A COMPUTATIONAL IMPLEMENTATION OF OPINION ANALYSIS: A CASE STUDY OF MALAYALAM POLITICAL TEXTS ON SOCIAL MEDIA

A dissertation submitted to the University of Hyderabad

In partial

Fulfillment of the requirements for the degree

of

Master of Philosophy

in

Applied Linguistics

By

FAITH T VARGHESE

Under the guidance of

Dr. K. PARAMESWARI

Center for Applied Linguistics and Translation Studies, School of Humanities
University of Hyderabad

Hyderabad 500046, India

December, 2018

A COMPUTATIONAL IMPLEMENTATION OF OPINION ANALYSIS: A CASE STUDY OF MALAYALAM POLITICAL TEXTS ON SOCIAL MEDIA

A Dissertation submitted to the University of Hyderabad

In partial

Fulfillment of the requirements of the degree

of

Master of Philosophy

in

Applied Linguistics

By

FAITH T VARGHESE

Under the guidance of

Dr. K. PARAMESWARI

Center for Applied Linguistics and Translation Studies, School of Humanities
University of Hyderabad

Hyderabad 500046, India

December, 2018

Centre for Applied Linguistics and Translation Studies
School of Humanities
University of Hyderabad
Hyderabad 500046, India

Declaration

I, Faith T Varghese, hereby declare that this dissertation entitled **"A Computational Implementation of Opinion Analysis: A Case Study of Malayalam Political Texts on Social Media"** is a bonafide research work submitted by me under the guidance and supervision of Dr. K. Parameswari. It is free from plagiarism. I hereby agree that my dissertation can be deposited in Shodhganga/INFLIBNET.

A report on plagiarism from the University Librarian is enclosed.

Hyderabad
Date: 31-12-2018

Signature of the candidate
Faith T Varghese
[17HAHL03]

Centre for Applied Linguistics and Translation Studies
School of Humanities
University of Hyderabad
Hyderabad 500046, India

CERTIFICATE

This is to certify that the dissertation titled **"A Computational Implementation of Opinion Analysis: A Case Study of Malayalam Political Texts on Social Media"** submitted by **Faith T Varghese** bearing registration number **17HAHL03** in partial fulfillment of the requirements for the award of Master of Philosophy in Applied Linguistics is a bonafide work carried out by him under my supervision and guidance.

This dissertation is free from plagiarism and has not been submitted in part or in full to this University or any other university or institution for the award of any degree or diploma.

Paper presentation in the following conference:

1. Faith T Varghese (2018, December). "Political Opinion Mining in Social Media: An Implementation in Malayalam." In the *40th International Conference of the Linguistic Society of India.* 05-07 December 2018 held at CIIL, Mysuru. Organized by the Linguistic Society of India, Pune, and the Central Institute of Indian Languages, Mysuru.

Further, the student has passed the following courses towards the fulfillment the coursework requirement for M.Phil.:

Course Code	Name	Credits	Pass/ Fail
AL701	Research Methodology	4.00	Pass
AL702	Current Trends in Applied Linguistics	4.00	Pass
AL721	Adv. Topics in Applied Linguistics	4.00	Pass
AL724	Functional Grammar Analysis	4.00	Pass

Dr. K. Parameswari
Supervisor

Head
CALTS

Dean
School of Humanities

Acknowledgment

I would like to take some time to thank all the people without whom, this dissertation would never have been possible. Although it is just my name on the cover, many people have contributed to the research in their own ways and I convey my heartfelt thanks to all of them.

First and foremost, I would like to thank my guide, Dr. K. Parameswari, for accepting to guide me. You have created an invaluable space for me throughout the research, to hew myself as a researcher in the best way possible. I greatly appreciate the freedom you have given me, to find my own path and for the sheer guidance and support, you have offered when needed. My sincere thanks to the advisory committee member, Dr. S. Arulmozi for all the support and encouragement through the tenure. All the insights provided by him, on academic works carried out so far was of great help and boosted my research in the healthiest manner.

I express my sincere gratitude to Prof. K. Rajyarama, the Head of the Department, CALTS, for providing me with the best opportunity and allowing me to undertake the research.

I would like to thank all the faculty members of the department for extending their availability, to clarify my doubts and for suggesting the corrections needed for my work. Thanking the non-teaching staff for playing a major role in easing me out with their complete co-operation for paper works.

My hearty thanks to Nikshep for his timely, immense help in correcting the codes and programming. Keerthana chechi's support and the valuable suggestions are to be remembered, while I list out my fellow researchers and roomie for being great critiques and nurturing my growth. I also remember my friends in EFLU (Sathya, Jaydeep, Ande, Raju, Atheena, Jobin, Anol,

Juli, Shivaji, Naresh) and HCU (Pattambi, Fahad Ikka, Manoj Ettan, Suju Ettan, Unni), never to forget my friends Vivek and Sainu for motivating me and being the pillars of my life.

A special thanks to the mess members for feeding us tasty food and making sure that all of us stay fit and healthy, which helped me to complete the dissertation on time.

Another special mention, Dundumol (pseudo name) for extending the solidarity with me for inexpressible needs.

Finally, and most importantly, I'm forever indebted to my family for their ceaseless moral support in all my endeavours. My parents, Aliyan and Chechi, for constantly encouraging me in pursuing my goals. Markedly, Aliyan, for providing timely solution by exhibiting his Word skills.

I sincerely apologize for the ones I have missed mentioning here. All your contributions are indeed respected and it is all of you who have taught me to choose the path less travelled by.

ABBREVIATIONS

ACC	Accusative case marker
ADJ	Adjective
ADV	Adverb
AP	Adverbial Participle
CONC	Concessive marker
COND	Conditional marker
COORD	Coordinator
DAT	Dative case marker
DEB	Debitive marker
FUT	Future Tense
GEN	Genitive case marker
IMP	Imperative marker
INFN	Infinitive marker
LOC	Locative case marker
NEG	Negation
NOM	Nominative case marker
NOML	Nominal
PAST	Past Tense
PRES	Present Tense
PROH	Prohibitive marker
QP	Quotative Participle
QTC	Cardinal Quantifiers
RP	Relative Participle

TRANSLITERATION SCHEMA

Roman	*a*	*ā*	*i*	*ī*	*u*	*ū*	*ṛ*	*e*	*ē*	*ai*	*o*	*ō*	*au*	*aṁ*	*aḥ*
WX	a	A	i	I	u	U	q	eV	e	E	oV	o	O	M	H
Malayalam	അ	ആ	ഇ	ഈ	ഉ	ഊ	ഋ	എ	ഏ	ഐ	ഒ	ഓ	ഔ	അം	അഃ

Roman	*ka*	*kha*	*ga*	*gha*	*ṅa*	*ca*	*cha*	*ja*	*jha*	*ña*	*ṭa*	*ṭha*	*ḍa*	*ḍha*	*ṇa*
WX	ka	Ka	ga	Ga	fa	ca	Ca	ja	Ja	Fa	ta	Ta	da	Da	Na
Malayalam	ക	ഖ	ഗ	ഘ	ങ	ച	ഛ	ജ	ഝ	ഞ	S	O	ഡ	ഢ	ണ

Roman	*ta*	*tha*	*da*	*dha*	*na*	*pa*	*pha*	*ba*	*bha*	*ma*	*ya*	*ra*	*la*	*va*	*śa*
WX	wa	Wa	xa	Xa	na	pa	Pa	ba	Ba	ma	ya	ra	la	va	Sa
Malayalam	ത	ഥ	ദ	ധ	ന	പ	ഫ	ബ	ഭ	മ	യ	ര	ല	വ	ശ

Roman	*ṣa*	*sa*	*ha*	*ḷa*	*ḻa*	*ṟa*	*ṉa*
WX	Ra	sa	ha	lYa	lYYa	rYa	nYa
Malayalam	ഷ	സ	ഹ	ള	ഴ	റ	ഩ

Roman	*ṇ*	*n*	*r*	*l*	*ḽ*	*k*
WX	N	n	r	l	lY	k
Malayalam	ൺ	ൻ	ർ	ൽ	ൾ	ൿ

TABLE OF CONTENTS

Chapter 1

Introduction

1.0. Introduction

Language is a human construct of shared experience and culture, which reflects the cultural values of the people who create and practice it. Speech, an expression of the thought of either conscious or unconscious reality of the society through language, is a medium to determine the norms and conventions pertaining in the language. Any form of expression of thought by an individual, at a point, can be termed as communication which is transformed by realization for the independent existence to share one's own self. These individualistic expressions are usually subjective in nature. In fact, at times, these expressions are objective for facts implied statements. The juxtaposition of acquired knowledge reveals the individual's thought and later relates to the motion of thoughts embodied in words. Thus, the creation of thoughts and expression of opinions are as old as the origin of a language and therefore it can be argued that 'opinion analysis' is not a new phenomenon. However, an automated opinion analyser is a new attempt in the field of Natural Language Processing (NLP) to track the opinions of the public in a particular field/ product. The paradigm shift of thought expressions from face-to-face interactions to technology-mediated interactions (as it attracts a larger audience) necessitates in building an automated opinion analyser for any language used.

Opinion analysis here refers to the literal sense of interpretation of opinions based on the subjectivity of texts. It does refer to the computational perspective of sentiment analysis except that it doesn't convey the feeling or emotion of any speaker/user. Opinion extraction is growing research to uncover the underlying meaning in the text with the aid of NLP tools. Opinions are constructed using opinion words to express positive or negative sentiments and

opinion analysis thus detects the opinion of the sentence/document and classifies it into positive, negative or neutral (Liu, 2012).

The quick accelerated growth of information technology has set forth to the emergence of a new public platform that enables the users to communicate, share views and opinions. These views are open to a large audience that could influence the choices of the readers. The abundance of views and opinions are expressed on social media through blogs, reviews, articles, forums, news, and comments. This is one significant reason for choosing social media and web-news as a platform for research in linguistics as it assists to improve the computational understanding of languages (Bučar et al., 2018).

The rise in the use of social platforms to express thoughts, opinions, and emotions have nurtured in parallel, the demand for the research in opinion analysis in order to extract the subjectivity in opinions. The democratic platforms that these social media providers have enabled the users to share their perspectives and thereby transforms the people to agree or disagree up on things such as political views, movie reviews, blogs etc. The task of opinion mining may help in identifying diverse opinions expressed by these platforms.

Implementation of opinion analysis in Indian languages is very recent that its emergence is noted by the end of the first decade of the 21st century and however, extensive research has not been reported in Malayalam since then. A major reason for the limited study in this field is the limited resources available on the web in Malayalam. For Malayalam, online news portals and social media are the only highly established platforms; online news portals' political affinity and user's political stake on each event shall provide ample political resources on the internet.

1.1. Importance of Opinion Analysis

The ability to comprehend and extract opinions from a social platform is considered a boon in the present era of advent social networks. This automated process of discerning or monitoring the opinions about a given subject, not only assists us in training the machine to associate certain inputs with the corresponding outputs but also spots the keywords to assess the stance of the consumer, to scan its polarity. Apart from this, tasks such as summarization of the multi-perspective questioning and answering, extraction of opinion-oriented information from various fields on the social networks and researches, and gathering media reports require sentence-level, phrase-level opinion analysis. It has a wide range of applications almost in every domain. The proliferation of commercial applications has been one of the major reasons for the flourishment of opinion analysis in the industrial field as well. This provides a strong motivation for research and offers many challenging research problems, which would have been tough to address, otherwise. Opinion analysis is right now the cynosure of social media research. Starting from the assessment of marketing the success of an ad campaign or new product launch, to determining the versions of a product or service that are popular, and even identifying the demographics of people's likes and dislikes particularly, is a contribution much needed. Another domain which has rendered its efficiency through opinion analysis is politics. Enabling the civilians or voters to elect their representatives, based on the statistics or reviews produced by opinion analysis is yet another example that could result in a revolutionary change. Hence, research in opinion analysis is not only creating an important impact on NLP, but also has a profound impact on management sciences, political science, economics, and social sciences as they are all substantially dependent on people's opinions. Opinion analysis also has a wide variety of application in summarizing reviews, classifying reviews, information system, market analysis, and decision making.

Therefore, opinion analysis or opinion mining is carried out to determine the attitude and opinions expressed in the text (Liu, 2015). Attitude is defined as a psychological tendency to evaluate a particular object or a thing with some amount of favor or aversion (Eagly & Chaiken, 1998). Opinion analysis classifies the theme of the content produced in the context into different levels based on the nature of its subjectivity or objectivity. There can be three ways of looking at the nature of the text.

i) Subjective opinion: Content that expresses personal opinion or feelings.
ii) Objective opinion: Content that describes facts or evidence without indicating any opinion.
iii) Objectivity in subjective opinion: Content that expresses facts implied personal or non-personal opinion.

1.2. Aims and Objectives of the Study

This research aims at developing an opinion mining tool for Malayalam that provides detailed subjectivity of a sentence in a document employing a simple rule-based lexicon model. Pang et al., (2002) and Turney (2002) describe document-level opinion analysis as the classification of the whole document into positive or negative, by considering the whole opinion expressed in the document. The sentence-level opinion analysis is a three-class-subjectivity classification of attitude expressed in a sentence into positive, negative or neutral. Sentiment in this current research connotes the underlying subjectivity of the sentence, determined by the polarity, that is whether the sentence signals a positive, negative or neutral attitude. The opinion of a sentence in a document may not be necessarily the sum of the polarity of whole word units present in the document (Turney & Littman, 2003).

Bučar et al., (2018) state that attitude changes are majorly associated with financial, economic and political reasons. The dominance of English on the web has influenced the rise

of resources in English and thus, much of the researches in this field are done in English. Potential of research and its development in the field of opinion analysis on minor[1] languages are dependent on the resources on the cyberspace. To understand the relationship between language and its use on the internet in an academic discipline, it is imperative to rely on the existing data. For Malayalam, availability of the resource for the study is a major concern. Financial texts, product reviews or film reviews are not extensively found on cyberspace in Malayalam. Political conflict or cooperation, widely expressed as user's statements and political affiliations of various news portals are potential politically relevant resources that exist in Malayalam.

This research envisages the following items:

a. Study and analyze the political resources available in cyberspace in Malayalam
b. Classify the political texts using linguistic cues and identify opinion expressed
c. Develop a tool for sentence level opinion analysis on political texts
d. Survey on the attitude of the public expressed opinions on various social platforms or news forums on various events aids to realize the social trend on events in Malayalam.

1.3. Malayalam in Political Cyberspace

The three major available media which are print, television, and internet have shown its own evolution from the early 2000s. The constant advancement of these media has brought out its space as an integral part in social, economic, political and cultural relations. The advent of the internet and its growing popularity has resulted in the steady decline of print and now, witnesses the transformation in the way of communication and knowledge shared. Though it is less known of the reason behind the shift from the traditional media to the internet, clutches

[1] Smaller languages estimated based on the available contents on online.

of online news media and social-web-platforms are strong evidence for the shift. Online editions of the 'print-news papers', instant broadcasts, active news updates clearly illustrate that print media has in a way transformed and occupied a space in the cyber world. Despite circulation declination over many decades, newspapers remain the news source of record (George, 2008). 'Decline' doesn't prove that cyber evolution has swallowed the traditional media, instead, it is to be understood that the space of existence of these media are occupied in different levels and in a wider perspective, the purpose and contributions of these media are never conflicting instead, they are interdependent.

As the space of online news media isn't a questionable fact, correspondingly, upper hand on the growth of communication on web platform should be also noted. Internet's growing popularity is not just because of the presence of online newspapers, it is majorly because of its substantial contribution of knowledge on national and international conditions and, progressive and dynamic development of social-public platforms for the purpose of communication. Language use and this technology development are very much directly proportional and interlinked.

Internet has visible dominance of English and it is viewed that the spread of English is as a result of an extension of globalization (Crystal, 2001, Fishman, 1998) whereas Phillipson, Skutnabb-kangas and others term it as 'linguistic imperialism' (Phillipson, 2001,2013, Skutnabb-Kangas, 2001). As per internet world stats[2], 25.4 percent of total internet users use English as content and w3techs[3] reports that 53.5 percent of the content available on the internet is English and surprisingly India is ranked the second position in the worldwide

[2] Internet World Stats is an International website that analyses internet usage and statistics.

[3] W3Techs (Web Technology Surveys) provides information about the usage of various types of technologies on the web.

internet users as per the data available on International Telecommunication Union (ITU)[4] 2017. This undoubtedly confirms the dominance of English over Indian languages on the Internet among Indian-internet users. Malayalam is ranked 76th position among the common languages based on the written content available on websites. 0.0012 percent of the total content available is written and maintained in Malayalam which cannot be considered insignificant as Malayalam on web is gaining its popularity in parallel with the growth in internet users and switching over to Malayalam from English among the present users is still a possibility where Forbes[5] reports that there is a recent tendency for internet users in the country to opt their native language instead of English.

Malayalam stepped into the cyber world from the time newspaper industry realized the threat in the rise of the internet. Malayalam newspaper *deepika* created the history by introducing its online daily in 1997 and later within a decade other prominent newspapers presented its evolved face on the new media. And years later, Malayalam gained its pace by users' acceptance and use in blogs and social platforms. Through this, it can be effortlessly argued that it is majorly political preferences, inclinations or opinions that are available on world wide web except some limited personal or public professional web portals and open-content encyclopedia. As the statistics stated above, the resources that exist on the new media is limited and its nullness is understood while it is compared to the contents available in English but resources on new media are always a growing phenomenon.

[4] International Telecommunication Union (ITU) is an agency of the United Nations (UN) for information and communication technologies. ITU ICT Facts and Figures 2017.

[5] Forbes is a business magazine published bi-weekly.

1.4. Political Discourse

The term discourse is anything that conceptualizes as a formal thought process, that is possibly expressed through language. It is also a social construct which brackets the statements delivered, pertaining to a specific domain. Discourse is also embedded as an activity that emerges out of power, as most of the spoken or written form of a document is done by those in control of institutions or communities such as media and politics, especially. In order to define political discourse, we need to examine the definition of politics first. The elements of language that metamorphoses a formal discourse as political, is due to the social context, which attributes a certain set of vocabulary as political and apolitical. The political text is not a lucid term covering a different variety of text types like speeches, political party meet, editorials or articles in the newspaper, parliamentary debates, etc. (Schäffner, 1997).

Typically, a political text can be persuasive, rationale, deceitful or coerce, for which words of the same nature can capture the essence of text effectively than a diplomatic language. Diplomatic language ranges between being completely generic (that has interestingly no affiliation to a specific domain, but ponders around all or many domains) and convincible set of vocabulary, to focus on the international relations, the interaction between the organisations which are apolitical, yet has international impacts. Though the differences between a normal language and a political language is thinned down, from a utilitarian point of view, usage of normal language is figured as a technique to captivate people into a domain specific language, as politics is nothing but the struggle to power. The semantic values in the words used for political texts are much affluent than the other language use. For example, all rapists are sentenced to death in the name of law, is an effective manifestation of a language through its properties pertaining to law and order. Sárosi-Márdirosz (2014) in *Problems Related to the Translation of Political Texts* states that “Political language forces us to reconstruct, through interpretation, those thoughts which are settled in the political text. This

reconstruction is a mental process through which we rebuild the text according to our knowledge in order to gain a better understanding".

To differentiate between a political text and any other text, let us look at elements that constitute to a political text for rhetorical purposes. Metaphors in political contexts, act as a persuasive tool. Theorists have observed that the necessity of having an effective communication in politics is closely associated with addressing the general public in order to nudge their consciousness of dormant symbolic themes. The metaphor also acts as a model to simplify the complicated information into a simpler and smaller pocketed information, which is better comprehendible by the public. Mio (1997) quotes Edelman to define metaphor as:

"…the pattern of perception to which people respond. To speak of deterrence and to strike capacity is to perceive war as a game; to speak of legalised murder is to perceive war as a slaughter of human beings; to speak of a struggle for democracy is to perceive war as a vaguely defined instrument for achieving an intensely sought objective. Each metaphor intensifies selected perceptions and ignores others, thereby helping one to concentrate upon desired consequences of favoured public policies and helping one to ignore their unwanted, unthinkable, or irrelevant premises and aftermaths. Each metaphor can be a subtle way of highlighting what one wants to believe and avoiding what one does not wish to face" (cited in Mio, 1997 :114).

This not only manifests that political metaphors are effective in stirring the emotions and bridging the gap between rational and irrational forms of persuasion. For example, a scientific metaphor is mostly analytical and fundamentally adheres itself to imply formulas and also to quantify specific relations and affirmations. In contrary, political metaphors, as mentioned earlier tend to stir up emotions and circumvent the logical aspects. Likewise, in literary discourse, the elements of metaphor, pun, metonymy are all a casual use and is thus considered a foregrounding of the language in itself. Foregrounding is a linguistic component which is intentionally and aesthetically distorted (Lodge, 2015). It is also important to note

that the components of a non-literary text do not rely on statistical frequencies alone. Instead, what differs a literary discourse from a non-literary discourse and makes it aesthetically relevant is the stability and organized nature of foregrounding and the relationship between the background and foreground of a text. However, only the foregrounded components of text are aesthetically relevant in a non-literary discourse. One of the major tools of euphemisms being metaphor and metonym, in most of the public speaking – particularly in newspaper headlines and political speeches, metonymies like 'Institution for The People Responsible, Place for Event'[6]are common usage. As mentioned in the case of metaphors already, metonyms also serve the same purpose, in a slightly different way. These euphemisms are much accredited to political discourse majorly, due to its abundance and frequent usage – frequent usage also denotes the quench for these techniques of speaking, in the field. Some of the root metaphors that can be mentioned for the political arena are 'organism'(something that should be seen as a whole and not as a part), 'machine' denotes working parts in a balanced form, 'disease' implies spreading of ideas and container which implies prevention of any spillovers. Excerpts of metaphors from news headlines- *Majority fear Vietnam will **fall** for communism, Public generosity **hit** by an immigrant wave.*[7]

1.5. Opinion Analysis

Opinion is a degree of likeness expressed towards or against anything as comments or discussions. These comments or discussions are not just limited to verbal forms. Visual forms, using symbols representing body languages or facial expressions posturing particular images manifest opinions which are identical or a degree higher than the verbally generated opinions. These visual forms are approximate mental images present in the notions of users or viewers which does not require any assistance of verbal forms to interpret opinions. Thus,

[6] E.g. ministry of home affairs or ministry of foreign affairs, a collective name of the institution. 26/11 for the Mumbai terrorist attack.

