

# A Rule Based System to Identify Negative Sentences from Text Documents

Amarjeet Kaur, M. Sasikumar

**Abstract**— the paper analyzes how the negative morphemes can change the semantic meaning of a phrase/sentence and the ways to identify it. In many cases, negation in a sentence/phrase signifies the presence of negative morphemes (grammatical morphemes which can be free or bound) in a sentence. The proposed approach is to check the suffix or prefix of phrase, antonym-of and presence of negative particle in the grammatically correct sentences. The proposed technique is divided into three tasks: first, to apply POS (parts-of-speech) tagging technique in the grammatically correct text documents and second, to identify negative morphemes in a sentence, and third, to count the frequency of negative morphemes.

**Index Terms**— linguistic unit, negative morphemes, POS tagging, rule based system, semantic meaning

## I. INTRODUCTION

Grammatical morphemes are not only expressed by words but by suffix, prefix or particles. Grammatical morphemes are classified into inflectional and derivational morphemes where negative morphemes falls under the category of derivational morphemes are our area of discussion in this paper. Negative morphemes can be free or bound, if we will consider these morphemes such as not, never, none, no, out, less, neither, nor etc. these comes under the category of free and bound negative morphemes both, but ‘un’, ‘dis’, ‘anti’, ‘non’, ‘mis’, ‘in’, ‘im’, ‘ir’, ‘il’ are negative bound morphemes which standalone can’t generate a meaningful unit.

A grammatical morpheme indirectly makes an impact on the semantic meaning of the sentence. In this paper, we will discuss about how the negative morphemes can change the semantic meaning of the whole sentence and proposed a rule based system to identify it.

Hypothesis is: negation signifies the presence of negative linguistic unit in a sentence. The occurrence of negative morphemes and frequency of it can change the semantic meaning of the sentence. Odd number of occurrences of negative linguistic unit in a sentence can negate a sentence where double negation can makes the sentence a positive according to the law of Boolean algebra, except in the cases

*Date on which the paper has submitted : Tuesday, January 2, 2018*

*Author<sup>1</sup> Amarjeet kaur is currently pursuing Ph.D. degree in computer science & technology, Usha Mittal Institute of Technology, SNDT Women’s University, Juhu-Tara Road, Mumbai, India.*

*Author<sup>2</sup> Dr. M Sasikumar is currently working as a Director, Center for Development of Advanced Computing (CDAC), Kharghar, Navi Mumbai, India.*

of correlative conjunction like neither-nor which makes a negative sentence.

In some of the exceptional cases, like in-built, in-born, in-genius, in-deed, il-luminate, im-pact where ‘in’, ‘il’, ‘im’ are negative prefix but the words are not negative words, this shows that a negative morphemes not always make the word/phrase negative.

The approach is to check the suffix or prefix of the verb phrase, prepositional phrase and conjunction phrase if it is negative, then it signifies the presence of negation in a sentence in which it is used. But not valid for noun (NP), adjective (ADJP) and adverb phrases (ADVP). For parts of speech like NP, ADJP and ADVP, we will check the ‘antonym of phrase’ which will signify its opposite meaning in a sentence. If we will consider the above mentioned examples then built, born, genius, deed, pact are all either noun or adjective phrases. So these words are not considered as negative words. The word/phrases should be a meaningful unit whose affixes are going to be checked like if we will consider the example of the word ‘understand’, here ‘un’ is a negative prefix but ‘derstand’ is not a meaningful word/phrase. So, in this case the above hypotheses are not valid.

## II. LITERATURE REVIEW

According to Longman dictionaries of language teaching and applied linguistics, negation is contradicting the meaning or part of meaning of a sentence [3]. Various notions have been given by the researchers and each notion signifies the way to define negation and negative sentence according to their perception. *Negative sentence* is an also a way of conveying that something is ‘not correct’ or ‘false’. In many cases, presence of negative morphemes in a sentence signifies negative phrase/sentence.

From early research work about grammatical morphemes to recent publications about the analysis of negative morphemes, a number of researchers analyzed about the negative linguistic units used in various languages like Chinese, Japanese, French and English also.

