Pakistan Languages and Humanities Review
www.plhr.org.pk

RESEARCH PAPER

Extracting Semantic Field of Joy and Sadness from the Movies Corpus: A Corpus-Based Study

Dr. Zafar Ullah*1   Dr. Muhammad Farooq Alam 2   Haseena Naz 3

1. Instructor, English Department, Virtual University, Islamabad, Pakistan
2. Assistant Professor, Department of English, National University of Modern Languages, Rawalpindi, Punjab, Pakistan
3. Research Associate, Department of English, National University of Modern Languages, Islamabad, Pakistan

PAPER INFO	ABSTRACT
	This study extracts the semantic field of joy and sadness from Movies Corpus focusing on the development of the trends and options to express the two moods of joy and sadness. Plutchik's Theory of Basic Emotions has been applied to extract the frequency of both emotions. Oxford Thesaurus and the corpus-lingual database of Movies Corpus have been used to generate data. From 200 million words of Movies Corpus, the two most occurring emotions have been scrutinized under synthesising inspections. The frequency, semantic drifts and linguistic variations for these two moods have been scrutinized under careful observation and calculation. Major findings reveal that the words 'pleasure, treat, joy and prize' are the most occurring options for happiness with the highest frequencies while the words 'grief, depression, misery and sadness' are the most suitable options for depicting the sad mood with the highest frequencies.

Introduction

This newly built movie Corpus (February, 2019) represents a variety of true depiction of emotions. It comprises more than 25,000+ movies from the 1930s to the contemporary era, and its size is more than 200 million words by the recent calculation. All movies are linked to their Internet Movie Database (IMDB) (Movie Corpus, n.d.) entry which elucidates that one can build customised virtual corpora using extensive metadata. It is a larger corpus than other existing corpora of informal English. The corpus also allows its users to look at linguistic variations over time and in different dialects. It also helps its users to record the possible synonyms of a word and its suitable usage in a language.

Movies are the depiction of real-life situations so we can get the impacts of these emotions by observing them in movies. From all the eight basic emotions, this research work is much focused on joy and sadness because concentrating on all of them is a very difficult task. Happiness and sadness are the parts and parcels of real life. Everyone frequently faces these two moods. This research work also employs linguistic ethnography to seek out where happiness and sadness lie in our daily lives. The current work has addressed how emotions of

happiness and sadness have been elucidated in Movies Corpus by deriving lists of synonyms of both the moods.

There are so many expressions which are not contemporary to their meanings. So many phrases and expressions are annotated with joy or sadness but they have contradictory meanings when they are analysed deeply. The current research work has been incorporated with such studies and analyses as well. The public is not aware of changing trends in such expressions so this research has also shed light on such problems. Various semantic analyses and studies incorporated in this paper reveal the happiness and sadness trajectories of a day, week, month or year and compare the socialness and human centeredness of happy descriptions versus sad descriptions. The study evaluates a corpus-based approach in a classification task and contrasts the wordlist with emotionally annotated wordlists produced by experimental focus groups. The paper concludes by offering a corpus inspired liveable recipe of happiness and sadness.

When we talk about synonyms, it directly refers to the semantic field of a word or a phrase. A semantic field can be defined as the family of a word to which it belongs. It can also be called a lexical field or domain which refers to the combination of a bunch of words with interrelated meanings and dominated under the same concept (Zhou, 2001). It has two folds of meaning. When the semantic field is inspected, we come to know about the paradigmatic as well as the syntagmatic meaning of a word or a phrase. It was Ferdinand de Saussure who established these two concepts of relations between words or more precisely between word senses. Paradigmatic relations are those in which a linguistic unit can be substitutable or it can be replaced. Syntagmatic relations can be explained in a way in which a unit contracts or expand with the changing surrounding units. This research work sheds light on the paradigmatic as well as the syntagmatic relations of happiness and sadness with other linguistic units.

Literature Review

There were many chronological works on the concerned study area in the history of linguistics. Some research findings have been mentioned here in order to identify, evaluate and examine the systematic, comprehensive, explicit and reproducible methods and techniques involved in the Corpus domain. There were different uniform corpora. It means that they were rigid to make changes which had frozen the learning process. Corpus of Contemporary American English (COCA) on the other hand had made a drastic change in order to make the learning process more fluent and progressed. Initial works on language change were actually done by the corpus study. It is comprised of 385 million words from the years 1990 to 2008. It is superior to other corpora due to its many advantages. Equal and proper classification of different genres of literature was prevalent in this corpus. The most important thing about this corpus is that it was updated every year with the successive addition of words and an equal amount of data. Due to these successive progressions, COCA was being consulted for diachronic studies within the English Language. It also helped in the analysis of phrasal verbs in American English (Davies, 2009).

In magazines, fiction, spoken, academic journals, social journals and newspapers, COCA had participated equally. So, it had an equal distribution of genre balancing in linguistic shifts and studies. It had made evolution in linguistic options with respect to time. COCA had exhibited interesting qualities to be cited as the first monitor of the English Language as it had

made an easy platform to examine diachronic linguistic shifts and evolutions in every genre of a language differently and equally (Davies, 2009).