[7]Example cited in newsframes under the title 'Archive for the metaphor category'

analyzing these images aren't any cumbersome task. Likewise, the presence of 'like' or 'dislike' buttons on social platforms or the existence of similar functional options has the same purpose of analysis. But the only visible difference is that though, these techniques have the same function, one that is verbal applies language for its fulfillment. Building a tool to mine subjectivity in verbal opinions is not as uncomplicated as framing one for visual opinions. It needs to theorize a structure from a complex natural language. Natural languages prove complicated for practical applications that ultimately no frameworks achieve hundred percent accuracy in a limited time frame. It requires research over research or continuous researches to identify precise technical solutions by problematizing language. Thus, extensive research on existing frameworks on opinion analysis enables to design a robust machine that addresses each and every technical issue pertaining to a natural language in the application stage.

Interpretation and classification of subjective opinions can be termed as opinion analysis. Subjective opinions on any aspects may differ in terms of the experience of a user, faith of people and perception acquired through experience or education or inherited through speech. For an instance, consider that two people have bought the same brand and model phone in a different time from different stores. Owe to the account of possibility, one got the product for a lesser price. There creates one possible way of change in the opinion. Now, consider that one user finds the product heavy or lesser battery life than promised by the company. In the above-mentioned instances on products, it's highly certain that the opinion of users differs. Same is the case with political news. An event or information could be reported differently by different people or media majorly based on their political affiliations or inclinations. In these cases, though opinion analysis cannot interpret the information or event, it interprets the user's subjective interpretation of the information or event.

A statement is said to be subjective if and only if the statement is opinionated. Thus, objective statements shall be eliminated from the concept of this study as these statements do not express any views, feelings, some personal feelings, views or beliefs.

Example of objective statements: - (1) *India has 29 states.*

In the example (1), the factual statement is expressed and hence, no opinion is identified.

Wiebe and Riloff (2005) termed 'subjectivity classification' for the task of bifurcating sentences into subjective and objective. As mentioned earlier that objective statements provide information about real events, facts which are not influenced by the personal views or emotions, they are termed as 'neutral' statements in opinion analysis.

Another kind of statements which have to be critically looked at is subjective statements that are objective in nature. Liu (2012) terms these statements as 'Fact-Implied Opinions'. These statements usually follow a nature where statements opinion would be wrapped in an objective statement.

Example (2) *I believe I can swim.*

(3) *I think I told you my name.*

Though the above statements are opinionated they do not express any views to any degree to classify them as positive or negative. This conveys an argument that not all subjective opinions can be interpreted as positive or negative. In this manner, statements which are fact-implied opinions are termed as 'neutral' statements in opinion analysis.

1.5.1. Types of Opinion Analysis

The field of subjectivity analysis operates differently in each domain. Application of subjectivity analysis can be traced mainly in four different domains (Bučar et al., 2018).

a) business/financial

b) film/movie review

c) product review

d) political

Each domain accommodates a specific nature of vocabulary that describes the semantic relations between concepts and texts and in each domain, opinion analysis employs a sophisticated method to describe semantic relations between concepts and domain-specific texts. For financial subjectivity analysis, financial domain texts are studied concerning the financial information on subjects such as stock market, investment, asset, property market, companies, etc. Lexicon developed for this method will have a structured vocabulary that emphasizes financial meaning. Considering movie review, the information for opinion lies on the elements like screenplay, cinematography, direction, music etc. and involves people like actor, cinematographer, director, screenwriter, etc. As the knowledge required for the domain demands for the movie, the semantic relations controlled by vocabulary shifts to the movie-based platform. Lexicon for the movie domain is to be identified and classified from the list of film reviews. As a reason for the rapid expansion of e-commerce, most of the existing work in subjectivity analysis is implemented on product reviews. Product reviews generally don't give much importance to the designer or manufacturer unlike movie reviews does for the director, cinematographer, etc. For product reviews, subjectivity is analyzed based on the features of a product. Each product may have different features like size, weight, quality, material, etc. The opinion of the users for a product would be inclusive of all its features. Thus, product review as different from financial provides a summarized opinion of the product adding-on to the feature-specific opinions such as specifying the interesting feature or uninteresting feature of the product. As mentioned above, for product subjective analysis,

training for the framework has to rely on the product-feature reviews. Political opinions address fewer features as against film or product reviews. Sentiment relations expressed through opinionated vocabulary in the political domain are associated with financial or political events. Lexicon for the political domain is compiled from the opinion-bearing texts in social media comments expressing political entities. Thus created lexicon fails remarkably if applied for the opinion generation in another semantic domain as the language is deeply rooted in the meaning beyond words.

Though the classification of subjectivity based on the semantic relations is analyzed in different domains, web-generated opinions are the only source for these analyses on the application level. This argument can be stated in other words that the research in the field of subjective analysis is supported by the expressions and opinions on web and demand of extensive research in this field exclusively depend upon the practical development in the practice of opinion sharing on the social media.

1.5.2. Levels of Analysis

Based on the problems of research on opinion analysis, Liu (2012) classifies opinion mining into majorly three levels.

1.5.2.1. Document Level Analysis

Document-level analysis classifies the opinion expressed in a document into its polarities. It considers the entire view present in the text and classifies into either positive sentiment or negative sentiment. For a document level classification to be carried out, the test document should be discussing exclusively on a particular topic and further, it should not have any comparative study on the topic. This level of analysis becomes insufficient for the assessment of multiple entities as it discusses multiple tokens. E.g. comparative movie reviews, product reviews. Document opinion analysis is considered as the traditional text classification and is

analyzed using any supervised machine learning algorithms. Though it is considered as the simplest analysis, an ample amount of problems is assigned to this level of classification. It does a general classification of the entire text using the opinion words present in the document and takes in a lot of negative opinions too into the analysis. It does a base task for the further sub-classification of the document. It operates on the basic assumption that the viewpoints expressed in a document is on a single product and expressed by a single person.

Its accuracy is insignificant in documents that contain both positive and negative entities. It performs well in product reviews since the opinion maker centers essentially one argument for the product which is either positive or negative. Any classification algorithm provides the basic requirement to perform this task.

1.5.2.2. Sentence Level Analysis

In this level analysis, each sentence in its semantic nature is classified into positive or negative or in addition, neutral. Document-level classification never produces a neutral sentiment. Nevertheless, sentence-level sentiment classifies into neutral since a few sentences in a document that are semantically relevant may not be carrying any opinion. Example (4) *I watched this movie yesterday*. The process of performance of both document and sentence level analysis is nearly similar. The only difference in practice is that sentence analysis investigates each and every sentence present in a document. Analysis becomes similar in a context where a document contains a single sentence. Wiebe et al. (1999) describe that despite classifying opinion sentences into positive, negative or neutral, sentences have to be classified based on the opinion-making which is called subjectivity classification. The difference that Wiebe (2000) points out is that subjective sentences may not always incorporate positive or negative sentiments, instead of it represents opinions, appraisals, evaluations, allegations, desires, beliefs, suspicions, speculations, and stances. These

subjective expressions probably at times don't express polarities. E.g. I told that I need a pen. As regard to the subjective statements, objective statements at times express opinions. E.g. the mobile got damaged in a week. Thus, it is necessary to classify the sentences into its subjectivity i.e. subjective or objective. Here objective sentences are classified as those sentences which are non-opinionated Liu (2015). Thus, subjectivity classification is a major process carried out prior to sentence level opinion analysis.

1.5.2.3. Aspect Level Analysis

Aspect level considers word or phrase as a single entity and determines the sentiment of it. Normally, holder talks about a topic that has different aspects and he/she makes different opinions about each aspect. Aspect-based opinion analysis (also called feature-based opinion analysis) is the research problem that focuses on the recognition of all sentiment expressions within a given document and the aspects to which they refer (Feldman, 2013).

Aspect or feature level classification is majorly implemented on product reviews that describe various features of the product. E.g. if a review written on a pen drive is taken into consideration, the main aspects that are possible to talk about is its shape, storage space, and speed. These features are the aspects of the pen drive and classifying this review into positive, negative or neutral would be meaningless as it conflates different features where some would be positive and the rest negative. Hence classification should be done based on the sentiment expressed on respective aspects. Liu (2015) illustrates two major methods; (a) aspect extraction and (b) aspect sentiment classification; that require deep NLP knowledge to carry out feature-level opinion analysis.

1.5.2.4. Aspect Extraction

Aspect extraction method extracts aspects and entities from the sentence. The main approach is to extract all noun phrases from the product review corpus. Those NPs that fills in the

measures of aspects and entities have to be recognized. From the noun phrase noise, a cleaning method has to be considered for the filtration. Frequently used method is the threshold frequency method i.e. noun phrases (NPs) that occur most frequently, above the threshold limit are considered as aspects and entities. Another method for the aspect extraction would be to apply a phrase dependency parser.

1.5.2.5. Aspect Sentiment Classification

This method classifies the extracted aspects into negative, positive or neutral. It requires a supervised learning method employed to carry out the task. The main approach for supervised learning is to extract those features that are dependent on the aspects extracted. Jiang et al. (2011) employ a syntactic parser to generate all words that are syntactically dependent on aspects.

Apart from the above three levels of analysis Feldman (2013) explains another level of analysis which is comparative opinion analysis.

1.5.2.6. Comparative Opinion Analysis

This kind of analysis is done on reviews that compare products. Some reviews do not provide a direct opinion about a product instead it compares with another kind and formulates the statement. The main task of this classification is to identify sentences which contain comparative opinions and extract the feature-based opinion. In English, this is identified mostly with comparative adverbs and adjectives, superlative adverbs and adjectives and some phrases.

E.g. Comparative and superlative adjectives: lighter, lightest.

Comparative and superlative adverbs: more, most.

Phrases include: number one, prefer, than, superior, inferior.

1.5.2.7. Sentiment Lexicon Acquisition

Sentiment lexical acquisition is the most substantial part of the opinion analysis. Acquisition can be accomplished using any three approaches namely manual approach, dictionary-based approach, and corpus-based approach. In a manual approach, the lexicon is classified manually and it is mostly domain specific as specific domain prefer a particular set of the lexicon. The dictionary-based approach requires a small set of opinionated words and these words are connected and elaborated with the help of WordNet. Rajendran & Soman (2017:119) explain WordNet as a nonlinear lexical structure based on semantic features of individual words. WordNet contains synonyms called synsets and antonyms. It provides the link between words based on the meaning each word shares. WordNets are independent and not domain specific. If domain-specific sentiment lexicon is preferred then the corpus-based approach is suggested.

1.5.3. Approaches in Opinion Analysis

Two main approaches: lexicon-based method and Machine learning technique are the widely practiced techniques to design an effective automated system that perform opinion analysis. A statistical model, another approach which is not widely used in the analysis make use of a balanced corpus of negative and positive texts to determine the polarity of texts. Pang and Lee (2004) term polarity dataset for the collection of processed negative and positive reviews. Documents are labelled with respect to the overall sentiment polarity and sentences are labelled with respect to the subjectivity status or polarity. (Pang et al., 2002). The machine learning approach employs various supervised learning system whereas the lexicon method applies SentiWordNet, dictionary-based approach, or a corpus-based approach to support opinion analysis. These methods generate lexicon based on the semantic orientation of the lexical items in the target language. SentiWordNet is a lexical resource specifically devised to

assist opinion analysis by assigning each synset of WordNet three numeral score which indicates positive, negative and objective (Baccianella et al., 2010). Wiebe and Riloff (2006) introduced two techniques to generate subjectivity lexicon for English which is dictionary approach and corpus-based approach. Mohammad et al. (2008) categorized antonyms as two types: gradable and productive. Gradable antonyms have word pairs where each positive lexical item has a negative counterpart e.g. hot-cold, good-bad whereas productive antonyms have orthographic affixes e.g. like-dislike, accurate-inaccurate, harm-harmless.

1.6. Methodology

This research aims at developing an opinion mining tool for Malayalam that provides detailed subjectivity [8]of a sentence in a document employing the feature-based model. The current work applies lexicon method with a corpus-based approach as against the semantic orientation (SO) method.

a) Corpus

One of the main goals of quantitative analysis in linguistics is data reduction, which is studied to summarize trends, capture the common aspects of a set of observations such as the average, standard deviation, and correlations among variables (Johnson, 2008). This analysis can be observed and realized through corpus-based research that involves the compilation of texts from several sources. The issue of the representativeness of the corpus is very important in corpus building. In many studies, representativeness is directly related to the ability to generalize the results of corpus investigation. As the analysis is expected to carry out in a political domain, the lexicon required to support opinion analysis is built from a corpus-based approach. Eighty-seven thousand three hundred and forty-seven (87347) sentences were

[8] Detailed subjectivity here means subjectivity classification that is classifying opinion expressed sentences into three classes (positive, negative, and neutral).

extracted from Malayalam online news websites majorly from SouthLive and Deshabhimani with the aid of Sketch Engine[9] in the period between 1st August 2018 and 30th September 2018.

b) Building Linguistic Cues for Opinion Analysis

Lexical resources for Malayalam are compiled manually from the built political corpus. Opinion carrying word classes are extracted from the corpus based on the frequency of occurrence. Cruse (1986) states that the meaning of a word is constituted by its contextual relations. Miller (1995) comments that choosing between alternative senses of a polysemous word is a matter of distinguishing between different sets of linguistic contexts in which the word form can be used to express the word sense. And addon to it Charles (1988) explains that it is language user's skill to distinguish word forms and use appropriate words to express meaning in the context. Highly frequent opinionated words in the political domain are extracted and are labelled with their corresponding parts of speech. The subjectivity word list is divided into two classes: positive and negative, based on their semantic orientation. Malayalam is an agglutinative language, i.e.word may contain multiple morphemes attached to the stem with distinct morpheme boundaries to form a multimorphemic word. As Malayalam undergoes a lot of inflectional and derivational processes, subjectivity word lists alone won't fulfill the requirement of opinion scaling. Along with the built-lexicon, it is also important to integrate suffixes that express negation to support the automated opinion analysis system as the suffixes in Malayalam possess a meaning that can alter the sense of opinion expressed lexical items.

[9] Sketch Engine is a corpus manager and text analysis software developed by Lexical Computing Limited since 2003

c) Polarity calculation and subjectivity classification

The built-lexicon along with the suffixes that express negation aids to bifurcate subjectivity words into positive and negative lists. Extracted sentences from the text are transferred into this model and if any opinionated word match with the subjectivity model, count one is added to the sentence either a positive count or a negative count depending on the word's polarity. A verb-based algorithm is built to check the composition of word classes pulled out from the sentence. After extracting the polarity counts of a sentence, the verb-based model determines the subjectivity of the sentence and classifies into negative, positive or neutral, opinionated sentence.

d) Evaluation: Strategy followed in opinion analysis- Precision and Recall, F-score

Empirical evaluation plays a pivotal role in assessing the performance of NLP tools (Goutte & Gaussier, 2005). To estimate the performance of the proposed model, this paper employs precision, recall, and F-score. Precision is a statistical measure to determine the ratio of observed true positives to the total number of positive observations. The recall is the ratio of observed true positives to the total number of observations. F-score is measured as the average score of precision and recall.

1.7. Chapter Outline

This research is framed on a computational perspective and the distribution of chapters are arranged on the order of study. The entire dissertation is divided into five chapters.

Chapter one is an introductory chapter discussing the importance of opinion analysis, explaining the aims and objectives of the research and formulating methodological reasons for the study.

Chapter two is a review of the available resources related to the study. This chapter attempts to contextualize the active researches in the field of opinion analysis and discusses the various approaches implemented for different kinds of analysis in different languages.

Chapter three explains the appropriate model built for the automated opinion classification. This chapter engages in reasoning the intention behind the selection of the model and theoretical explanations for the potential features preferred for the study.

Chapter four discusses the implementation and evaluation of the model. It includes the quantitative analysis of the performance of the built system and discussions of the results.

Chapter five concludes the dissertation stating some observations and extends the possibility for further experiments in the field of opinion analysis.

Chapter 2

Literature Review

2.0. Introduction

The objective behind the development of opinion analysis was purely business in the early stages. The purpose was to develop a reliable system that compares and predicts the attitude and demands of the market through language. Language is considered an effective tool for this, as word's usage indicates the polarity of the word (Lehrer, 1974). Polarity signals the direction of semantic deviation of a word from its lexical field (Lehrer, 1974). The field of opinion analysis sprouted in the early 2000s with the interest in predicting market and customer needs. Das, Sanjiv Ranjan & Mike Y. Chen (2001) developed algorithms to compute the investor opinion of management announcements, press releases, third-party news, and regulatory changes. Morinaga et al. (2002) introduced a framework that collects opinions available on the web about the target products and attempts to predict the product reputations.

A widely used technique to perform opinion analysis is a machine learning approach. Tang et al. (2009) try to address four different problems predominating in this research community, namely, subjectivity classification, word sentiment classification, document sentiment classification, and opinion extraction. This article discusses two issues on testing texts based on the manually classified document sentiment. One is extracting feature where only a part of the meaning of a word supplies to polarity and the other is domain specificity. Document sentiment trained in one domain cannot be used to test another domain. This article employs a similarity approach, naive Bayes classifier and multiple naive Bayes classifier for subjectivity classification. Moilanen et al. (2010) propose a quasi-compositional sentiment learning and parsing framework which is well suited for classification across words, phrases, and

sentences. This article accounts on a forth polarity ‘sentiment reverse potential’ other than the three sentiment polarities (positive, negative and neutral). The analysis takes a sentence in linear order and phrase is considered as a separate feature. The article argues that compositional sentiment reduces the feature space by stating an example [evil wars]$^{(-)}$ instead of taking evil and wars as two negative features, it is counted as a single feature.

Limitation of a statistical model is that language is never perceived as a social fact, behavior, nor an abstract object. Ferdinand de Saussure (1916) terms language as a social fact as he believed insights on language could be achieved from the thoughts of a language user. Skinner (1957) states that language user is conditioned to respond in the patterns found in the language. Katz (1981) argues language as an abstract object explaining language existence is independent of the existence of mind and language do not occupy a position in space and time. Machine learning equates language to numbers to perform opinion analysis without looking into the insights of language.

Lexicon-based semantic orientation method is another popular method pertaining to this field. Maite Taboda, Julian Brooke et.al (2011) implement a lexicon-based approach to extract opinion from the text. For this method of sentiment analysis, they have hand-tagged 400 text corpus of reviews on a scale ranging from -5 to +5. Semantic-orientation calculator (SO-CAL) approach is incorporated into this method to annotate words with their respective polarity and strength. That means words are assigned positive or negative values through semantic-orientation calculator based on its usage across domains. PD Turney and ML Littman (2002) coined the term SO in the report “Unsupervised learning of semantic orientation from a hundred-billion-word corpus” presented at national research council Canada. The scale is calculated based on the strength of desirability (positive semantic orientation) and undesirability (negative semantic orientation). In this chapter, a detailed

review of opinion analysis implemented in English, Asian languages, Indian languages, and Malayalam is discussed.

2.1. Opinion Analysis in English

Experiments on online blogs, for the classification of subjectivity and polarity of opinions using features like verb class information and Wikipedia dictionary, is the primary focus of the article 'using verbs and adjectives to automatically classify blog sentiment' by Chesley et al. (2006). Opinions of words, to figure out if it falls into positive or negative sides of an opinion, are extracted from verb class information, whereas the polarity of adjectives is extracted from their entries in the Wikipedia dictionary. The experiment is not domain specific and does not only analyze one particular topic but is designed to ponder all the topics in the blog and is classified as objective, subjective-positive or subjective-negative. The accuracy of the polarity of the adjectives is given out to be 90.9%, and the accuracy of the polarity of two classifications of verb classes are 89.3% and 92.1% respectively. Accuracies of the initial classifiers are 72.4% for objective posts, 84.2% for positive posts and 80.3% for negative posts. It is to be noted that the experiment has represented a substantially higher accuracy than the expected baseline classifications.

The 76 test files contained around 3460 tokens of adjectives, along with 836 types of adjectives. Among the 836 types of adjectives provided, 10.5% of the set, i.e., around 88 were assigned polarity. Apart from the first category of lexicon-level results give above, the accuracy of polarity in the Wiktionary method has also been observed as plausible and the figure is 90.9%. Test dataset consisted of 29 objective posts, 25 positive posts and 22 negative posts, for which the accuracy rates yielded were 72.4%, 84.2%, and 80.3% respectively. In case of verb class information, the doubting verbs had an accuracy of 25/28 i.e. 89.3%, which was higher than the accuracy of positive mental affecting verbs which are

marked to be 31/34 i.e. 91.2%. As a further extension of this experiment, the aim of the researcher is to improvise on the recall and not to deplete on the precision of the manually jotted list of adjectives with known polarity.

A challenge that was faced during the experiment was to uncover the ways of classifying a blog into determining a certain kind of orientation for the already existing lexical orientation in the database. As already mentioned, this experiment is not domain specific and is thus looking forward to coalescing the research to a domain specific or topic-classifying search engine for the online web and blogs. Advancement through the addition of word tokens, a textual feature that classifies opinions with the inclusion of more adjectives list of known polarity. Constructive was of imbibing negative opinions into the analyzer or classifier is also promised to be taken care of. Mainly, the research also wants to inspect the facets of rhetorical structure and linguistic features in the blogs or comments that express opinions.

