



# Method for Emotion Corpus Validation from the Consensual Identification of Patterns in Alzheimer's Patients

Pablo Gómez<sup>(✉)</sup>, Alexandra González-Eras<sup>(✉)</sup>,  
and Pablo Torres-Carrión<sup>(✉)</sup>

Universidad Técnica Particular de Loja, San Cayetano Alto Loja, Loja, Ecuador  
{pfgomez1, acgonzalez, pvtorres}@utpl.edu.ec

**Abstract.** The present research proposes a method for the construction of a corpus for the early detection of Alzheimer's, by identifying basic patterns of emotions on video, from the collection of patient information in the form of videos, analysis and identification of emotions based on facial expressions and finally validation by two statistical measures: weighted Kappa and Kappa. Applying the method on a corpus of 40 videos, an average score of 0.60 obtained in the Kappa index and 0.67 in the weighted Kappa index, which indicates a good agreement among the observers, and provides encouraging results for the use of the corpus in automatic learning, on the patterns of emotions that allow the detection of Alzheimer's disease.

**Keywords:** Emotion analysis Alzheimer's disease · Kappa index  
Weighted Kappa index · Alzheimer's detection

## 1 Introduction

Research related to mental illness is gaining importance for society. The World Health Organization indicates that “it is essential to conduct studies to adults over 60 years of age because they are more vulnerable to suffering a type of dementia at a certain point in their lives” [1]. Alzheimer's is a type of dementia that in the elderly population causes difficulties in the development of activities, directly affecting memory, thinking, orientation, and learning ability and how to express themselves. In that sense, observation of the emotions that Alzheimer's patients express when they carry out the activities affected by the disease, is a viable diagnostic option, instead of expensive and invasive methods that are currently the only way to obtain information about the disease.

Consequently, it is necessary to undertake the development of reliable sources about the emotions, expressed by Alzheimer patients in their daily activities, with the following purposes:

---

The original version of this chapter was revised: Third author name has been updated. The erratum to this chapter is available at [https://doi.org/10.1007/978-3-319-73450-7\\_110](https://doi.org/10.1007/978-3-319-73450-7_110)

1. To use them as input in the development of other investigations, oriented towards the creation of environments and technological applications that promote the quality of life of both people with the disease and their caregivers.
2. To establish non-invasive methods for collecting disease information from patients, to identify patterns that allow the study of disease progression.

To date, there are no corpus or datasets that identify basic emotions of people with Alzheimer's, so the purpose of this project is precisely build a corpus of videos of Alzheimer's patients, to identify patterns of emotions, which allows to research on the disease.

In the present work, we propose a method for creating a dataset of videos of Alzheimer's patients, in whom emotions are recognized and validated by statistical methods, ensuring the consistency and reliability of identification. For the construction of the dataset, using interview videos to analyze the emotions experienced by patients, two observers responsible for labeling the emotions manually analyzed each video and independently, using in both cases a window time of twenty seconds. Then, Kappa and weighted Kappa index are used to represent the proportion of agreement observed after eliminating the agreement by chance; thus establishing the reliability of the corpus and ensuring its use in other research studies related to the Alzheimer's disease.

## 2 Related Works

There are previous studies, where corpus constructed through observer annotation and the use of the Kappa index for the study of pharmacological substances and their drug-drug interactions, with results between 0.55 and 0.72 [2]. In the analysis of textual information, a corpus of medical phrases constructed from 10000 abstracts annotated by experts, validated by the Kappa index with an average score of 0.62 indicating a substantial agreement [3].

In other works the recognition of patterns of emotions based on polarity in videos of Alzheimer's patients is performed by manual annotation by experts, and the use of human emotion detection (HER) software; achieving an agreement between annotations, according to the weighted Kappa index of 0.7 [4]. In [5] CSIS, a software that collects information from cancer medical reports of 710 patients, of which 179 have been evaluated by two independent pathologists to measure the agreement between the human expert and the software.

In conclusion, the use of Kappa and Kappa weighted indexes is considered adequate for the treatment of subjective evaluation results of emotion patterns, obtained from observation of annotators [15].

## 3 Methodology

The proposed methodology allows the identification of emotions through videos with Alzheimer patients for the construction of a corpus of videos. For this reason, we have worked with four phases: Phase I part of the research design, Phase II video analysis and Phase III methods are used statistics for corpus validation (Fig. 1).

