International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Designing of a Novel Framework for Marathi Natural Language Processing: MR-LIWC2015

Saroj Date[1], Sachin N. Deshmukh[2], Ryan Boyd[3], Ashwini Ashokkumar[4], James W. Pennebaker[5]

Submitted: 01/11/2023 **Revised**: 20/12/2023 **Accepted**: 02/01/2024

Abstract: The role of linguistic analysis in understanding human behaviour, emotions, and psychological states has gained significant prominence in various domains, including psychology, social sciences, and computational linguistics. The Linguistic Inquiry and Word Count (LIWC) is a widely used tool, developed by American social psychologist James W. Pennebaker and team of the University of Texas, Austin, enables automated linguistic analysis of text. This analysis provides insights into psychological and emotional dimensions. However, its applicability has been mainly restricted to English and a few other languages, limiting its usage in multilingual contexts. Originally developed in English, it has been adapted to several other languages like German, Dutch, Spanish, Chinese, Turkish, French, etc. However, this tool is not yet available for Marathi language- a major language spoken by people of Maharashtra, India. This paper presents a novel framework for the development and evaluation of a Marathi translation of the LIWC dictionary, aiming to expand its utility to the Marathi speaking population. The development process of Marathi version of LIWC is based on English LIWC-2015. The work is unique since it is the first LIWC translation for any Indian language. The development of Marathi version of LIWC includes several steps like initial translation and wildcard(*) expansion, dictionary expansion , linguistic analysis , wordlist development ,cultural adaptation ,wordlist validation process , refinement phase , equivalence research, addition of summary variables and wrap-up final dictionary in official LIWC format. The evaluation of the Marathi LIWC is conducted on a diverse dataset of Marathi text samples, encompassing social media posts, speech transcripts, blogs, short stories and book summaries. The performance of the translated dictionary is assessed based on its ability to accurately capture linguistic features, emotional tones, and psychological constructs present in the Marathi language. To evaluate the effectiveness of the Marathi LIWC, a diverse dataset of Marathi texts was analyzed using both the original English LIWC and the newly developed Marathi LIWC. The results of the evaluation demonstrate that the Marathi LIWC maintains its alignment with the original LIWC's underlying linguistic and psychological dimensions while catering to the specifics of the Marathi language. The translated dictionary exhibited promising reliability and validity in capturing linguistic and psychological features within Marathi texts.

Keywords: *English LIWC, LIWC, Marathi, Marathi LIWC, Marathi translation, NLP, Natural language processing, Sentiment analysis, translation, translation procedure, translation process*

1. Introduction

For the linguistic, psychological and emotional analysis of text data, Linguistic Inquiry and Word Count (LIWC) have proven to be a valuable lexical resource. It has gained popularity for its ability to assess psychological and emotional aspects by analyzing text samples. American social psychologist James W. Pennebaker and team of the University of Texas, Austin, developed this lexical resource called as LIWC for English language text analysis [1]. It is a widely-used resource that enables academicians and researchers to analyze written texts. The English version of LIWC has played a significant role in various fields, including psychology, social sciences, marketing, computational linguistics, and similar other fields. Originally developed in English, it has been translated to several other languages like German, Dutch, Spanish, Chinese, Turkish, and French. However, this tool is not yet available for Marathi language- a major language spoken by people of Maharashtra, India. This paper aims to describe the process of translation and evaluation of a Marathi version of LIWC-2015 , officially called as MR-LIWC2015 dictionary, which would enable researchers to analyze Marathi language texts. As analyzed from the literature survey, it is the first LIWC translation for any Indian language.

The rest of the paper is structured as follows. Related work describing LIWC construction to other languages is discussed in Section 2. Section 3 describes the detailed process of building a Marathi version of LIWC-2015 dictionary. Section 4 illustrates experimental work using

[1] *Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhajinagar, Maharashtra, India*
ORCID ID: 0000-0003-3393-884X
[2] *Dr. Babasaheb Ambedkar Marathwada University, Chh. Sambhajinagar, Maharashtra, India*
ORCID ID: 0000-0002-1212-3118
[3] *University of Texas at Austin, USA*
ORCID ID: 0000-0002-1876-6050
[4] *University of Texas at Austin, USA*
[5] *University of Texas at Austin, USA*
ORCID ID: 0000-0001-9091-214X
** Corresponding Author Email: saroj.date@gmail.com*

Marathi LIWC-2015. Section 5 is the conclusions and future scope.

2. Literature Review

The English LIWC dictionary has been translated into many languages all over the world, since it’s inception. These translations are based on English version of LIWC2001, LIWC2007 and LIWC2015. However, there is no publicly available translation of LIWC into Marathi. LIWC translation into other languages may give cross-cultural psychological findings [2]. The dictionaries are now available in fourteen languages, however not all of them have been tested and validated to the same level as the original English dictionary [3].

Carvalho et al. and Filho et al. have developed LIWC lexicon for Brazilian Portuguese version of LIWC 2007 & 2015 which makes use of 2007 and 2015 version of LIWC program respectively [4][5]. Huang et al. constructed and published the Chinese LIWC dictionary and establishes its reliability and validity [6]. Boot et al. and van Wissen & Boot presented the Dutch translation of LIWC 2007 to analyze Dutch language texts [7][8]. Piolat et al. translated English version of LIWC to French language [9]. Meier et al. and Wolf et al. introduced the DE-LIWC2015, a version of German adaption of LIWC. The work aimed to create an upgrade to the previous version of the German LIWC adaption that corresponded to the LIWC2015 properties [10][11]. Agosti & Rellini developed Italian version of LIWC [12]. Igarashi et al. translated Linguistic Inquiry and Word Count Dictionary 2015 (LIWC2015) to a Japanese version for performing NLP- natural language processing and cross-cultural research [13]. Goksøyr developed 2007 version of the Norwegian LIWC dictionary and manual [3]. Dudău& Sava translated and published Romanian version of LIWC2015 known as Ro-LIWC2015, and contributed to the field of multilingual analytic study [14]. Kailer & Chung developed and evaluated Russian version of LIWC [15]. Bjekić et al. created and validated a new adaptation of LIWC software for the Serbian language (LIWCser) [16]. Ramírez-Esparza et al. developed a Spanish version of LIWC2001 [17]. Müderrisoğlu translated LIWC to Turkish version [3]. Zasiekin et al build Ukrainian LIWC-2015 dictionaries based on the English version of LIWC-2015 Ukrainian. [18].

The summary of these translations is shown in Table 1.