Similarly, the Corpus of Historical American English (COHA), a contemporary to COCA was an axis to study morphology, syntax, lexis, phraseology, semantics, American culture and society. Language changes occurred with time and their simultaneous studies are very crucial for the progress of the language which was also accompanied in COHA with the other genre balancing. Other American, English and text corpora were unable to examine the changes which COHA had made accessible and understandable. It was actually a distinctive and absolute manifesto for language changes. Between the texts dating from the 1810s and 2000s, COHA had experienced almost more than 400 million words which were taken from 100,000 works. In late 2010, COHA was annotated, sketched and instigated. COHA had perpetuated a historical record of genre balancing in ten years which is from 1810 to 2010. It maintained a frequency change of words with the passage of time which helps to compare words. Engrossing insights into social and cultural changes in the history of words could be kept by the frequency changes recorded by COHA. This unique quality had constructed a dominancy of COHA over other contemporary text media like the Oxford English Dictionary which was unable to provide information of frequency over time. It had allocated the attestation and origination of a word or a phrase (Davies, 2012).

Lexical, morphological, phraseological, syntactic and semantic changes were documented with the successive involvement of COHA. Changes dealing with the form and concept were termed lexical changes. Different word formation patterns were included in morphological evidence. Grammatical constructions with the transfer and rotation of words were made by the wings of syntactic fashions. The variation in the meaning of a word or in a phrase was made under the umbrella term of semantic changes. These changes were easily marked in COHA and in its contemporary COCA as well. This unique quality has made learning operations more effective and progressive (Davies, 2012).

The application of split infinitives had made more prominent during the era of 1940s and 1980s. The model was 'to boldly speak'. This formulation was an example of prescriptive syntactic changes. Both the prescriptive and descriptive syntactic changes had been seen by the literary lenses of COHA. The excessive usage of 'have to +verb' was an instance of descriptive alternations. Models were 'we have to clean' and 'I have to wash'. These models were an increasing trend while post-verbal negation was a decreasing trend. Models were 'We need not worry' and 'She need not mention'. Rotation in the meaning of a word or phrase was prevalent in COHA "A word is known by the company it keeps" is congruous for the corpus as its axis. These changes were observed over time. For example, the word 'gay' was first collocated as attractive, brilliant and jolly in the 1880s, while it was collocated as heterosexual and bisexual in the 1980s. The word 'women' was collocated as cultivated, pious, abandoned and noble in 1800s while it was collocated as catholic, working class and middle class in the 1900s (Davies, 2012).

In an article on the semantic drift, it was mentioned that it was a difficult and mystifying task to discern the semantic alternations in early modern English. The nonce words and phrases were difficult to mark in the word list. This difficulty was solved by two wed based diachronic corpora, Lexicons of Early Modern English (LEME) and Early English Books Online (EEBO). These two corpora were being used for the coalition and lemmatization of meaning-making process. Due to the no survival of any monolingual dictionary of the time, there was no record of innovative words of the Shakespearean period. That's why the instances

and models of diachronic drifts were dissipated from the documents. They are not found in the current history of words. There were some words which exemplify the dare of finding semantic alternations. In Shakespeare's play Macbeth, the word 'pricking' was applied to 'thumb of witch'. In Titus Andronicus, the word 'Aron' was used for the villain's name. the word 'acting' was used to explain the dream of the assassination conspiracy against Caesar seen by Brutus. The challenge of semantic alternations of innovative words subsists if corpus users are dependent on the analysis of literary text for the placement of semantic gravitation (Lancashire, 2012).

Numerous works had been done on emotion corpora. International Survey on Emotion Antecedents and Reactions dataset (ISEAR) was the first remarkable toil in this domain (Scherer and Wallbott, 1994). It had got 7000 responses from the partakers. In order to extricate collocation features the ISEAR dataset was used. The partakers were being asked to describe the situation in which they had faced certain emotions.

It is a fact that it is difficult to model emotions in a conversation but it is a rewarding and productive task with limitless implementations. A system cannot be considered as good if it does not consider every word in a conversation, its grammatical structure and syntactic deviations such as negations, exclamations, embedded words and phrases and the form of expressions and sentences. There is always an impact of the first utterance with the later one (Collier, 2014) but there is a lack of distinctive emotional expression which can be easily noticed. There were other corpora which could not solve such problems so Emotional Movie Transcript Corpus (EMTC) was introduced in order to observe, distinguish and solve such complications.

It was based on a multilabel annotating scheme following Plutchik's Theory of emotions. Instead of only providing text with different annotator schemes also donated the annotators with movie clips. In a multilabel corpus study in the movie domain, Plutchik proposed his theory on eight emotions (anger, fear, sadness, disgust, surprise, anticipation, acceptance and joy) in movies in which he had explained that many works limit the labels of emotions or they only allow annotators to label one emotion per utterance.