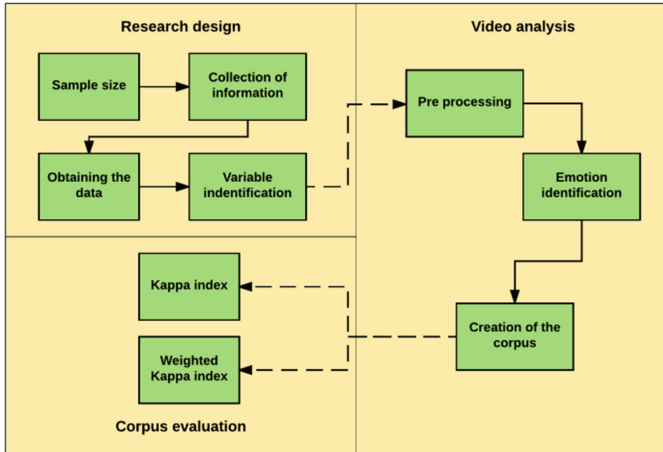


Fig. 1. Corpus development methodology.

### 3.1 Phase I: Research Design

The research design establishes a set of activities to follow sequentially to achieve the research objective [13]. With this purpose, activities defined such as sample size, collection of information, data collection and identification of variables, which detailed below.

#### 3.1.1 Sample Size

The sample comprise 40 videos, corresponding to interviews of elderly people belonging to care centers, private individuals and interviews with the elderly published on YouTube. The distribution of the videos according to their source detailed below.

#### Physical sources

- Centro del Día San José (Loja): 4
- Centro del Adulto mayor (Catamayo): 2
- Centro del Adulto mayor (Vilcabamba): 2
- Fundación del Adulto mayor (Quito): 2
- Ciudad de Loja: 1
- Ciudad de San Pedro de la Bendita: 1

#### Digital sources

- YouTube: 28

#### 3.1.2 Collection of Information

In order to create the conditions for the collection, criteria established for each source of information. In the case of physical sources, a team is formed by, the interviewer, a medical or psychology professional and a camera operator. On the side of the interviewee, elderly people considered according to their age, their disease level and the written consent given by the patient or a family member to participate in the experiment.

Additionally, requirements are set for the preparation of the interview site, such as a family space for the interviewee, with adequate lighting and acoustics and the presence of the caregiver or relative with whom the elder has a greater affinity.

For the digital sources, the search criteria are “interview Alzheimer’s people”, obtaining 3990 results; and “Alzheimer interviews”, obtaining 530.00 results. Then two scorers select videos that comply with requirements such as videos with free license (that is to say published in the platform to be used without any restriction on the part of the author), with a duration of 2 and 6 min, and in which the interviewee appears in frontal position. As a result, a number of 28 valid videos obtained.

### 3.1.3 Obtaining the Data

In the case of physical sources, an interview conducted with each person for a duration of five minutes, where the interviewer used a questionnaire based on the Folstein Test, in order to obtain visual answers to the questions raised and thus detect the expressions of the interviewee. Table 1 presents the structure of the questionnaire, where it is observed that the selected classifications of the test: general information, temporal orientation and spatial orientation are sufficient to evaluate the cognitive level of patients and provide a brief analysis of their mental state [6]; also that the number of questions selected is appropriate to the duration of the videos.

**Table 1.** Questionnaire structure for interviews.

Questionnaire	Answers
Informative data	What’s your name?
	How old are you?
	What year were you born?
	What are your children’s names?
	What are their siblings named?
Temporary orientation	What year are we in?
	On what day of the week?
	What month of the year?
Spatial orientation	In what place do we find ourselves?
	On what floor?
	What city?
	In what country?

For the digital sources, videos were downloaded from the YouTube platform, following criteria such as: quality of the interview, i.e. the responses of the interviewee are similar to those obtained with the questions of the questionnaire; and, the quality of the video file that must be high, in order to clearly appreciate the face and expressions of the interview.








### 3.1.4 Variable Identification

For the analysis process, the following groups of variables are established:

**Informative.** That in the case of physical sources gather personal information of the interviewee, such as name, age, geographical location, among others; while in digital sources, in addition to those mentioned data on origin and copyright.