Some of the research works are, a method based on Turney’s model [8], for inferring the Sentiment orientation of a word from its statistical association with strongly-polarized words and morphemes in Chinese. Extracting morphemes to improve sentiment analysis for the Chinese fund reviews, by replacing words by morphemes can improve system capability in Chinese language environment [1]. The study of negation and negative morphemes played an important role in the analysis of natural and formal linguistic system.

### III. METHODOLOGY

The approach used for the identification of the negation in a sentence is to check the presence of negative bound/free morphemes by analyzing:

- Affixes of verb phrase (VP), prepositional phrase (PP) and conjunction phrase (CP).
- Antonym-of noun (NP), adjective (ADJP) and adverb phrases (ADVP).
- Presence of negative particles like not, none, never etc. in a sentence.

The occurrences of negative affixes and particles in a sentence are checked to conclude the semantic meaning of a sentence.

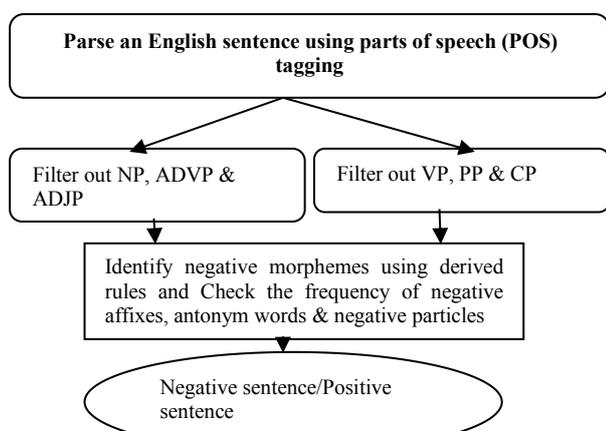


Figure 1: The Broad Approach

Table 1: List of some negative prefix with example words

	Negative prefix	Wordlist
1	Mis	Mislead, misplace, misunderstood, mistranslate, misinform, misconduct, mistaken, misdiagnose, misspell, mistrust
2	Non	Nonsense, nonfiction, nonexistent, nonentity, nonresident, nonstop
3	Un	Undo, unequal, unusual, unkind, uncomfortable, unsuccessful, uninformed, unknown, unnatural, unfriendly, unwilling, unpleasant, unclear, unafraid, unstable, uncertain, untouched
4	Dis	Disagree, disappear, dishonest, disloyal, dissimilar, discontented, disaffected
5	Mal	Malfunction, malnutrition
6	Anti	Anti-against, antiwar, antisocial, anti-ragging
7	No	Nothing, no one
8	In	Inactive, invisible, inadequate, incomplete, instability
9	Im	Impossible, improper, impolite, imbalance, immature, impatient, imperfect
10	De	Deactivate, deform, decode, deconstruct, decontaminate, Decrease, derail, demythologize, decommission, debug,

	Negative prefix	Wordlist
		defrost, devalue
11	il	Illegal, illogical, illiterate

Table 2: List of some negative suffix with example words

	Negative suffix	Wordlist
1	Not	Do+not Does+not Can+not Did+ not
2	Out	without
3	Less	Useless, Harmless, Careless, Friendless, Tireless, ageless, beardless, shameless, Blameless, breathless, Childless, Cloudless, Fearless, Hopeless, Helpless, Homeless, Jobless, painless, sleepless, restless, worthless

Table 3: Negative morphemes with positive meaning

Inbuilt	Irradiate	Illustrious	Inside
Inborn	Impact	Intoxicate	Import
Ingenious	Irruption	Inflection	insert
income	illuminate	harmless	
Inflammable	Indoor	Inrush	
Inhabitable	Inlet	Inset	
Priceless	Inlay	insight	
Input	inroad	Into	

Example sentences consist of negative particle:

- Today, i am *not* going to school.
- She is *not* at home.
- I do *not* see him.
- I *never* go there.
- *None* of them were good.
- *No* children were present in the building.