Examining the experiment, the researcher notes that sentences beginning with *despite* and *although* opinionizes the existing polarity in the subordinate clause and the complement polarity in the main clause. It strongly suggests that the inspecting and imbibing the subordinate clause and examining the rhetorical structure amalgamate the best with techniques like lexical information for extracting the opinions in blogs.

In the article ‘Twitter as a corpus for sentiment analysis and opinion mining’, Pak, A. and Paroubek, P, (2010) have concentrated on the most popular microblogging platform in the present days, Twitter. Corpus from Twitter has been analyzed to derive the outcome of multiple opinions of the public, through opinion analysis. The article emphasizes to demonstrate the method of automatically collecting corpus for multiple purposes to ease the process of opinion analysis or opinion mining. Ability to perform linguistic analysis of the collected corpus and explain discovered phenomena, using the corpus, was the major need to

build a sentiment classifier that is capable of determining positive, negative and neutral opinions for a document. Experimental evaluations of the proposed techniques in this article are claimed to be efficient and give out a better performance than the previously proposed methods. This research deals majorly with English, however, the article also claims to have been designed to use the proposed technique with other languages. The dataset collected for this research were the comments that ranged from personal opinions to political statements. Among the machine learning techniques like SVM and CRF used, the Sentiment analyzer built using the multinomial Naïve Bayes Classifier worked best in this research. Two Bayes classifiers are trained, which use different features: the presence of n-grams and part-of-speech distribution information. N-gram based classifier to evaluate the presence of an n-gram in the post as a binary feature and the classifier based on POS distribution estimates the probability of POS-tags presence within different sets of texts and uses it to calculate posterior probability. Although POS is dependent on the n-grams, the assumption is made for conditional independence of n-gram features and POS information for simplification of calculation. Results were mentioned as favorable to bigrams as it has provided a good balance between coverage (unigrams) and the ability to capture the sentiment expression patterns (trigrams). To examine the impact of dataset size on the performance of the system classifier, F- measure has been used and the accuracy for the same is complimented.

To conclude from the observations of this research, the authors use syntactic structures to describe emotions or state facts. It is also stated that POS-tags may be strong indicators of emotional texts. For further research advancements, plan to collect a multilingual corpus of Twitter data and compare the characteristics of the corpus across different languages, has been proposed by this article. Also, a plan to use the collected data in order to build a multilingual sentiment classifier has been mentioned.

In the article ‘learning word vectors for sentiment analysis’, the model presented by Maas et al. (2011) use a mix of unsupervised and supervised techniques in order to learn word vectors that capture semantic term–document information along with rich opinion content. The proposed model is said to leverage both continuous and multi-dimensional opinion information in addition to the non-opinion annotations. The model is also instantiated to utilize the document-level sentiment polarity annotations that are present in many online documents. The proposed model is evaluated using widely popular sentiment and subjectivity corpora and has claimed to have out-performed several previously acquainted methods of sentiment classification. A large dataset of movie reviews is also introduced in this article, affirming to set a robust benchmark of work in this domain.

As mentioned earlier, this article presents a vector space model that imbibes word representations which captures semantic and sentiment information. The article justifies the probabilistic theoretical foundation as a technique suggested for word vector induction that acts as an alternative to the most commonly used factorization-based technique, which consists of a huge number of matrices. The vector space model is thus compared to the log-bilinear model, which has followed a recent success in using similar techniques for language models and is also related to probabilistic latent topic models. It is also observed that the model is designed in such a way that the topical components of the model aim to capture word representations instead of latent topics. The author also claims to have fared well in the experiments, which performed better than LDA, a model that latent topics directly.

The unsupervised model was further extended to inculcate sentiment information and demonstrated the outcome of how the extended model can heft the sentiment-labeled text available online, to yield an outcome of word representations that garners both sentiment and semantic relations. The research also demonstrates the utility of the word representations that are captured by both sentiment and semantic relations, on two tasks of sentiment

classification. Sentiment classification was done using the collected datasets, having further plans to enlarge the boundary of datasets for future extended research purposes. In order to collect words that have semantic similarities, a probabilistic model of documents that learns the word representations. It has been mentioned that this very process does not require labeled data, also it shares its foundation with the probabilistic topic models like the “Latent Dirichlet Allocation”. The elements of sentiment in the designed model is said to use the sentiment annotations to filter words that express all similar sentiments, in order to have similar representations within the model. Alternating maximization is a proposed parameter for learning the joint objective functions used in the research and is even proposed for detailed further research in the future. Among the word representations, Latent Semantic Analysis, Latent Dirichlet Allocation, and Weighting Variant were tested, out of which LDA was the preferred model. The collection of datasets included Pang and Lee Movie Reviews and the movie reviews from IMDB.

O'Leary (2011) carries out research on blog mining. He explains several sets of blogs that are to be analyzed based on the choices. The choices are a small selected set of blogs, a random set of blogs, all available blogs, blogs of a particular type, blogs from a particular time period, or an experimental set of blogs. He states that blogs provide opinion, sentiment, and information about a range of issues. He identifies opinion words in blogs and classifies into positive and negative verbs and adjectives. This article suggests a domain-specific analysis in order to improve the quality of the analysis.

Feldman (2013) discusses the problems in the techniques used for the sentiment analysis. He states that sentiment lexicon is the most important aspect needed for sentiment analysis algorithms. He describes some of the major applications of sentiment analysis and discusses some limitations and among those, the major issue is the lack of knowledge in the classification methods to analyze compositional sentiment.

2.2. Opinion Analysis in Asian Languages

The article ‘lexical based sentiment analysis – verb, adverb, and negation’ (Shamsudin et al., 2016) introduces us to a lexical based method in classifying opinions of Facebook comments into Malay. Term Counting and Term Counting Average are two types of lexical based techniques that are implemented in order to classify the opinions of Facebook comments. POS is also being taken into account for the analysis. The method Pre-processing process is involved to deal with the noisy texts in data. Term Counting is found to have been working better for adjectives and adverbs, while Term Counting Average has been performing better for verbs and negation words.

Taking a look at the accuracy of different POS combinations used in TC and TCAvg methods in this research, it is observed that the accuracy of TC and TCAvg, is based on the usage of the Adjectives as their base source, and on the original data, Adj and pre-processed data Adj. It is noted that the TC and TCAvg methods that have been applied to pre-processed data, has produced a better result when compared to the original. The accuracy skimmed by the methods given above shows that data pre-processing is essential in putting forth a quality result. When comparing the results of Adjective POS combinations on the pre-processed data, the Adj + Neg the greatest accuracy was given out for both the combination of methods. It is also to be observed that, in contrast, Adj shows the lowest accuracy among the two. Comparing the results of both the Verb POS combinations, it is evident that the Verb + Neg combination embraces the highest accuracy, as the Verb + Adv combination has projected the lowest accuracy. To be comprehensive, the Verb + Neg combination of TCAvg has the highest accuracy of 52.12% while the Verb + Adv combination of TC has the least accuracy of 6.36%. The research also promises a further advancement in ameliorating the active scoring methods to make use of each word, given for the analysis and to improve the reliability of the methods in use.

The article, ‘deeper sentiment analysis using machine translation technology’ by Hiroshi et al. (2004), proposes a high-precision opinion analysis system, in a low production cost using an already existing transfer-based machine translation engine. The process includes translation of a text leading to opinionizing the same. Transfer-based translation system has been divided by the researcher into three parts, namely: a source language syntactic parser, a bilingual transfer which handles the syntactic tree structures, and a target language generator. The very technique of this type of translation is considered highly complex, as there can be a glitch due to a huge number of similar patterns of combinations in this operation. The aim of this article, to generate a high-precision opinion unit is at farce if the full syntactic parsing works right in opinion extraction. For this reason, the researcher uses a top-down pattern which matches the tree structures than complete parsing, in order to find every single opinion fragment that is essentially a part of the opinion unit. The three parsing patterns proposed by the researcher are principle patterns, auxiliary patterns, and nominal patterns.

The results of the experiment put forth that the precision of the opinion polarity was higher than it is usually for the conventional methods, and the opinion units designed by the researchers were not superfluous and were more productive than when the naive predicate-argument structure was used. It is confessed in the article that they have exploited a lot of advantages of deep analysis, keeping in mind the cost-effective technique to be developed. So that, many of the existing or the upcoming techniques of machine translation can be used by many naturally with regard to the extraction of opinion units as a form of translation. The experiment leaves us with a conclusion that most of the techniques mentioned and studied here, for machine translation, like word sense disambiguation, anaphora resolution, and automatic pattern extraction from corpora can enhance the future researches on opinion analysis or other NLP tasks. Therefore, they leave us with a note that this particular work is the first step towards the intersection point for shallow and wide NLP, with deep NLP.

2.3. Opinion Analysis in Indian Languages

Opinion analysis is done on Hindi in the article titled ‘a framing study on sentiment analysis of Hindi language using machine learning’ (Sheetal et al., 2018). The emphasis is not just on the creation of information, but also on the classification of the sentiment in Hindi. The main aim of the research is to determine the attitude of a sentence, pertaining to a specific domain or as simple as a general contextual polarity for all unspecific domains. Early applications laid by Turney and Pang for detecting the polarity of product and movie reviews in Hindi, is referred by the author, to expand their research. The data mining and machine learning techniques are carried out to churn out the positive and negative polarity of sentiments in Hindi news. Furthermore, these approaches were designed to analyze the merits and demerits across different genres f sentiment classification in all domains. The problems or issues in working with the user-generated contents like movie reviews and news in Hindi is discussed in detail.

Datasets were collected from social networking sites and the application of the same is to be inculcated in business intelligence like marketing; cross-domain applications like sociology, psychology, and administration; feedback and recommendation systems. The datasets included a large number of Hindi news sentences from the Hindi news websites. All the pre-processed data are then considered the desired dataset for input and processed using algorithms. Experiments conducted on movie reviews were based on the datasets from websites that contain Hindi reviews. The inputs are then classified into different chunks, with reference to their polarity assessed by the machine, in order to run a comparison to the analysis made by human judgment. There were three evaluation measures used, on the basis of which the performance of the system is computed, they are precision, recall, and accuracy. The result as mentioned in the article is that the best k was found for instance, at 850 and

accuracy found was about 0.664041994751 at the given dataset (3000 Hindi News Sentences).

Opinion analysis is done on the tweets in three languages, namely, Hindi, Bengali, and Tamil in the article ‘shared task on opinion analysis in Indian languages; (sail) tweets - an overview’ by Patra et al. (2015). The article claims to be the first research or study over opinion analysis on Twitter, in Indian languages. Positive, negative and neutral polarity was taken into account for the tweets in each language, under the constrained and unconstrained systems. The ranking system of six teams was based on the accuracy attained through the systems, maximum accuracy achieved for Bengali, Hindi, and Tamil were 43.2%, 55.67%, and 39.28% respectively. This article also makes a strong statement on the use of SAIL-2015, i.e., Sentiment Analysis for Indian Languages, as being beneficial to Indian researchers working on automatic opinion analysis for each of their own regional languages in order to extract relevant data. The prime objective of SAIL-2015 is described as gathering research scholars, experts, and practitioners from this area, to discuss, collaborate and instigate the research on opinion analysis, especially for Indian languages. The research is also said to involve the technique of research-creation, sharing of data and collaboration for further advancements.

Training and test datasets were collected from twitter over a period of three months. The monolingual corpus for each of the languages mentioned above was collected manually on different topics. A word frequency list was prepared to remove the stop words and had a thorough check if each word exists in the frequency list in Twitter. There have been over 2000 tweets collected from each language, and the method of implementation used was the TWITTER4J2, a Java supporter of Twitter API to download the tweets. The duplicate tweets were removed manually. Happy smileys were normalized before considering it into count

from each language, where usage of smiley in Tamil was found more than that of Bengali or Hindi.

It is observed in this research that the maximum accuracy achieved for Bengali, among the four teams which submitted the results is 43.2 % by the team IITTUDA. For Hindi, among the six teams that submitted the results, AMRITA-CEN achieved the maximum accuracy of 55.67 %. Tamil had the least number of participants participating, out of which AMRITA-CEN has observed to have achieved the maximum accuracy of 39.28 %. Most of the teams that have participated have seen to have used the SentiWordNet system, that was developed for a constrained system. Notably, teams have also used techniques or methods like hashtags, retweet, TF-IDF scores of n-grams, links, question marks, exclamatory marks, smiley lists and SentiWordNet for the task of analyzing the opinions. These allotted teams have used several well-known supervised classification algorithms like Decision Tree, Naïve Bayes, Multinomial Naïve Bayes and Support Vector Machines (SVMs). It is observed by the researcher that, the accuracies of the unconstrained systems are less, compared to the constrained systems. The main reason as suspected and mentioned by the researcher is the unavailability of basic NLP tools like POS taggers and NER specifically for Indian language tweets. The reason behind inefficient results, in this case, is judged as the accuracy systems for Indian language tweets which are lesser compared to the systems for English tweets. It is a chain of actions, connected to the scarcity of opinion lexicons for Indian languages and Indian language tweets. However, there is a good number of opinion lexicons available for Hindi, Bengali and Tamil, collected as plain texts but not specialized for tweets. The reason being, in case of tweets, there are many variations in spellings. Also, acronyms and emoticons that make the opinion analysis a more difficult task and indeed challenging, when compared to other tasks on the conventional sentiment analysis techniques. In most of the cases, it is difficult to collect the monolingual Indian language tweets because, in most of the

cases the tweets are written in English scripts and it should also be taken into consideration that, tweets are code mixed. The annotation of such monolingual tweets, based on sentiment expressions, requires the involvement of manpower and time is the conclusion that this article leaves us with.

Kaur, J., & Saini, J. R. (2014) carry out research on opinion mining in Indo-Aryan, Dravidian, and Tibeto-Burman language families. It does a survey on opinion mining tasks performed in various languages under the above-mentioned families. It finds out that most of the opinion mining requires WordNet and support vector machine is widely used for the classification in most of the Indian languages. It is observed that among all language families, Dravidian languages show higher performance.

Cross-lingual sentiment analysis for Indian languages: Marathi and Hindi using linked WordNet is dealt with in the article of Balamurali et al. (2012). They explain cross-lingual sentiment analysis as "the task of predicting the polarity of the opinion expressed in a text in a language L_{test}using a classifier trained on the corpus of another language L_{train}". Marathi WordNet is created from the Hindi WordNet and this approach provides an accuracy of 72% and 84%for Hindi and Marathi respectively.

Bansal, Naman el al. (2013) performs a task of sentiment analysis in Hindi movie reviews and attains an accuracy of 64% by training 150 labeled reviews. For the data mining process, the article uses deep belief networks.

2.4. Opinion Analysis in Dravidian Languages

In the work 'enhanced sentiment classification of Telugu text using ml techniques' (Mukku, S. S.; Choudhary, N.; and Mamidi, R., 2016), the researchers have tried to classify the polarity (negative, positive and neutral opinions) of Telugu sentences using various Machine Learning Techniques like Naive Bayes, Logistic Regression, Support Vector Machines,

Multi-Layer Perceptron, Neural Network, Decision Trees and Random Forest. There are two models built for classifying the opinions into two tasks: the first one is a binary task of classification of opinions, where the opinions are divided into positive and negative polarities only, whereas, a ternary task of classification of opinions are conducted to divide the opinions into positive, negative and neutral polarities. Algorithm and formulations are provided for the same, within explanation.

The corpus consisted of 7,21,785 raw sentences from Telugu, which was collected from ILCI- Indian Languages Corpora Initiative which was used for generating sentence vectors by training Doc2vec model. Each Annotated corpus consisting of Telugu sentences were then attached to a corresponding polarity tag. The results are given as follows: among the tested 1644 sentences, 1068 were found positives, 219 negatives and 357 neutral sentences. These sentences were made use of, for training, testing and evaluating the classifier models. Source for the corpus was the raw data taken from the Telugu Newspapers. It is mentioned that the collected raw data from newspapers was first annotated by two native Telugu speakers separately and then the data was merged by a third native speaker, who also validated it simultaneously. The inter-annotator agreement that consists of annotation of the three polarity tags, positive, negative and neutral was done using Cohens´ kappa coefficient. Annotation consistency in k value was observed and noted to be 0.92, in perfect agreement.

Researchers have converted the annotated data of Telugu sentences into 200- dimension feature sentence vectors. Doc2vec tool was used for the process of conversion that is a python module, provided by Gensim. For the experiment, the 5-fold cross-validation method is used to perform the experiment four times in order to improve the validity of results. Division of randomly chosen sentences into parts are made and performed in the training step. In the observation, it was found that the binary classification, that includes Random Forest, Logistic Regression and Support Vector Machines yielded good results. Among all of these, the

Random Forest classifier has been observed to be preferred and convenient, as it is easily understandable, and under ternary classifier Logistic Regression yielded better results. This approach has claimed to have no restrictions for any specific domain. However, it has been concluded that small modifications in the pre-processing, would be sufficient to use this algorithmic formulation in different domains or languages.

Highlighting the absence of formidable techniques of reviewing or classification of opinion in Indian languages like Kannada, Kumar et al. (2015) in the article 'analysis of users' sentiments from Kannada web documents' aim to develop algorithms for the same, in order to apply on the opinions expressed in Kannada websites. A dataset of both positive and negative keyword list was developed with manual identification of the reviews, translations were done using Google translate to build a list of negative words used in the windows algorithm, in Kannada. Kannada POS tagger software was used to implement and analyze adjectives through Turney's algorithm. Experiment on the datasets was done in the sentence level approach and was experimented by splitting the opinions into individual sentences. Apart from this, a few more machine learning algorithms like J48, Random Tree, ADT Tree, Breadth First, Naïve Bayes and Support Vector Machine in the Weka software were tried and the obtained results were compared with the semantic methods as well.

As a database, the researcher has collected 182 positive Kannada opinions and 105 negative Kannada opinions, and the above algorithms were applied to analyze the results. Sizes of 3, 5 and 7 are taken for the Negative-Window Algorithm and for Sentence Analysis Algorithm, the first, middle and last sentence was taken as significant sentences and the baseline algorithm was tried for a few more special cases. The reviews collected were mainly for broad and yet specific domains like automobiles, health and body care products like soaps, shampoo, electronic items like TV, mobiles, movies, songs, websites, TV programs, famous people, etc. Especially the reviews on commercial products were collected and analyzed.

It is found that the results of the baseline method have performed better among the other approaches mentioned. Considering the window algorithm applied, it is observed that window 3 is more accurate when compared to window 5 and window 7. When it comes to the sentence-based approach, the significant sentence has fared well, compared to the other three sentence-based approaches. It is also evident that POS Kannada, Turney Kannada methods were a failure when compared to the POS English, Turney English Pattern methods. Another finding in the research is that the classification of positive reviews in Kannada web was more accurate than the classification of negative reviews in the same. Among all the machine learning methods experimented, all the algorithms performed decently and it is to be noted that Naïve Bayes gave the better result, compared to other techniques in Weka, in terms of accuracy. The results inferred are based on learning algorithms, which is based on the training set that falls under the supervised learning methods.

2.5. Opinion Analysis in Malayalam

The article 'SentiMa - Sentiment Extraction for Malayalam' (Nair et al., 2014) propounds a rule-based approach for opinion analysis from Malayalam movie reviews. The rule-based approach that has been suggested by the researcher for extracting the opinion analysis in Malayalam is the Negation-Rule, that has claimed to have achieved 85% of accuracy. For the implementation of this approach, the collection of a set of corpora from a specific domain, majorly on film reviews from plenty of sources like blogs, magazines, and newspapers. Further down, these collected corpora go through the process of tokenization, before directly applying the negation-rule approach. Based on the frequency of occurrence of a specific word, opinions of those words are added to analyze the set of words using negation-rule. Opinions of words which are not both negative and positive will be marked neutral based on the pre-fed opinion data. After the process going through word-by-word analysis, the next stage is to analyze the meaning of the entire sentence, with the collective meaning of the

analyzed words. The designed program is in python and the input and output of the analysis done in Unicode notation. It is also claimed that the analyzer is also capable of opinionizing smileys used in the comments, is also a pre-fed data on a certain opinion. As a future work proposed in the article, the researcher has approached for other machine learning techniques like SVM, CRF, Maximum Entropy etc, through which implementation of opinions is expected yield advanced results from now.

Nair, Deepu et al. (2015) employs machine learning techniques to mine the opinion from Malayalam film reviews. It does a sentence level sentiment analysis on reviews with the aid of machine learning techniques like Support Vector Machine and Conditional Random Fields (CRF). 3000 tokens are used for the experiment and result is that Support Vector Machines provides better accuracy.

Mohandas et al. (2012) research focus on domain-specific sentence-level mood extraction from Malayalam text. Research suggests two methods of sentiment analysis: machine learning method and semantic orientation method. The task is carried out using a semantic orientation method using PMI-IR (pointwise mutual information retrieval) algorithm.

Chapter 3

Feature-Based Classification for Opinion Analysis in Malayalam

3.0. Introduction

In this chapter, the feature-based classification of linguistic features for opinion analysis in Malayalam is discussed. From the previous researches, it is understood that the lexicon-based method is the most effectively practiced method in this field (Turney & Littman, 2002; Taboada, et al., 2011; Chesley, et al., 2006). This chapter focuses on various linguistic features used to extract information for the opinion classification based on lexical-based method. Classifying opinions is invariably challenging for certain extent if the nature of the language that we work on is unclear. Though lexical feature aids in determining polarity, considering Malayalam being agglutinative, understanding certain linguistic features which are encoded as inflection and derivation is crucial in opinion classification. In the present research, we have attempted the feature-based classification, wherein lexical features, functional features, textual features, and parts-of-speech (POS) information shall provide the required information for classifying sentence.

