# How Humans Analyse Lexical Indicators of Sentiments- A Cognitive Analysis Using Reaction-Time

*Marimuthu K and Sobha, Lalitha Devi*

AU-KBC RESEARCH CENTRE, MIT Campus of Anna University, Chrompet, Chennai, India

marimuthuk@live.com,sobha@au-kbc.org

ABSTRACT

In this paper, we try to understand how the human cognition identifies various sentiments expressed by different lexical indicators of sentiments in opinion sentences. We use the psychological index, Reaction Time (RT) for the analysis of various lexical indicators required for understanding the sentiment polarity. The test bed was developed using linguistic categories of lexical indicators of sentiments and selected sentences which have various levels of sentiments. Experimental results indicate that variations in syntactic categories of the lexical indicators influence the thought in deciding sentiments at varied levels. The results from this work is to be used for fine tuning machine learning algorithms which are used for sentiment analysis and it can also be used in the development of real time applications such as educational tools to better educate students, particularly those with neurocognitive disorders.

KEYWORDS: Cognition, Syntactic Categories, Reaction-Time, Lexical Indicators, Sentiment Analysis

*Proceedings of the 2nd Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2012)*, pages 81–90,
COLING 2012, Mumbai, December 2012.

## 1 Introduction

Sentiment analysis and opinion mining have gained importance in research for the last several years, giving emphasis on classification of opinions in movie and product reviews. Sentiments are expressed in a text through two ways, 1) explicitly marked lexical indicators and 2) implicitly carried out through non-evaluative and non-visibly subjective statements such as sarcasm. Sentiment analysis is a process to identify the opinion of a statement. The analysis is done by various disciplines such as linguistics, cognitive science, computational linguistics etc.

Our goal is to find the level of cognition associated with various syntactic categories of lexical indicators of sentiments in opinion sentences. Generally, the syntactic categories of words vary depending on the context in which they appear in a sentence. And specifically, in sentiments, the lexical indicators can have different Part-Of-Speeches (POS) as the sentence construction may vary depending on the reviewers' writing style. On analysing various sentiments, we found that lexical indicators commonly associate themselves with four syntactic categories i.e. adjective(ADJ), adverb(ADV), noun(N), and verb(V). In our study, we consider only these lexical indicators which bring out sentiments of statements. Analysis further brought in that is lexical indicators inherently intensifies the sentiments at varied levels. Consider the following examples.

1. *one of the greatest family-oriented fantasy-adventure movies.*

Here, "*greatest*" which is an adjective acts as a positive sentiment-indicating word.

2. *unfortunately, the story and the actors are served with a hack script.*

Here, "*hack script*" which is a noun acts as negative sentiment-indicating words.

Each of the above statements has a lexical indicator which acts as a stimulus for deciding the polarity of the snippet. But, the level of cognition required to identify and comprehend the stimulus varies with various syntactic categories among various participants. So, we concentrate to find this varied level of cognition using our psychological experimentation.

## 2 Literature Survey

Sentiment analysis is a thrust area in computational linguistics and different approaches such as heuristics based, linguistic rules based, statistical based, machine learning based, and cognitive methods, are used to classify sentiments.

At linguistics level, sentiments can be extracted from a sentence using various approaches like lexicon based approach, exploiting morphological features, semantic orientation of individual words etc. One typical example is a contextual intensifier. (Polanyi and Zaenen, 2004) defined contextual intensifiers as lexical items that weaken or strengthen the base valence of the term modified. The work by (Benamara et al., 2006) determine the importance of syntactic category combinations in opinions. They suggest that adjective and adverb combinations are better than adjectives alone in determining the strength of subjective expressions within a sentiment sentence.

At Rule-based level, polarity prediction depends mainly on hand-coded rules. Class Sequential Rules (CSR) had been studied in the work of (Luke K.W. Tan, 2011) and generalized polarity prediction rules were introduced that allows polarity rules to be applied across different domains.

Statistical approaches involve implementation of machine learning algorithms for sentiment classification. Performance of three machine learning methods (Naïve Bayes, Maximum Entropy classification, and Support Vector Machines) for sentiment classification of movie reviews had been analysed by (Pang et al., 2002). They concluded that these methods do not perform as well on sentiment classification as on traditional topic based categorization. Most prior work on the specific problem of categorizing expressed opinionated text had focussed on binary classification i.e. positive vs. negative(Turney, 2002; Pang et al., 2002; Dave et al., 2003; Yu et al., 2003).