Example sentences of double negation:

Double negation can occur in the sentence in many ways like negative particle and negative affixes or antonym of the word in the sentence, together makes the sentence positive but some exceptional cases are still there like correlative conjunction etc.

1. The performance was *not bad*.  
Or  
The performance was good.
2. I am not going nowhere.  
Or  
I am going somewhere

Example of correlative conjunction violates the rule of double negation like:

- I found pen *neither* on the table *nor* inside the drawer.

A. Rules derived

Any sentence(s) consists of phrases/words where phrases can be noun phrase(NP), adjective phrase(ADJP), adverb phrase(ADVP), verb phrase (VP), prepositional (PP), conjunction (CP) etc. and a negative sentence with some negative morphemes in addition with all parts of speech.

Some rules are derived to identify the negative sentences:

**Rule 1:** For the entire Verb Phrase (VP), Prepositional (PP), Conjunction (CP) in the sentence(s), identify and calculate frequency of negative morphemes, particles and antonym

If ((suffix-of (VP, PP, CP) == Not, Out, Less) or

(Prefix-of (VP, PP, CP) == No, Mis, Non, Un, Dis, Mal, Anti, Not, In, Im, De))

**Rule2:** For all the noun phrase (NP), adjective phrase (ADJP), adverb phrase (ADVP), in the sentence(s), check

If (antonym-of (NP, ADVP, ADJP) == True)

**Rule 3:** For negative- particle in the sentence(s), check

If (constituent-of(s) == no, never, none, not)

**Rule 4:** For correlative conjunction in the sentence(s), check

If (constituent-of(s) == neither-nor, neither, nor)

**Frequency Check:**

If (frequency of negative morphemes, particles and antonym == ODD number)

*it's a negative sentence*

Else

*it's a positive sentence*

IV. IMPLEMENTATION

The rules are derived (please refer section III.A) to identify negative sentences are implemented using PROLOG programming language.

A. The Data

The samples are descriptive answers written by the students in the examinations. Samples of 200 descriptive answers are collected. The question is: “What is distributed computing system?”

Assuming that, descriptive answers are grammatically correct and filtered with spell checker.

Let’s consider three samples of students’ descriptive answers:

**Sentence (s1):** A distributed computing system is *not* a collection of *dependent* computers linked by a computer network that appear to the users of the system as a single coherent system.

**Sentence (s2):** A distributed computing system is *not* a collection of *independent* computers linked by a computer network that appear to the users of the system as a single coherent system.

**Sentence (s3):** A distributed computing system is a collection of dependent computers linked by a computer network that appear to the users of the system as a single coherent system.

B. Implementing Rule Based System

**For sentence s1,**

**Step 1:** Parse an English sentence (s1) using parts of speech tagging.

(s2 (NP A distributed computing system)  
(VP is not  
(NP (NP a collection)  
(PP of  
(NP (NP (NP dependent computers)  
(VP linked  
(PP by  
(NP a computer network))))))  
  
(SBAR (WHNP that)  
(S (VP appear))))))  
(PP to (NP (NP the users)  
(PP of  
(NP the system))))  
(PP as  
(NP a single coherent system))))

**Step 2:** FILTER out the phrases and Check the occurrences of negative affixes, antonym words & negative particles

NP, ADJP, ADVP	VP, PP, CP
A distributed computing system	Is not
a collection	Of
dependent computers	Linked
a computer network	By
the users	Appear
the system	To
a single coherent system	Of
	as

Rules 2 and 3 are applicable in s1.

