the recognition of GFEs that are used as grammatical markers from expressions in Brazilian Sign Language, as defined by Quadros and Karnopp (2004) and Brito (1995)[3]. There are eight types of grammatical markers in the Libras system:

- **WH-question**: generally used for questions with WHO, WHAT, WHEN, WHERE, HOW and WHY;
- **yes/no question**: used when asking a question to which there is a "yes" or "no" answer;
- **doubt question**: this is not a "true" question since an answer is not expected. However, it is used to emphasize the information that will be supplied;
- **topic**: one of the sentence's constituents is displaced to the beginning of the sentence;
- **negation**: used in negative sentences;
- **assertion**: used when making assertions;
- **conditional clause**: used in subordinate sentence to indicate a prerequisite to the main sentence;
- **focus**: used to highlight new information into the speech pattern;
- **relative clause**: used to provide more information about something.

By way of illustration, Table 1 and the Figure 1 show the effect of GFEs in a sign sentence and some examples of GFEs, respectively. By adopting a similar graphic scheme to that employed by Kacorri (2013), Table 1 provides three sentences with the same sequence of manual signs. However, the sentences are open to different interpretations depending on the accompanying facial expressions. Notice in this example it is essential that the GFEs are carried out in coordination with specific manual signs to allow the interpretation of the sign sentences. Table 1 shows examples *questions*. Notice that each GFE is formed by movements of facial features and head tilts/motions.

Table 1: Three different sentences with the same sequence of manual signs. The use of GFEs combined with a sign or set of signs gives the desired meaning to each sentence.

Meaning (in English)	**Grammatical Facial Expressions**		
John likes Mary.	assertion		
John, does he like Mary?	topic	yes/no question	
John dos not like Mary.	—	negation	
Libras signs	*John*	*likes*	*Mary*

(a) (b) (c)

Figure 1: Example of WH-question (a), yes/no question (b) and doubt question (c).

3 Related Works

As stated by Caridakis, Asteriadis and Karpouzis (2011), most of the approaches that have addressed the question of sign language recognition, have disregarded the need for an analysis of facial expressions. A good review of this topic is provided by Ong and Ranganath (2005). However, it was possible to find some papers in which the authors explore this feature of sign language.

The objective of Ari, Uyar, and Akarun (2008) is to establish a framework for face tracking and facial expression recognition. These authors used a set of expressions applied in sign language to test their approach. The sign language recognition task cannot be regarded as fully completed if the non-manual signs do not fulfil the objectives of the recognition. The exploration of these sign language components was undertaken by the authors in (von Agris, Knorr, and Kraiss 2008), (Kelly et al. 2009), (Michael, Metaxas, and Neidle 2009) and (Nguyen and Ranganath 2012). An interesting point related to sign language expression is the correlation between manual and non-manual

[2]From the standpoint of Psychology of Human Relations, in 1977 Ekman and Friesen (1977) proposed the existence of six emotions in the area of facial expression, which were believed to be a comprehensive means of understanding human reactions and feelings: happiness, surprise, anger, disgust, fear and sadness; although a neutral face is generally used as an initial benchmark for making comparison with the other feelings. These expressions, called Affective Facial Expressions, have the same meaning for deaf people and co-occur with GFEs in the expression of sign language.

[3]The studies (Quadros and Karnopp 2004) and (Brito 1995) for Libras are similar to studies of Liddell (1978) for American Sign Language and Sutton-Spence and Woll (1999) for British Signs.

signs during the speech, and how the two types of signs co-occur. These points are discussed by the authors in (Krnoul, Hruz, and Campr 2010) and (Kostakis, Papapetrou, and Hollmén 2011).

There is a serious drawback to sign language recognition which is that different conditions are used to make comparisons by the various research studies. Generally, each research group prepares a special and exclusive dataset and carries out its experiments on these datasets. Moreover, the quality measures applied in each approach are often different. Thus, only indirect and superficial comparisons can be made to show the similarities between the various studies. Table 2 shows characteristics that were taken from the papers listed in this section to make comparisons between them. The last line in this table refers to the present paper.