At the cognitive level, (Ignacio Serrano et al., 2009) had tried to simulate high level cognitive processes in human mind. Their model relied on semantic neural network to build a cognitive model for human reading. Further, it is well known from (Saul Sternberg, 2004) that human reading is a process of sequential perception over time during which the brain builds mental images and inferences which are reorganized and updated until the end of the text. While reading a text, these mental images will help people to relate similar texts, extract, and classify them. The dependence between cognitive linguistics and sentiments, its metaphor and prototypical scenario, and various sentiments or emotions briefed in (Ignacio Serrano et al., 2009).

In this work, our main aim is to understand how various syntactic categories influence sentiment prediction and we find this using RT index. The rest of the paper is organized as follows. Section 3 discusses reaction time opinion mining experiment. Section 4 explains results, graphical representation, and comparison of various syntactic categories. Section 5 focuses on the inferences drawn from results. Section 6 explains major problems encountered in our experiment. In section 7 we give a detailed discussion on results and in final section we conclude by giving future work directions.

## 3 Reaction Time Opinion Mining Experiment

### 3.1 Definition

Reaction Time (RT; also called response time or latency) is the time taken to complete a task by a human. Specific to our goal, RT is the total time taken by a participant to read an opinion sentence, interpret the sentiment polarity and record the choice. In general, there are two parameters in this experiment, Recognition RT and Choice RT. Recognition RT is the time in which the subjects should respond and Choice RT is the time in which the subjects have to select a response from a set of possible responses. In our experiment, each obtained RT represents a combination of Recognition and Choice RT.

### 3.2 Input Data Description

For our experiment, we had used the publicly available Pang et al. movie review corpus. The data set had 5331 positive-polarity snippets and 5331 negative-polarity snippets. The data is clean i.e. contained only English language text. From this dataset, we took 1000 unique snippets i.e. 500 from positive-polarity and 500 from negative-polarity category. Since we consider only four syntactic categories (ADJ, ADV, N, and V), each syntactic category will have 125 unique positive and negative snippets. Based on the POS of lexical indicator of snippets, each snippet (positive or negative) in input dataset is manually classified into one of the four categories until we reach a total count of 250 snippets (125 positive and 125 negative) for each category.

## 3.3 Set data Preparation and Representation

In each category, snippets are manually marked either as simple or complex opinion based on the number of words. A set of 20 opinions is prepared for every participant. In order to maintain a constant measuring factor among various participants' RT values, and to provide a blend of varying difficulty level opinions and also to avoid mere guessing of sentiment polarity, six different techniques are followed while forming an opinion set. 1) First, each set has equal number of simple and complex opinions from positive and negative category and none of the same category opinions are displayed to participants in a follow-up fashion. 2) Second, the count of all syntactic categories is maintained at a fixed ratio so that Mean and SD measurements are not biased. Hence, for a set with 20 snippets, each category's snippet count will be 5 i.e. 5ADJs, 5ADVs, 5Ns, 5Vs. 3) Third, the sentiment polarity count is also maintained at a fixed ratio to maintain a balance between both polarity categories. So, in a set with 20 snippets, each polarity's count will be 10. 4) Fourth, none of the same polarity snippets are displayed in a consecutive manner throughout the test. This is to avoid mere guessing of sentiment polarities. 5) Fifth, snippets are jumbled in a random fashion so that no two snippets of same syntactic category follow one another. 6) Sixth, snippets in a particular set will not be repeated in any other set.

## 3.4 Experimental Setup and RT Measurement

The system design is an important factor in RT measurement and its importance is emphasized in (Saul Stemberg, 2004). While designing the user interfaces of RT system, stimulus design considerations specified in (Saul Stemberg, 2004) such as large displays, minimized noise etc., had been strictly followed. This is to confirm that these factors should not make the user uncomfortable during the test thereby affecting RTs in an adverse manner which is not desirable. To accurately measure RT, we also strictly adhered to the following design considerations. At any given moment during the testing time, only one sentiment snippet is shown at the top of the webpage along with a running timer at top right corner of the page. The polarity choice buttons are always placed nearer to the end of snippets to attenuate any millisecond delay that will be caused when moving the cursor away from snippets towards the buttons. The cut-off time for answering each snippet in question is *15s* after which the timer will expire. Providing cut-off time is to measure the precise RT which can be set depending upon the task. It is also to attenuate the effect of overtime which otherwise would make the final graph skewed. The timer runs separately for each snippet. So, the participants can take as much time as needed before navigating to next snippet but they are not advised to do so. The number of snippets per participant is limited to 20 so that they will not get bored which otherwise will affect the RT adversely (Saul Sternberg, 2004). We developed a web-based system to collect and record response time.