Table 1. LIWC Dictionary Translations based on English version

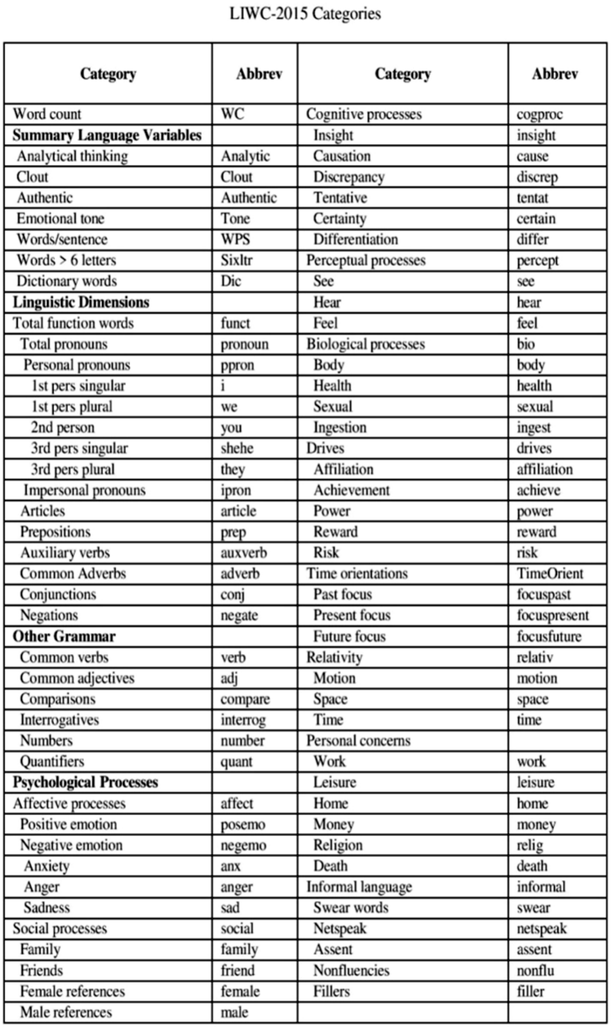
Sr. No.	LIWC Translation to other language
1	Brazilian Portuguese (Carvalho et al., 2019; Filho et al., 2013)
2	Chinese (Huang et al., 2012)
3	Dutch (Boot et al., 2017; van Wissen& Boot, 2017)
4	French (Piolat et al., 2011)
5	German (Meier et al., 2018; Wolf et al., 2008)
6	Italian (Agosti&Rellini, 2007)
7	Japanese (Igarashi et al., 2021)
8	Norwegian (Goksøyr, 2019)
9	Romanian (Dudău& Sava, 2020)
10	Russian (Kailer& Chung, 2011)
11	Serbian (Bjekić et al., 2014)
12	Spanish (Ramírez-Esparza et al., 2007)
13	Turkish (Müderrisoğlu, 2012)
14	Ukrainian (Zasiekin et al., 2018)

3. Methodology for Building Marathi Version of LIWC-2015

This section describes how LIWC works and calculates the scores of linguistic dimensions. Following this, the detailed process to translate English LIWC-2015 dictionary into Marathi is discussed.

3.1 Working of Linguistic Enquiry and Word Count Software (LIWC)

LIWC scans given text file/ files and compares each word from input files to a list of dictionary words, calculating the proportion of total words in the text that match each of the dictionary categories. For example, if LIWC used the built-in LIWC-15 lexicon to analyze a single speech comprising 1,000 words, it may discover that 50 of those phrases are associated to positive emotions and 10 words are related to affiliation. These figures would be converted to percentages by LIWC: 5.0% positive feeling and 1.0% affiliation [3]. Marathi Version of LIWC is built on the same logic as that of English LIWC. For every analyzed text file in LIWC2015, about 90 output variables are generated as one row of data into an output file. The list of output variables is given in Figure 1. It calculates various dimensions score, as described following this figure.



LIWC-2015 Categories

Category	Abbrev	Category	Abbrev
Word count	WC	Cognitive processes	cogproc
Summary Language Variables		Insight	insight
Analytical thinking	Analytic	Causation	cause
Clout	Clout	Discrepancy	discrep
Authentic	Authentic	Tentative	tentat
Emotional tone	Tone	Certainty	certain
Words/sentence	WPS	Differentiation	differ
Words > 6 letters	Sixltr	Perceptual processes	percept
Dictionary words	Dic	See	see
Linguistic Dimensions		Hear	hear
Total function words	funct	Feel	feel
Total pronouns	pronoun	Biological processes	bio
Personal pronouns	ppron	Body	body
1st pers singular	i	Health	health
1st pers plural	we	Sexual	sexual
2nd person	you	Ingestion	ingest
3rd pers singular	shehe	Drives	drives
3rd pers plural	they	Affiliation	affiliation
Impersonal pronouns	ipron	Achievement	achieve
Articles	article	Power	power
Prepositions	prep	Reward	reward
Auxiliary verbs	auxverb	Risk	risk
Common Adverbs	adverb	Time orientations	TimeOrient
Conjunctions	conj	Past focus	focuspast
Negations	negate	Present focus	focuspresent
Other Grammar		Future focus	focusfuture
Common verbs	verb	Relativity	relativ
Common adjectives	adj	Motion	motion
Comparisons	compare	Space	space
Interrogatives	interrog	Time	time
Numbers	number	Personal concerns	
Quantifiers	quant	Work	work
Psychological Processes		Leisure	leisure
Affective processes	affect	Home	home
Positive emotion	posemo	Money	money
Negative emotion	negemo	Religion	relig
Anxiety	anx	Death	death
Anger	anger	Informal language	informal
Sadness	sad	Swear words	swear
Social processes	social	Netspeak	netspeak
Family	family	Assent	assent
Friends	friend	Nonfluencies	nonflu
Female references	female	Fillers	filler
Male references	male		

Fig 1. List of LIWC-2015 output variables

Score calculation by Marathi LIWC Program

LIWC calculates scores for linguistic dimensions using a set of linguistic rules and a dictionary containing words categorized into different dimensions.

- ***Start LIWC Application for text processing*:** As LIWC application is started, the predefined LWC dictionary runs in the background.
- ***Resultant Score storage preparation*:** Initialize an empty dictionary to store dimension scores, initialize all dimensions scores as zero.
- ***Text Preparation*:** Load the input text files. The input text files are preprocessed to remove punctuation, special characters, and formatting. The text is split into individual words or tokens.
- ***Dictionary Lookup*:** Each token is looked up in the LIWC dictionary to determine which linguistic dimensions are associated with it. The dictionary contains entries for words along with their associated dimensions and scores.
- ***Dimension Assignment and Scoring*:** For each token, LIWC identifies the dimensions associated with that token based on the dictionary lookup. The software assigns the token's corresponding dimensions with the scores specified in the dictionary.
- ***Score Aggregation*:** The scores for each dimension are accumulated based on the tokens that were assigned to those dimensions. If a token is associated with multiple dimensions, its score is added to the scores of each relevant dimension.
- ***Normalization*:** The aggregated scores are normalized to ensure that they sum up to a particular value (e.g., 100) or to bring them onto a common scale. Normalization helps in comparing and interpreting the dimension scores effectively.

Higher scores for a particular dimension indicate a greater presence of linguistic elements related to that dimension in the text.