3.1. Feature-based classification

Chesley et al., (2006) employ three features which are lexical features, POS features and textual features to determine and classify blog sentiment in English. Apart from these features, this study also considers using the functional features i.e. various morpho-syntactic information encoded on lexical items in terms of derivation and inflection as these features carry features relevant for opinion analysis.

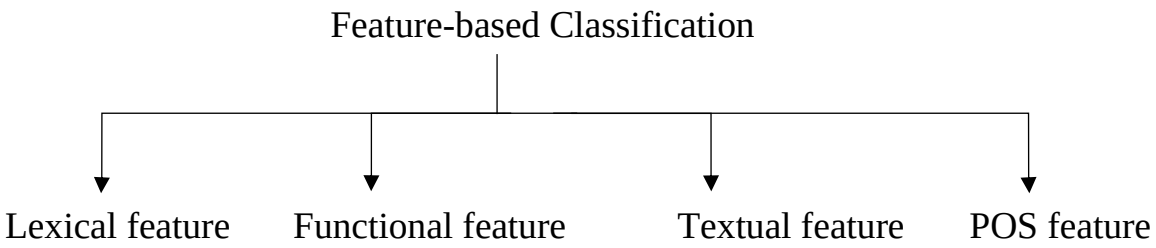


Fig.1: Feature-based classification

The feature-based classification is explained in the following section with examples from Malayalam.

3.1.1. Lexical feature

Lexical categories such as verbs, adjectives and adverbs are considered in this study as they are the potential opinion carriers (Subrahmanian, & Reforgiato, 2008). Hatzivassiloglou, & McKeown (1997) work majorly depended on adjectives to examine the sentence subjectivity. Hatzivassiloglou & Wiebe (2000) focus on the polar verbs and verb classes to determine the sentence subjectivity.

Adverbs express the degree of likeness of the opinion. They usually do not negate the opinion as against the adjectives, but add the intensity to the opinion.

In the present study, the lexical categories such as verbs, adjectives and adverbs are considered to extract opinion from a given sentence. Other lexical categories like nouns, pronouns and numbers are not taken for the current study as they do not supply any opinion information in a sentence. This research has adopted two methods to study the lexical polarity.

a) verb class information from Levin's *English verb classes and alternations* (Levin, 1993) and 'verb classes' stated by Chesley et al. (2006) and

b) corpus-based lexicon approach for verbs, adjectives and adverbs based on WordNet (Miller, George A et al., 1990).

3.1.1.1. Verbs

Verbs are the major category in deciding the opinion in a sentence. More than classifying text into subjective or objective they express the strength of desirability and undesirability in an opinion. To study the lexical polarity the verb classes are divided into subjective verb classes and objective verb classes based on its semantic orientation. Among subjective verb classes, a certain group of verbs expresses positive orientation and certain other groups of verbs shows negative orientation. Based on the semantic features of verbs, a verb-based model is built and two verb classes are identified which are polarity verb classes and non-polarity verb classes. Polarity verb classes are divided into positive and negative verb classes whereas Non-polarity verb classes express the neutral opinion and are divided into objective verb classes and facts-implied subjective verb classes as shown in figure 2.

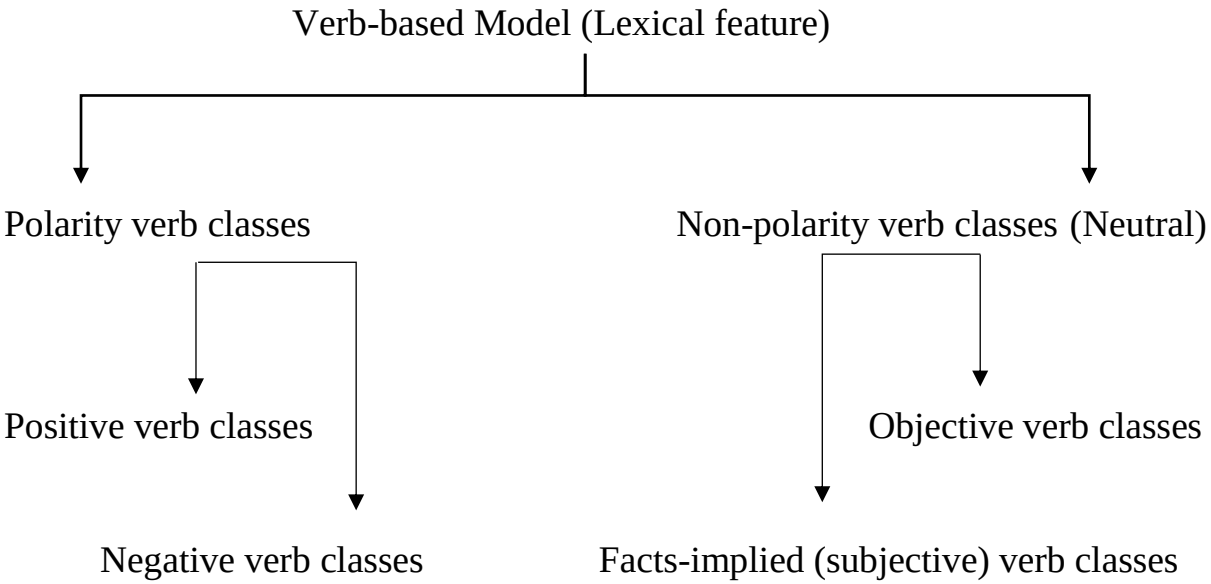


Fig.2: Classification of verb-based model

Levin (1993) classifies English verbs on a principle that the syntactic behavior of a verb can be traced from its meaning. Chesley et al. (2006) classifies verbs and integrate polarity into the verb class and implemented in automated blog sentiment. Verb classes such as *roll verbs, run verbs, asserting verbs*, etc. fall into the category of objective verb class whereas *declare*

verbs, suggesting verbs, mental sense, etc. classify facts-implied verb class. Positive verb classes include *praising verbs, declare verbs, obtain verbs,* etc. and negative verb classes include *abusive verbs, remove verbs, steal verbs, negative admire-type psych-verbs,* etc. Table.1 provides a detailed description of verb classes.

Verb-based model (Lexical Feature)			
Polarity Verb classes		Non-polarity Verb classes	
Positive verb classes	Negative verb classes	Objective verb classes	Facts-implied (subjective)
create verbs, obtain verbs, verbs of selection, positive social interaction verbs, amuse verbs, positive admire-type psych-verbs, rescue verbs, approve verbs, praise verbs, positive mental affecting verbs, appoint verbs, positive judgement verbs, etc.	remove verbs, destroy verbs, cheat verbs, steal verbs, negative social interaction verbs, break verbs, hit verbs, cut verbs, killing verbs, suffocate verbs, negative mental affecting verbs, bang verbs, negative admire-type psych-verbs, negative judgement verbs, abusive verbs, accuse verbs, negative mental affecting verbs, etc.	asserting verbs, roll verbs, run verbs, eat verbs, drive verbs, motion verbs, special verbs, etc.	mental sensing verbs, declare verbs, suggestion verbs, etc.

Table.1: Examples for verb-based model

Since polarity verb classes express positive and negative opinions, they are classified and stored in a database whereas, non-polarity verb classes do not express an opinion and they are considered neutral opinion verbs. Verbs other than positive and negative classes are considered neutral and they are not specifically stored in a database.

Examples for positive verb classes and negative verb classes in Malayalam are discussed below.

(i) Table. 2 provides some examples of positive verb classes in Malayalam

Verb class	Transliteration	Gloss
Fulfilling Verbs	*niṟavēṟṟuka* *pūrttiyākkuka*	'to fulfill' 'to accomplish'
Psych-verbs	*abhinandikkuka* *ādarikkuka*	'to appreciate' 'to honour'
Social interaction	*sahakarikkuka* *pariharikkuka* *sammatikkuka* *aṅgīkarikkuka*	'to cooperate' 'to solve' 'to agree' 'to approve'
Rescue verbs	*rakṣappeṭuttuka* *mōcippikkuka* *sahāyikkuka*	'to rescue' 'to set free' 'to help'

Table.2: Positive verb classes in Malayalam

(ii) Table. 3 provides some examples of negative verb classes in Malayalam.

Verb class	Example	Gloss
Psych-verbs	*pīḍippikkuka* *apamānikkuka* *ākṣēpikkuka*	'to harass' 'to insult' 'to reprimand'
Banish verbs	*ṟaddākkuka* *vilakkuka*	'to ban' 'to forbid'
Social interaction	*avagaṇikkuka* *anādarikkuka* *bhīṣaṇippeṭuttuka* *viyōjikkuka*	'to disregard' 'to disrespect' 'to threaten' 'to disagree'

Social interaction	*visammatikkuka*	'to disagree'
Killing verbs	*kolappeṭuttuka*	'to murder'

Table.3: Negative verb classes in Malayalam

Sentences (1) and (2) carries positive verbs.

(1) *ākramaṇaṁ* *ceṟukk-ān* *maikhiḷ* *pōlīsine* *sahāyi-ccu*
attack resist-INFN Michael police help-PAST
'Michael helped the police in resisting the attack.'

(2) *kiṇaṟṟil* *cāṭ-iy-a* *yuvāvin-e* *pōlīs* *rakṣapeṭutt-i*
well fall-PAST-RP man-ACC police save-PAST
'Police saved the man who fell in the well.'

In the above sentences (1 and 2), verbs *sahāyichu* 'helped' and *rakṣapeṭutti* 'saved' belong to the positive verb classes and hence the sentences express a positive sense.

Sentences (3) and (4) carries negative verbs

(3) *sunil pi iḷayiṭatti-nṟe* *ōphīs* *ākrami-ccu*
Sunil P Ilayidom-GEN office attack-PAST
'Sunil P Ilayidom's office was attacked.'

(4) *kēsumāyi* *bandhappeṭṭa* *kūṭutal* *vivaraṇṇaḷ* *paṇkuvaykk-ān*
case related more information share-INFN
pōlīs *visammati-ccu*
police refuse-PAST
'Police refused to share more information about the case.'

In the sentences (3) and (4), verbs *ākramichu* 'attacked' and *visammatichu* 'refused' belong to the negative verb class and hence they express negative sense in the context.

(5) *el. ḍi. klarkk niyamanaṁ pi.es.si. r̲addŭ ceyt-u*
L.D. clerk appointment P.S.C. ban do-PAST
'P.S.C. banned the appointment of L.D. clerk.'

The sentence (5) expresses negative sense in the context because of the presence of nominal conjunct verb '*r̲addŭ ceyt-u'*.

3.1.1.2. Adjectives

Along with verbs, adjectives do play a role in opinion making. Adjectives function as an attribute of a noun and they play crucial role in opinion making.

(6) *te̲rr̲āya pravartti ceyy-arutŭ*
wrong acts do-NEG.IMP
'Don't do wrong acts.'

(7) *avan vaḷare nalla tīrumānaṁ eṭu-ttu*
he very good decision take-PAST
'He took a very good decision.'

Sentence (6) and (7) shows the importance of adjectives in opinion making. In sentence (6), the verb carries the negative opinion and the adjective *te̲rr̲āya* (wrong) could shift the overall opinion of the sentence into positive. The combination of verb's semantics along with adjectives play a crucial role in identifying the overall opinion of a sentence. '*nalla'* (good) deepens the positive response in the sentence (7).

3.1.1.3. Adverbs

Adverbs modify primarily the meaning of verb other than adjectives and other adverbs. It may not shift the opinion instead it retains the opinion of the verb but intensifies the degree of the opinion verbs.

(8) *avan* *kaṣṭiccŭ* *rakṣapeṭṭ-u*

he narrowly escape-PAST

'He narrowly escaped.'

(9) *avan* *ēr̲r̲avum* *vēgaṁ* *ōṭ-i*

he extremely fast run-PAST

'He ran extremely fast.'

In sentence (8) the adverb *kaṣṭiccŭ* and in the sentence (9) *ēr̲r̲avum* and *vēgaṁ* intensify the polarity carried out by the verb.

3.1.1.4. Opinion Lexicon

Lexical resources for this study are extracted from IndoWordNet (Bhattacharyya, 2017) as it contains concepts and synsets for the concepts. Table.4 gives a detailed description of opinion lexicon built for this research. The political corpus containing eighty-seven thousand and three hundred and forty-seven (87347) sentences that are extracted from Malayalam online news websites is used in this study. The frequent lexical items in the corpus that show polarity i.e. positive and negative are extracted and potential synonyms of these extracted lexical items are obtained from Malayalam WordNet.

Malayalam opinion lexicon is built with a classification of adjectives, adverbs, and verbs with their polarity. The selected positive and negative opinionated words in the political context reflect their certain desirability or undesirability in the political field by their usage. The functional approach emphasizes the fact that text and context are inseparable, instead, they are conflating and interdependent (Labov 1972, Halliday 1978 and Bernstein 1970). Meaning of an utterance is shared and stored experience of the speech community and it is stored in the system of context. Any kind of social interaction especially, communication happens if and only if the speaker and listener have shared knowledge and communication becomes

impossible without it (Searle 1969). Thus, any text or utterance is functional or socially constructed. Meaning of any utterance is produced when it is expressed in a context. Hence, the built Malayalam lexicon has a certain way of expressing its meaning in a political context which may differ its meaning in other domains. So, the lexical items present in the Malayalam lexicon are political dependent words that perform effectively to infer viewpoint of the sentence shaped in the political field.

Lexical Feature	No. of positive words	No. of negative words	Total No. of words
Verb	516	408	924
Conjunct Verb	41	21	62
Adjective	58	20	78
Adverb	19	8	27
Total	634	457	1091

Table.4: Size of Opinion Lexicon

3.1.2. Functional Feature

In Malayalam, functional features are the morphological elements that contain grammatical functions in a sentence. These features are the exponents of various morpho-syntactic information that word carries. In this study, functional features of verbs are studied to identify different functional features which express positive or negative semantic orientation. The copula verbs such as *ākŭ* 'to be' and *uṇṭŭ* 'to be' have their corresponding negative forms: *alla* and *illa* respectively are attached to any constituent to convert the expression negative.

Finite and non-finite forms of the verb express negation through suffixes as negative markers. Figure.3 gives the description of verb conjugation of Malayalam.

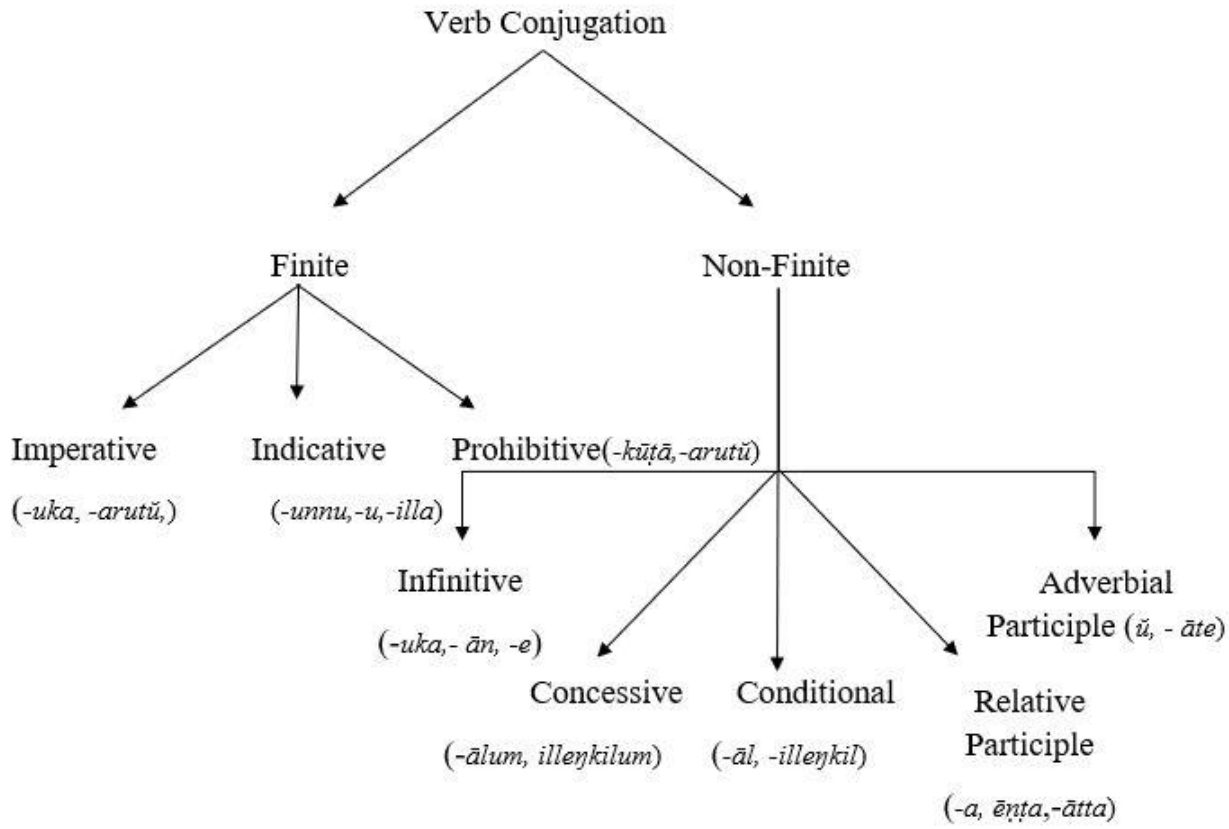


Fig.3: Verb conjugation in Malayalam

This section describes the functional features of the verb conjugation with examples.

3.1.2.1. Finite

Finite verbs in Malayalam carry features such as imperative, indicative and prohibitive.

3.1.2.1.1. Imperative Form

Imperatives are expressed with second person singular and plural pronouns. Imperative forms can be divided into negative and positive based on its sense of expression. Imperative has polite forms in both positive and negative mood.

(i) Positive imperative

For some given verbs, the positive imperative forms are listed below in table 5.

I	II	III	Gloss
vā	*varū*	*varuka*	'come'
nōkhŭ	*nōkhū*	*nōkhuka*	'look'
ceyyŭ	*ceyyū*	*ceyyuka*	'do'

Table.5: Imperatives in Malayalam

The first class (I) imperatives show the non-polite form. The second class (II) imperatives express polite form but a degree lesser than the class (III) imperative. The third class (III) with infinitive form of the verb, having *-uka* suffix is the formal and polite form of the imperatives. Examples (10), (11) and (12) show three classes of imperatives in Malayalam.

(10) *nī* *paṟa*
you say-IMP (class I)
'You say'

(11) *ni* *paṟay-ū*
you say-IMP (class II)
'You say'

(12) *nī* *paṟay-uka*
you say-IMP (class III)
'You say'

(ii) Negative Imperative

Negative imperative verbs are constructed by attaching *-arutŭ,* to the verb stem.

(13) *ā* *satyaṁ* *viśvasikk-arutŭ*
that truth believe-NEG.IMP
'Don't believe that truth.'

Negative imperatives do express politeness in Malayalam when the suffix-*allē* is attached to the verb stem.

(14) *ā* *satyaṁ* *viśvasikk-allē*
that truth believe-NEG.IMP
'Please don't believe that truth.'

Another polite negative form is -*aṇṭa*, the debitive form attached to the verb stem.

(15) *nī* *onnum* *paṟay-aṇṭa*
you nothing say-NEG-DEB
'You need not to say anything.'

3.1.2.1.2. Prohibitive

The prohibitive form is formed by adding -*kūṭā* after an adverbial participle. Prohibition is also expressed by the marker -*arutŭ*.

(16) *ni* *onnum* *paṟaññŭ-kūṭā*
you nothing say.AP-PROH
'You should not say anything.'

(17) *nī* *paṟay-arutŭ*
you say-NEG.PROH
'You should not say.'

3.1.2.1.3. Indicative Form

Indicative is expressed in three tenses: past, present and future. Negative indicative is realized by adding –*illa* to all the positive forms. The past form is realized in two ways, one class of verb with -*i* ending and another class of verb with –*u* ending.

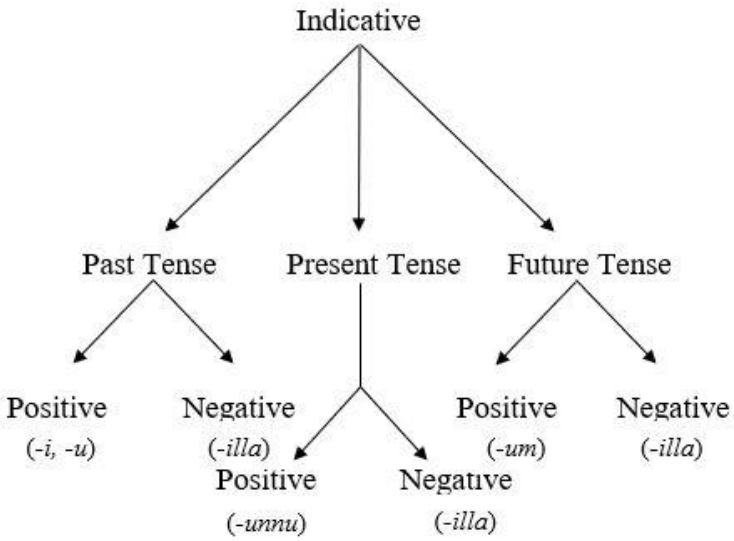


Fig.4: Indicative forms in Malayalam

1. Past tense

(i) Positive form

Past tense is formed by adding past tense marker *-i* or *-u* to the verb stem.

(18) *kōṭati* *śikṣicc-u*
court punish-PAST
'The court punished.'

(19) *ñān* *ā* *satyaṁ* *viśvasicc-u*
I that truth believe-PAST
'I believed that truth'

(ii) Negative form (by adding *–illa* marker to the past tense form.)

(20) *kōṭati* *śikṣi-cc-illa*
court punish-PAST-NEG
'The court did not punish.'