A participant begins the test by reading the rules. Then s/he enters the testing session and starts answering the choices for all 20 snippets in the given set. The RT values of each snippet will be automatically recorded in a database which will be retrieved later for further analysis. This procedure is repeated for all 50 participants and the corresponding RT values are recorded. Ideal state of a participant is a condition in which s/he mentally reacts normally under normal circumstances and is also devoid of any serious physical or mental disorder that degrades Intelligent Quotient (IQ) level.

### 3.5 Evaluation

We calculated mean and SD values for our statistical RT analysis. Prior to calculation of these values, we have considered three cases of RT values i.e. *Raw case, Correct case*, and *Wrong case*. *Raw case* contains RTs of both correctly predicted and wrongly predicted opinions i.e. true positive, false positive, true negative, and false negative. *Correct case* contains only the RTs of correctly predicted opinions out of the given set i.e. true positive and true negative. *Wrong case* contains only the RTs of wrongly predicted opinions out of the given set i.e. false positive and false negative.

## 4 Experimental Results

The calculated Mean and SD of various RT values are tabulated here. The measured RT values are in centiseconds (cs). 10millisecond=1cs

| **Syntactic Category** | **Positive Opinion** | | **Negative Opinion** | |
|---|---|---|---|---|
| | **SD** | **Mean** | **SD** | **Mean** |
| Adjective | 128.796 | 350.89 | 204.619 | 415.10 |
| Adverb | 181.545 | 383.50 | 211.351 | 432.58 |
| Noun | 190.602 | 454.84 | 238.067 | 506.73 |
| Verb | 180.740 | 450.97 | 219.593 | 455.77 |

TABLE 1 – Mean & SD values for Correct Case.

### 4.1 Graphical Representation of RT for various Syntactic Categories

In all the graphs depicted here, only some sample snippets in each case are plotted in x-axis and the corresponding atomic RT values are plotted in y-axis. For a given syntactic category, the atomic RT comparison is done only with snippets of similar difficulty category i.e. simple positive vs. simple negative and complex positive vs. complex negative.

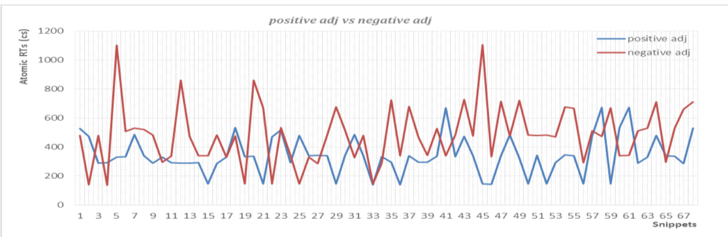


FIGURE 1 *Snippets vs. RTs* for Correct Case (pos-adj & neg-adj comparison)

From Fig.1, we can infer that there is an appreciable variation in RT for each positive and negative adjective snippet.

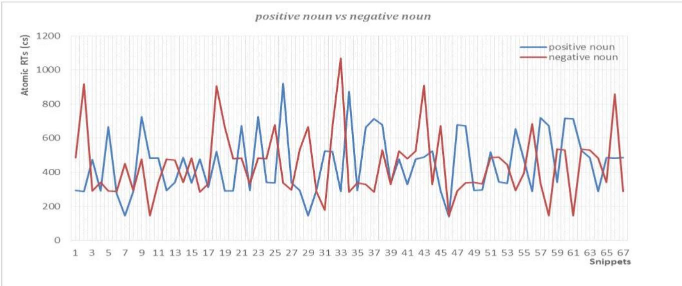


FIGURE 2 *Snippets vs. RTs* for Correct Case (pos-noun & neg-noun comparison)

The Correct Case Noun chart (Fig.2) indicates not many differences in RT values of positive and negative noun snippets. But slight variation exists which further suggests some participants struggled with positive opinions while most others struggled with negative opinions for this category.