Pseudo code for score calculation by Marathi LIWC Program

```
//  LIWC dictionary containing words/phrases associated with each dimension
LIWC_Dictionary = {
   Dimension1: ["word1", "word2", ...],
   Dimension2: ["word3", "word4", ...],
   …..
   …..
Dimension n: ["word  ", "word ", ...]
}
// Initialize an empty dictionary to store dimension scores
dimension_scores = {
    Dimension1: 0,
    Dimension2: 0,
…..
….
    Dimension n: 0,
  }
//Tokenize the input text into words or tokens
input_text = "..."
tokens = tokenize(input_text)
// Iterate through each token in the input text
for token in tokens:
//Check if the token is present in the LIWC dictionary
for dimension, word_list in LIWC_Dictionary.items():
if token in word_list:
dimension_ dimension scores[] += 1
//Normalize dimension scores by the total number of tokens
total_tokens = len(tokens)
for dimension in dimension_scores:
normalizedScores[dimension] = (dimension_scores / total Score) * 100
```

3.2 Translation Process of English LIWC to Marathi Version

Translating English LIWC-2015 to Marathi LIWC involves creating and evaluating Marathi word lists for the various linguistic dimensions that the English LIWC covers. It’s a semi-automatic approach involving Python programming language scripts, use of appropriate tools for tasks like data scraping, POS tagging etc. We closely studied the research articles (mentioned in section 2 of this paper) on LIWC dictionary translations. By referring other researcher's work, a step-by-step process to develop Marathi version of LIWC-2015 is designed. The complete overview of the translation process is presented here.

- ***Translation and wildcard(*) expansion****:* For the development of Marathi LIWC-2015, English LIWC-2015 version is used as a starting point. All words from the English LIWC 2015 dictionary are directly translated into Marathi using Google Translate. The purpose of this step is to create a basic working version of a new Marathi dictionary that will be used in later stages of the work. English LIWC contains several words ending in an asterisk (*).It is used to denote English words in a text that begins with the preceding string. All these words are expanded to see their original form.

- ***Dictionary expansion****:* The basic version of Marathi Dictionary is expanded by adding more words from IndoWordnet and supplemented with synonyms of the words. Also more words were added by collecting a large corpus of Marathi text data. The corpus includes a diverse range of topics like Marathi blogs, short stories, news articles, social media posts, books. This is with the aim of improve the Dictionary. IndowordNet is basically a linked lexical knowledge base. It consists of WordNets of eighteen scheduled languages of India, including Marathi Language. It is created by Indian Institute of Technology Bombay, Maharashtra, India [19]. The duplicate words are removed and remaining words are put in alphabetical order.

- ***Linguistic analysis*** *:* Next, linguistic analysis tools such as part-of-speech taggers are used to extract linguistic features from the corpus. This step involves identifying the specific linguistic categories that are relevant for Marathi text analysis.

- ***Wordlist development****:* Once the relevant linguistic categories have been identified, next step is to develop a Marathi language dictionary that includes words corresponding to each category.

- ***Cultural adaptation:*** Cultural adaptation is a crucial aspect of the translation process. Marathi version of LIWC was modified to accommodate the linguistic and cultural aspects specific to Marathi language. Marathi has its unique grammatical rules, sentence structures, and vocabulary, which must be appropriately incorporated into the dictionary. Validation team members with a deep understanding of Marathi linguistics contribute to this adaptation process.

- ***Wordlist validation process:*** Once the Marathi wordlist for LIWC dictionary has been created, it is important to refine and validate it. The process involved assessing inter-rater agreement to confirm that the Marathi version of LIWC accurately captures the intended linguistic and psychological dimensions. The translated Marathi LIWC version is evaluated by a panel of three members who are proficient in Marathi language from Marathi Language Department. They review the dictionary entries to verify their accuracy, cultural relevance, and appropriateness for Marathi text analysis. In order to best guarantee the efficiency of Marathi version of LIWC dictionary, each lexical item were checked and validated manually. The whole wordlist was validated by the panel members. In order for a word to remain in a given category, a majority of members had to agree on its inclusion. Words for which members could not decide on appropriate category placement were removed from the dictionary. According to the suggestions given by them, the dictionary is again revised. To calculate the inter-rater agreement, Fleiss' kappa statistical measure is used. This contrasts with other kappas such as Cohen's kappa, which only work when assessing the agreement between not more than two raters or the intra-rater reliability and agreement [20]. The Fleiss Kappa score for Marathi LIWC was calculated to indicate inter-rater agreement. It is defined as

$$K = (Po - Pe)/(1 - Pe) \qquad (1)$$

The degree of agreement that is attainable above chance is calculated as $1 - Pe$. The term $Po - Pe$ indicates the degree of agreement actually achieved above chance. If $K = 1$, it shows that the raters are in complete agreement. If there is no agreement among the raters (other than what would be expected by chance) then $K \leq 0$. Following Table 1 shows the range for inter-rater agreement for Fleiss Kappa calculations.

Table 2. Meaning of Fleiss Kappa range

k-Value	Meaning
0.8 < k < =1	Perfect agreement
0.6 < k < =0.8	Substantial agreement
0.4 < k < = 0.6	Moderate agreement
0.2 < k < = 0.4	Fair agreement
0 < = k < = 0.2	Slight agreement
k < 0	Poor agreement

For Marathi LIWC, agreement score is 0.96, Po Value is 0.95 and Pe value is 0.92. The calculation for Fleiss Kappa gives the score as 0.375, which shows fair agreement.

- ***Refinement phase**:* After the preceding rounds were completed, we partially repeated them in an iterative method to catch any potential errors or oversights in the dictionary construction process.
- ***Equivalence research**:* In order to check the equivalence between Marathi and English LIWC dictionaries, we collected the parallel corpora which exist in digital form in both English and Marathi. It aims to develop information-processing tools to facilitate human-machine interaction in Indian languages. English and Marathi versions of all texts are analyzed using English and Marathi LIWC dictionary respectively.
- ***Addition of summary variables:*** Once Marathi-LIWC2015 dictionary is finalized, four summary variables as that of in English LIWC2015 version has to be integrated into the Marathi LIWC system: Analytical thinking, Clout, Authenticity, and Emotional Tone. It is important to highlight that the summary variables are the only non-transparent dimensions in the LIWC output and that these variables have to be transformed to percentiles based on standardized scores from large Marathi text samples.
- ***Wrap-up final dictionary in official LIWC format**:* The dictionary must be saved in a format compatible with the official LIWC application once the translation and validation procedures are finished.

4. Experimental Work using Marathi version of LIWC-2015

This section describes the experimental work carried out using Marathi Version of LIWC. As discussed in Section 3, Marathi version of LIWC has been developed for Marathi text processing. It contains about 50,000 Marathi lexicons. The performance evaluation of Marathi version of LIWC is presented here.