**Annotations.** They refer to the recognition of the emotion expressed by the interviewee at a specific time of the interview. The annotations are made on segments of the video

**Table 2.** Description of basic emotions used in identification.

Expression	Emotion	Description
	Joy	Express joy in your face when has a sense of well-being and security. Feeling cheerful is a positive emotion.
	Surprise	The person reflects in his face startled, astonished. This basic emotion arises when we are not prepared and something comes up suddenly.
	Sadness	The patient expresses in his face solitude, pessimism in general sadness manifested by loss or rejection.
	Fear	When it is in danger, what it produces anxiety, insecurity and uncertainty.
	Anger	Emotion that a person experiences when you feel harmed or when we see that someone does something unjust to another person.
	Disguss	Emotion that produces escape or avoidance responses.
	Contempt	Emotion that expresses disrespect to another person through a deal derogatory.

(windows of 20 s), which are analyzed to detect emotions according to the descriptions proposed by [14] and detailed in Table 2 [7]. It also establishes the definition of “Neutral”, for those expressions that are not within those mentioned in the table.

### 3.2 Phase II: Video Analysis

Video analysis performed both on physical and digital sources, including the following steps:

**Pre-processing.** Where each video was edited, in order to eliminate those temporary spaces, where the interviewer does not have a conversation with the interviewee. Figure 2 presents the development of this process, using VideoPad<sup>1</sup> tool. In this article, the interviewee’s face was covered, to protect his identity.

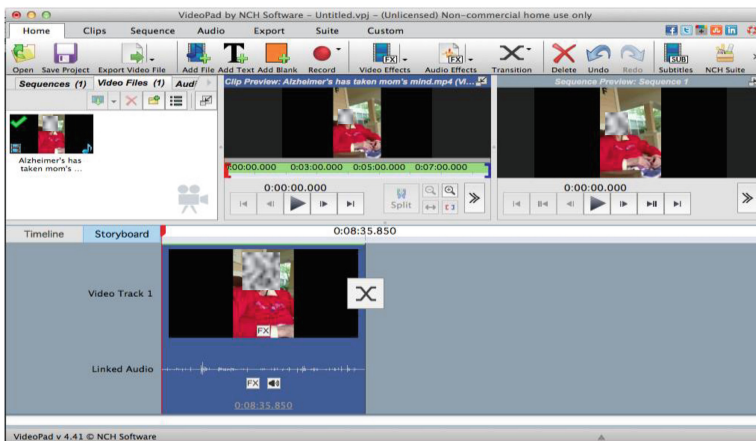


Fig. 2. Video pre-processing

**Identification of Emotions.** The pre-processed videos analyzed by two observers who for each video establish temporary windows of 20 s, within which they identify the emotions based on the facial expressions of the interviewee, according to the descriptions presented in Table 2. In case of no recognize the expression; the observer classifies it as “Neutral”. Each observer performs this process individually, which ends when the observer has checked every second of the video of all the established temporary windows, and has labeled with an emotion type all existing expressions.

<sup>1</sup> Video Pad, available online on <http://www.nchsoftware.com/videopad/index.html>.

**Creation of the Corpus.** The corpus includes a set of multidimensional characteristics, collected in the different phases of the methodology, and explained below:

- A collection of digital files of interview videos was established, tagged with a unique identifier.
- A general register is established, where each row corresponds to a video, which is assigned a unique identifier, along with the informative variables.
- For each observer a second record was created, where each row corresponds to a video, recognized by its identifier and each cell contains the emotions found during the identification process.
- Each observer creates a third record, where each row corresponds to a video, each column corresponds to one type of emotion and each cell stores the frequency of occurrence of emotion in the video.

### 3.3 Fase III: Corpus Evaluation

The evaluation of the corpus allows measuring the level of agreement of the annotations made by the observers, for this, the following steps were established:

- **Contingency Matrix.** Constructed by each video, where, the observations of the two observers were crossed in order to obtain the matches between them. Figure 3 presents a contingency matrix, where it was observed that, for each type of emotion, the annotations made by Observer 1 are located in the rows, while the observer 2 annotations are located in the columns; the diagonal of the matrix represents the number of times that, for a type of emotion, the two observers agree.