According to the same article, emotions are not mutually exclusive (Plutchik, 2001; Russell, 2003). In many cases, people may experience many emotions at the same time (Choe et al., 2013). Therefore, the study of corpus for emotion analysis should be multilabel. Limiting the labels of emotions may decrease the problems but it can cause trouble for annotators to provide correct judgements. There are many annotating schemes when inter-annotator agreements were discussed and the word-embedding of the corpus. Some emotions are the primary ones and are simple while some are the mixtures of primary emotions and form complex ones. Both the types of simple and complex emotions are intense differently. In this corpus, they had set 48 emotion labels. Sometimes one utterance can carry zero emotion and it is annotated with 'none' so its intensity would also be considered as none. Sometimes an utterance can have more than one emotional label. This corpus capitulated a better agreement score to the multiclass.

In the same work, according to (Arimoto and Okanoya, 2016; Kruger, 2005), it was proved that in any online communication method, the interpretation and conveyance of emotions are seemed difficult for the users than face to face communications. It is due to the limitations in communication modality.

In an article on opinion and emotion in movies, E. Mouka and his colleagues suggested inter-annotator agreement schemes and translation-oriented experimentations that how a word or phrase or any expression is being transferred from the textual source word to the target language through audio-visual data. Audio-visual data was first designated in order to conduct translation-oriented research. All the types of texts grounded by thoughts, opinions, beliefs, goals, judgements, feelings, emotions and evaluations were given a platform to annotate and to be translated into the target speech which made this corpus an excellent pool to transfer all types of texts from textual word to the target language is the more casual way.

In order to show opinions and emotions through these annotations, news articles, news titles and a corpus of real-life self-expressed emotions were selected according to their topic, time, realistic approach and their content. Various textual genres had been studied under the influence of the Emotiblog annotation model (Boldrini, 2010).

For the interesting polarity to be defined, Appraisal theory was proposed in order to set the nature of emotion (Martin and White, 2005). According to this theory, emotions were given positive, negative and neutral charges in consonance with their nature. For example, emotions of joy, happiness and satisfaction were considered to have a positive polarity. Anger, fear and sadness were marked as negative polarity while emotions like surprise and anticipations were considered to have neutral polarity.

In an article, the concept of 'incidental vocabulary learning' was introduced which was a corpus driven study of the potential for vocabulary learning and suggested that how one could have an effective vocabulary learning process by watching movies. According to this concept, one could improve his vocabulary by watching movies with subtitles. Watching movies with subtitles had a higher learning potential rate than watching movies without subtitles (d'Ydewalle & Pavakanun, 1997). It makes the second language learning process more progressive and casual. In order to prove the effectiveness of subtitles, Neuman and Koskinen (1992) measured incidental second language (L2) vocabulary learning by giving the students in grade 7th and 8th readings, providing them with movies with subtitles, and television scripts. As a result, they found that the students who were provided with subtitles scored higher than those who only relied on the transcript of the programs. It proved that the aural and visual input provided to the students encouraged them in the effective learning process.

A collection of movie plot synopses and tags were kept in a corpus in order to get feedback from the audience. Folksonomy (Vander Wal, 2005) introduced which is actually a process of gathering feedback from the target community. It is also known as social tagging. The community actually tag by calling the movies fantastic, horrifying, funny etc. These tags were taken as the actual keywords to correlate the plot synopses of movies with the tags. It also helped to discover the social interest and involvement in the movie domain (Borne 2013; Lambiotte and Ausloos, 2006; Li, 2008; Szomszor, 2007). This corpus entails the list of words which the community had tagged movies according to their nature. In recommendation systems like IMDb1 and MovieLens2, user-generated tags contributed to different features of movies. These tags had literally benefitted the recommendation and feedback engine. The recommendations and 71 fine-grained tags given by the consumers are listed in this corpus (MPST) with having 14,828 plot synopsis of movies which was outlined by the association between tags and the movie plot synopsis.

In a nutshell, all the above-mentioned corpora were good enough to deal with the trends and variations in different genres of literature. They dealt with the linguistic variations on the basis of genre balancing and they also had performed on emotion analysis but they were unable to find emotion analysis on Movies Corpus because it is the latest corpus and no work has been done on it. The ongoing research work is dealt with the manifestation of happiness and sadness and the evolving trends in expressing them in Movies Corpus which help the script writers, psychologists, teachers and learners to comprehend them properly.

Material and Methods

Theoretical Framework

In order to get a general emotional response, Plutchik (2001) proposed a psycho-evolutionary classification approach to emotions. There are eight basic emotions according to this theory of emotion; anger, fear, sadness, disgust, surprise, anticipation, acceptance and joy which were set under the observation of different Corpus-based works in order to analyse joy and sadness in Movies Corpus. Joy and sadness are contradictory pairs that Plutchik has proposed. The research took the two opposites as it is always an interesting phenomenon to examine one stance or feature by examining its opposite first. So the two opposing 'joy and sadness' had been much focused on in the current research work administered by Plutchik's Theory of emotions.