Following experiments are performed using Marathi version of LIWC:

- Performance evaluation of Marathi Version of LIWC
- Quantifying the Internal Consistency Reliability of Marathi Version of LIWC-2015
- Comparing English LIWC2015 with Marathi LIWC2015

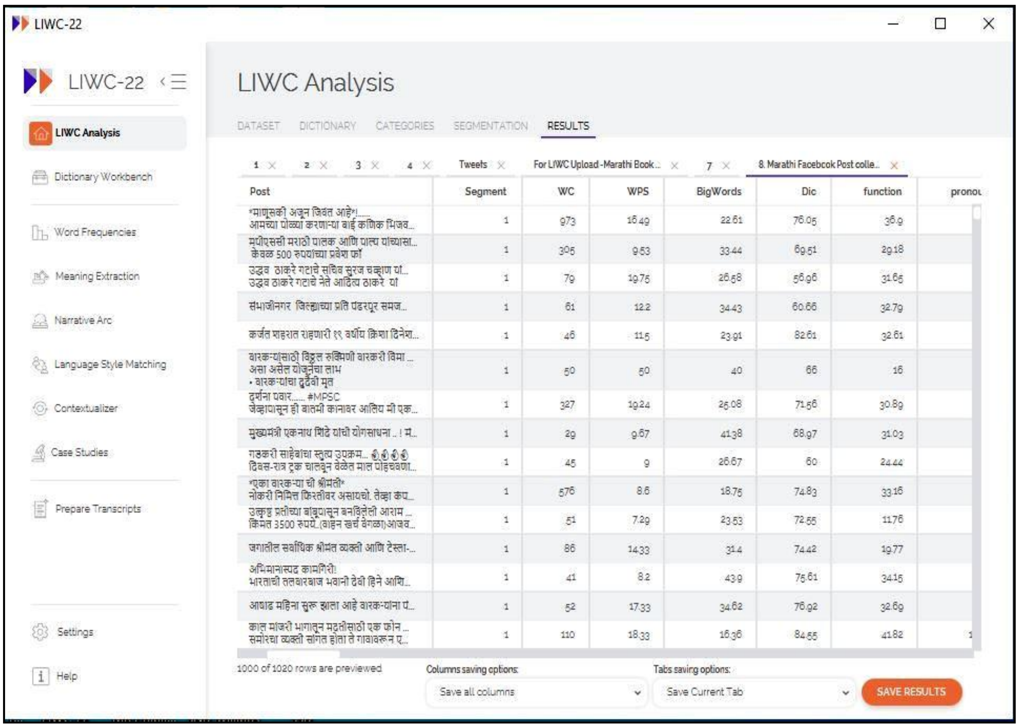
4.1 Performance evaluation of Marathi Version of LIWC

In order to check how much percentage of words Marathi LIWC captures, Marathi corpus of 2,087 Marathi text files was created. It covers texts from multiple domains like short stories, blogs, tweets, etc. The details of the corpus are given in Table 3.

Table 3. Marathi corpus of 7,23,918 Words

Marathi Corpus domain	Text files in each domain	Total Word Count	Percent of words captured by Marathi LIWC Dictionary
Marathi Short Stories	150	50,237	73.95 %
Speech Transcripts (Mann Ki Baat-Marathi Anuvad)	61	1,98,799	76.85 %
Marathi Blogs	200	1,57,482	73.79 %
Marathi News Articles	100	20,925	72.09 %
Marathi Tweets	1070	24,386	68.03 %
Marathi Book Summary	206	65,087	72.03 %
Marathi interviews	100	1,86,542	76.09 %
Facebook posts in Marathi	200	20,460	70.48 %
Total	**2087**	**7,23,918**	

All these text files were processed using Marathi LIWC. The screenshot of input and output are shown in Figure 2 and 3. LIWC software automatically analyzes the content of Marathi text files and calculates how people use different categories of words that they have used in their written communication like speeches, social media posts, summaries, blogs, applications, stories, etc. Technically, this software identifies and analyses about 90 dimensions of given files (.doc, docx, .pdf, .xls,.xlsx,.csv) within a very few seconds of time. Data obtained are easily transferable to files suitable for statistical analysis like Microsoft Excel, Statistical Package for the Social Sciences (SPSS).



LIWC Analysis

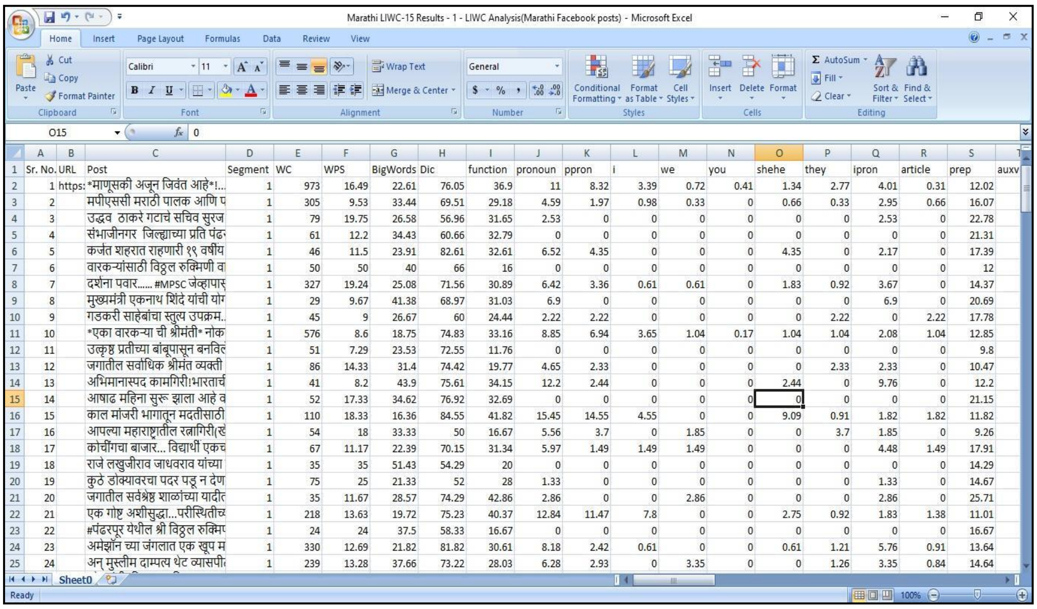
Post	Segment	WC	WPS	BigWords	Dic	function	pronou
माणूसकी अजून जिवंत आहे!...... आमच्या पोळ्या करणाऱ्या बाई कणिक भिजव...	1	973	16.49	22.61	76.05	36.9	
मपीएससी मराठी पालक आणि पाल्य यांच्यासा... केवळ 500 रुपयांच्या प्रवेश फॉ	1	305	9.53	33.44	69.51	29.18	
उद्धव ठाकरे गटाचे सचिव सुरज चव्हाण यां... उद्धव ठाकरे गटाचे नेते आदित्य ठाकरे यां	1	79	19.75	26.58	56.96	31.65	
संभाजीनगर जिल्ह्याच्या प्रति पंढरपूर समज...	1	61	12.2	34.43	60.66	32.79	
कर्जत शहरात राहणारी १९ वर्षीय क्रिषा दिनेश...	1	46	11.5	23.91	82.61	32.61	
वारकऱ्यांसाठी विठ्ठल रुक्मिणी वारकरी विमा ... असा असेल योजनेचा लाभ • वारकऱ्यांचा दुर्दैवी मृत	1	50	50	40	66	16	
दर्शना पवार...... #MPSC जेव्हापासून ही बातमी कानावर आलिय मी एक...	1	327	19.24	25.08	71.56	30.89	
मुख्यमंत्री एकनाथ शिंदे यांची योगसाधना .. ! मं...	1	29	9.67	41.38	68.97	31.03	
गडकरी साहेबांचा स्तुत्य उपक्रम... 👏👏👏👏 दिवस-रात्र ट्रक चालवून वेळेत माल पोहचवणा...	1	45	9	26.67	60	24.44	
एका वारकऱ्या ची श्रीमंती नोकरी निमित्त फिरतीवर असायचो. तेव्हा कप...	1	576	8.6	18.75	74.83	33.16	
उत्कृष्ठ प्रतीच्या बांबूपासून बनविलेली आराम ... किमत 3500 रुपये.(वाहन खर्च वेगळा)आजव...	1	51	7.29	23.53	72.55	11.76	
जगातील सर्वाधिक श्रीमंत व्यक्ती आणि टेस्ला-...	1	86	14.33	31.4	74.42	19.77	
अभिमानास्पद कामगिरी! भारताची तलवारबाज भवानी देवी हिने आशि...	1	41	8.2	43.9	75.61	34.15	
आषाढ महिना सुरू झाला आहे वारकऱ्यांना पं...	1	52	17.33	34.62	76.92	32.69	
काल मांजरी भागातून मदतीसाठी एक फोन ... समोरचा व्यक्ती सांगत होता ते गावावरून ए...	1	110	18.33	16.36	84.55	41.82	1