		Observador2								
		joy	sadness	anger	fear	surprise	disgust	contempt	neutral	total
Observador1	joy	43	1	0	0	0	0	0	1	45
	sadness	3	30	3	0	0	0	0	0	36
	anger	4	0	29	0	0	0	0	0	33
	fear	0	5	0	6	0	0	0	0	11
	surprise	0	0	0	0	2	0	0	4	2
	disgust	0	0	0	0	0	0	0	0	0
	contempt	0	0	0	0	0	0	0	0	0
	neutral	2	0	0	0	0	0	0	9	11
	total	52	36	32	6	2	0	0	10	138

Fig. 3. Contingency matrix for the case of Video01

- **Kappa Index Calculation.** Cohen (1960) introduces the Kappa coefficient, which represents the ratio according to or after eliminating the agreement by pure chance. To calculate the Kappa coefficient in any problem of nominal scale between two judges, there are two relevant quantities, expressed in Eq. 1 [8].

$$kappa = \frac{\sum Po - \sum Pe}{1 - \sum Pe} \tag{1}$$

Where:

**Po.** Indicate the proportion of units in which observers agreed. Equation 2 represents the proportion of agreements observed.

$$Po = \sum_{i=1}^k \frac{P_{ij}}{N} \quad (2)$$

**Pe.** It is the proportion of units from which an agreement is expected by chance (3).

$$Pe = \sum_{i=1}^k \frac{P_{i+} * P_{j+}}{N^2} \quad (3)$$

- **Weighted Kappa Index Calculation.** Cohen (1968) introduced a Kappa extension called a weighted Kappa statistic that is denoted by the symbol Kw, is a measure in which weights are assigned in each disagreement [9]. To calculate the weighted Kappa statistic the same as part of the Kappa index equation as shown below:

$$K_w = \frac{Po_w - Pe_w}{1 - Pe_w} \quad (4)$$

Where:

**Po<sub>w</sub>** It is the weighted observed agreement given by the product sum of all the records in the contingency table by the weights assigned to each record and divided by the total number of subjects N being evaluated (5).

$$Po_w = \sum_{i=1}^k \sum_{j=1}^k w_{ij} \left( \frac{nij}{N} \right) \quad (5)$$

**Pe<sub>w</sub>.** Agreement proportion expected by weighted pure chance, given by the product of the marginal borders of the contingency table, corresponding to the cells that have the same weight, these products are added and multiplied by the weight. The same procedure is performed with all the cells of the table, to then obtain a total value and divide it by the square of the total number of subjects N.

$$Pe_w = \sum_{i=1}^k \sum_{j=1}^k w_{ij} \left( \frac{Ai}{N} \right) \left( \frac{Bi}{N} \right) \quad (6)$$

- **Allocation of Quadratic Weights.** The assignment of the weights was considered using a linear system and later a system that is the most used called quadratic or biquadratic system [10, 11] (7).

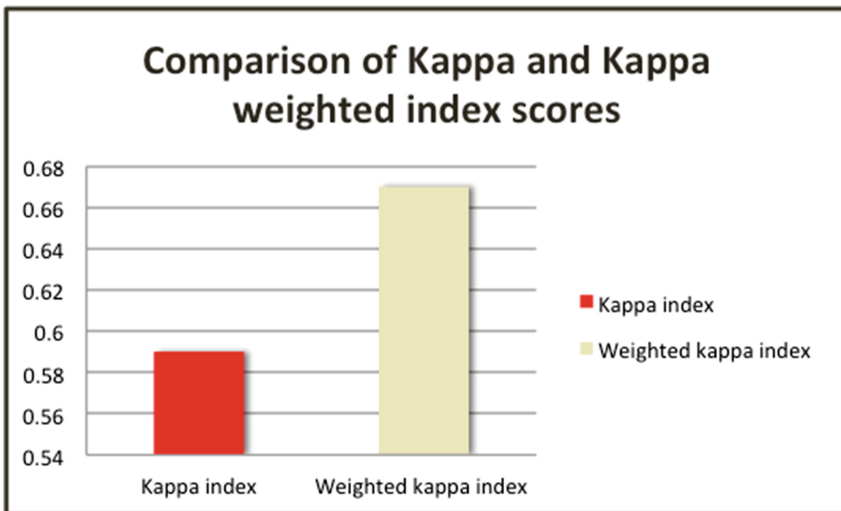


$$w = 1 - \frac{|i - j|^2}{k - 1} \quad (7)$$

- **Valuation Scale.** Landis and Koch (1977) proposed a Kappa index rating scale divided into six classifications to facilitate their interpretation. The values of  $k$  go from 0.00 to 1.00, being 0.00 the value where there is more disagreement and 1.00 the interval where there is greater agreement among the evaluators. Their classification indicates that the Kappa index can be Poor, when their value is (0.00), Light, when their value oscillates between 0.01 to 0.20, Acceptable (0.21 to 0.40), Moderate (0.41–0.60), Good (0.61 to 0.80) and Very Good (0.81 to 1.00) [12].