Plutchik proposed this theory on the basis of shreds of evidence from many scholarly works. These emotions have both the dimensions of positivity and negativity and at the same time, they elucidate their strengths. Some emotions carry more strength which means they are stronger and heavy in use while some are mild emotions and some are weak according to the frequency of their effects. Some emotions are primary which could be the root words while some are secondary which are actually the combinations of primary emotions or the ones which are seldom used. So the frequency changes according to its usage and demand.

Research Design

This research work is themed on both qualitative and quantitative based study of emotions especially the most occurring two emotions are joy and sadness in Movies Corpus. Qualitative in the sense that the study opted for these words to deeply understand the difference as well as the sameness of meanings of synonyms to describe the two moods and also had elucidated the impact of these synonyms in different contexts that these synonyms carry diverse meanings. Such evaluations had alluded to diversified linguistic options for the same word or idea. Quantitative in a sense that it is a corpus-based study in which the amount of data is huge and widespread.

The current work had been analysed the emotions of joy and sadness in the Movies Corpus in the light of Plutchik's Theory (Plutchik, 2001). To analyse the two moods of happiness and sadness, their synonyms were first taken by Oxford Thesaurus which were put on the searching box of Movies Corpus one by one in order to enlist them with their frequencies then later they were set on the frequency order. The more frequency of a word shows frequent usage of the word.

Subsequently, the elaborative study according to the context provided by the Movies Corpus made it understandable that there were 34 synonyms for happiness and 15 for sadness

in the Movies corpus. Their strength, their nature and their effect varied accordingly. The diversified options for happy and sad moods were traced by the expanded study of the corpus.

Data Required and Sampling

The current research work required a corpus-based study on two basic emotions 'joy and sadness', different articles on happiness and sadness were required, and synonyms of joy and sadness were generated from www.thesaurus.com. These synonyms were taken as samples for the inspection. The words were then put on the search bar in order to elucidate their dimension of use and frequency. The frequency list was framed for both the moods separately. The snipping tool was used in order to capture records of words for the required moods. Then the words were exported with the help of a snipping tool which was later analysed through a qualitative interpretive approach.

Results and Discussion

Happiness and sadness are the parts and parcels of our life. People try to voice and gesture them in many ways and expressions. Both the formal and informal conversations have an impact on these two moods but informal languages and conversations are indebted to emotions more than formal ones. So, an inspection of conversations in movies aids more interest in interestingly examining the two moods.

Data Analysis for Emotion of Joy

There were 34 synonyms of joy found in the Movies Corpus. Their meanings are traced by Oxford thesaurus and their frequencies are calculated by searching them one by one on the search bar in Movies Corpus. Frequency changes indicate the intensity of using a word for a suitable purpose. Greater frequency shows its greater strength and vice versa. It demonstrates that the word with greater strength would be the strongest and it complements the concerned mood more nicely than others. Every person has his or her own way of expressing emotions and one can have these 34 linguistic options in order to express our feeling of joy which were found out by Movies Corpus. As movies are the depiction of real life so it is related to daily life.

Table 1
Semantic Field of Emotion of Joy

No.	Emotion Word	Frequency	No.	Emotion Word	Frequency
1	Pleasure	16346	18	Gaiety	112
2	Treat	9093	19	Merriment	103
3	Joy	6718	20	Gladness	97
4	Happiness	5785	21	Elation	93
5	Prize	4310	22	Exhilaration	75
6	Cheer	3635	23	Buoyancy	58
7	Blessing	2216	24	Jubilation	53
8	Satisfaction	1435	25	Exaltation	26
9	Delight	1155	26	Exultation	22
10	Bliss	988	27	Cheerfulness	17
11	Ecstasy	839	28	Jollity	10
12	Enjoyment	341	29	Joyfulness	6
13	Felicity	330	30	Joviality	3

14	Glee	285	31	Joyousness	1
15	Rapture	251	32	light-heartedness	1
16	Gratification	154	33	Jocundity	0
17	Contentment	135	34	Blithesomeness	0

Table 1 demonstrates the descending order of frequency of semantic field for the word 'joy' ranges from 16346-0. It means that the word 'pleasure' is the most suitable word to fittingly express a happy mood. The word 'treat' is added up to the second number in the list, 'joy' third and so on to express joy in Movies Corpus. There are some words in which frequency rates were zero. For example, 'jocundity' and 'blithesomeness'. These are the synonyms of joy but are not set up in Movies Corpus. The list is figured also with words which seldom occurred. For instance, 'joyousness' and 'light-heartedness'. The lesser the frequency the lesser would be its strength and use.

The research has taken random screenshots of conversations from Movies Corpus which are mentioned in the figures and helped to examine the semantic field of the two emotion words comprehensively. These emotions directly have associations with human feelings which are traced in the given figures. For example, joy and pleasure have an association with 'love', anticipation with 'optimism', trust with 'hope', and satisfaction with 'submission'.