There are 2 types of negation in the sentence s1:

1. ‘not’ particle
2. ‘antonym- of’ NP

Frequency of negative morphemes is 2, so the double negation

will signify a positive sentence using rule of Boolean algebra,  
Hence its positive sentence

**For sentence s2,**

**Step 1:** Parse an English sentence s2 using parts of speech tagging.

(S (NP A distributed computing system)  
(VP is not  
(NP (NP a collection)  
(PP of  
(NP (NP (NP independent computers)  
(VP linked  
(PP by  
(NP a computer network))))  
(SBAR (WHNP that  
(S (VP appears)))))))))  
(PP to  
(NP (NP the users)  
(PP of  
(NP the system))))  
(PP as  
(NP a single coherent system))))

**Step 2:** FILTER out the phrases and Check the occurrences of negative affixes, antonym words & negative particles

NP, ADJP, ADVP	VP, PP, CP
A distributed computing system	Is not
a collection	Of
independent computers	Linked
a computer network	By
the users	Appear
the system	To
a single coherent system	Of
	as

Rule 3 is applicable in sentence s2.

Types of negation in the sentence s2 are:

1. 'not' particle

Occurrence of negative particle in odd number of times will signify a negative sentence using rule of Boolean algebra, hence its negative sentence.

**For sentence s3,**

**Step1:** Parse an English sentence using parts of speech tagging.

(S (NP A distributed computing system)  
(VP is  
(NP (NP a collection)  
(PP of  
(NP (NP (NP dependent computers)  
(VP linked  
(PP by  
(NP a computer network))))  
(SBAR (WHNP that  
(S (VP appears)))))))))  
(PP to (NP (NP the users)

(PP of  
(NP the system)))  
(PP as (NP a single coherent system))))

**Step2:** FILTER out the phrases and Check the occurrences of negative affixes, antonym words & negative particles

NP, ADJP, ADVP	VP, PP, CP
A distributed computing system	Is
a collection	Of
dependent computers	Linked
a computer network	By
the users	Appear
the system	To
a single coherent system	Of
	as

Rule 2 is applicable in sentence s3.

Type of negation in the sentence s3:

1. 'antonym-of' NP

Occurrence of negation in odd number of times will signify a negative sentence using rule of Boolean algebra, hence its negative sentence.

Let's consider some more examples s4, s5, s6 and s7.

Sentence (s4): Distributed computing system involves multiple processors accessing *without* shared memory.

Sentence (s5): Distributed computing system involves multiple processors accessing *no* shared memory.

Sentence (s6): Distributed computing system involves multiple processors accessing *unshared* memory.

Sentence (s7): Distributed computing system involves multiple processors accessing *without unshared* memory.

Sentence s4 has prepositional phrase (with) 'out' as a suffix which negates the whole sentence. Sentence s5 has 'no' negative particle which changes its semantic meaning, s6 has antonym of 'shared' in its sentence which makes it a negative sentence. Sentence s7 has negative suffix 'out' and antonym of 'shared' both, hence it's a positive sentence.

V. RESULTS

The hypotheses are validated by considering 200 samples of sentences (students' descriptive answers). The research study shows that the presence of negative morphemes and its frequency actually changes the semantic meaning and polarity of the whole sentence. Samples of 200 sentences are collected and considered for implementation, out of which 20 sentences are violating the rules. It has been observed through implementation of the rule based system that the accuracy of the system is 90%. The derived rules can identify negative sentences.

VI. CONCLUSION & FUTURE SCOPE

A negative sentence in a textual document can be identified

by presence of negative morphemes, negative particles and antonym in a sentence. The occurrence and its frequency help us in detecting the negative sentence. In many cases presence of opposite meaning word also indicates the negation in a sentence. In further work, various types of negation in a textual document can be identified and implemented.

#### ACKNOWLEDGMENT

We would like to thank Dr. Sanjay S. Pawar, Principal, Usha Mittal Institute of Technology, SNDT Women's University, Mumbai, for guidance and support. This research was supported by The Department of Science & Technology, Ministry of Science & Technology, under Women Scientist Scheme (WOS-A) – SR/WOS-A/ET-1064/2014(G)