4 Machine Learning Techniques

Machine Learning is characterized by the development of methods and techniques that can be employed to implement inductive learning. In inductive learning, the hypothesis (partitions and functions) is determined from datasets – the bigger the dataset is, the more complex the resulting hypothesis can be (Russell and Norvig 2009). This type of learning can be achieved by supervised or unsupervised methods. In the supervised methods, (the modality applied in the present paper), the technique (or iterated algorithm) adjusts the parameters to minimize an error function. A well-known technique that implements inductive supervised learning is the neural network Multilayer Perceptron (MLP). The MLP is a feedforward and multilayer neural network architecture, usually trained with the also well-known backpropagation method (or generalized delta rule). In the experiments discussed in this paper, the backpropagation method is implemented using the gradient descent. For further information about the MLP neural network, see (Haykin 2008).

5 Grammatical Facial Expression Recognition

In this section we describe how the problem of GFE recognition has been modeled and create a proof of concept scope to support the experiments and analytical results.

Definition Problem

In this paper, a facial expression $FE \in \{FE_1, FE_2, ..., FE_{FEm}\}$ is a set of 3-dimensional datapoints $\{p_1, p_2, ..., p_{pm}\}$, taken from a human face, using the Microsoft Kinect™sensor[4]. These datapoints are composed of the following: x, y-coordinates, which are positions in pixels in an image captured by a Kinect RGB camera; z-coordinate, which is a depth measure, given in millimetres and captured by a Kinect infra-red camera. One FE carries out one or more semantic functions in a sign language, and then it can be called GFE. The semantic functions considered in this paper and the correlated facial expressions are described in Table 3.

Table 3: Semantic Functions X Facial Expressions Characteristics.

Semantic Functions	Eyebrows	Eyes	Mouth	Head
WH-question	↑			↑
Yes/no question	↑			↓
Doubt question	↓	∗	∗	⊖
Topic		⋄		↓
Negation	↓		∩	↔
Assertion				↕
Conditional clause	As in yes/no question			
Focus	As in topic			
Relative clause	↑			

↑ – upward head; ↓ – downward head
↕ – up and downward head; ↔ – left and rightward head
∗ – compressed mouth; ⋄ – open mouth; ∩ – downward mouth
⊕ – aproximation; ⊖ – detachment

The problem addressed in this paper is modeled as a binary classification task, in which an FE can be classified as either a specific GFE or as a neutral FE. The classification model is based on an analysis of FE characteristics, represented by measures stored in a vector representation.

Data representation

In the experiments discussed in this paper, 17 (x, y, z)–datapoints have been taken from the image of a human face to represent the face that corresponds to an FE. The software application captures around 27 frames per second and stores the RGB image and the 17 (x, y, z)–datapoints related to each frame. Figure 2 shows examples of GFEs with 17 (x, y)–datapoints. Notice that there are changes in the relative position between the points in different GFEs.

(a) (b) (c)

Figure 2: Example of GFEs with 17 (x, y)–points: (a) WH-question; (b) Yes/no question; (c) negative.

The raw data acquired by the sensor, are pre-processed to obtain measures that represent the facial features, and as a result, the FE. Thus, for each FE, two sets of measures are obtained: $D = \{d_1, d_2, ...d_{dm}\}$ and $A = \{a_1, a_2, ..., a_{am}\}$ are, the respective set of distances and set of angles between the pairs of points that describe the face. The distances and angles used in the experiments are illustrated in Figure 3.

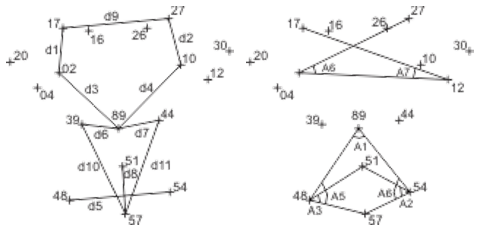
After the pre-processing, the information about distances and angles is arranged in a vector representa-

[4]The data acquisition was conducted with the aid of Microsoft Tracking Software Development Kit for Kinect for Windows (Face Tracking SDK). With functions provided by such a SDK we are able to identify $\{x, y, z\}$-points automatically and also use a prediction procedure in order to minimize noise

Table 2: Comparative summary among the related works.