Figure 1. Data for Pleasure

Figure 2. Data for Treat

By examining the Movies Corpus, this study is concerned with the frequency of a word and its semantic field. A semantic field establishes dimensions of meaning with the surrounding words a word keeps. It is also concerned with the paradigmatic and syntagmatic meanings of syntax. Figure 2 illustrates that the word 'treat' is used in many dimensions regarding the context in which it is used. As it is observed that it is the second suitable word to show someone's joy. It is also used in order to show some act of behaviour. For example, in

the line taken from figure 2 ', The Gay Divorce' 'The lady in question is very sensitive and you must treat her accordingly' the word 'treat' is used to show someone's behaviour. While the line from 'Evelyn Prentice' 'That is because they are having a special treat-ice cream?'. In this line, the concerned word is taken as a noun that is a celebration.

1	1994	US/CA	Imaginary Crimes	A	B	C	Margaret's mother doing that either. He's missing out on life's greatest joy. My business may suffer but all I need to do is look over here
2	2015	US/CA	Inside Out	A	B	C	- Oh-oh what do we do? Guys, th-this... You pretend to be Joy. Wouldn't it be great to be back out on the ice? Oh
3	1940	US/CA	Remember the Night	A	B	C	much trouble. Trouble? Why, bless you, child, it's a joy to have you here. You might have sent a telegram, John Sargent.
4	1973	Misc	Lady Snowblood	A	B	C	that I can write it anymore. But... Chapter Four: The House of Joy, The Final Hell. " Charity Masquerade " May I have this dance?
5	2015	US/CA	Bachelors	A	B	C	, even though Floyd does all the work and Joy takes all the credit. Joy really cracks the whip on Floyd, too. We've seen her grab him
6	2011	UK/IE	Life in a Day	A	B	C	be a long day, as most are. He's always a bundle of joy when he wakes up. Come on, come get in the shower. Get
7	2012	US/CA	Abducted: The Carlin...	A	B	C	a chance. Let's make - - make up for lost time. - JOY: I was on such a high when I was first reunited with my daughter
8	1943	US/CA	It Ain't Hay	A	B	C	old lapel Wrap your dreams in brocade Put on your Sunday clothes And join the joy parade Open up your heart and sing the sunbeam serenade Light
9	1999	UK/IE	An Ideal Husband	A	B	C	has done. And yet?... And yet?... I have never known such joy as when I'm with him. I've never felt so... free as
10	1958	UK/IE	Hell, Heaven or Hoboken	A	B	C	43 One moment, Colonel, it is for you, Sir Thank you Any joy? You have How do you manage it? I finally got to the right

Figure 3. Data for Joy

In figure 3, the word 'joy' is used in different contexts. Importantly, Joy is the primary opposite of sadness but found as the third most suitable word to express a happy mood. 'He is missing out of life's greatest joy'. This line displays the feeling of happiness of the person. It is not necessary for a word to stick with its meaning. It changes when it collocates with different words. For example, it can also be used as an acronym in order to show some positive gestures which can be seen in the line 'Let's make up for a lost time. JOY (Just Open Yourself). When a word combines with different words its syntax changes automatically changing the meaning.

31	2004	Misc	Alexander	A	B	C	blood. Zeus is your father. Oh, I understand she brings you some happiness but how can she help you? You must kno
32	1943	US/CA	I Walked with a Zombie	A	B	C	. Yea, Lord... pity them who are dead... and give peace and happiness to the living.
33	1952	US/CA	Limelight	A	B	C	get away. That house in the country, where we could have peace and happiness. " Happiness. " it's the first time I've e
34	1937	US/CA	You Only Live Once	A	B	C	do matter. Come on, Bonnie, be a good sport, wish us happiness. I do, darling, but you haven't a Chinaman's chance.
35	2008	US/CA	August	A	B	C	successful futures. We'll be bound together so that our happiness can become your happiness. Happiness? Dear boy,
36	2014	UK/IE	That Day We Sang	A	B	C	Come hear the noise # Hear those marching feet # We're heading down to Happiness Street # We've nothing to lose
37	2006	UK/IE	As You Like It	A	B	C	to make me sad. And to travel for it too. Good day and happiness, dear Rosalind. Nay, then, God by you, and you talk
38	1988	US/CA	Dangerous Liaisons	A	B	C	have to live without her. Tell her, her love was the only real happiness that I have ever known. Will you do that for me
39	1941	US/CA	Ball of Fire	A	B	C	future father-in-law over the telephone. But before even considering entrusting your daughter's future happiness to r
40	1968	US/CA	The Little Drummer Boy	A	B	C	when they heard it, began to dance to its sound. But Aaron's happiness was to end all too soon, for one night, bandits