1000 of 1020 rows are previewed

Columns saving options: Save all columns

Tabs saving options: Save Current Tab

Fig 2. LIWC Screenshot: Input files processing



O15 | fx | 0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Sr. No.	URL	Post	Segment	WC	WPS	BigWords	Dic	function	pronoun	ppron	i	we	you	shehe	they	ipron	article	prep	auxv
2	1	https:	*माणूसकी अजून जिवंत आहे*!...	1	973	16.49	22.61	76.05	36.9	11	8.32	3.39	0.72	0.41	1.34	2.77	4.01	0.31	12.02	
3	2		मपीएससी मराठी पालक आणि प	1	305	9.53	33.44	69.51	29.18	4.59	1.97	0.98	0.33	0	0.66	0.33	2.95	0.66	16.07	
4	3		उद्धव ठाकरे गटाचे सचिव सुरज	1	79	19.75	26.58	56.96	31.65	2.53	0	0	0	0	0	0	2.53	0	22.78	
5	4		संभाजीनगर जिल्ह्याच्या प्रति पंढ	1	61	12.2	34.43	60.66	32.79	0	0	0	0	0	0	0	0	0	21.31	
6	5		कर्जत शहरात राहणारी १९ वर्षीय	1	46	11.5	23.91	82.61	32.61	6.52	4.35	0	0	0	4.35	0	2.17	0	17.39	
7	6		वारकऱ्यांसाठी विठ्ठल रुक्मिणी वा	1	50	50	40	66	16	0	0	0	0	0	0	0	0	0	12	
8	7		दर्शना पवार...... #MPSC जेव्हापास	1	327	19.24	25.08	71.56	30.89	6.42	3.36	0.61	0.61	0	1.83	0.92	3.67	0	14.37	
9	8		मुख्यमंत्री एकनाथ शिंदे यांची योग	1	29	9.67	41.38	68.97	31.03	6.9	0	0	0	0	0	0	6.9	0	20.69	
10	9		गडकरी साहेबांचा स्तुत्य उपक्रम..	1	45	9	26.67	60	24.44	2.22	2.22	0	0	0	0	2.22	0	2.22	17.78	
11	10		*एका वारकऱ्या ची श्रीमंती* नोक	1	576	8.6	18.75	74.83	33.16	8.85	6.94	3.65	1.04	0.17	1.04	1.04	2.08	1.04	12.85	
12	11		उत्कृष्ठ प्रतीच्या बांबूपासून बनवि	1	51	7.29	23.53	72.55	11.76	0	0	0	0	0	0	0	0	0	9.8	
13	12		जगातील सर्वाधिक श्रीमंत व्यक्ती	1	86	14.33	31.4	74.42	19.77	4.65	2.33	0	0	0	0	2.33	2.33	0	10.47	
14	13		अभिमानास्पद कामगिरी!भारताच	1	41	8.2	43.9	75.61	34.15	12.2	2.44	0	0	0	2.44	0	9.76	0	12.2	
15	14		आषाढ महिना सुरू झाला आहे व	1	52	17.33	34.62	76.92	32.69	0	0	0	0	0	0	0	0	0	21.15	
16	15		काल मांजरी भागातून मदतीसाठी	1	110	18.33	16.36	84.55	41.82	15.45	14.55	4.55	0	0	9.09	0.91	1.82	1.82	11.82	
17	16		आपल्या महाराष्ट्रातील रत्नागिरी(खे	1	54	18	33.33	50	16.67	5.56	3.7	0	1.85	0	0	3.7	1.85	0	9.26	
18	17		कोचींगचा बाजार... विद्यार्थी एकच	1	67	11.17	22.39	70.15	31.34	5.97	1.49	1.49	1.49	0	0	0	4.48	1.49	17.91	
19	18		राजे लखुजीराव जाधवराव यांच्या	1	35	35	51.43	54.29	20	0	0	0	0	0	0	0	0	0	14.29	
20	19		कुठे डोक्यावरचा पदर पडू न देण	1	75	25	21.33	52	28	1.33	0	0	0	0	0	0	1.33	0	14.67	
21	20		जगातील सर्वश्रेष्ठ शाळांच्या यादीत	1	35	11.67	28.57	74.29	42.86	2.86	0	0	2.86	0	0	0	2.86	0	25.71	
22	21		एक गोष्ट अशीसुद्धा...परीस्थितीच	1	218	13.63	19.72	75.23	40.37	12.84	11.47	7.8	0	0	2.75	0.92	1.83	1.38	11.01	
23	22		#पंढरपूर येथील श्री विठ्ठल रुक्मिण	1	24	24	37.5	58.33	16.67	0	0	0	0	0	0	0	0	0	16.67	
24	23		अमेझॉन च्या जंगलात एक खूप म	1	330	12.69	21.82	81.82	30.61	8.18	2.42	0.61	0	0	0.61	1.21	5.76	0.91	13.64	
25	24		अन् मुस्लीम दाम्पत्य थेट व्यासपी	1	239	13.28	37.66	73.22	28.03	6.28	2.93	0	3.35	0	0	1.26	3.35	0.84	14.64	

Fig 3. LIWC Screenshot: Output file

The domainwise results of Marathi LIWC processing are shown

in Figure 4. It has been observed that, it captured 72% of dictionary words in average.

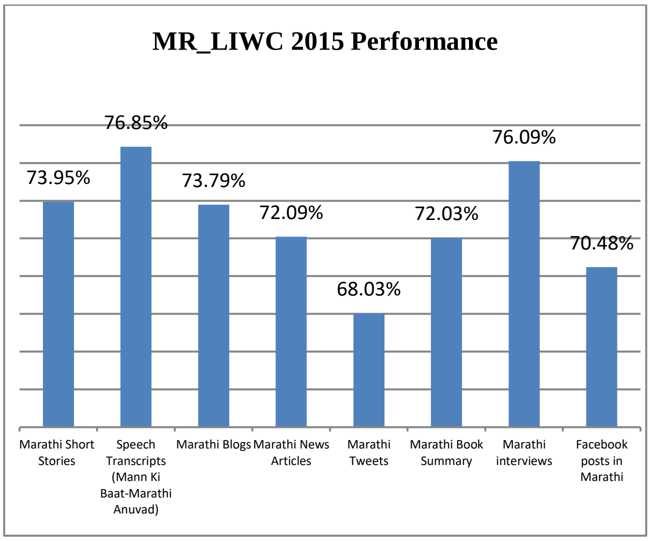


Fig 4. Performance evaluation of Marathi LIWC

4.2 Quantifying the Internal Consistency Reliability of Marathi Version of LIWC-2015

Another experiment has been carried out to quantify the internal consistency reliability of Marathi Version of LIWC-2015. Table 4 shows details of text files, used for this analysis.