Paper	Sign Language	Dataset	Data Type	# instance	Recognition Technique	Evaluation	Quality (max)
1	Turkish	from third-parts	sentences	132	(SVM) Support Vector Machine	5-Cross Validation	classification rate (0.90)
2	German	own	sentences	780	(HMM) Hidden Markov Models	Leave-one--Out	recognition rate (0.87)
3	Irish	own	utterances	160	HMM		accuracy and reliability (0.95 0.93)
4	American	ASLLRP*	utterances short narratives	400 15	SVM	Cross Validation	precision, recall accuracy (0.90 1 0.95)
5	Czech	UWB-07-SLR-P corpus**	signs	200	–	–	–
6	American	ASLLRP***	utterances	873	–	–	–
7	American	own	sentences	297	SVM / HMM	Holdout	recognition rate (0.87)
8	Brazilian	own	sentences	225	Multilayer Perceptron	Holdout	f-score (0.91)

1- (Ari, Uyar, and Akarun 2008); 2- (von Agris, Knorr, and Kraiss 2008); 3- (Kelly et al. 2009); 4- (Michael, Metaxas, and Neidle 2009); 5- (Krnoul, Hruz, and Campr 2010); 6- (Kostakis, Papapetrou, and Hollmén 2011); 7 - (Nguyen and Ranganath 2012); 8- **This paper**.
* - http://www.bu.edu/asllrp/SignStream/; ** - www.elra.info; *** - http://www.bu.edu/asllrp/cslgr/



(a) (b)

Figure 3: In the experiments 11 distances and 7 angles were used, as shown in (a) and (b) respectively.

tion of temporal-space. The temporal information is represented through a "window" procedure that is employed in the sequence of frames that compose a video. Let w be the size of the window of the frames (the number of frames included in a window), then the temporal vector representation for two consecutive windows is $v_1 = \{measures\ of\ frame\ 1; ...; measures\ of\ frame\ w\}$, $v_2 = \{measures\ of\ frame\ 2; ...; measures\ of\ frame\ w+1\}$, and so on. In the experiments, w varies in $[1..FE_{lenght}/2]$, where FE_{lenght} is the number of frames needed to perform a FE.

Following this temporal representation, two characteristic vectors were planned that could be used in the experiments: **Representation 1**, which is given by

$$v_1 = \{d_1^1, .., d_{dm}^1, a_1^1, .., a_{am}^1; ...; d_1^w, .., d_{dm}^w, a_1^w, .., a_{am}^w\},$$
$$v_2 = \{d_1^2, .., d_{dm}^2, a_1^2, .., a_{am}^2; ...; d_1^{w+1}, .., d_{dm}^{w+1}, a_1^{w+1}, .., a_{am}^{w+1}\},$$

and so on, where distances d and angles a were calculated using 17 (x, y)-datapoints; and **Representation 2**, which is given by

$$v_1 = \{d_1^1, .., d_{dm}^1, a_1^1, .., a_{am}^1, z_1^1, .., z_{17}^1; ...; d_1^w, .., d_{dm}^w, a_1^w, .., a_{am}^w, z_1^w, .., z_{17}^w\},$$
$$v_2 = \{d_1^2, .., d_{dm}^2, a_1^2, .., a_{am}^2, z_1^2, .., z_{17}^2; ...; d_1^{w+1}, .., d_{dm}^{w+1}, a_1^{w+1}, .., a_{am}^{w+1}, z_1^{w+1}, .., z_{17}^{w+1}\},$$

and so on, where z is the depth information of each datapoint.

Datasets

The data consists of streams of sign language to indicate the performance achieved by a person fluent in a Libras (called here the *user*), when forming sentences that represent specific grammatical markers. Since the objective of the experiment was to analyze the complexity of the GFE recognition problem, each sentence was carefully chosen so that its execution in Libras did not require gestures that could cause occlusion of the face during the image acquisition process[5]. The choice of sentences was made with the support of a Libras and Linguistic expert (see the sentences in Table 4).