(21) *ñān* *ā* *satyaṁ* *viśvasi-cc-illa*
I that truth believe-PAST-NEG
'I did not believe that truth.'

2. Present tense

(i) Positive form

Indicative in present tense is formed by attaching –*unnu* to the verb stem.

(22) *kōṭati* *śikṣikk-unnu*
court punish-PRES
'The court is punishing.'

(23) *ñān* *ā* *satyaṁ* *viśvasikk-unnu*
I that truth believe-PRES
'I believe that truth.'

(ii) Negative form (by adding –*illa* marker to the present tense form.)

(24) *kōṭati* *śikṣikk-unn-illa*
court punish-PRES-NEG
'The court is not punishing.'

(25) *ñān* *ā* *satyaṁ* *viśvasikk-unn-illa*
I that truth believe-PRES-NEG
'I am not believing that truth.'

3. Future tense

(i) Positive form

The future form is realized by attaching -*um* suffix to the verb stem.

(26) *kōṭati* *śikṣikk-um*
court punish-FUT
'The court will punish.'

(27) *ñān* *ā* *satyaṁ* *viśvasikk-um*
I that truth believe-FUT
'I will believe that truth.'

(ii) Negative form (by adding *–illa* marker to the future tense form.)

(28) *kōṭati* *śikṣi-kk-illa*
court punish-FUT-NEG
'The court will not punish.'

(29) *ñān* *ā* *satyaṁ* *viśvasi-kk-illa*
I that truth believe-FUT-NEG
'I will not believe that truth.'

The verb *illa* indicates negation in two ways: i) functioning as the main verb and ii) combined to a lexical verb (Nair, 2012) and the verb *alla* expresses negation in i) functioning as the main verb. See example (30) and (31).

(30a) *avan* *kuṭṭi* *alla*
he boy be-NEG.PRES
'he is not a boy.'

(30b) *avan* *vīṭṭil* *illa*
he home be-NEG.PRES
'he is not at home.'

(31) *kōṭati* *śikṣi-cc-illa*
court punish-PAST-NEG
'The court did not punish.'

In the example (30a), '*alla'* functions as constituent negation, negating the boy.

In the example (30b), '*illa'* functions as existential negation, negating the existence of 'he' at home.

In the example (31), '*illa'* functions as a negation attached to a lexical verb.

3.1.2.2. Non-Finite

Non-finite forms of verbs in Malayalam are infinitives, concessives, conditionals, relative participles and adverbial participles. This section explains non-finite forms of verbs with examples in Malayalam.

3.1.2.2.1. Infinitive

Infinitive has three different forms, they are i)-*uka* which is the base form for many verbs, ii) verb stem + -*e* suffix which is also called verbal participle iii) verb stem + ān / uvān which is called purposive infinitive (Asher, 1997).

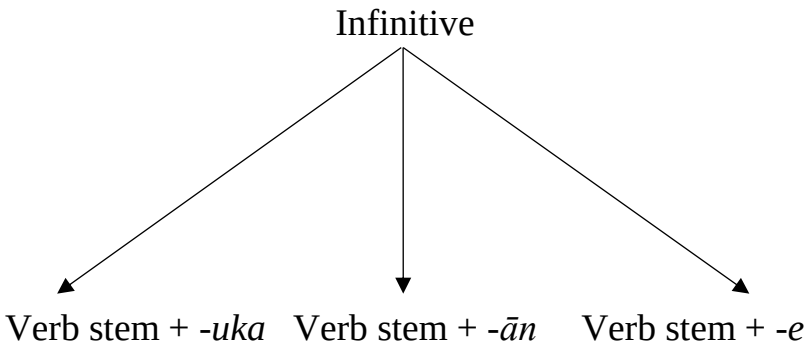


Fig.5: Infinitive forms in Malayalam

(32a) Verb stem + -*uka* is one form of the imperfective, it occurs before 'be' forms, future forms and it also used before coordinating suffixes (Asher,1997; Nair,2012).

kōṭati	*śikṣikk-uka*	*illa*
court	punish-INFN	be.NEG

'The court does not punish.'

(32b) The infinitive form 'Verb stem + -*ān'* express purpose.

kōṭati	*śikṣikk-ān*	*tīrumānicc-u*
court	punish-INFN	decide-PAST

'The court decided to punish.'

(32c) The infinitive form 'Verb stem + *-e*' express simultaneity.

ellārum	*nōkk-i*	*nilkkē*	*avan*	*vīṇ-u*
all	look-AP	stand-INFN	he	fall-PAST

'While everyone was looking, he fell.'

3.1.2.2.2. Conditional Form

The conditional form is expressed using two markers: *-āl* and *-eŋkil* occurring in the clause-final position. *-āl* is attached to the adverbial participle whereas *-eŋkil* is attached to the finite form of a verb.

i) Verb stem + *-āl* in clause-final position.

(33)

śampaḷaṁ	*koṭutt-āl*	*jōlikkār*	*jōlicceyy-um*
salary	give-COND	employees	work-FUT

'If the salary is given, the employees will work.'

ii) Verb stem + *-eŋkil* (allows a wide range of possibilities (Asher, 1997)).

(34)

śampaḷaṁ	*koṭukk-um- eŋkil*	*jōlikkār*	*jōlicceyy-um*
salary	give-FUT-COND	employees	work-FUT

'(May be) If the salary is given, the employees will work.'

The negative form of the conditional marker is obtained in two ways: (i) adding the suffix *-ātirunnāl* to the verb stem and (ii) adding the suffix *-illeŋkil* to the verb stem.

i) Verb stem + *-ātirunnāl*

(35)

śampaḷaṁ	*koṭukk-ātirunnāl*	*jōlikkār*	*rājivekk-um*
salary	give-NEG.COND	employees	resign-FUT

'If the salary isn't given, the employees will resign.'

ii) Verb stem + *-illeŋkil*

(36) *śampaḷaṁ* *koṭutt-illeŋkil* *jōlikkār* *rājivekk-um*

salary give-NEG-COND employees resign-FUT

‘(May be) If the salary isn’t given, the employees will resign.’

3.1.2.2.3. Concessive Form

Concession can be expressed using *-ālum* and-*eŋkilum* suffixes. The concession clause is similar to the conditional clause; the difference is present only in the level of probability. The concession form is the combination of conditional form + *-um*. In the concessive clause, the speaker is sure of the action whereas, for conditional clause, the speaker is unsure of the action.

i) Verb stem + *-ālum*

(37) *saṁsthānaṁ* *paṟaññ-ālum* *kēndraṁ* *nammaḷe* *anukūlikk-illa*

state tell-CONC centre us favor-NEG

‘Even if the state tells, the centre won’t favor us.’

ii) Verb stem + -eŋkilum

(38) *saṁsthānaṁ* *paṟaññ-eŋkilum* *kēndraṁ* *nammaḷe* *anukūlicc-illa*

state tell-CONC centre us favor-PAST-NEG

‘Even though the state told, the centre didn’t favor us.’

The negative form of the concessive marker is formed by adding a negative marker *-illa* to the *-eŋkilum* suffix.

(39) *avan* *vann-illeŋkilum* *nammaḷ* *kāḷi* *jayikkum*

he come-NEG.CONC we game win

‘Even if he doesn’t come, we will win the game.’

3.1.2.2.4. Relative Participle

The relative participle is realized in three tense forms: past, present and future. The negative sense of the relative participle is not marked for tense.

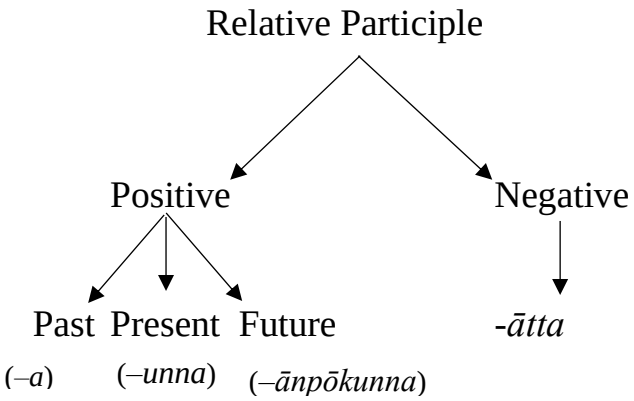


Fig.6: Relative participle in Malayalam

i) The suffix–*a* is used as a past relative participle form

(40) *viśvasi-cca* *kuṭṭi*
believe-PAST.RP boy
'The boy who belived.'

ii) The suffix–*unna* is used as a present relative participle form

(41) *viśvasi-kkunna* *kuṭṭi*
believe-PRES.RP boy
'The boy who is believing.'

iii) The suffix *-ānpōkunna* is used as a future relative participle form

(42) *viśvasi-kkānpōkunna* *kuṭṭi*
believe-FUT.RP boy
'The boy who will believe.'

iv) To express obligation in past tense form.

(43) *nammaḷ* *vijayi-kkēṇṭiyirunna* *tiraññeṭupp*
we win-DEB.PAST.RP election
'Election that we should have won.'

v) To express obligation, verb stem + relative participle form *-ēṇṭa*

(44) *nammaḷ* *vijayikk-ēṇṭa* *tiraññeṭupp*
we win-DEB-RP election
'Election that we should win.'

vi) Negative relative participle, verb stem + *-ātta*

(45) *uttaraṁ* *paṟay-ātta* *kuṭṭi*
answer say-NEG-RP child
'The child who doesn't / don't / won't answer.'

3.1.2.2.5. Adverbial Participle

The adverbial participle is realized with a past tense morpheme *-i* or *–ŭ* for class I and class II verbs respectively (Asher, 1997). Though it carries a tense marker, the tense is expressed by the verb in the main clause.

(46) *avan* *uttaraṁ* *paṟa-ññŭ* *naṭann-u*
he answer say-AP walk-PAST
'He walked answering.'

i) *– āte* suffix is attached to a verb stem to produce negative participle form

(47) *uttaraṁ* *paṟay-āte* *avan* *naṭann-u*
answer say-NEG-AP he walk-PAST
'He walked without answering.'

ii) Adverbial participle marked for aspect with the suffix *-āññŭ +iṭṭŭ*

(48) *avan* *eḻut-āññiṭṭŭ* *ñān* *nirbandhicc-u*
he write-NEG.AP I compel-PAST
'I compelled as he didn't write.'

3.1.2.3. Multiple Negation

Asher (1997) observes four ways in which multiple negations occurs in a sentence. In the case of double negation, the overall effect of negation draws a positive sense.

i) Finite verb + *-āt* which can accommodate negative marker *-illa*

(49) *avan* *var-āt-irunn-illa*
he come- NEG-PAST-NEG
Lit. 'He didn't not come' i.e. 'He came.'

(50) *kōṭati* *śikṣikk-āt-irikk-illa*
court punish-NEG-FUT-NEG
Lit. 'The court won't not be punishing.' i.e. 'The court will punish.'

ii) Quotative Participle + *-illa*, after negation

(51) *avan* *var-illa* *enn-illa*
he come.FUT-NEG QP-NEG
'It is not the case that he won't come.'

(52) *kōṭati* *śikṣikk-illa* *enn-illa*
court punish.FUT-NEG QP-NEG
'It is not the case that the court will not punish.'

iii) Cleft constructions where negative marker in the nominalized verb form and negative marker *-alla* attached to the other constituent.

(53) *avan* *alla* *pōk-āt-tatŭ*
He be-NEG go-NEG-NOML
'He is not the one who doesn't go.'

(54) *kōṭati* *alla* *śikṣikk-āt-tatŭ*
court be-NEG punish-NEG-NOML
'The court is not the one that doesn't punish.'

iv) Infinitive marked subordinate clause and negative marker *-alla* attached to the other constituent.

(55) *avanŭ* *pōk-āt-irikk-ān* *ka̲liy-illa*
he.DAT go-NEG-be-INF can-NEG
Lit. 'He can't not go' i.e. 'He has to go.'

(56) *kōṭati-kkŭ* *śikṣikk- āt-irikk-ān* *ka̲liy-illa*
court-DAT punish-NEG-INF can-NEG
Lit. 'The court cannot not punish' i.e. 'The court has to punish.'

3.1.3. Textual Feature

Textual features are surface-level cues that often suggest in understanding subjective tests. E.g. Punctuation marks. Earlier researches on subjectivity identification have referred punctuation marks and sentence position features for the subjectivity classification study (Wiebe et al., 1999) whereas the recent researches give more weight on lexical features than textual features (Wiebe et al., 2005). It is unclear whether the textual features prove useful in classification and to a limited extent question mark or exclamation mark may serve the purpose of extracting the expression in the sentence. Apparently, exclamation marks used to

emphasize the content or question marks expressing doubts, queries or irony may assist in classifying the sentence as subjective or objective.

3.1.4. POS Feature

Most of the NLP tasks require careful information of parts of speech (POS). Wilson et al. (2005) in the article 'recognizing contextual polarity in phrase-level sentiment analysis' employs parts of speech information as a feature in opinion classification. POS helps to understand the structure of any language. This research uses the adjective, adverb and verb combinations to classify the opinion where verb of the main clause carries the opinion and adjectives and adverbs express the quality of the attitude in the events. To count on each word in the sentence, it requires a tagger. The task of assigning each word of a corpus with appropriate parts of speech tag in the context of appearance is called Parts of Speech tagger.

(57) *oru* *bhāṣa*

one language

POS Tagged example is

1 *oru* QTC

2 *bhāṣa* NN

Thus, this study uses a combination of lexical features, functional features, textual features and POS features to identify the opinion of a sentence.

Chapter 4

Implementation and Evaluation

4.0. Introduction

The feature-based model described in the previous chapter is utilized to build an automated opinion analyzer on the Python platform. This chapter mainly includes the algorithm that explains techniques used in the model and the procedures performed is illustrated through a flow chart. Following, it discusses the challenges and limitations of the model. Testing the performance is crucial for the built model as it predicts the goodness of the model. The system performance is estimated by calculating the precision, recall, and the F-score This chapter in overall, discusses the implementation of the feature-based model and its evaluation.

4.1. Corpus, Preprocessing and Tokenization

Corpus, a large systematic collection of naturally-occurring texts stored in a machine-readable form (Meyer, 2002), is designed to represent the textual domain that defines the linguistic patterns of a language. One of the main goals of quantitative analysis in linguistics is data reduction, which is studied to summarize trends, capture the common aspects of a set of observations such as the average, standard deviation, and correlations among variables (Johnson, 2008). This analysis can be observed and realized through corpus-based research that involves the compilation of texts from several sources. The issue of the representativeness of the corpus is very important in corpus building. In many studies, representativeness is directly related to the ability to generalize the results of corpus investigation.

The first task of this research is to collect appropriate political domain-specific Malayalam text. Web crawling is the method executed to raise up texts from an online newspaper. Texts were extracted from Malayalam online news website majorly from *SouthLive* and *deshabhimani* with the aid of Sketch Engine[10]: a corpus development and management tool. Corpus is built compiled of randomly sampled online newspaper articles and is not biased to any specific fields of sections in the newspaper. Unprocessed data is collected for the corpus with Unicode encoding and readable in text format. The raw text data extracted for the corpus are processed to discard noise in the data. This is cleaned using a regular expression with the help of 're' package in NLTK library. White spaces are removed with the help of substitution method employing regular expression and 're' package. Verification of the accuracy of text corpus, especially in terms of spelling, is very important for the reliability and validity of the corpus. It is very important to guard against losing linguistic features during the storage and cleaning process. The preprocessed input text which is in Unicode format is converted to WX notation, a non-diacritic notation for processing and is again converted to Roman, the diacritic notation for the display. The input text is tokenized using the split command.

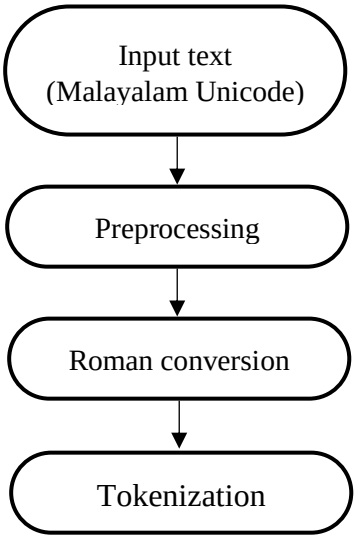


Fig.7: Flowchart for tokenization

[10]Sketch engine is a corpus tool to create and manage text corpus. https://www.sketchengine.eu.

4.2. Algorithm

The algorithm for building opinion analysis is explained below.

1. Input document
2. Sentence split
3. Roman conversion and tokenize
4. ***Open*** built lexicon (containing lists of positive and negative verbs, adjectives, adverbs after stemming)
5. ***Regex*** Remove exceptional words (to reduce error)
6. ***Re.match***POS (verbs, adjectives and adverbs)

7. ***If*** negative suffix match with any word ***&&if*** word == positive
8. ***Then*** count == Neg_verb+1
9. ***else:***
10. ***If*** negative suffix match with any word ***&&if*** word == negative
11. ***Then*** count ==Pos_verb+1
12. ***If*** double negative suffix match with any word ***&&if*** word == positive
13. ***Then*** count == Pos_verb+1
14. ***else:***
15. ***If*** double negative suffix match with any word ***&&if*** word == negative
16. ***Then*** count ==Neg_verb+1
17. ***Re.match*** with lexical feature
18. ***If*** word == positive verb
19. ***Then*** count == pos_verb+1
20. ***else if*** word == Negative verb
21. ***Then*** count == Neg_verb+1
22. ***If*** word ==positive adjective
23. ***Then*** count == Pos_Adj+1

24. ***else if*** word == negative adjective

25. ***Then*** count == neg_adj+1

26. ***If*** pos_verb count > 0 ***&&*** other counts 0

27. ***Print*** ‘positive’

28. ***If*** neg_verb count > 0 ***&&*** other counts 0

29. ***Print*** ‘negative’

30. ***If*** positive count == 0 ***&&*** negative count 0

31. ***Print*** ‘neutral’

32. ***If*** neg_adj == 1 ***&&*** neg_verb ==1

33. ***Print*** ‘positive’

34. ***else if*** pos_adj == 1 ***&&*** neg_verb ==1

35. ***Print*** ‘negative’

36. ***else if*** pos_adj == 1 ***&&*** pos_verb ==1

37. ***Print*** ‘positive’

38. ***else if*** neg_adj == 1 ***&&*** pos_verb ==1

39. ***Print*** ‘negative’

40. ***If*** adverb_count = 1 ***&&*** pos_verb ==1

41. ***Print***‘positive’

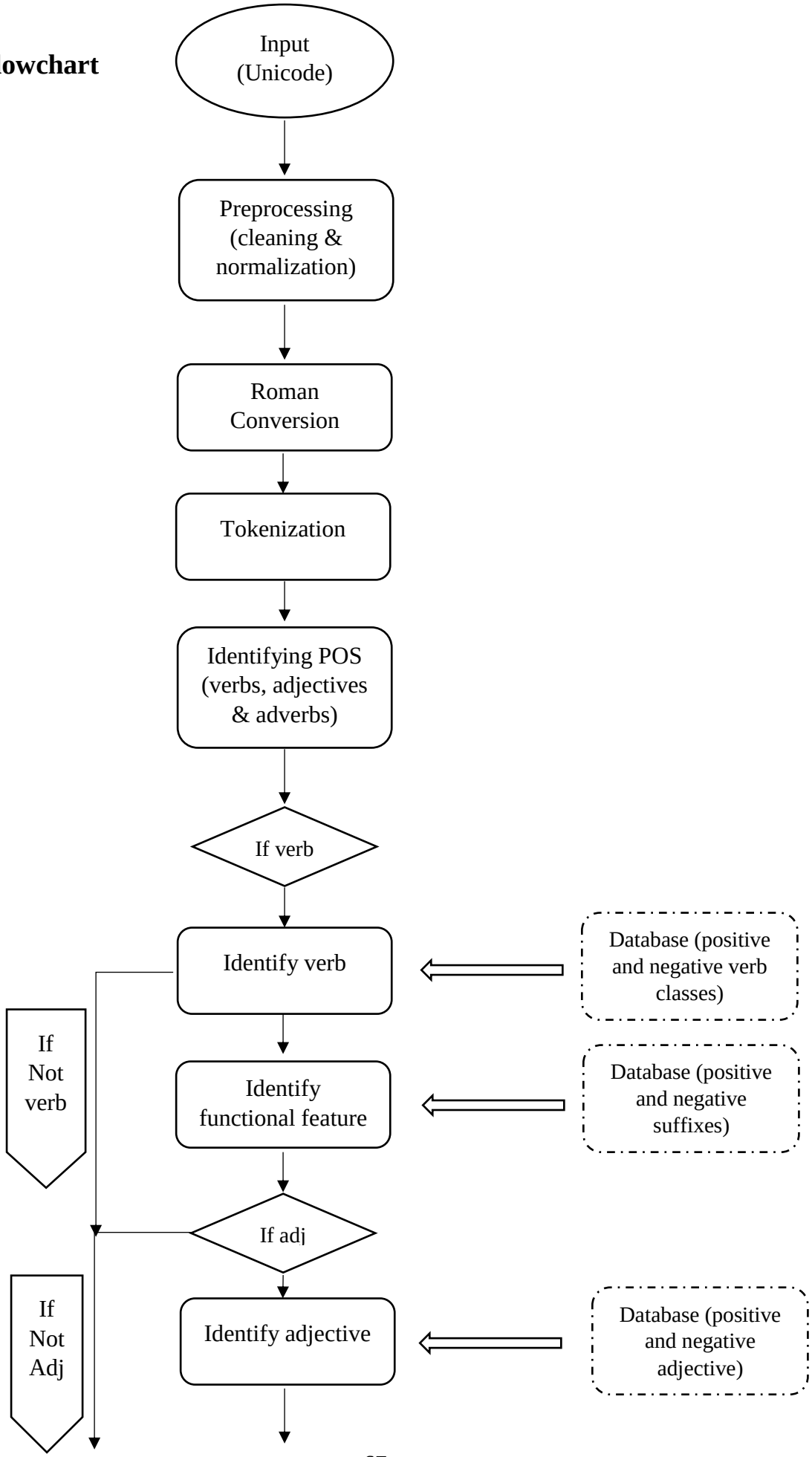
42. ***else if*** adverb_count =1 ***&&*** neg_verb ==1