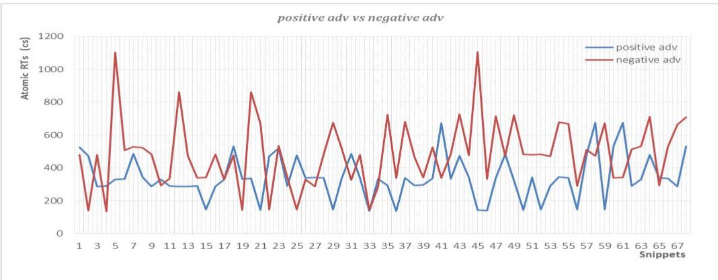


FIGURE 3 *Snippets vs. RTs* for Correct Case (pos-adv. & neg-adv. comparison)

Fig.3 graph clearly shows the observable differences in the RT values. On comparing this with Noun chart (Fig.2), the curve in this graph shows a clear difference in atomic RT values which also suggests its difficulty level. Analysing Fig.4 yields the inference that there is not much variation in RT with some snippets but considerable difference still with other snippets.

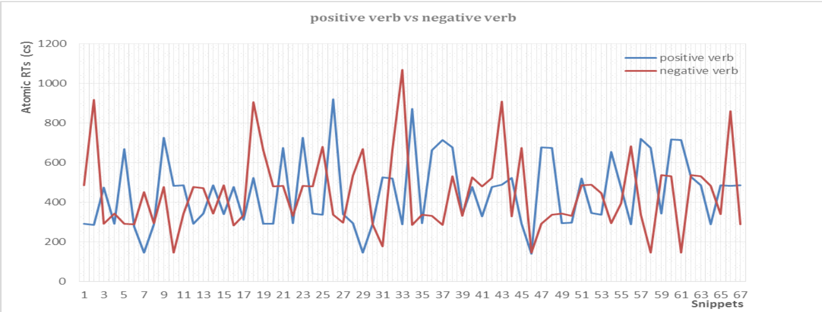


FIGURE 4 *Snippets vs. RTs* for Correct Case (pos-verb & neg-verb comparison)

## 5 Inferences

Interesting inferences and conclusions are derived from the analysis of above graphs and the Mean RT values. Irrespective of syntactic category, negative polarities are more difficult to predict than positive polarities i.e. participants take more time (slow RT) to recognize negative polarity than positive polarity. This is concluded by comparing Mean, and atomic RT values of positive and negative snippets of equal difficulty level (simple-simple or complex-complex opinion). In both positive and negative category opinions, polarity prediction is relatively easy when lexical indicators in a snippet has an *ADJ* syntactic category than the one with an adverbial or other category.The above inference implies that brain's perception is quick in polarity identification when sentiments contain the syntactic category *ADJs*. But, it is relatively slow for other syntactic categories with the highest level of difficulty (slower RT) corresponding to *NOUN* category. It is also evident from Wrong Case RT measure that people commit mistakes relatively often in the case where *ADVs* and *NOUNs* serves as lexical indicators (sentiment-indicating words) in positive opinions and *ADVs and VERBs* in negative opinions. This implies that people are easily deceived by the usage of *negated adverbial and verb* category than other negated syntactic categories.

## 6 Participants and Problems faced

The present study had been experimented among Indian students who learned English as a second language and were almost at graduation level (age group range 20-23). They faced difficulties mainly because of second language phenomenon. Particularly, they had struggled due to the usage of hard vocabulary and movie jargon words in sentiments. To get an insight of the difficulty level, consider the following opinion snippets,

3. *"a screenplay more ingeniously constructed than " memento " " (ingeniously-deceiving and hard vocabulary)*

## 7 Discussion

In rare cases, participants missed polarity detection within allotted time. The actual reason is not clear and to detect that further investigations are essential. We tried to find the reason by seeking feedback from test taking population. In that, they had expressed their difficulty in understanding the semantics of highly complex and jargoned nature of the movie reviews within 15s. In an effort to study and mitigate this problem, trained RT test is conducted. Initially, some participants are trained with some sample set of snippets for polarity identification. Then, the RT values of these participants are measured for a variety of different set of snippets. On comparing the trained test RT values with previously obtained RT values, we found that time taken for every snippet had been slightly reduced (quick RT). This reduction in RT is due to the training tests taken. One of the four truths mentioned in (Saul Sternberg, 2004) states that RT diminishes with practice. So, for the precise measurement of RT values, factors such as mock tests, training, giving hints etc. should be carefully considered when designing an RT system.