The dataset consists of 225 videos recorded in five different recording sessions carried out with the user. In each session, one performance of each sentence was recorded. The description of the dataset is given in Table5. There are 15255 frames arranged in the 225 videos, and each frame was labeled by a human coder (a person fluent in Libras) to allow the GFE classification models to be established and validated. The videos were recorded with one person, and the labeling was carried out for another person to avoid any bias in the labeling process.

6 Experiments and Results

The experiments were carried out to determine the effectiveness of a Machine Learning approach in the recognition of

[5]Further study of the sensor's robustness or the treatment of occlusions is beyond the scope of this paper.

Table 4: Set of sentences that compose the dataset.

WH-questions	**Yes/no questions**
When did Waine pay?	Did Waine buy a car?
Why did Waine pay?	Is this yours?
What is this?	Did you graduate?
How do you do that?	Do you like me?
Where do you live?	Do you go away?
Doubt questions	**Topics**
Did Waine buy A CAR?	University ...
	I study at "anonymous"!
Is this YOURS?	Fruits ... I like pineapple!
Did you GRADUATE?	My work ...
	I work with technology!
Do you like ME?	Computers ...
	I have a notebook!
Do you GO AWAY?	Sport ... I like volleyball!
Assertions	**Negatives**
I go!	I don't go!
I want it!	I didn't do anything!
I like it!	I never have been in jail!
I bought that!	I don't like it!
I work there!	I don't have that!
Conditional clauses	**Focus**
If rain, I don't go!	It was WAYNE who did that!
If you miss, you lose.	I like BLUE.
If you don't want, he wants.	It was Wayne who pay for it!
If you don't buy, he wants.	The bike is BROKEN.
If it's sunny, I go to the beach.	YOU are wrong.
Relative clauses	
The girl who fell from bike? ... She is in the hospital!	
The *"anonymous"* university? ... It is located in *"anonymous"*!	
That enterprise? ... Its business is tecnology!	
Waine, who is Lucas's friend, is graduated in Pedagogy!	
Celi, the deaf school, is located in "anonymous"!	

GFEs, when modeled as a binary classification task. The experiments were conducted by means of the supervised neural networks (MLP), implemented in the Matlab®Neural Toolbox. A gradient descent backpropagation procedure was used to train the MLP models using: ten hidden neurons; constant learning training during the training, with its value varying in the $\{1, 0.5, 0.25, 0.125, 0.065, 0.031, 0.01\}$ set through the different training sessions.

Three disjoint sets were used as training, validation and test sets to control the stopping conditions for training and meet the training and testing requirements of each model. The training set was composed of 12 sentences, the validation set was composed of 2 sentences and 6 sentences composed the test set in the experiments. Since the number of frames that form each sentence and the number of frames that correspond to a GFE in a sentence vary in the dataset, the dataset is unbalanced; thus the F-score measure was chosen as the principal way of assessing the quality of the models. However, the recall and precision measures are also shown.

All of the MLP architectures were employed with each data representation; this took account of all possible window sizes, and involved adopting a user-dependent approach (i.e. the training, validation and test sets consisted of data from a specific user). The performance results are listed in Table 6.

Table 5: Description of the dataset. The symbols (+ frames) and (− frames) indicate, respectively, the number of frames belonging to GFEs, and the number of frames belonging to neutral FEs.

GFE	+ **frames**	− **frames**	% of + **frames**
WH-question	643	962	0.40
Yes/no question	734	841	0.46
Doubt question	1100	421	0.72
Topic	510	1789	0.22
Focus	446	863	0.34
Negative	568	596	0.48
Assertion	541	644	0.45
Conditional clause	448	1486	0.23
Relative clause	981	1682	0.36

In general, the results show that the problem of GFE recognition problem is easier to solve by using **temporal information**, since the F-scores obtained in the experiments were higher when the windows of the frames were used. As was expected, there was a need to assess the suitability of using features taken from the motions so that the FEs in a sign language could be analyzed. This was largely due to the importance of the head motions or head tilting in characterizing the sense of the grammatical sentence.