43. ***Print***‘negative’

44. ***If*** positive count == negative count

45. ***Print***‘negative’

4.3. Flowchart



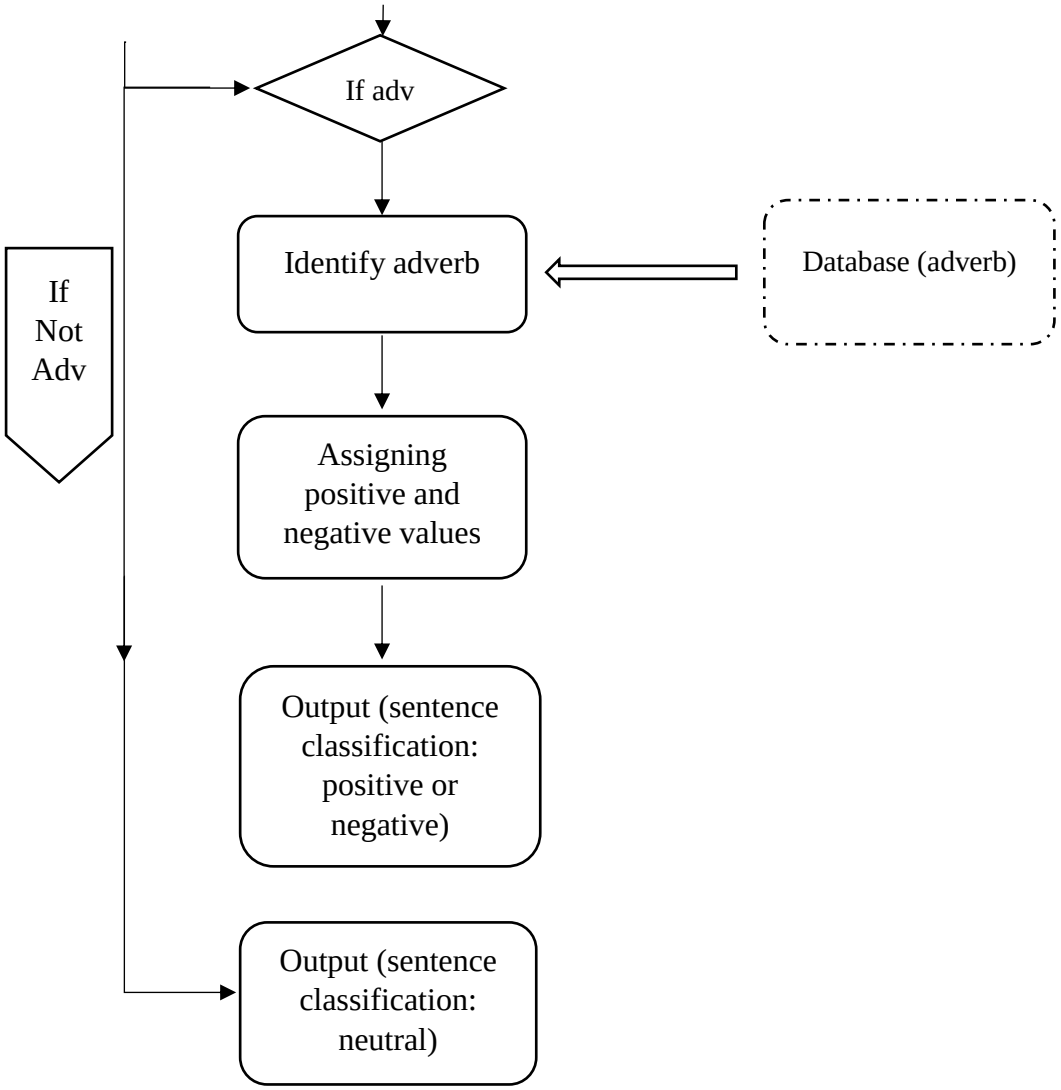


Fig.8: Flowchart for opinion analysis in Malayalam

4.4. Issues

4.4.1. Wrong POS Identification

Error in POS identification leads to error in opinion analysis. When POS tagger fails to correctly identify the lexical category, it leads to wrong identification.

(1)	*kollattŭ*	*āṇŭ*	*enṟe*	*vīṭŭ*
	Kollam-LOC	be-PRES	I-GEN	Home
	'My home is in Kollam.'			

(2)	*avan*	*koll-ān*	*pōy-i*
	he	kill-INFN	go-PAST
	'He went to kill.'		

In the sentence (1), ‘Kollam’ (place) is a noun whereas in the sentence (2), *koll-ān* is a verb which means ‘to kill’. The verb stem *koll* is stored in the negative verb class and hence it wrongly identifies kollam as a negative verb.

(3) *briṭṭīṣ* *gavaṇmen̲r* *vijaymalyaye* *indyakkŭ* *kaimāṟ-um*
British government Vijaymallya India hand over-FUT
‘British government will hand over Vijay Mallya to India.’

(4) *avan* *ā* *parīkṣaṇaṁ* *vijayicc-u*
he that experiment win-PAST
‘He won that experiment.’

In the sentence (3), ‘Vijay’ (name) is a noun whereas in the sentence (2), vijayicc-u is a verb which means ‘won’. The verb stem *vijay*is stored in the positive verb class and hence it wrongly identifies *vijay* as a positive verb in the sentence (1).

4.4.2. Ambiguous Words

The presence of homonyms in the input text creates ambiguity in the classification.

(5) *apēkṣikk-uka* -to plea, to apply
apply/plea-INFN

(6) *natakk-um* -will happen, will walk
walk/happen-FUT

4.4.3. Wrong Suffix Identification

Suffixes are found to be ambiguous in certain contexts which lead to wrong opinion analysis.

i) Verb stem+ Negative debitive marker -*aṇṭa*

(7) *enikkŭ* *itŭ* *vēṇṭa*
I-DAT this need.NEG
‘I don’t need this.’

ii) Verb stem + Relative participle + noun

(8) *ñān* *kāṇ-ēṇṭa* *kuṭṭi*
I see-DEB-RP boy
'The boy whom I have to see.'

In the sentence (7) and (8), the suffixes attached to the verb *vēṇ/vēṇuka* and *kāṇuka* are identical. In the case of sentence (7), the verb *vēṇ/vēṇuka* expresses its negation as *vēṇṭa* and the negative debitive marker stored in the suffix database is *-aṇṭa*. As a reason, the system fails to identify *-ēṇṭa* as a negative suffix and moreover *-ēṇṭa* is stored as a relative participle marker. Hence, the sentence (7) may not be classified as negative.

(9) *avan* *ārum* *aṟiy-āte* *ā* *bāg* *mōṣṭich-u*
he anyone know-NEG-AP that bag steal-PAST
'He stole that bag without anyone's notice.'

(10) *bampar* *ōphar* *kūṭāte* *niravadhi* *vismaya* *sammānaŋŋaḷum* *nēṭ- ū*
Bumper offer without more exciting prize get-IMP
'Get more exciting prizes in addition to the bumper offer.'

In the sentence (9), the adverbial participle negates the verb whereas in the sentence (10), *kūṭāte* which is a post position, meaning 'without' produce the sense of 'in addition' in the sentence. The reason behind this is that in Malayalam, one way to express 'addition' is, a noun in the nominative case + *kūṭāte.*

4.4.4. POS Feature

Lexical items derived from verbs: - Positive and negative verb classes are stored in the database after stemming. Thus, any lexical category derived from those verbs present in the database may result in the wrong stem identification and result in a false count.

E.g. *ākramikkuka* - ‘to attack’

ākramaṇaṁ - ‘attack’

pīdanaṁ - ‘harassment’

pīḍippikkuka - ‘to harass’

sahāyikkuka - ‘to help’

sahāyaṁ - ‘help’

4.5. Results and Error-Analysis

In this study thousand and ninety-one (1091) lexical items and functional items that express negation were assigned polarity labels to assist the automated polarity classification. From the built Malayalam corpus, five thousand and four sentences (5004) were extracted to evaluate the performance of the opinion analysis system. Sentences extracted are manually classified into positive, negative, and neutral (GOLD standard).

The input given in in Malayalam Unicode format is converted to WX notation for testing and output generates the sentence classification. Figure.9 shows an example of the generated output.

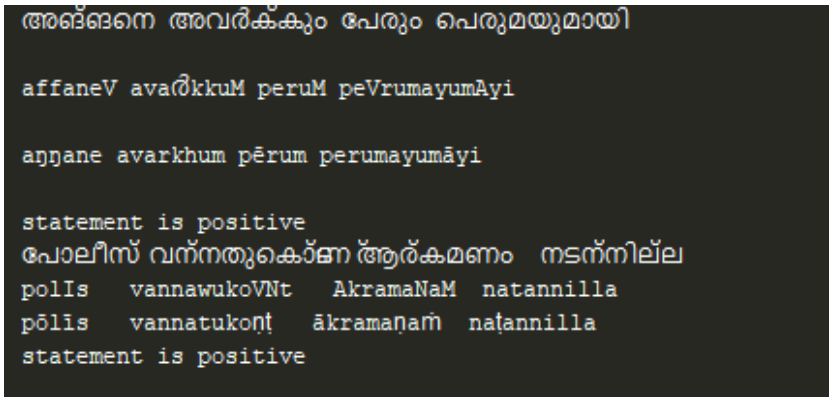


Fig.9: Sample output

Results of a few are discussed below:

(1) *aŋŋane* *avarkkum* *pēr-um* *perumay-um* *āyi*
Thus they name-COORD fame-COORD be-PAST
'Thus, they too became famous.'
OUTPUT: statement is positive

The sentence (1) is classified positive because of the presence of positive adjective '*peruma*' in the sentence.

(2) *iyāḷe* *kaṇṭ-āl* *ētorāḷum* *karut-uka* *it* *oru* *puruṣan* *āṇu-ennā*
he see-COND anyone think-INFN this a man be-QP
'If anyone sees him, they will think it is a man.'
OUTPUT: statement is neutral

In sentence (2), the verb '*karutuka'* (think) does not fall into positive or negative verb classes. So, the sentence is classified as neutral.

(3) *vr̥nda* *kārāṭṭ-in̲re* *abhiprāyaṁ* *a̲riy-ān* *tālparyapeṭ-unnu*
Brinda Karat-GEN opinion know-INFN interest-PRES
'I am interested to know the opinion of Brinda Karat.'
OUTPUT: statement is positive

Sentence (3) is classified positive because of the presence of the positive verb '*tālparyapeṭunnu'*.

(4) *ñaṅṅaḷ* *vanitā* *matilin-e* *etirkk-um*
we woman wall-ACC oppose-FUT
'We will oppose the women's wall.'
OUTPUT: statement is negative

Sentence (4) is classified negative because of the presence of the negative verb '*etirkk-um'* (will oppose).

(5) *ñaṅṅaḷ* *vanitā* *matilin-e* *anukūlikk-um*

we woman wall-ACC support-FUT

'We will support the women's wall.'

OUTPUT: statement is positive

Sentence (5) is classified positive because of the presence of the positive verb '*anukūlikk-um'* (will support).

Precision is used to measure the performance while recall is used to measure the positive rate and F-score, the average of precision and recall, measures the accuracy.

Precision is calculated as the ration of observed true positives to the total number of positives.

$$\text{Precision} = \frac{\text{TruePositive}}{\text{True Positive+False Postive}} * 100$$

Recall is calculated as the ratio of observed true positives to the total number of observations.

$$\text{Recall} = \frac{\text{TruePositive}}{\text{True Positive+False Negative}} * 100$$

F-score is measured as the harmonic average score of precision and recall.

$$\text{F-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision+Recall}}$$

For overall sentences,

True positive = 4643

True positive +False positive = 4990

False negative = 361

True positive + False negative = 5004

Precision = 93.04%, Recall = 92.78%, F-score = 1.7264/1.8582*100 = 92.90%

Positive, negative and neutral sentences from the overall output is analyzed and the results are given below.

(i) For Positive identification

True positive = 1628

True positive + false negative = 1743

True positive + false positive = 1647

Precision = 98.84%, Recall = 93.40%, F-score = 1.8463 /1.9224*100 = 96.04%

(ii) For Negative identification

True positive = 1623

True positive + false negative = 1784

True positive + false positive = 1926

Precision = 84.26%, Recall = 90.97%, F-score = 1.5330 /1.7523*100 = 87.48%

(iii) For neutral identification

True positive = 1392

True positive + false negative = 1477

True positive + false positive = 1421

Precision = 97.95%, Recall = 94.24%, F-score = 1.8461 /1.9219*100 = 96.05%

Classification			
Texts	Precision	Recall	F-score
Overall	93.04%	92.78%	92.90%
Positive	98.84%	93.40%	96.04%
Negative	84.26%	90.97%	87.48%
Neutral	97.95%	94.24%	96.05%

Table.6: Results

In consideration of five thousand and four (5004) sentences, four thousand six hundred and forty-three (4643) were correctly classified out of four thousand nine hundred and ninety (4990) sentences that provided output. Three hundred and sixty-one sentences (361) were incorrectly classified. For positive texts, out of one thousand seven hundred forty-three sentences (1743) one thousand six hundred and twenty-eight (1628) sentences were correctly classified. For negative texts, out of one thousand seven hundred and eighty-four (1784) sentences, one thousand six hundred twenty-three (1623) sentences were correctly classified. For neutral texts, out of one thousand four hundred seventy-seven (1477) sentences, one thousand three hundred ninety-two (1392) sentences were correctly classified.

Chapter 5

Conclusion

The emergence of new resources aids in promising advancement in this research through empirical study. Apparently, research in this field depicts the correlation between opinion and language use. This research experimented feature-based classification to analyze the opinion present in the Malayalam political texts. Utilizing four features namely lexical, functional, textual and POS features, this study developed a phase of not classifying the statements into objective or subjective instead classifying the statements as positive, negative or neutral. In the field of opinion analysis, one question to be answered is the choice of selection of an appropriate model for the opinion classification. This research tried to address the very question and reached a conclusion that it could be feature-based method which can provide better accuracy in classification. The model can be further improved by building a large opinion lexicon database and by adding other features that can address opinion expressions.

In this study, eighty-seven thousand three hundred and forty-seven (87347) Malayalam sentences with political content were selected from news web resources in the period between 1stAugust 2018 and 30th September 2018. From the collected corpus, five thousand and four sentences (5004) were manually termed or classified into negative, positive and neutral based on its content. Test on the overall classified data which was the subject to assess the accuracy resulted in producing a precision of93.04 percent, recall of 92.78 percent and f-score of 92.90 percent for the model. Test on the manually classified negative sentences resulted in measuring 84.26 percent for precision, 90.97 percent for recall and 87.48 percent for f-measure. Test on the manually classified positive sentences yielded a precision of 98.84 percent, recall of 93.40 percent and f-measure of 96.04 percent. Test on the manually

classified neutral sentences resulted in producing 97.95 percent precision, 94.24 percent recall and 96.05 percent f-measure. The low precision, recall and f-score for negative classified sentences is that the lexical and functional features built to classify negative texts are so powerful that it even classifies certain positive and neutral texts as negative texts. The major reason behind this phenomenon is the wrong stem identification either due to morphological derivation or due to the existence of some lexical item which exhibits similar kind of stem as of verbs. This system elicits better precision, recall and f-score on positive and neutral texts. This shows that the features used to classify positive or neutral texts have a great influence in predicting correctly. The predominance of negative opinion is very evident that the presence of any positive and negative lexical item in the same clause will result in a negative statement. The existence of double negatives in a clause will turn out to a positive meaning. Negative adjective and negative verb in a clause shall yield a positive statement and the same way the presence of two negative opinionated verbs in a clause will produce a positive meaning. Adverbs, though never engage in a meaning shift, deepens the degree of meaning in any statement.

As the results shown, the system can be further improved in improvising the negative and positive verb list, adjectives which contributes to the opinion analysis. Further the preprocessing modules require improvisation. Using Morphological analyzer can be further useful in identifying the suffixes which may prevent the wrong suffix identification. Similarly, the POS tagger requires improvisation to get better results in lexical category identification.

Furthermore, the same technique can be applied to analyze the opinion in a document. Analysis of sentences in a document led to a significant discovery that every document follows a specific pattern in establishing the premeditated opinion. If the content has to be on positive news, then the document begins on a positive note and the content may have a mix of

positive and neutral statements with least negative statements. In the case of negative opinionated documents, the highlighted opinion would be invariably negative with the least positive statements. Presence of positive statements in such documents is noticed if the content is a follow-up of any other previous stories and it is witnessed majorly not in the beginning statements. One major point is very coherent that the premeditated opinion is majorly emphasized.