## Conclusion & Future Work

The level of cognition associated with various syntactic categories is found. The comparative analysis of various syntactic categories had been done and valuable inferences were drawn. We also arrived at a representation of difficulty level for the considered syntactic categories. It is evident from the results that *adjective* category requires very less RT than other considered syntactic categories. So, *adjective* category will serve as a better stimulus (quick RT) than *adverb or noun or verb*. This finding can be incorporated in the development of better educational tools to better educate students particularly those with neurocognitive disorders. Future work will focus on incorporating the findings of this work into machine learning algorithms which can then be used for automated sentiment classification task. This may help to improve the accuracy of sentiment prediction which will make these algorithms intelligent and also fast in sentiment classification.

## Acknowledgments

We thank Department of Computer Science, Cornell University for providing the movie review dataset and the student participants who had helped us with RT test. We also thank the reviewers for their insightful comments.

## References

Bo Pang and Lillian Lee. (2008). *Opinion Mining and Sentiment Analysis,* Foundations and Trends in Information Retrieval Vol. 2, Nos. 12, 1135

Bo Pang and Lillian Lee, Shivakumar Vaithyanathan. (2002). *Thumbs up? Sentiment Classification using Machine Learning Techniques*, Proceedings of EMNLP 2002, pp. 79-86

Farah Benamara, Carmine Cesarano, Antonio Picariello, Diego Reforgioto, VS Subrahmanian. (2006). *Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone*, International Conference on Weblogs and Social Media, Boulder, Colorado, U.S.A

http://www.sbl-site.org/publications/article.aspx?articleId=660

Ignacio Serrano, J., Dolores del Castillo, M., and Iglesias, A. (2009). *Dealing with written* *language semantics by a connectionist model of cognitive reading,* Neurocomputing 72, 713725.

Judith Tonhauser. (2000). *An Approach to Polarity Sensitivity and Negative Concord by Lexical Underspecification,* Proceedings of the 7th International HPSG Conference, UC Berkeley.

Luke K.W. Tan, Jin-Cheon Na and Yin-Leng Theng, Kuiyu Chang. (2011). *Phrase-level sentiment polarity classification using rule-based typed dependencies*, In proceedings of SWSM'11, China.

Michael Israel. (1996). *Polarity Sensitivity as Lexical Semantics,* In Linguistics and Philosophy, pp. 619-666.

Miller, C. A. and G. H. Poll. (2009). *Response time in adults with a history of language difficulties*. Journal of Communication Disorders 42(5): 365-379

Paula Chesley, Bruce Vincent, Li Xu, and Rohini K. Srihari. (2006). *Using Verbs and Adjectives to Automatically Classify Blog Sentiment.* National Conference on Artificial Intelligence-Spring Symposium, AAAI Press, pp. 27-29.

Perfetti, C.A. (1999). *Comprehending written language: a blue print of the reader*, in: Brown, C., Hagoort, P. (Eds.), The Neurocognition of Language, Oxford University Press, Oxford, pp. 167208.

Pyhala R, Lahti J, Heinonen K, Pesonen A.K., Strang-Karlsson S, Hovi P, Jarvenpaa A.L., Eriksson J.G., Andersson S, Kajantie E,Raikkonen K. (2011). *Neurocognitive abilities in young adults with very low birth weight*, Neurology, December 6, 201177:2052-2060

Robert Whelan. (2008). *Effective analysis of reaction time data*, The Psychological Record, 58, 475-482.

Sternberg, S. (1969). *The discovery of processing stages: Extensions of Donders' method. In W.G. Koster (Ed.), Attention and performance II. Acta Psychologica, 30, 276-315*

Saul Sternberg. (2004). *Reaction-Time Experimentation, Proseminar in Psychological Methods*, Psychology 600-301, spring semester.

Yang Mu, Dacheng Tao. (2010). *Biologically inspired feature manifold for gait recognition,* Neurocomputing 73, 895902.