Figure 4. Data for Happiness

7	2009	US/CA	Saint John of Las Vegas	A	B	C	Nice Day ". I had to get it. And it's a big prize, too, if you get " Have ", " A ", "
8	1963	US/CA	Shock Corridor	A	B	C	newspaper greats. And this long corridor is the magic highway... to the Pulitzer Prize. Maybe I'm looking at the killer right now. Three witnesses - Stuar
9	2004	US/CA	Closer	A	B	C	, I'm allowed to flirt. To prize my money from me. To prize your money from you, I may do or say as I please. -
10	2012	UK/IE	Ill Manors	A	B	C	and now you want the baby back with half the money? What a fucking prize cunt or what? Look at her. Done up very well. Yeah,
11	1986	UK/IE	Slaughter High	A	B	C	got it. Boy, would you look at Joey? He never collected that prize, did he? No, they took him straight to the hospital. He
12	1982	UK/IE	Trail of the Pink Pa...	A	B	C	- I suppose I do. - Exactly. The Pink Panther. Such a prize he could never resist. He'd be disappointed. The Pink Panther is in
13	1956	US/CA	Forever, Darling	A	B	C	Susan, that's the way you're heading. Well, you're no prize. You have a few faults of your own, if you must know.
14	1997	US/CA	Jack Frost	A	B	C	the competition today? - Oh sure, how else you gon na win first prize. But I want you all wrapped up. Okay? And the tree trimming
15	1933	US/CA	Doctor Bull	A	B	C	's your thermometer. Wasn't any good, was she? She won theJersey prize at the fair. - What, that thing? - Yes, that thing
16	1935	US/CA	The Farmer Takes a Wife	A	B	C	ONLY FARM PEOPLE SENT THEIR COOKING, SO I THOUGHT IF I COULD WIN A PRIZE WITH MY PRESERVES, IT'D PROVE TO DAN THAT CANAL FOLK CAN
17	1948	US/CA	The Treasure of the ...	A	B	C	what you mean? - You say it's the winner? - 200 peso prize. Where? - That's my number. - Sure enough. Just look

Figure 5. Data for Prize

The word 'prize' is also a suitable word to express the feeling of joy. One can enjoy a happy mood by getting a prize or reward. From figure 5, in entry 8, the word 'prize' is used as a proper known. In entry 13, the expression 'you are no prize' is somehow idiomatic that the person is not difficult to understand. So the word 'prize' can also be used in order to show some understanding, but these types of expressions are not in much use. All the other

synonyms can also have different interpretations so one can get different dimensions and choices in order to show positive gestures.

Data Analysis for Emotion of Sadness

Likewise, there are also various synonyms in order to express the feeling of sadness. Human life is a mixture of happy and sad moments so everyone has their own way of expressing them. Both of them occur side by side but every person wants to be happy more than to be sad so the synonyms for the feeling of joy are inspected more than the synonyms for sad mood in Movies Corpus. It means that expressions of happy moods are practised more in real life. Consequently, the words expressing happiness are more in frequency than the words of sadness. The 15 words in table 2 represent a sad mood.

Table 2
Semantic Field of Emotion of Sadness

No.	Emotion Word	Frequency	No.	Emotion Word	Frequency
1	Grief	2391	9	Despondency	16
2	Misery	2137	10	Dejection	4
3	Depression	1772	11	Dolour	3
4	Sorrow	1726	12	Gloominess	1
5	Sadness	1250	13	Sorrowfulness	1
6	Melancholy	471	14	Despiritedness	0
7	Gloom	323	15	Dejectedness	0
8	Unhappiness	281			

The findings of Table 2 illustrate that there are 15 words to express sad emotional state and they differ with their frequency. Some words are more suitable so they occurred in greater frequency and vice versa. The study finds a descending order of the semantic field of word sadness with respect to their use and demands ranging from 2137-0. The word with the highest frequency means that it is the most suitable and most occurring word in order to complement the sad mood. Similar to the synonyms of happy emotions, synonyms of sad emotions are also associated with human feelings. Misery and sadness hold and are associated with 'pessimism', disgust and anticipation with 'cynicism', anger with 'contempt' and also with 'dominance'. It demonstrates that the words taken for emotion analysis are universal to the human feelings but there were also words which are associated with human feelings and are not found in Movies Corpus. For example, 'Dejectedness and dispiritedness' drew the lowest frequency of 0. These associations of emotions with human feelings are illustrated in the given data.

Figure 6. Data for Grief

Grief is the intense opposite of ecstasy. While inspecting Figure No. 6, we come to know that grief has five stages that entail 'denial, anger, bargaining, depression and acceptance'. One thing that seems interesting here is that 'acceptance' is one of the stages of 'grief' which people usually take with a positive connotation. Entry 6 alludes to 'grief' with a positive denotation that it is coupled with the word good. 'good grief, what a suggestion'. Usually, we do not take grief as good but in Movies Corpus we find its positive annotation too.

Figure 7. Data for Misery

Misery is the second most occurring word to show sad feelings. It is used in many ways to annotate the mood of sadness, but in entry 3, it is capitalized to emphasise more misery and to show the strength and power misery has on someone's life. It is also shown in figure 7 that the word misery is used from 1980 to 2014 for the same connotation.