The analysis of the experimental results that were obtained with the use of a **unique frame** to represent the data, showed indications of the complexity of the GFE. In these experiments, the GFEs that were strongly characterized by eyebrow movements (they are colored *gray* in Table 6) and the GFE *negative* were poorly recognized by the MLP models. Furthermore, the usage of depth information led to worse results. On the other hand, in the cases of GFEs where the performances of the models were considered good (F-score above 0.75), the inclusion of depth information slightly improved the results.

The complex pattern of GFE and the improvement obtained with the depth information, remained in the experiment when the **windows of frames** were applied; with the exception of the sentences that had *doubt questions*, where the depth information did not lead to any improvements.

Finally, with regard to the **size of the windows**, it could be observed that in most of the cases in which good results were obtained by means of depth information, smaller windows were needed to represent the data and achieve a suitable recognition.

In summarizing the analytical results of the experiments, the following assertions can be made: **(a)** temporal information is essential to solve the target problem; **(b)** the combination of the measurements of distances and angles, depth information and temporal information represented by windows of frames allows good recognition models to be established for specific types of GFEs; **(c)** the grammatical markers for *WH-questions*, *doubt questions*, *assertions*, *topic* and *focus* has a lower complexity level when the MLP capabilities are taken into account; (d) the grammatical markers to *yes/no questions*, *conditional clauses*, *relative clauses* and *negative* have a higher level of complexity .

Table 6: Experiments Results: F-scores – recall – precision and windows size (between brackets) for $w > 1$.

GFE	Representation 1			Representation 2		
	$w = 1$	$w = 2..FE_{lenght}/2$		$w = 1$	$w = 2..FE_{lenght}/2$	
WH-questions	**0.77** – 0.70 – 0.84	**0.84** – 0.81 – 0.88	(9)	**0.80** – 0.75 – 0.85	**0.87** – 0.80 – 0.96	**(6)**
Yes/no questions	0.73 – 0.59 – 0.96	**0.83** – 0.73 – 0.98	(3)	0.49 – 0.33 – 0.92	0.63 – 0.48 – 0.93	(2)
Doubt questions	**0.84** – 0.94 – 0.76	**0.89** – 0.92 – 0.86	(6)	0.44 – 0.29 – 0.87	**0.82** – 0.80 – 0.85	**(5)**
Topics	**0.80** – 0.75 – 0.85	**0.89** – 0.85 – 0.92	(6)	**0.82** – 0.74 – 0.92	**0.90** – 0.85 – 0.95	**(4)**
Negative	0.44 – 0.33 – 0.66	0.69 – 0.96 – 0.54	(3)	0.06 – 0.03 – 0.67	0.54 – 0.56 – 0.48	(10)
Assertion	**0.76** – 0.62 – 0.98	**0.87** – 0.79 – 0.96	(4)	**0.83** – 0.81 – 0.86	**0.89** – 0.90 – 0.88	**(2)**
Conditional clauses	0.65 – 0.50 – 0.91	0.68 – 0.55 – 0.89	(4)	0.39 – 0.25 – 0.91	0.51 – 0.36 – 0.91	(3)
Focus	**0.88** – 0.82 – 0.94	**0.91** – 0.89 – 0.94	(2)	**0.91** – 0.74 – 0.92	**0.91** – 0.88 – 0.94	(2)
Relative clauses	0.59 – 0.42 – 0.98	0.67 – 0.50 – 0.99	(4)	0.43 – 0.27 – 0.95	**0.77** – 0.67 – 0.91	(6)

7 Final Considerations

This paper has undertaken a study of the GFEs used in Brazilian Sign Language recognition by employing a Machine Learning technique. Our experiments have covered nine types of sentences with distinct grammatical sense, and the recognition problem was modeled by means of a set of binary classification tasks. As far as could be corroborated from the results of our bibliographic review, this is the first study in the automated recognition of GFEs in Libras. Thus, our conclusions about the complexity of the expressions represent an unpublished contribution. Moreover, although it is impossible to make direct comparisons in giving the results – since different datasets and metrics are employed in the correlated literature – the results obtained in our experiments are in accordance with those found in related work.