Table 4. Marathi corpus of 7,23,918 Words with Mean and SD

Corpus	No.of Text files	Word Count	Mean(SD)
Marathi Short Stories	150	50,237	334.91 (236.67)
Speech Transcripts (Mann Ki Baat-Marathi Anuvad)	61	1,98,799	3259 (499.72)
Marathi Blogs	200	1,57,482	787.41 (270.99)
Marathi News Articles	100	20,925	209.25 (121.49)
Marathi Tweets	1070	24,386	22.81 (13.61)
Marathi Book Summary	206	65,087	315.95 (138.03)
Marathi interviews	100	1,86,542	1865.42 (1416.94)
Facebook posts in Marathi	200	20,460	102.30 (129.17)
	2087	**7,23,918**	**871.28 (346.47)**

Note: Each corpus is composed of a number of text files as shown above, with approximately 100 to 5,000 words.

To calculate the internal consistency reliability, Cronbach Alpha and Kuder-Richardson Formula (KR-20) are used as a statistical measures. The results of the evaluation are shown in Table 5.

As shown in above table 5, it is inferred that the translated dictionary exhibited promising reliability and validity in capturing linguistic and psychological features within Marathi texts.

4.3 Comparing English LIWC2015 with Marathi LIWC2015

Marathi version of LIWC is compared with the existing English version of LIWC. For this a parallel corpora of English-Marathi Text Files was used. It contains 1,525 text files from various domains like Marathi news, tweets, etc. The results of the evaluation are shown in Table 6.

As sown in above table 6, to evaluate the effectiveness of the Marathi LIWC, a diverse dataset of Marathi texts was analyzed using both the original English LIWC and the newly developed Marathi LIWC. The results of the evaluation demonstrate that the Marathi LIWC maintains its alignment with the original LIWC's underlying linguistic and psychological dimensions while catering to the specifics of the Marathi language. The comparison results indicate that 35% of the categories are highly correlated with pearson correlation score as 0.5 and above.

Most correlated Ten Categories: negate (0.85), i (0.79),we (0.75), number (0.73), ppron (0.73), affiliation (0.67), pronoun (0.66), female (0.65), percept (0.61), and see (0.61)

Less correlated Ten Categories: sexual (0.15), swear (0.14), feel (0.14), home (0.11), nonflu (0.05), reward (0.05), article (0.04), netspeak (0.03), informal (0.03), and filler (0.00).

5. Conclusions and Future Scope

In this paper, we described the process of developing a Marathi version of LIWC-2015 dictionary. Through rigorous procedure explained in section 3, we developed a version of Marathi LIWC, MR-LIWC2015, with more than 50,000 lexicons. Also we carried out the experimental work using Marathi version of LIWC. We analyzed the performance of Marathi Version of LIWC. It is observed that it captured 72% of the dictionary words. Through experimental work of 4.2, it is inferred that the translated dictionary exhibited promising reliability and validity in capturing linguistic and psychological features within Marathi texts. To evaluate the effectiveness of the Marathi LIWC, a diverse dataset of Marathi texts was analyzed using both the original English LIWC and the newly developed Marathi LIWC. The results of the evaluation demonstrate that the Marathi LIWC maintains its alignment with the original LIWC's underlying linguistic and psychological dimensions while catering to the specifics of the Marathi language.

The limitations of English LIWC are also applicable to Marathi version. Despite its widespread use, LIWC has some limitations. It relies on a predefined dictionary of words and may not capture new words which are not included in the dictionary. Another issue is that sarcasm and irony may be missed by LIWC since they involve using words with opposite meanings. LIWC program only counts words and does not take into account the context in which words are used. This may cause some bias in the results, especially for words that have multiple meanings.

Future research could focus on optimizing and expanding the dictionary of Marathi LIWC to include more lexicons in various linguistic categories and exploring its application in different domains, such as sentiment analysis, mental health research, and social media analysis. As this lexical resource is extensively validated, it is likely that it will become an increasingly important tool for researchers and practitioners working with Marathi language texts.

Table 5. Quantifying the Reliability of Marathi LIWC-2015 dimensions

Marathi LIWC -2015 Language Dimensions and Internal Consistency Reliability					
Category	**Abbrev.**	**Description/Most frequently used exemplars**	**Words in each category**	**Cronbach's Alpha (Raw)**	**Kuder-Richardson Formula 20 (KR-20)**
Linguistic Dimensions					
Total Function Words	function	आम्ही,आहे,जर-तर,तथापि,नको,होते	477	0.78	0.98
Total Pronouns	pronoun	आमचे,आम्हाला,ती,प्रत्येकजण,मी,आम्ही	108	0.66	0.94
Personal Pronoun	ppron	ती,तुम्ही,तू,ते	71	0.72	0.92
1st person singular	i	मला,माझे,मी	16	0.69	0.82
1st person plural	we	आपण,आपले,आम्ही	19	0.6	0.86
2nd person	you	तुझे,तुम्ही,तुला,तू	16	0.47	0.63
3rd person singular	shehe	तिचे,तिला,त्याचा,तो	19	0.63	0.78
3rd person plural	they	ते,त्यांचे,त्यांना	8	0.6	0.8
Impersonal pronoun	ipron	इतर,जी,ते,हे	37	0.22	0.85
Articles	article	एक	1	-0.01	0.09
Prepositions	prep	*आधी, *कडे, *खाली	124	0.61	0.93
Auxiliary Verbs	auxverb	असते,आहे,नको	107	0.3	0.92
Adverbs	adverb	सर्वदा,वारंवार,अविरत	1,136	0.67	0.96