#### REFERENCES

- [1] Bo Wang and Xian Feng Liu (2010), "Extracting morphemes to improve sentiment analysis for Chinese fund review", IEEE Youth Conference on Information Computing and Telecommunications (YC-ICT), pp. 226 – 229
- [2] Daniel G. Morrow (1986), "Grammatical Morphemes and Conceptual Structure in Discourse Processing", Cognitive Science-A Multidisciplinary Journal, Vol. 10, no. 4, pp. 423-455
- [3] Jack C. Richards, and Richard W. Schmidt (2010), "Longman dictionaries of language teaching and applied linguistics", Publisher: Longman, 4th edition, pp.354
- [4] Lawrence R Horn, and Yasuhiko Kato (2000), "Introduction: Negation and Polarity at the Millennium", Available: [https://www.researchgate.net/publication/255592030\\_Introduction\\_Negation\\_and\\_Polarity\\_at\\_the\\_Millennium](https://www.researchgate.net/publication/255592030_Introduction_Negation_and_Polarity_at_the_Millennium)
- [5] Raymond W. M. Yuen, Terence Y. W. Chan, Tom B. Y. Lai, O. Y. Kwong, and Benjamin K. Y. T'sou (2004), "Morpheme-based derivation of bipolar semantic orientation of Chinese words", *ACM-Digital Library, COLING '04 Proceedings of the 20th international conference on Computational Linguistics*, Article No. 1008
- [6] Agarwal B., and Mittal N.(2014), "Semantic Feature Clustering for Sentiment Analysis of English Reviews", *IETE Journal of Research*, Vol. 60, no. 6, pp. 414-422
- [7] Turney P.D (2002), "Thumbs up or Thumbs down? Semantic Orientation Applied to Unsupervised Classification of Reviews", *Proceedings of the Association for Computational Linguistics*, University of Pennsylvania, Philadelphia, PA, USA
- [8] Turney P.D., and Littman M.L. (2003), "Measuring Praise and Criticism: Inference of Semantic Orientation from Association", *ACM Transactions on Information System (TOIS)*, Vol. 21, no. 4, pp. 315-346
- [9] Porter M. F. (1980), "An algorithm for suffix stripping", *Program*, Vol. 14 Issue: 3, pp.130-137
- [10] Balyan A., Dev A., Kumari R (2016)., and Agrawal S.S., "Labelling of Hindi Speech", *IETE Journal of Research*, Vol. 62, no. 2, pp. 146-153

- [11] "English for Students", Available: <http://www.english-for-students.com/Negative-Prefixes.html>
- [12] "Grammatical Morphemes", Available: <http://www.departments.bucknell.edu/linguistics/lectures/gramorph.html>
- [13] "Ling110ChapterV:Structure1", Available: [www.sfu.ca/~ramccall/Ling110ChV.pdf](http://www.sfu.ca/~ramccall/Ling110ChV.pdf)
- [14] Davy Temperley, Daniel Sleator, and John Lafferty, "Link Grammar", Available: <http://www.link.cs.cmu.edu/link/submit-sentence-4.html>



Amarjeet Kaur, received the B.E. degree in computer science & engineering from Pt. Ravishankar Shukla University, Raipur in 2006 and the M.E. degree in computer technology & application from Chattisgarh Swami Vivekanand Technical University, Bilhail in 2009.

He is currently pursuing the Ph.D. degree in computer science & technology at SNDT Women's University, Mumbai. From 2006 to 2009, she was a lecturer with the Shri Shankaracharya College of Engineering & Technology, Bilhail. From 2010 to 2013, she was an assistant professor with the MGM college of engineering, Mumbai. Her research interest includes the development of automated system for the evaluation of students' descriptive answers and to work in a domain of data science.

Her awards and honors include the Women Scientist Fellowship (Department of Science & Technology, Ministry of Science & Technology, and New Delhi). She was also awarded with Gold Medal in graduation and secured a top rank in university during post graduation. Her research paper titled "A comparative analysis of various approaches for automated assessment of descriptive answers" got a best research paper award in IEEE international conference 2017.



M. Sasikumar is currently Associate Director at C-DAC Mumbai. He heads the educational technology and knowledge based systems groups there. He has guided 5 PhD theses. He is main author of a book on parallel processing published by Prentice Hall. He has about 100 publications in various national and international forums. He is an alumnus of IIT Chennai, IISc Bangalore and BITS

Pilani. He is also an adjunct professor at IIT Bombay.