Appendix I

Python script

```
def main ():
    file = open("D:/Input.txt","r",encoding='utf-8')
    lines = file.readlines()
    file.close()
    for lin in lines:
        print(lin)
        mal_wx = wxc.convert(lin)
        print(mal_wx)
        abb_dict = {'K': 'kh','kk': 'kk','G': 'gh','ff': 'ṅ','cc': 'ch','J':
'jh','F': 'ñ','t': 'ṭ','T': 'ṭṭ','d': 'ḍ','D': 'ḍḍ','N': 'ṇ','M': 'm','w':
't','W': 'th','x': 'd','X': 'dh','P': 'ph','B': 'bh','ൽ': 'l','ർ':
'r','ൻ':'ṇ','ൺ':'n','ൾ':'ḷ','S': 'ś','R': 'ṣ','nY': 'ṉ','lY': 'ḷ','lYY':
'ḻ','rY': 'ṟ','A': 'ā','I': 'ī','U': 'ū','q': 'r̥','eV': 'e','e': 'ē','E':
'ai','oV': 'o','o': 'ō','O': 'au','aM': 'aṁ','aH':'aḥ','f':'ŋ',}
        New_dic = re.compile("(%s)" % "|".join(map(re.escape, abb_dict.keys())))
        roman = New_dic.sub(lambda
mo:abb_dict.get(mo.group(1),mo.group(1)),mal_wx)
        print(roman)
        words = re.split (r'\s', mal_wx)
        a = 0,b = 0,c = 0,x = 0,y = 0,z = 0
        path = "C:/Users/faith/Music/M.Phil/pos^neg_review/opinion_lexicon/name"
        file_1 = open(os.path.join(path, "Negative_adj"), "r")
        file_2 = open(os.path.join(path, "Positive_adj"), "r")
        file_3 = open(os.path.join(path, "Negative_adverb"), "r")
        file_4 = open(os.path.join(path, "Positive_adverb"), "r")
        file_5 = open(os.path.join(path, "Negative_verb"), "r")
        file_6 = open(os.path.join(path, "Positive_verb"), "r")
        file_7 = open(os.path.join(path, "Positive_conjunct_verb"),"r")
        file_8 = open(os.path.join(path, "Negative_conjunct_verb"),"r")
        ex = 0
        while ex<len (words):
            exceptions = str(words[ex])
```

```
            if re.findall
(r'\bvaruwwAn|\bvaruwwan|\bmisPirYrY|\biruNta|\bmisorYi|\bwiNta|\bsawyan|\bwrillar
|\bpaScAwwala|\bvalYav|\bkurYavilaffAt|\bpaScAwwalamAkkAmeVnn\b|saMrakRikkappeVteN
ta|\bceVyyeNtaw|\bkaruwi|\bkaruwunnu|\burYappuvaruwwAn|\bkaNta|\bvaruwwan|\bpaScAw
walavuM|\bmisrYrYar|\bvillej|\bwutaffiyava|\bwutaffiya\b|\bjilla|\bvijayan|\bvijay
aṁ|\bvaruwiyil|\bnINta|\bwAmasiyAweV|\bkoVllaM|\bkaNakkileVtukkeNta',
exceptions):

                words.remove (exceptions)

                ex =ex-1,ex=ex+1

        h=0

        while h<len (words):

            exception = str(words[h])

            if re.findall (r'\w*Awir\w*illa\b', exception):

                    words.remove (exception)

                    z=z+1

                    file_1 = open(os.path.join(path, "Negative_adj"), "r")

                    file_3 = open(os.path.join(path, "Negative_adverb"), "r")

                    file_5 = open(os.path.join(path, "Negative_verb"), "r")

                    for line11 in file_1:

                        check6 = (len (line11)-1)

                        if re.match(exception[:check6], line11) and if (len
(line11)-2)<len (exception):

                            y=y+1,z=z-1

                    for line22 in file_3:

                        check66 = (len (line22)-1)

                        if re.match(exception[:check66], line22) and if (len
(line22)-2)<len (exception):

                            y=y+1,z=z-1

                    for line33 in file_5:

                        check666 = (len (line33)-1)

                        if re.match(exception[:check666], line33) and if (len
(line33)-2)<len (exception):

                            y=y+1,z=z-1,h=h-1,h=h+1

        f=0

        while f<len (words):

            exception1 = str(words[f])

            if re.findall (r'\w*illAwwaw', exception1):

                z=z+1

                words.remove (exception1)
```

```
                file_1 = open(os.path.join(path, "Negative_adj"), "r")), "r")
                file_3 = open(os.path.join(path, "Negative_adverb"), "r")
                file_5 = open(os.path.join(path, "Negative_verb"), "r")
                for line7 in file_1:
                    check7 = (len (line7)-1)
                    if re.match(exception1[:check7], line7) and if (len (line7)-
2)<len (exception1):
                for line77 in file_3:
                    check77 = (len (line77)-1)
                    if re.match(exception1[:check77], line77) and if (len
(line77)-2)<len (exception1):
                        y=y+1,z=z-1
                for line777 in file_5:
                    check777 = (len (line777)-1)
                    if re.match(exception1[:check777], line777) and if (len
(line777)-2)<len (exception1):
                        y=y+1,z=z-1,f=f-1,f=f+1
        mm=0
        while mm<len (words):
            exception2 = str(words[mm])
            if re.findall (r'\w*aruwAyirunnu\b', exception2):
                words.remove (exception2)
                z=z+1,mm =mm-1,mm=mm+1
        exp=0
        while exp<len (words):
            exception2 = str(words[exp])
            if re.findall
(r'\bnaRtaparihAraM\b|\bpravarwwikkeNta|\bupayogappeVtuwweNtawuNt|\bnalla\b|\bnall
aw\b|\bnallapilYlYa\b|\bkurYrYamarYrYaw\w*', exception2):
                words.remove (exception2)
                z=z+1,exp =exp-1,exp=exp+1
        e=0
        while e<len (words):
            exception3 = str(words[e])
            if re.findall
(r'\w*illAyirunneVfkil|\w*illeVfkiluM|\w*alleVfkiluM|\w*allAyirunneVfkil',
exception3):
                words.remove (exception3)
```

```
                z=z+1,e=e-1,e=e+1
        d=0
        while d<len (words):
            exception4 = str(words[d])
            if re.findall (r'\w*irunneVfkil|\beVwireV\b', exception4):
                words.remove (exception4)
                y=y+1,d =d-1,d=d+1


        m=0
        while m<len (words):
            exception5 = str(words[m])
            if re.findall
(r'\w*illAwe|\w*illAwwa|\w*illa|\w*alla|\w*aruw|\w*Awir|\w*Awwa|\w*Nte|\w*illeVfki
l|\w*eVwireV\b|\w*AweV', exception5):
                words.remove (exception5)
                y=y+1
                file_1 = open(os.path.join(path, "Negative_adj"), "r")
                file_3 = open(os.path.join(path, "Negative_adverb"), "r")
                file_5 = open(os.path.join(path, "Negative_verb"), "r")
                file_8 = open(os.path.join(path, "Negative_conjunct_verb"),"r")


                for line8 in file_1:
                        check8 = (len (line8)-1)
                        if re.match(exception5[:check8], line8) and if (len
(line8)-2)<len (exception5):
                            z=z+1,y=y-1
                for line88 in file_3:
                        check88 = (len (line88)-1)
                        if re.match(exception5[:check88], line88) and if (len
(line88)-2)<len (exception5):
                            z=z+1,y=y-1
                for line888 in file_5:
                        check888 = (len (line888)-1)
                        if re.match(exception5[:check888], line888) and if (len
(line888)-2)<len (exception5):
                            z=z+1,y=y-1
                for line8888 in file_8:
```

```
                        check8888 = (len (line8888)-1)
                        if re.match(exception5[:check8888], line8888) and if (len
(line8888)-2)<len (exception5):
                            z=z+1,y=y-1,m=m-1,m=m+1
        file_1 = open(os.path.join(path, "Negative_adj"), "r")
        file_2 = open(os.path.join(path, "Positive_adj"), "r")
        file_3 = open(os.path.join(path, "Negative_adverb"), "r")
        file_4 = open(os.path.join(path, "Positive_adverb"), "r")
        file_5 = open(os.path.join(path, "Negative_verb"), "r")
        file_6 = open(os.path.join(path, "Positive_verb"), "r")
        file_7 = open(os.path.join(path, "Positive_conjunct_verb"),"r")
        file_8 = open(os.path.join(path, "Negative_conjunct_verb"),"r")

        for line46 in file_8:
            nm=0
            while nm< len (words):
                exception46 = str(words[nm])
                check4 = (len (line46)-1)
                if re.match(words [nm][:check4], line46) and if (len (line46)-
2)<len (words[nm]):
                    words.remove (exception46)
                    y=y+1,nm=nm+1
        for line59 in file_7:
            po=0
            while po< len (words):
                exception59 = str(words[po])
                check5 = (len (line59)-1)
                if re.match(words [po][:check5], line59) and if (len (line59)-
2)<len (words[po]):
                    words.remove (exception59)
                    z=z+1,po=po+1
        for lines in file_1:
            k=0
            while k< len (words):
                check0 = (len (lines)-1)
                if re.match(words [k][:check0], lines) and if (len (lines)-2)<len
(words[k]):
```

```
                    a=a+1,k=k+1
        for line1 in file_2:
            i=0
            while i< len (words):
                check1 = (len (line1)-1)
                if re.match(words [i][:check1], line1) and if (len (line1)-2)<len
(words[i]):
                    b=b+1,i=i+1
        for line2 in file_3:
            j=0
            while j< len (words):
                check2 = (len (line2)-1)
                if re.match(words [j][:check2], line2) and if (len (line2)-2)<len
(words[j]):
                    c=c+1,j=j+1
        for line3 in file_4:
            l=0
            while l< len (words):
                check3 = (len (line3)-1)
                if re.match(words [l][:check3], line3) and if (len (line3)-2)<len
(words[l]):
                    x=x+1,l=l+1
        for line4 in file_5:
            n=0
            while n< len (words):
                check4 = (len (line4)-1)
                if re.match(words [n][:check4], line4) and if (len (line4)-2)<len
(words[n]):
                    y=y+1,n=n+1
        for line5 in file_6:
            o=0
            while o< len (words):
                check5 = (len (line5)-1)
                if re.match(words [o][:check5], line5) and if (len (line5)-2)<len
(words[o]):
                    z=z+1,o=o+1
```

```
if y+z+x==4 and y==1 and a==0 and b ==0 and c==0 and x==1 and z==2:
        print ("statement is positive")
if y+z+x==6 and y==1 and a==0 and b ==0 and c==0 and x==1 and z==4:
        print ("statement is positive")
if y+z==7 and y==1 and a==0 and b ==0 and c==0 and x==0 and z==6:
        print ("statement is positive")
if z+x==3 and y==0 and a==0 and b ==0 and c==0 and x==1 and z==2:
        print ("statement is positive")
if y+a==2 and y==1 and a== 1 and b ==0 and c==0 and x==0 and z==0:
        print ("statement is positive")
if y+a==2 and y==1 and a== 1 and b ==0 and c==0 and x==0 and z==2:
        print ("statement is positive")
if y+a==2 and y==1 and a== 1 and b ==0 and c==0 and x==0 and z==1:
        print ("statement is positive")
if y+a==3 and y==1 and a== 2 and b ==0 and c==0 and x==0 and z==0:
        print ("statement is NEGATIVE")
if b+x+z==3 and y==0 and a== 0 and b ==1 and c==0 and x==1 and z==1:
        print ("statement is positive")
if b+x==2 and y==0 and a== 0 and b ==1 and c==0 and x==1 and z==0:
        print ("statement is positive")
if b+z==2 and y==0 and a== 0 and b ==1 and c==0 and x==0 and z==1:
        print ("statement is positive")
if x+z==2 and y==0 and a== 0 and b ==0 and c==0 and x==1 and z==1:
        print ("statement is positive")
if x+y==3 and y==2 and a== 0 and b ==0 and c==0 and x==1 and z==0:
        print ("statement is NEGATIVE")
if x+y==5 and y==4 and a== 0 and b ==0 and c==0 and x==1 and z==0:
        print ("statement is NEGATIVE")
if z+a==2 and z==1 and a== 1 and b ==0 and c==0 and x==0 and y==0:
        print ("statement is negative")
if y+a==3 and z==0 and a== 1 and b ==0 and c==0 and x==0 and y==2:
        print ("statement is negative")
if y+c==2 and y==1 and c== 1 and b ==0 and a==0 and x==0 and z==0:
        print ("statement is negative")
if b+y+z==5 and y==1 and c== 0 and b ==1 and a==0 and x==0 and z==3:
```

```
        print ("statement is negative")
if  y==0 and c== 1 and b ==0 and a==0 and x==0 and z==0:
        print ("statement is negative")
if  y==4 and c== 0 and b ==0 and a==0 and x==0 and z==0:
        print ("statement is negative")
if  y==0 and c== 0 and b ==0 and a==1 and x==0 and z==0:
        print ("statement is negative")
if y+c+z==3 and y==1 and c== 1 and b ==0 and a==0 and x==0 and z==1:
        print ("statement is negative")
if z==1 and y==0 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is positive")
if z==2 and y==0 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is positive")
if z==3 and y==0 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is positive")
if z==4 and y==0 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is positive")
if z==5 and y==0 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is positive")
if z==0 and y==5 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y==4 and z==3 and y==1 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y==1 and z==2 and y==-1 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y==6 and z==1 and y==5 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y+b==4 and z==1 and y==2 and a== 0 and b ==1 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y==5 and z==1 and y==4 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y==5 and z==3 and y==2 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y==5 and z==4 and y==1 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
```

```
if a+y==4 and z==0 and y==3 and a== 1 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if a+y+z==5 and z==2 and y==2 and a== 1 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y==5 and z==2 and y==3 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if z+y==4 and z==2 and y==2 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is NEGATIVE")
if y==1 and z==0 and a== 0 and b ==0 and c==0 and x==0:
        print ("statement is negative")
if y+z==2 and y==1 and z==1 and a== 0 and b ==0 and c==0 and x==0:
        print("statement is negative")
if y+z==3 and y==1 and z==2 and a== 0 and b ==0 and c==0 and x==0:
        print("statement is negative")
if x+y+z==3 and y==1 and z==1 and a== 0 and b ==0 and c==0 and x==1:
        print("statement is negative")
if x+y+z==6 and y==1 and z==3 and a== 0 and b ==0 and c==0 and x==2:
        print("statement is POSITIVE")
if x+y==2 and y==1 and z==0 and a== 0 and b ==0 and c==0 and x==1:
        print("statement is negative")
if y==0 and z==0 and a== 0 and b ==0 and c==0 and x==1:
        print("statement is positive")
if y==0 and z==0 and a== 0 and b ==1 and c==0 and x==0:
        print("statement is positive")
if y+b+z==6 and z==4 and a== 0 and b ==1 and c==0 and x==0 and y==1:
        print("statement is positive")
if b+z==3 and z==2 and a== 0 and b ==1 and c==0 and x==0 and y==0:
        print("statement is positive")
if b+z==4 and z==3 and a== 0 and b ==1 and c==0 and x==0 and y==0:
        print("statement is positive")
if b+z==4 and z==1 and a== 0 and b ==3 and c==0 and x==0 and y==0:
        print("statement is positive")
if x+z==4 and z==3 and a== 0 and b ==0 and c==0 and x==1 and y==0:
        print("statement is positive")
if y+z==4 and y==3 and z==1 and a== 0 and b ==0 and c==0 and x==0:
```

```
                print("statement is negative")
        if y+z==6 and y==3 and z==3 and a== 0 and b ==0 and c==0 and x==0:
                print("statement is negative")
        if y+z==8 and y==4 and z==4 and a== 0 and b ==0 and c==0 and x==0:
                print("statement is negative")
        if y+z==6 and y==2 and z==4 and a== 0 and b ==0 and c==0 and x==0:
                print("statement is negative")
        if z==0 and y==0 and a== 0 and b ==0 and c==0 and x==0:
                print("statement is neutral")
        if y+z+a+b+c+x==2 and y==2 and z==0 and a== 0 and b ==0 and c==0 and x==0:
                print("statement is positive")
        if y==3 and z==0 and a== 0 and b ==0 and c==0 and x==0:
                print("statement is negative")
        if y+z==3 and y==2 and z==1 and a== 0 and b ==0 and c==0 and x==0:
                print("statement is negative")
        if y+z==6 and y==4 and z==2 and a== 0 and b ==0 and c==0 and x==0:
                print("statement is negative")

main()
```

Appendix II

List of lexical items used to build opinion lexicon for the study

Positive Adjective	
nalla	nayanasuBagamAya
mikacca	AnanxakaramAya
sampUrNNa	SreRtamAya
kUtuwa	oVnnAnwaramAya
valiya	mqxuvAya
paryApwamAya	komalYamAy
samqxXamAya	nerwwa
sAmarwWyamulYlYa	suBagamAya
kalYYivulYlYa	sUkRmamAya
AsvAxyamAya	pariRkqwamAya
wqpwikaramAya	SuxXamAya
vExagxXyamulYlYa	wIkRaNamAya
uwwamamAya	rasakaramAya
anuyojyamAya	BaMgiyAy
BakwiyulYlYa	manojFamAya
guNakaramAya	sunxaramAya
BaxramAya	nermmayAya
uwakunna	uwkqRtamAya
saxguNamulYlYa	nermmayAyi
guNavawwAya	nirxxoRamAya
koVlYlYAvunna	anyUnamAya
viSuxXiyulYlYa	kanaMkurYaFFa
nanmayulYlYa	guNamulYlYa
manoharamAya	cAwuryawwoteV
alYYakulYlYa	viSiRTamAya
ramaNIyamAya	SreRTamAya
sanxaryamulYlYa	nilavAramulYlYa
hqxyamAya	menmayulYlYa
karNNasuKapraxamAya	uwkqRtamAya

Negative Adjective
neVgarYrYIv
cIwwa
weVrYrYAya
vargIya
ApawkaramAy
krUra
piriy
vyAja
jIrNNicca
wAnwonniyAya
moSamAya
saxAcAraviruxXamAya
ayogyamAya
cIwwayAya
koVlYlYaruwAwwa
ketulYlYa
prayojanaSUnyamAya
xarBAgyakaramAya
vexanAkaramAya
sanwoRakaramallAwwa

Positive Verb			
SraxXapiti	valYarww	oVnnAyi	prakatippi
xruwagawi	uNarww	wilYakkawwilAN	nilani
anveR	jayi	puwukka	virYrYu
BAgya	munnottA	navIkari	eVtuwwu
sammawi	samIkqwam	viniyogi	prerippi
aMgIkari	Ayi	upayogappeV	paxXawiyi
yoji	utamAyA	upayogi	minusappeV
poVruwwappeV	utamAyA	manasilA	pravaci
katAkRi	puRpi	prayojanamA	munkUtti
pUrwwiyA	SoBi	Arjji	wayyA
saPalamA	aBivqxX	karasWamA	paricayappeV
nirvvahi	nirNNayi	weVreVFFeVtuww	avawari
SupArSa	sanxarSi	svAyawwamA	sammAni
paripUrNNamAkk	mohippi	sampAxi	awijIvi
grahikk	pAttilAkku	cittappeV	accaticc
sahqxa	pAlikk	arYivuNtA	natapatikk
sAxXyam	anusari	upaxeSi	natapatiyeV
vIkRaNa	AGoRi	natappilvaru	sakary
rucikara	prakIrwwi	uwkkqRtamA	sajjIkari
upaxeRtA	viSeRi	veVlYippeV	vAff
janaprI	koVNtAt	prawyakRappeV	purogami
pariSram	ArppuvilYi	veVlYivA	prawikari
rakRApravarwwan	AhlAxi	rakRikk	punaHkramIkari
praswAvana	prasixX	rakRicc	eVwwi
sWAna	viSaxamAkk	rakRapeV	sWirIkari
karakayar	alYYukkakarYrY	anuRTi	Agrahi
saviReSa	vqwwiyAkk	kAwwusUkRi	SuxXIkari
svapna	SuciyAkk	ummavay	sugamamA
prawipAxicc	SucIkari	cuMbi	lalYiwamA
vyakwiwva	weVlYi	manassilA	punaHsWApi

Sramicc	oVnniccu	nIwikari	AsUwraNaMceV
anuyojya	saMBari	spaRtamAkk	oVwwunokk
anuvaxi	svarUpi	xqDIkari	parIkRi
garavamarhikk	saMyojippi	XAraNayA	erYrYuvAff
pinwuNa	kUticce	XAraNayuNtA	ayacc
eVntri	AraMBi	weVraFFeVtuww	elpi
natappAkk	wutaff	weVraFFeVtukk	laByamA
xuriwASvAsa	praSaMsi	ulYkkoV	punaHsaMGati
parihasicc	veVlYippeVtt	aXikArappeVtuww	alYYiccupaNi
koVtukk	veVlYippeVtuww	kalYYivu	mataffiv
koVtuww	yoji	SakwamAkk	matakkiww
prawIkRa	uNtAkk	balappeVtuww	parAmarSi
rakRappeV	rUpIkari	XEryappeV	weVrYrYuwiruww
rUpIkar	ekopippi	prAwsAhAppi	sUcippi
paripUrNNam	SraxXi	sAkRyappeV	uxAhari
PrI	mulYYumippi	urYappukoVtu	reKappeVtuww
AnukUlya	sanXi	rakRAnatapatikalYeVtu	punarAviRkkari
varxXi	urYappAkk	salkari	anusmari
saMBarikk	sWApi	mohippi	ormmappeV
net	wIrccappeVtuww	wulyamA	wiriccukoVtu
nirYaverYrY	cerYww	pariSoXi	viSvasi
nirYaverY	Gatippi	eVlYuppam	ASvasi
sampAxikk	kUttiyojippi	pariSIli	SAnwamA
uwwejippi	kIlYYatakk	aByasi	ceVrYuppamA
uwsAhappeVt	saMrakRi	paricayi	suKappeV
Sari	paripAli	vAwsalya	natannu
pukalYYww	nirmmi	karuNa	vyakwamA
Axari	pUrNNamA	anumoxi	BexappeVtuww
nataww	ulYkoVNt	XanasahAyaMceV	wiraFFeVtuww
svIkari	pariNami	ayay	nItt
pUji	ekopi	wutakkami	SuSrURi

ArAXi	wiruww	anunaya	mawsari
munnerY	pracarippi	nirvahi	purYawwuvit
nerawweVy	pakRaww	saPalIkari	oVrumikk
uyarwwa	parihari	wqpwippeV	parihAraM
maXura	parigaNi	alafkari	iRta
awBuwa	janippi	eVwwiccukoVtu	sAXyam
omanapperilarYiyappeV	ASleRi	unnawivaru	ekIkaraNam
cavittupati	omani	yogyamA	poVtipoVti
wamASa	lAlYikk	yogyawanet	nirmicc
upakari	lAlYicc	uwsAhi	hqxyam
Agraham	paripoRippi	aBivqxXi	praKyAp
AgrahaM	kaNNaFci	puRtivay	janAXipawy
prayojanappeV	wIrumAn	munnileVww	vanvijaya
parasyamAkk	ceVrYuwwuni	viSuxX	lId
prakASi	prawiroXi	nanma	raNtAmawweV
uyarnnu	pAliccu	vijayicc	raNtAM
svanwamA	niyogi	vijayikk	kyApeVyin
oPar	Ananxi	Bexagawi	sammawa
Aktivekt	sanwoRi	ulpAxippi	sWirIkaraNa
ilYav	sanwoRa	kramIkari	ulYppeVtuww
AnUkUlya	ullasi	niyanwri	keVttippatu
karuwwAN	AhlAxi	valYYikANi	ceVrYukku
aBimAn	boxXyappeVtuw	ekIkari	ceVrYuww
nirxeSi	boXyappeVtt	cerccay	AvaSya
peVruma	niyami	sahAyikk	vijayam
anumox	kaNtupiti	sahAyicc	natapatiyuN
vEki	SraxXa	kUttikkoVNtuv	AveSa
guNaxoRi	niScayi	bahumAn	mawipp
anuSAsi	vikasippi	sawkari	sneha
urYappikk	SakwippeV	nannA	kaNteVwwi
bahumAn	poRippi	cirikk	vikasana

awyAvaS	nirxxeSi	ciricc	anumawi
urYappicc	valYYikAtt	prasaMgi	pariRkAraff
samAhari	kaNtukitt	mocippi	karSanamAkk
SeKari	carccaceV	svawanwramA	upakArapraxa
wuNay	praxarSippi	iRtappeV	puraskAram
sahakari	kANikk	iRtamA	arhi
nalk	kANiccu	Sravi	varXiPP
kUtticcer	kramappeVtuww	prowsAhi	sawya
oVnnAkk	viwaraNaMceV	pracoxi	valYYiwwiri
awBuwappeV	vIwiccukoVtu	wAwparyam	pitiyil
meVccappeV	laBi	AkarRi	lakRya
pariRkari	eVlYYunnelppi	kRaNi	suvyakwam
aBivqxXamAkk	svAgawaMceV	kurYavunikaww	nanxi
viSaxIkari	vilamawi	vismayi	sampannan
vivari	anukUlicc	maXuri	surakR
jIvippi	anukUlikk	katamappeVtuww	poRaka
prawIkRikk	sajIvamA	sAXi	saviS
aBinanxi	wiriccarYi	natappilA	ulppAx
Asvaxi	arYi	samarppi	sqRti
anumaXi	grahi	rUpappeV	anivArya
uwwejippi	cumawalaye	janiccu	paricayappeV
sevi	wutakkaM	varakk	avawari
kEvari	vyApippikk	varacc	sammAni
atuppi	Barame	pafkeVtu	awijIvi