## 4 Results

Applying the Kappa and weighted Kappa index to the sample of the 40 videos showed that the mean of the Kappa index is 0.60 and the mean of the weighted Kappa index is 0.67. Therefore, it concluded that the index concordance force Kappa is Moderate, and the strength of concordance of the weighted Kappa index is good. By applying the weighted Kappa index and assigning the quadratic weights, we can conclude that the value of the weighted Kappa index increases relative to the Kappa index (Fig. 4).



**Fig. 4.** Comparison of Kappa and Kappa weighted index scores

It is also observed that the results of the Kappa index reflect that the largest number of videos in the corpus have a concordance in the annotations, located in the Moderate (0.6) and Good (0.8) levels of the Landis scale and Koch (Fig. 5).

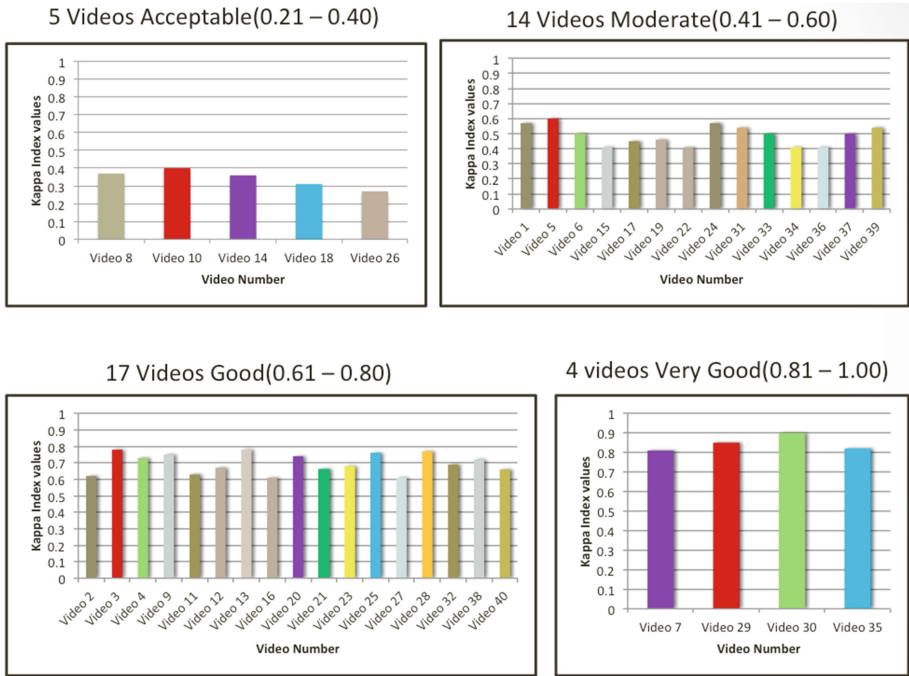


Fig. 5. Kappa index results

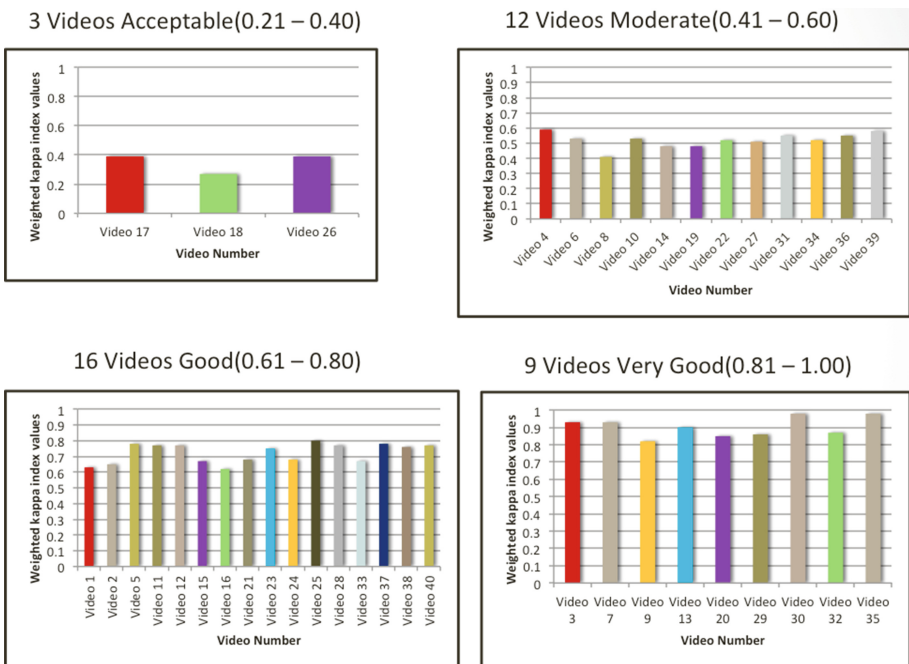


Fig. 6. Weighted Kappa index results

It is also observed that the results of the weighted Kappa index reflect that the largest number of videos in the corpus have a concordance in the annotations, located in the Moderate (0.6) and Good (0.8) levels of the Landis scale and Koch (Fig. 6).

## 5 Conclusions and Future Works

The present work presents a method for corpus validation, identifying basic patterns of emotions on video and evaluating by weighted Kappa and Kappa. Applying the method on a corpus of 40 videos, we obtain levels of agreement of Moderate and Good in the scale of Landis and Koch, which provides encouraging results for the use of the corpus in automatic learning, on the patterns of emotions related with emotions in Alzheimer's patients.

The following steps are directed towards the extension of the experiment to another collection of videos to obtain a corpus with a greater level of agreement, allowing experimentation in the recognition of emotion patterns and the use of corpus in research on detection and learning of Alzheimer's disease characteristics.