Conjunctions	conj	आणि,किंवा,पण	42	0.37	0.83
Negation	negate	नका,नको,नाही	30	0.31	0.68
Common Verbs	verb	अनुभवणे,ऐकणे,कळणे	8,713	0.73	0.98
Common Adjectives	adj	अमूल्य,उर्वरित,क्रूर	9,932	0.7	0.98
Comparison	compare	अतिसुंदर,अत्याधुनिक	380	0.37	0.91
Interrogative	interrog	कधी,कसे,का,किती	103	0.39	0.76
Number	number	आठ,चार,त्रेसष्ट,सोळा	486	0.64	0.85
Quantitative	quant	अंतिम,अनेक,चतुर्थ	239	0.29	0.87
Psychological Processes					
Affective processes	affect	अत्यानंद,अनादरनीय	2745	0.59	0.97
Positive Emotion	posemo	आनंद,आवडणे,पवित्र	1426	0.59	0.97
Negative Emotion	negemo	अडथळा,किंचाळणे,चिंता	1328	0.43	0.86
Anxiety	anx	अकस्मात,अस्वस्थ,दुःखी	398	0.25	0.73
Anger	anger	अन्याय,निंदनीय,राग	513	0.44	0.78
Sadness	sad	खिन्न,दुःखद,निराश	192	0.32	0.59
Social Processes	social	आई,आप्तेष्ट,कुटुंब	1129	0.72	0.96
Family	family	आईबाबा,कन्या,गृहस्थ	419	0.57	0.85
Friends	friend	बालमित्र,मित्र,मैत्रिणी	69	0.21	0.66
Female references	female	कुमारी,तरुणी,लेखिका	310	0.68	0.84
Male references	male	भाऊ,युवक,शिक्षक	309	0.6	0.84
Cognitive processes	cogproc	इच्छिणे,कल्पना_करणे	744	0.66	0.97
Insight	insight	अंतर्दृष्टी,एकाग्रता,कल्पना	317	0.38	0.92
Causation	cause	अनुमान,नियंत्रण,रुकार	154	0.2	0.84
Discrepancy	discrep	अनिश्चित,उणीव,खेद	49	0.16	0.73
Tentative	tentat	अंदाज,अनिश्चित,अस्पष्ट	113	0.28	0.8
Certainty	certain	अचूक,आत्मविश्वास,खरे	109	0.27	0.86
Difference	differ	फरक,भिन्न,अपवाद	64	0.47	0.81
Perceptual processes	percept	अंधार,उग्र,ऐकणे,कुजट	512	0.37	0.87
See	see	गडद,चकाकी,जांभळा	157	0.2	0.78
Hear	hear	आवाज,ऐकणे,श्रवणीय	105	0.29	0.71

Feel	feel	उबदार,गारठणे,वेदना	150	0.14	0.57
Biological Processes	bio	अंधत्व,अन्नपचन,आहार	1756	0.35	0.89
Body	body	अवयव,चरबी,ज्ञानेंद्रिये	216	0.11	0.63
Health	health	अपचन,क्षयरोग,जखम	591	0.52	0.86
Sexual	sexual	कामुकता,गर्भधारणा	90	0.4	0.44
Ingestion	ingest	अन्नपदार्थ,खारट,तहान	646	0.3	0.79
Drives	drives	अग्रगण्य,अधिकारी,उच्चता	880	0.57	0.98
Affiliation	affiliation	एकत्र,नातेवाईक,संगती	234	0.67	0.93
Achievement	achiev	अत्युत्तम,इतिहास_रचणे	525	0.4	0.94
Power	power	अत्युच्चपद,अधिकारी,तज्ञ	878	0.49	0.93
Reward	reward	अभिनंदन,पूर्तता,बक्षीस	239	0.34	0.91
Risk	risk	अनिर्णीत,असुरक्षित,	280	0.27	0.75
Time Orientation					
Focus past	focuspast	अडवले,घटले,जळाले	1260	0.71	0.95
Focus Present	focuspresent	ऐकतो,थांबा,पहा,बोलतो	325	0.41	0.93
Focus Future	focusfuture	आगामी,आशादायक	63	0.37	0.89
Relativity	relativ	अनुक्रम,अलीकडील	653	0.55	0.96
Motion	motion	उलथापालथ,गतिमान	241	0.31	0.83
Space	space	क्षेत्र,जागा,ठिकाण	138	0.23	0.88
Time	time	अखंडित,अनंत,अहोरात्र	372	0.47	0.92
Personal Concerns					
Work	work	संशोधन,अभ्यास	2,098	0.65	0.95
Leisure	leisure	आराम,कॅफेटेरिया,गप्पा	280	0.43	0.74
Home	home	घरसामान,टेलिव्हिजन	215	0.08	0.49
Money	money	खरेदी-विक्री,चलन,देणगी	360	0.64	0.78
Religion	relig	अध्यात्म,आत्मा,ऋग्वेद	2298	0.4	0.91
Death	death	अंतिमयात्रा,आत्महत्या	175	0.31	0.54
Informal Language	informal	अरे,कृपया,नमस्ते	224	0.44	0.87
Swear words	swear	नालायक,बदमाश,मुर्ख	15	-0.03	-0.03

Netspeak	netspeak	इन्स्टाग्राम,हॅलो,टेलिग्राम	126	0.15	0.79
Assent	assent	खरंच,बरोबर,ठीक_आहे	19	0.18	0.65
Nonfluencies	nonflu	अरेव्वा,हं,मस्त	45	0.43	0.5
Filler	filler	असो,अस्सं	34	0.55	0.84

Table 6. Comparison of English LIWC2015 with Marathi LIWC2015

LIWC Dimension	Output Label	EN-LIWC2015		MR-LIWC2015		EN-LIWC/ MR-LIWC Correlation
		Mean	**SD**	**Mean**	**SD**	
Summary Variables						
Word count	WC	91.88	148.21	77.65	124.90	**0.99**
Words per sentence	WPS	20.15	12.65	20.90	12.85	**0.57**
Big words	BigWords	24.72	12.25	33.17	15.37	**0.48**
Dictionary words	Dic	77.00	15.99	68.96	13.97	**0.50**
Linguistic Dimensions						
Total Function Words	function	42.88	13.80	30.85	12.43	**0.52**
Total Pronouns	pronoun	7.60	6.79	5.00	5.29	**0.66**
Personal Pronoun	ppron	4.05	5.15	2.90	4.15	**0.73**
1st person singular	i	0.76	2.46	0.76	2.10	**0.79**
1st person plural	we	0.83	2.43	1.10	2.73	**0.75**
2nd person	you	0.97	2.75	0.25	1.25	**0.51**
3rd person singular	shehe	1.03	2.15	0.56	1.53	**0.49**
3rd person plural	they	0.47	1.49	0.83	1.77	**0.37**
Impersonal pronoun	ipron	3.55	3.84	2.38	3.29	**0.46**
Articles	article	7.84	5.52	0.39	1.37	**0.04**
Prepositions	prep	13.95	6.40	16.51	9.92	**0.34**
Auxiliary Verbs	auxverb	7.15	5.86	4.96	4.99	**0.47**
Adverbs	adverb	2.84	3.38	6.70	6.02	**0.37**
Conjunctions	conj	4.26	3.81	3.93	4.16	**0.53**
Negation	negate	1.06	2.20	1.01	2.28	**0.85**