Negative Verb			
allA	rakwarURi	pinmArY	katam
illA	wakarkk	anaXikqwa	moSamA
koVlla	balAwsaMga	parAwi	moSaM
oVlYYivAkk	wakarww	weVrYrYixXari	prawisanXi

prakopi	purYawwAkki	arocaka	xusWiwi
anwya	bAXyawa	aXArmika	oVliccu
parAjaya	niroX	laMGana	arAjakawva
AkramaNa	cinwAkkulYYapp	viruxXam	marikk
kasrYrYadi	kaRtappAt	apamAnakar	maricc
BIRaNi	kaluRiwamA	watay	vIlYYca
prakoppicc	rUkRa	wataFF	polIsif
hani	prawiReX	aniRta	valiccilYYacc
patiyirYaff	Binnippi	valYYipilYYa	prawikkUttil
itiv	atakkiBari	pilYYa	koVla
eVwir	kuwwerYrY	kulsiwa	saspeVnd
rAji	wiroXAn	gUDAlocana	wolppi
weVrYrY	xurupayoga	keVNi	kalarpp
peti	laMKi	SikR	pinwiripp
uffimaricc	rYaxxAkk	valYYakk	guruwarAvasWayil
xAruNAnwy	vettayAti	wakar	malinIkaraNa
porYal	Aropi	wall	maraNa
apamAna	awqpwi	vexanippi	hawaBAgya
xAruNa	aparyApw	praSana	malinIkari
safkIrNa	vissamaw	aviSvAsa	vAyumalinIkar
walYarn	lajjAkara	weVrYrYixX	ottisa
avagaNi	aparimiwa	SalyapeVtuww	guruwara
anAxar	muRicc	vilapi	viRamAlinya
kuprasixX	ApawGattaww	karay	mis
vaFcicc	katannupiti	karaFF	praSna
watasa	nilakk	mutakk	warkk
prawikAra	nilacc	awikram	pArSvaPala

nirasi	anAsW	cURaNa	prawyAGAwa
koVlYYiFF	xurlaBa	pIdippi	vinASa
vilakk	rEp	vIlYYcc	alasipp
laMGi	oVlYippi	hIna	poVlYlYi
kUttabalAwsaMga	vaRalYAkk	vqwwihIna	worYrYu
xurbalappeV	naRta	itarcc	veVtiyerYrY
vixveRa	Binnaw	atiweVrYrY	prawicer
pIdanaww	Binni	arYasrYrY	kurukku
niReX	xoRa	AyuXamA	rAjivacc
visammaw	vimarS	alYYimawi	xuriwa
viparIwa	vaRalYawwar	vittunilkku	walYlYu
marxicc	sammarxx	lahari	oVlYYiya
nirYuww	ASafkAjanak	vittuninn	kApatya
kurYrYa	AkRep	patikkappurYa	BinnABiprAy
dilIrYrY	marxx	pAkappilYYa	muffimari
xuRicc	nASa	apakata	pulivAl
Awmahawya	xuryoga	poVliFF	rYepp
vaXaSikRa	poVttiwweVrYi	prayAsam	ceVrYrYawwar
upaxravi	vEkippi	veVlYlYaMkutipp	pirYupirY
koVnn	xuHsahana	poVlYiccatukk	xuSAtya
mqwaxehaM	nirbanXippi	lajj	mArkkatamuRt
saMSayikkawwakk	piticcuparYa	banXiyAkk	pitivASS
piriFF	wiriccati	ayogya	Baya
vettayA	pinvali	Akramikk	xuHSIl
roRaBariwa	upekRi	Akramicc	viyoj
arakRiwAvas	ASafka	verpiri	AropaN
vilYlYalel	bahiRkari	xuranwa	baliyAtAN

vivAxam	attimarYi	alakRya	pilYYacc
wiraskari	aXikRepi	asAXyam	pilYYakk
narakayAwana	veVrYuMvAkka	oVliccupoyi	jIrNiccA
uruki	lajjAkara	kurYav	viRamA
vexan	itapeVtal	kurYayu	woVlYYikk
walYarww	kulYYapp	kurYaFF	woVlYYicc
awikram	ninxicc	AGAwa	XikkArapar
safkata	xuRkara	bAXikk	arYasrYrY
naSippi	xuRikk	bAXicc	ApawBIwi
oticcu	avawAlYawwilA	cIFFu	nyUnawacU
walYYaya	walYlYi	veVlluvilYi	vipawwA
walYYaFF	walYlYuka	katabAXyawa	ninxikk
kUppukuwwi	awirUkRa	muRikk	apAyasWi
baliyAtA	vyAmoha	kuRTamA	eVwirparYa

Positive Conjunct Verb
mikacca prakatan
xuranwa nivAraN
mAppu parY
surakRa SakwamA
munnarYiyipp nalk
parAwi nalk
sammAnaM laBi
raNtAM sWAna
oVnnAM sWAna
mUnnAM sWAna
mikacca prawikaraNa
parYayAn parYrY

Negative Conjunct Verb
arYasrYrY ceV
balaprayogaM nata
neriteNti vann
vilakk erppeVtu
SakwamAyi prawiReX
plAsrYrYik malinIkaraNaM
nalkunnawin vilakk
jIvan naRta
swrI viruxX
parAwi nalk
parasyamAyi kalleVrYi
xoRakaramAyi bAXi

paTanaM nata
natapati svIkari
kotawi walYlYi
raMgaww vann
paTanaffalY sUcippi
wolppicca praNayaM
kUtuwal PIccar
sanwoRa vArwwa
valiya prawIkRa
valiya oVrYrYakkakRi
lakRyaM va
paTanaffalY sUcippi
rUpa mutakki
mikacca netta
prawIkRayoteV kAwwiru
Mikacca prakatana
paramAvaXi sqRti
hAjarAkkAnAN nirxxeSa
svIkaraNamAN laBi
SupArSa ceVyw
nilapAt vyakwamAkk
golYukal neti
avakASa saMrakR
avakASaM saMrakR
SraxXeyamAya kArya
SraxXa piti
SakwamAya nilapAt
plAn ceVyy
praswAvana cUNtikkAtt

kasrYrYadiyil AvaSyappe
SakwamAyi eVwir
SakwamAya prawiReX
SakwamAya natapati
vilakk erppeVtuww
moSaM anuBava
BIRaNi nerit
ceVrYiya kAryam
valiya itiv

Positive Adverb
valYare
wikaccuM
kurYeV
awIva
awyanwaM
vegaM
utaneV
awivegaM
peVtteVnn
poVtunnaneV
valYareVyaXikaM
affeyarYrYaM
wIrccayAyuM
erYrYavuM
kUtuwa
vallAw
alpaM
mAwiri

Negative Adverb
wAlYYeV
eVwwAweV
kaRticc
prayAsappeVtt
oVruviXaM
FeVruffi
kurYav
niSSafkamA

References

Asher, R.E. 2013. *Malayalam*. London: Routledge.

Baccianella, S; Esuli, A; and Sebastiani, F. 2010. Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In *Lrec*. 10(2010). 2200-2204.

Balamurali, A. R.; Aditya Joshi; and Pushpak Bhattacharyya. 2012. Cross-lingual sentiment analysis for Indian languages using linked wordnets. *Proceedings of COLING 2012: Posters*. 73-82.

Bansal, Naman; Umair Z. Ahmed; and A. Mukherjee. 2013. Sentiment analysis in Hindi. *Department of Computer Science and Engineering, Indian Institute of Technology, Kanpur, India*. 1-10.

Baxi, Abhishek. 2018. Tech Startups, Take Note: More Indians Access The Internet In Their Native Language Than In English. *Forbes*. 24 October 2018. Online: https://www.forbes.com/sites/baxiabhishek/2018/03/29/more-indians-access-the-internet-in-their-native-language-than-in-english/#45555f9f4a03.

Bernstein, B. 1970. A sociolinguistic approach to socialization: With some reference to educability. *Language and poverty*. 25-61.

Bhattacharyya, P. 2017. IndoWordNet. *The WordNet in Indian Languages*. 1-18. Singapore: Springer.

Bučar, J; Žnidaršič, M; and Povh, J. 2018. Annotated news corpora and a lexicon for sentiment analysis in Slovene. *Language Resources and Evaluation*. 52(3). 895-919.

Charles, W. G.1988. The categorization of sentential contexts. *Psycholinguistic Research*.17(5) 403-411.

Chesley, P.; Vincent, B.; Xu, L.; and Srihari, R. K. 2006. Using verbs and adjectives to automatically classify blog sentiment. *Training*, 580(263), 233.

Cruse, D.A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Das, Sanjiv R.; and Mike Y. Chen. 2001. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management science*. 53(9). 1375-1388.

David, C. 2001. *Language and the Internet*. Cambridge: CUP.

Eagly, A. H.; and Chaiken, S. 1998. Attitude structure and function. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.). *The handbook of social psychology*. 269-322. New York: McGraw-Hill.

Edelman, M. 2013. *Politics as symbolic action: Mass arousal and quiescence*. Amsterdam: Elsevier.

Feldman, R. 2013. Techniques and applications for sentiment analysis. *Communications of the ACM*. 56(4). 82-89.

Fishman, J. A. 1998. The new linguistic order. *Foreign policy*. 26-40.

George, K.M. 1972. *Western influence on Malayalam language and literature*. New Delhi: Sahitya Akademi.

George, L. M. 2008. The Internet and the market for daily newspapers. *The BE Journal of Economic Analysis & Policy*. 8(1).

Goutte, C.; and Gaussier, E. 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European Conference on Information Retrieval*. 345-359. Berlin: Springer.

Halliday, M. A. K. 1978. *Language as social semiotic: The social interpretation of language and meaning*. Maryland: University Park Press.

Hatzivassiloglou, V. and McKeown, K.R. 1997. Predicting the semantic orientation of adjectives. In Proceedings of the 35th annual meeting of the association for computational linguistics and eighth conference of the European chapter of the association for computational linguistics. 174-181. *Association for Computational Linguistics*.

Hiroshi, K.; Tetsuya, N.; and Hideo, W. 2004. Deeper sentiment analysis using machine translation technology. In Proceedings of the 20th international conference on Computational Linguistics. *Association for Computational Linguistics*. 494.

International Telecommunication Union; .2017. *ITU ICT Facts and Figures.* 15 October 2018. Online: https://www.itu.int/en/about/Pages/default.aspx.

Internetworldstats. 2017. *Internetworldstats.* 15 October 2018. Online: https://www.internetworldstats.com.

Jayaseelan K.A. 2004. The Serial Verb Construction in Malayalam. In: Dayal V., Mahajan A. (eds) *Clause Structure in South Asian Languages. Studies in Natural Language and Linguistic Theory*, vol 61. Dordrecht: Springer.

Jiang, Long.; Mo Yu.; Ming, Zhou; Xiaohua, Liu.; and Tiejun, Zhao. 2011. Target-Dependent Twitter Sentiment Classification. *In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*.

Johnson, Keith. 2008. *Quantitative methods in linguistics*. Malden: Blackwell Pub.

Katz, J. 1981. *Language and other abstract objects*. Oxford: Basil Blackwell.

Kaur, J.; and Jatinderkumar R. Saini. 2014. A study and analysis of opinion mining research in Indo-Aryan, Dravidian and Tibeto-Burman Language families. *Int. J. Data Mining Emerg*. 4(2). 53-60.

Kumar, K. A.; Rajasimha, N., Reddy; M., Rajanarayana, A.; and Nadgir, K. 2015. Analysis of users' Sentiments from Kannada Web Documents. *Procedia Computer Science*. 54.247-256.

Labov, W. 1972. *Sociolinguistic patterns* (No. 4). Pennsylvania: University of Pennsylvania Press.

Lehrer, Adrienne. 1974. Semantic Fields and Lexical Structure. *American Elsevier*.

Levin, B. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago press.

Liu, B. 2012. Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*. 5(1). 1-167.

Liu, Bing. 2015. *Sentiment analysis mining opinions, sentiments, and emotions*. New York, NY: Cambridge University Press.

Lodge, D. 2015. *The modes of modern writing: Metaphor, metonymy, and the typology of modern literature*. London: Bloomsbury Publishing.

Maas, A. L.; Daly, R. E.; Pham, P. T.; Huang, D.; Ng, A. Y.; and Potts, C. 2011. Learning word vectors for sentiment analysis. In Proceedings of the 49th annual meeting of the association for computational linguistics: *Human language technologies*. *Association for Computational Linguistics*. 1.142-150.

Meyer, Charles F. 2002. *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge University Press.

Miller, G. A. .1995. WordNet: a lexical database for English. *Communications of the ACM*. 38(11), 39-41.

Miller, G. A. 1995. WordNet: a lexical database for English. *Communications of the ACM*. 38(11). 39-41.

Mio, J. S. 1997. Metaphor and politics. *Metaphor and symbol*. 12(2). 113-133.

Mohammad, Saif; Dorr Bonnie; and Hirst Graeme. 2008. Computing Word-Pair Antonymy. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*.

Mohandas, Neethu; Janardhanan P. S; and V. Govindaru. 2012. Domain specific sentence level mood extraction from Malayalam text. In *Advances in Computing and Communications (ICACC)*. 78-81.

Moilanen, Karo.; Stephen, Pulman.; and Yue, Zhang.; 2010. Packed feelings and ordered sentiments: Sentiment parsing with quasi-compositional polarity sequencing and compression. In *Proceedings of the 1st Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*. 36-43.

Morinaga, Satoshi; Kenji Yamanishi; Kenji Tateishi; and Toshikazu Fukushima. 2002. Mining product reputations on the web. *In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. 341-349.

Mukku, S. S.; Choudhary, N.; and Mamidi, R. 2016. Enhanced Sentiment Classification of Telugu Text using ML Techniques. In *SAAIP@ IJCAI* .29-34.

Nair, Deepu S; Jisha P. Jayan; and Elizabeth Sherly. 2014. SentiMa-sentiment extraction for Malayalam. In *Advances in Computing, Communications and Informatics (ICACCI)*. 1719-1723.

Nair, Deepu S; Jisha P. Jayan; R. R. Rajeev; and Elizabeth Sherly. 2015. Sentiment Analysis of Malayalam film review using machine learning techniques. In *Advances in Computing, Communications and Informatics*. 2381-2384.

Nair, R.S.S. 2012. *A grammar of Malayalam*. Language in India.www.languageinindia.com.

NewsFrames. 2013. Archive for the 'Metaphor' Category. *NewsFrames*. 18 October 2018. Online: https://newsframes.wordpress.com/category/metaphor.

O'Leary, D.E. 2011. Blog mining-review and extensions: "From each according to his opinion". *Decision Support Systems*. 51(4). 821-830.

Pak, A.; and Paroubek, P. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Language Resources and Evaluation Conference*. 10. 1320-1326.

Pang, B.; and Lee, L. 2004. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd annual meeting on Association for Computational Linguistics*. *Association for Computational Linguistics*. 271-285.

Pang, Bo; Lillian Lee; and Shivakumar Vaithyanathan. 2002. Thumbs Up? Sentiment Classification Using Machine Learning Techniques. *In Proceedings of Conference on Empirical Methods in Natural Language Processing*.

Patra, B. G.; Das, D.; Das, A.; and Prasath, R. 2015. Shared task on sentiment analysis in indian languages (sail) tweets-an overview. *In International Conference on Mining Intelligence and Knowledge Exploration.* 650-655. Cham: Springer.

Phillipson, R. 2013. *Linguistic imperialism continued*. Abingdon: Routledge.

Phillipson, R.; and Skutnabb-Kangas, T. 2001. Linguistic imperialism. *Concise encyclopaedia of sociolinguistics*. Oxford: Pergamon Press.

Rajendran, S.; and Soman, K. P. 2017. Malayalam WordNet. The WordNet in Indian Languages. 119-145. Singapore: Springer.

Riloff, E., Wiebe, J.; and Phillips, W. 2005. Exploiting subjectivity classification to improve information extraction. In *Proceedings of the national conference on artificial intelligence*. 20(3), .1106.

Riloff, E.; Patwardhan, S.; and Wiebe, J. 2006. Feature subsumption for opinion analysis. In *Proceedings of the 2006 conference on empirical methods in natural language processing*. *Association for Computational Linguistics*. 440-448.

Sárosi-Márdirosz, K. 2014. Problems related to the translation of political texts. *Acta Universitatis Sapientiae, Philologica*. 6(2). 159-180.

Saussure, F. D. 1983. *Course in General Linguistics*. 1916. Trans. Roy Harris. London: Duckworth.

Schäffner, Christina. 1997. Strategies of Translating political texts. Essay. In, *text typology and translation*, vol. 26, 119–125. Amsterdam: John Benjamins Publishing.

Searle, J.R. 1969. *Speech acts: An essay in the philosophy of language* (Vol. 626). Cambridge: Cambridge university press.

Shamsudin, N. F.; Basiron, H.; and Sa'aya, Z. 2016. Lexical Based Sentiment Analysis-Verb, Adverb & Negation. *Journal of Telecommunication, Electronic and Computer Engineering*. 8(2), 161-166.

Sheetal, Sharma.; S, K, Bharti.; Raj, Kumar, Goel. 2018. A Frame Study on Sentiment Analysis of Hindi Language Using Machine Learning. *International Journal of Trend in Scientific Research and Development*. 2(4). Retrieved from URL: http://www.ijtsrd.com/papers/ijtsrd14397.pdf

Skinner, B. F. 1957. *Century psychology series. Verbal behavior*. East Norwalk, CT, US: Appleton-Century-Crofts.

Subrahmanian, V. S., & Reforgiato, D. 2008. AVA: Adjective-verb-adverb combinations for sentiment analysis. *IEEE Intelligent Systems*. 23(4). 43-50.

Taboada, M; Brooke, J; Tofiloski, M; Voll, K; and Stede, M. 2011. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.

Tlabang, Huifeng; Songbo Tan; and Xueqi Cheng. 2009. A survey on sentiment detection of reviews. *Expert Systems with Applications*. 36(7). 10760-10773.

Trosborg, A. ed. 1997. *Text typology and translation* (Vol. 26). Amsterdam: John Benjamins Publishing.

Turney, P.D. and Littman, M.L., 2002. Unsupervised learning of semantic orientation from a hundred-billion-word corpus. *arXiv preprint cs/0212012*.

Turney, Peter D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *In Proceedings of Annual Meeting of the Association for Computational Linguistics*.

Turney, Peter D; and Michael L. Littman. 2003. Measuring praise and criticism. *ACM Transactions on Information Systems*. 21(4). 315–346.

V. Hatzivassiloglou; & J. Wiebe. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *Proceedings of the International Conference on Computational Linguistics*.

W3Techs. 2018. *World Wide Web Technology Surveys*. 15 October 2018. Online: https://w3techs.com.

Wiebe, Janyce. 2000. Learning Subjective Adjectives from Corpora. *In Proceedings of National Conference on Artificial Intelligence*.

Wiebe, Janyce; Rebecca F Bruce; and Thomas P. O'Hara. 1999. Development and Use of a Gold-Standard Data Set for Subjectivity Classifications. *In Proceedings of the Association for Computational Linguistics*.

Wilson, T., Wiebe, J.; and Hoffmann, P. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In Proceedings of the *conference on human language technology and empirical methods in natural language processing*. *Association for Computational Linguistics*. 347-354.

भारतीय भाषा संस्थान

मानव संसाधन विकास मंत्रालय, भारत सरकार,
मानसगंगोत्री, मैसूरु

CENTRAL INSTITUTE OF INDIAN LANGUAGES

Ministry of Human Resource Development, Government of India
Manasagangothri, Mysuru – 570006

&

Linguistic Society of India

Deccan College Post Graduate & Research
Institute, Pune.

लिंग्विस्टिक सोसाइटी ऑफ इंडिया

डेक्कन कॉलेज स्नातकोत्तर एवं शोध संस्थान, पुणे

Certificate

This is to certify that Mr/Ms/Dr/Prof ***Faith T Varghese*** has participated in ***40th International Conference of the Linguistic Society of India*** held at the Central Institute of Indian Languages, Mysore from 05-07 December 2018. The title of his/her presentation is ***Political Opinion Mining in Social Media: An Implementation in Malayalam***

(Tariq Khan)
Coordinator, ICOLSI-40

(Panchanan Mohanty)
President, LSI

(D. G. Rao)
Director, CIIL

A Computational Implementation of Opinion Analysis: A Case Study of Malayalam Political Texts on Social Media

by Faith T Varghese

Submission date: 14-Dec-2018 04:57PM (UTC+0530)
Submission ID: 1056990534
File name: A_Computational_Implementation_of_Opinion_Analysis.pdf (593.77K)
Word count: 17479
Character count: 95970

A Computational Implementation of Opinion Analysis: A Case Study of Malayalam Political Texts on Social Media

ORIGINALITY REPORT

2%	1%	1%	0%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	8springer.com Internet Source	<1%
2	Kumar, K. M. Anil, N. Rajasimha, Manovikas Reddy, A. Rajanarayana, and Kewal Nadgir. "Analysis of users' Sentiments from Kannada Web Documents", Procedia Computer Science, 2015. Publication	<1%
3	Shatabda, Swakkhar. "HMMBinder: DNA-Binding Protein Prediction Using HMM Profile Based Features.(Research Article)(Report)", BioMed Research International Publication	<1%
4	"Recent Developments in Intelligent Information and Database Systems", Springer Nature America, Inc, 2016 Publication	<1%
5	Submitted to Cranfield University Student Paper	<1%

6	www.dcc.ufrj.br Internet Source	<1%
7	aclweb.org Internet Source	<1%
8	"Mining Intelligence and Knowledge Exploration", Springer Nature America, Inc, 2015 Publication	<1%
9	acl.eldoc.ub.rug.nl Internet Source	<1%
10	Bing Liu. "Web Data Mining", Springer Nature America, Inc, 2011 Publication	<1%
11	www.cs.pitt.edu Internet Source	<1%
12	"Social Computing and Social Media. Applications and Analytics", Springer Nature, 2017 Publication	<1%
13	Submitted to Queen Mary and Westfield College Student Paper	<1%
14	Submitted to University College London Student Paper	<1%
	Submitted to Rochester Institute of Technology	

15	Student Paper	<1%
16	www.ijaerd.co.in Internet Source	<1%
17	Son Trinh, Luu Nguyen, Minh Vo, Phuc Do. "Chapter 23 Lexicon-Based Sentiment Analysis of Facebook Comments in Vietnamese Language", Springer Nature, 2016 Publication	<1%
18	"Computational Intelligence: Theories, Applications and Future Directions - Volume II", Springer Nature America, Inc, 2019 Publication	<1%
19	scholarbank.nus.edu.sg Internet Source	<1%
20	library2.usask.ca Internet Source	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		