## References

1. OMS: Alzheimer's report. <http://www.who.int/mediacentre/factsheets/fs362/es/7>
2. Herrero-Zazo, M., Segura-Bedmar, I., Martínez, P., Declerck, T.: The DDI corpus: an annotated corpus with pharmacological substances and drug–drug interactions. *J. Biomed. Inf.* **46**(5), 914–920 (2013)
3. Kim, S.N., Martínez, D., Cavedon, L., Yencken, L.: Automatic classification of sentences to support evidence based medicine. *BMC Bioinform.* **12**(2), 1 (2011)
4. Narváez, M.: Análisis y reconocimiento de la expresión facial de la emoción en video de personas con demencia. <http://dspace.utpl.edu.ec/handle/123456789/15600>
5. McCowan, I.A., Moore, D.C., Nguyen, A.N., Bowman, R.V., Clarke, B.E., Duhig, E.E., Fry, M.J.: Collection of cáncer stage data by classifying free-text medical reports. *J. Am. Med. Inf. Assoc.* **14**(6), 736–745 (2007)
6. Folstein, M., Folstein, S., McHugh, P.: A practical state method for. *J. Psychiatr. Res.* **12**, 189–198 (1975)
7. Ekman, P., Friesen, W.V.: *Facial Action Coding System*. Consulting Psychologists Press, Palo Alto (1977)
8. Cohen, J.: A coefficient of agreement for nominal scale. *Educ. Psychol. Meas.* **20**, 37–46 (1960)
9. Cohen, J.: Weighted Kappa: nominal scale agreement provision for scaled disagreement or partial credit. *Psychol. Bull.* **70**(4), 213 (1968)
10. Ruiz, A., Morillo, L.: *Epidemiología clínica aplicada: Investigación clínica aplicada*. In: Bogotá: Médica Internacional (2004)
11. Díaz, I., Sidorov, G., Suárez, S. (n.d.): *Creación y evaluación de un diccionario marcado con emociones y ponderado para el español* (2004)

12. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. *Biometrics* **33**(1), 159–174 (1977). <https://doi.org/10.2307/2529310>
13. Bravo, S., Bravo, R.S.: *Técnicas de investigación social: teoría y ejercicios*. Thomson (2003)
14. Ekman, P., Friesen, W.V.: *Unmasking the Face: A Guide to Recognizing Emotions from Facial Clues*. Malor Books, Cambridge (2003)
15. Manning, C.D., Hinrich, S.: *Foundations of statistical Natural Language Processing*. MIT Press, Cambridge (1999)