Figure 8. Data for Depression

The word 'depression' is more probably related to the human mind that when one's mind is overburdened with some sort of threshold, it gets depressed and scrutinizes negative feelings. Depression is also used in many ways in order to allude to sad moods. In entry 7, 'our main export is crippling depression' the word depression is used as a name of a clinical disorder in which one becomes a part of a lifelong struggle. It is proven how these emotion-annotated words are used in clinical psychology in order to pinpoint any disorder. And in entry 9, it is capitalized because depression here is showing the 'Great Depression Era' of Franklin D. Roosevelt when the United States was suffering from depression. It alludes that these emotion-annotated words have great importance even to represent the state of any state.

Figure 9. Data for Sorrow

In the very first entry it is used as the name of the song 'The Fabric of Sorrow'. It is used here with a positive connotation. It means that the song is fabricated with sorrow and this song of sorrow is becoming a source of sublimation to the hearer, 'the Fabric of Sorrow beseech thee'. Entry 4, 'But the world still carries on Love must always change to sorrow' exhibits a beautiful truth of life that no matter how loved we are we always have to replace it with sorrow one day. This line is depicting the reality of life which helps the human psyche to keep calm and peace in every hard and soft tinge of time. It is also inspected how these emotion-annotated words are beautifully jotted down to illustrate the realities of human life.

Figure 10. Data for Sadness

Figure 10 entails those expressions and phrases which exhibit sad moods. It means that in these entries the word 'sadness' is used for what it is expected to exhibit. The word 'sadness' is being used in so many dimensions. On one occasion sadness is considered that much great that the only remedy to get rid of it is to die. It is entailed in entry 6, 'if you want to keep sadness away, what better remedy than death. Here the power of negative emotion is shown greatly which can be effective on the human psyche negatively the way it is expected for. It is also demonstrated that one utterance can have more than one emotion at the same time or it can have no emotion rather. It simply means that if annotated with the multilabel scheme, it yields a better agreement score with the multiclass.

Figures taken for both the moods have better coverage of informal, real-life conversations taken from Movies Corpus, which have important implications for the quality and quantity of the data overall. While inspecting the concrete examples the research focused on two phenomena-collocates and syntax. The figures provide data for both the lower frequency constructions as well as the higher frequency constructions. All the figures show a number of different collocates that occur with a node word. When a node word collocates with different words its syntax changes and automatically changes the meaning.

The research also finds out that the words commenting on happy moods are positively polarized while the words complementing sad moods are most of the time negatively polarized. It is the human psyche to drive more towards the positively polarized things so the data for the semantic field of joy is inspected more than the data for the semantic field of sadness. Therefore, the frequencies for the words of joy are greater than the frequencies for the words depicting sadness. It simply means that the greater the positive polarity of a word the greater would be its frequency.

Conclusion

Inspections and findings have evidenced that there are 34 synonyms for the word happiness and 15 synonyms for sadness in Movies Corpus. Of 34 synonyms, 'pleasure, treat, joy and prize' are the most occurring words. It means that they are more appropriate to exhibit a happy mood. They are used differently in different contexts. In the same way, 'grief, misery, depression and sorrow' are the most occurring synonyms which we can call the more suited expressions to express a sad mood. Their frequencies are examined carefully from 1930 to 2018.

Various dialects are entailed in the Movies Corpus in order to examine 200 million words. These words are sampled from 25,000+ movies. These are inspected under the variations of dialects from 6 countries.

It has been easier to examine these two moods through the free online access of corpus which is an amalgamation of words, phrases and ideas from all the genres of literature and social sciences. It has been introduced by Mark Davies, a professor of linguistics at Brigham Young University. With the help of this online access, the researcher comes to know how native speakers speak and write these two moods (joy and sadness) in daily conversations. The frequency according to their usage has also been summed up. The historical and dialectal variations while exhibiting these two moods come out on the sheet due to the inspections done on Movies Corpus. It has become possible to lay out all the possible entries of synonyms of happiness and sadness on one page only because of the intense study on this corpus otherwise it would be a very hard and laborious task to find out all the possible entries of these moods by analysing all the texts.

By inspecting the Movies Corpus, the study becomes acquainted with the outcome that it is not necessary for a word to follow the same trend of meaning that it had been following for years. A word is always context-dependent in that it portrays according to the context it is surrounded by remaining within the boundary of its semantic field. Words can also be collocated with words which seem odd when coupled together. For example, from the above entries, the word 'good' is collocated with 'grief' which seems odd. Grief can never be good to average men.