The next stage in our research will include the following: (a) improving our dataset so that it can include sentences formed by more than one user. Initial experiments in this line of research indicated that, although the MLP performance has been lower, it is possible to calibrate the classifier to use it with, at least, two users; (b) using data obtained in different sessions, so that be possible to study variations in GFEs caused by different moods, or disposal or contexts. In this sense, also could be possible to analyze the co-occurrence of GFEs and Affective Facial Expressions; (c) exploring the data representation to meet the needs of research and thus obtain better results for the more complex GFEs; (d) analyzing our results by using the evaluation metrics commonly used by Linguistic researchers, as discussed in (Madeo, Lima, and Peres 2013). This type of analysis could base further explanations about the performances obtained for each grammatical expressions.

References

Ari, I.; Uyar, A.; and Akarun, L. 2008. Facial feature tracking and expression recognition for sign language. In *23rd Int. Symp. on Computer and Information Sciences*, 1–6. IEEE.

Battison, R. 1974. Phonological deletion in american sign language. *Sign language studies* 5(1974):1–14.

Brito, L. F. 1995. *Por uma gramática de línguas de sinais*. Tempo Brasileiro.

Caridakis, G.; Asteriadis, S.; and Karpouzis, K. 2011. Non-manual cues in automatic sign language recognition. In *4th Int. Conf. on Pervasive Technologies Related to Assistive Environments*, 43. ACM.

Ekman, P., and Friesen, W. V. 1977. Facial action coding system.

Haykin, S. O. 2008. *Neural Networks and Learning Machines*. Prentice Hall, 3 edition.

Kacorri, H. 2013. Models of linguistic facial expressions for american sign language animation. *Accessibility and Computing* (105):19–23.

Kelly, D.; Reilly Delannoy, J.; Mc Donald, J.; and Markham, C. 2009. A framework for continuous multimodal sign language recognition. In *Int. Conf. on Multimodal interfaces*, 351–358. ACM.

Kostakis, O.; Papapetrou, P.; and Hollmén, J. 2011. Distance measure for querying sequences of temporal intervals. In *4th Int. Conf. on Pervasive Technologies Related to Assistive Environments*, 40. ACM.

Krnoul, Z.; Hruz, M.; and Campr, P. 2010. Correlation analysis of facial features and sign gestures. In *10th Int. Conf. on Signal Processing*, 732–735. IEEE.

Liddell, S. K. 1978. Nonmanual signals and relative clauses in american sign language. *Understanding language through sign language research* 59–90.

Madeo, R. C. B.; Lima, C. A. M.; and Peres, S. M. 2013. Gesture unit segmentation using support vector machines: segmenting gestures from rest positions. In *Proc. of Symposium of Applied Computing*, 46–52. ACM.

Michael, N.; Metaxas, D.; and Neidle, C. 2009. Spatial and temporal pyramids for grammatical expression recognition of american sign language. In *11th Int. Conf. on Computers and Accessibility*, 75–82. ACM.

Nguyen, T. D., and Ranganath, S. 2012. Facial expressions in american sign language: Tracking and recognition. *Pattern Recognition* 45(5):1877–1891.

Ong, Sylvie, C. W., and Ranganath, S. 2005. Automatic sign janguage analysis: a survey and the future beyond lexical meaning. *Trans. on Pattern Analysis and Machine Intelligence* 27(6):873–891.

Quadros, R. M. d., and Karnopp, L. B. 2004. Língua de sinais brasileira: estudos lingüísticos. *Porto Alegre: Artmed* 1:222.

Russell, S., and Norvig, P. 2009. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 3 edition.

Stokoe, W. 1960. Sign language structures.

Sutton-Spence, R., and Woll, B. 1999. *The linguistics of British Sign Language: an introduction*. Cambridge University Press.

von Agris, U.; Knorr, M.; and Kraiss, K.-F. 2008. The significance of facial features for automatic sign language recognition. In *8th Int. Conf. on Automatic Face & Gesture Recognition*, 1–6. IEEE.