Looking at the corpus studies the learners not only come to know the number and frequency of a specific word or phrase, rather they also come to know the qualitative inscriptions and linguistic options of the particular term. The greater the frequency the greater would be the usage of that word. It is an entailment of the investigation that it is to enhance the vocabulary and information bar of the learner. Another important feature of this academic work is knowing the suitable and diversified options for informal usage of these two moods. Informal usage is much more important in the way that a person most of the time is involved in informal conversations in a day. Teachers, scriptwriters, and learners can have more options to make their works more captivating than before. It is obvious that one has to deal with many dialectal thresholds by being a part of a miscellaneous society. This scanning of corpus work makes it easier to understand such variations and choices. More importantly, this innovative work is significant for the field of psychology because psychology deals with the human psyche, feelings and behaviours which are directly associated with emotion-annotated words and expressions. By perceiving the number and frequency of a word one can get a more suitable option to express them fully and at the same time one can get to know about the intense emotion, mild emotion and primary emotion so leaving room to use emotion words accordingly.

It has been interesting to know that the human psyche is more inclined toward positive things that is why the research finds out a greater number of synonyms for the word joy than the synonyms for the word sadness in the Movies Corpus. It means that people try to gesture happiness more than sadness because everybody wants to be happy more than being sad.

References

Borne, K. (2013). Collaborative annotation for scientific data discovery and reuse. *Bulletin of the American Society for Information Science and Technology*, *39*(4), 44-45. doi:10.1002/bult.2013.1720390414

Collier, G., & Collier, G. J. (2014). *Emotional expression*. Psychology Press. doi:10.4324/9781315802411

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and linguistic computing*, *25*(4), 447-464.

Davies, M. (2012). Expanding horizons in historical linguistics with the 400-million word Corpus of Historical American English. *Corpora*, *7*(2), 121-157. doi:10.3366/cor.2012.0024

d'Ydewalle, G., & Pavakanun, U. (1997). Could enjoying a movie lead to language acquisition?. In *New horizons in media psychology* (pp. 145-155). VS Verlag für Sozialwissenschaften, Wiesbaden. doi:10.1007/978-3-663-10899-3_10

Gao, C., & Xu, B. (2013). The application of semantic field theory to English vocabulary learning. *Theory and Practice in Language studies*, *3*(11), 2030. doi:10.4304/tpls.3.11.2030-2035

Kruger, J., Epley, N., Parker, J., & Ng, Z. W. (2005). Egocentrism over e-mail: Can we communicate as well as we think?. *Journal of personality and social psychology*, *89*(6), 925. doi:10.1037/0022-3514.89.6.925

Lambiotte, R., & Ausloos, M. (2006, May). Collaborative tagging as a tripartite network. In *International Conference on Computational Science* (pp. 1114-1117). Springer, Berlin, Heidelberg. doi:10.1007/11758532_152

Lancashire, I. (2012). Semantic drift in Shakespeare, and Early Modern English full-text corpora. In *English Corpus Linguistics: Crossing Paths* (pp. 181-196). Brill. doi:10.1163/9789401207935_010

Li, X., Guo, L., & Zhao, Y. E. (2008, April). Tag-based social interest discovery. In *Proceedings of the 17th international conference on World Wide Web* (pp. 675-684). doi:10.1145/1367497.1367589

Martin, J. R., & White, P. R. R. (2005). Enacting Appraisal: Text Analysis. In *The Language of Evaluation* (pp. 210-261). Palgrave Macmillan, London. doi:10.1057/9780230511910_5

Neuman, S. B., & Koskinen, P. (1992). Captioned television as comprehensible input: Effects of incidental word learning from context for language minority students. *Reading research quarterly*, 95-106. doi:10.2307/747835

Palmer, F. R., & Frank Robert, P. (1981). *Semantics*. Cambridge university press.

Plutchik, R. (1983). Emotions in early development: A psychoevolutionary approach. In *Emotions in early development* (pp. 221-257). Academic Press. doi:10.1016/b978-0-12-558702-0.50014-5

Plutchik, R. (2001). 'The Nature of Emotions'. *American Scientist 89*(4), 344. doi:10.1511/2001.28.739

Polanyi, L., & Zaenen, A. (2006). Contextual valence shifters. In *Computing attitude and affect in text: Theory and applications* (pp. 1-10). Springer, Dordrecht. doi:10.1007/1-4020-4102-0_1

Rambund, M. G. (2012). *Basic Semantics*. Editorial UNED.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological review*, *110*(1), 145-172. doi:10.1037//0033-295x.110.1.145

Sandamirskaya, Y. (2013). Report about the 35th annual meeting of the cognitive science society. *Cognitive Processing 14*(4), 443-444. doi:10.1007/s10339-013-0579-8

Scherer, K. R., & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of personality and social psychology*, *66*(2), 310-328. doi:10.1037//0022-3514.66.2.310

Topal, K., & Ozsoyoglu, G. (2016, August). Movie review analysis: Emotion analysis of IMDb movie reviews. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 1170-1176). IEEE. doi:10.1109/asonam.2016.7752387

Webb, S. (2010). A corpus driven study of the potential for vocabulary learning through watching movies. *International Journal of Corpus Linguistics*, *15*(4), 497-519. doi:10.1075/ijcl.15.4.03web