Common Verbs	verb	13.31	8.53	12.23	8.28	**0.60**
Common Adjectives	adj	3.45	4.39	10.09	7.43	**0.33**
Comparison	compare	1.31	2.15	1.24	2.31	**0.30**
Interrogative	interrog	1.34	2.44	1.01	2.89	**0.57**
Number	number	1.63	3.08	2.21	4.95	**0.73**
Quantitative	quant	1.64	2.97	1.54	2.72	**0.56**
Psychological Processes						
Affective processes	affect	5.44	6.73	4.82	5.87	**0.34**
Positive Emotion	posemo	3.65	6.15	3.22	5.24	**0.35**
Negative Emotion	negemo	1.76	3.05	1.59	3.22	**0.47**
Anxiety	anx	0.26	1.14	0.59	1.85	**0.20**
Anger	anger	0.65	2.05	0.97	2.61	**0.44**
Sadness	sad	0.33	1.51	0.31	1.16	**0.28**
Social Processes	social	8.83	7.39	5.68	5.93	**0.50**
Family	family	0.48	1.70	0.75	2.13	**0.47**
Friends	friend	0.22	1.27	0.18	1.08	**0.60**
Female references	female	0.53	1.74	0.61	1.92	**0.65**
Male references	male	1.16	2.42	1.70	3.47	**0.21**
Cognitive processes	cogproc	7.56	6.62	6.49	5.93	**0.52**
Insight	insight	1.61	2.59	1.89	2.94	**0.35**
Causation	cause	1.21	2.48	1.19	2.80	**0.30**
Discrepancy	discrep	1.06	2.17	0.99	2.18	**0.34**
Tentative	tentat	1.24	2.29	1.13	2.36	**0.45**
Certainty	certain	1.30	2.79	0.95	2.36	**0.56**
Difference	differ	2.01	3.14	1.86	3.19	**0.50**
Perceptual processes	percept	1.63	3.20	0.95	2.37	**0.61**
See	see	0.67	1.83	0.41	1.58	**0.61**
Hear	hear	0.51	1.72	0.30	1.11	**0.49**
Feel	feel	0.34	1.21	0.13	0.74	**0.14**
Biological Processes	bio	2.06	3.53	1.60	3.99	**0.38**
Body	body	0.48	1.45	0.15	0.90	**0.34**
Health	health	0.97	2.59	0.94	2.81	**0.43**

Sexual	sexual	0.04	0.36	0.08	0.58	**0.15**
Ingestion	ingest	0.51	1.90	0.46	2.75	**0.41**
Drives	drives	9.52	8.02	6.15	6.45	**0.43**
Affiliation	affiliation	2.29	3.74	2.10	3.58	**0.67**
Achievement	achiev	1.49	2.81	0.87	2.63	**0.30**
Power	power	4.52	5.52	3.77	5.71	**0.47**
Reward	reward	1.47	4.19	0.71	2.03	**0.05**
Risk	risk	0.63	1.83	0.51	1.78	**0.20**
Time Orientation						
Focus past	focuspast	3.29	3.93	3.42	4.19	**0.58**
Focus Present	focuspresent	8.70	7.03	4.89	5.33	**0.60**
Focus Future	focusfuture	1.12	2.38	0.96	2.21	**0.30**
Relativity	relativ	13.07	8.60	4.69	5.14	**0.42**
Motion	motion	1.83	3.25	0.87	2.07	**0.16**
Space	space	6.75	5.84	1.30	2.83	**0.27**
Time	time	4.70	5.73	2.08	3.34	**0.41**
Personal Concerns						
Work	work	4.30	5.47	4.99	6.76	**0.28**
Leisure	leisure	1.18	2.44	0.36	1.59	**0.39**
Home	home	0.59	2.26	0.11	1.07	**0.11**
Money	money	0.97	2.90	0.48	1.57	**0.40**
Religion	relig	1.00	2.72	1.12	3.66	**0.26**
Death	death	0.31	1.31	0.31	1.59	**0.55**
Informal Language	informal	0.16	0.73	2.34	5.67	**0.03**
Swear words	swear	0.02	0.22	0.02	0.24	**0.14**
Netspeak	netspeak	0.04	0.37	1.01	5.19	**0.03**
Assent	assent	0.05	0.39	0.13	0.62	**0.31**
Nonfluencies	nonflu	0.05	0.38	0.29	1.21	**0.05**
Filler	filler	0.00	0.03	1.98	3.43	**0.00**

References:

[1] Pennebaker, J. W., Francis, M. E., & Booth, R. J., "Linguistic inquiry and word count: LIWC 2001", Mahway*: Lawrence Erlbaum Associates*, *71*(2001), 2001.

[2] Chung, C. K., & Pennebaker, J. W., "Linguistic inquiry and word count (LIWC): pronounced "Luke,"... and other useful facts", In Applied natural language processing: Identification, investigation and resolution, 2012, (pp. 206-229). IGI Global.

[3] Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W., "The development and

psychometric properties of LIWC-22”, *Austin, TX: University of Texas at Austin*, 2022, 1-47.

[4] Carvalho, F., Rodrigues, R. G., Santos, G., Cruz, P., Ferrari, L., & Guedes, G. P., “Evaluating the Brazilian Portuguese version of the 2015 LIWC Lexicon with sentiment analysis in social networks”, In Anais do VIII Brazilian Workshop on Social Network Analysis and Mining, 2019, (pp. 24-34). SBC.

[5] Balage Filho, P., Pardo, T. A. S., & Aluísio, S., “An evaluation of the Brazilian Portuguese LIWC dictionary for sentiment analysis”, In Proceedings of the 9th Brazilian Symposium in Information and Human Language Technology,2013

[6] Huang Jinlan, Chung, C. K., Hui, N., Lin Yizheng, Xie Yitai, Lam, B. C., ... & Pennebaker, J. W. , “The Development of the Chinese Linguistic Inquiry and Word Count Dictionary]”, Chinese Journal of Psychology, 2012, 54(2), 185-201

[7] Boot, P., Zijlstra, H., & Geenen, R, “The Dutch translation of the linguistic inquiry and word count (LIWC) 2007 dictionary”, Dutch Journal of Applied Linguistics, 2017, 6(1), 65-76.

[8] Van Wissen, L., & Boot, P.,”An electronic translation of the LIWC Dictionary into Dutch”, In Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference (pp. 703-715). Brno: Lexical Computing., 2017

[9] Piolat, A., Booth, R., Chung, C. K., Davids, M., & Pennebaker, J. W., “The French dictionary for LIWC: Modalities of construction and examples of use| La version franaise du dictionnaire pour le LIWC:”, modalités de construction et exemples d'utilisation, 2011

[10] Meier, T., Boyd, R. L., Pennebaker, J. W., Mehl, M. R., Martin, M., Wolf, M., & Horn, A. B., ““LIWC auf Deutsch”: The development, psychometrics, and introduction of DE-LIWC2015”, PsyArXiv, (a)., 2019

[11] Wolf, M., Horn, A. B., Mehl, M. R., Haug, S., Pennebaker, J. W., & Kordy, H.,”Computergestützte quantitative textanalyse: äquivalenz und robustheit der deutschen version des linguistic inquiry and word count”, Diagnostica,, 2008, 54(2), 85-98.

[12] Agosti, A., & Rellini, A., “The Italian liwc dictionary”, *Austin, TX: LIWC. Net* ,2007

[13] Igarashi, T., Okuda, S., & Sasahara, K., “Development of the Japanese Version of the Linguistic Inquiry and Word Count Dictionary 2015”, Frontiers in psychology, 2022, 13, 841534

[14] Dudău, D. P., & Sava, F. A.,”The development and validation of the Romanian version of Linguistic Inquiry and Word Count 2015 (Ro-LIWC2015)”, Current Psychology, 2022, 41(6), 3597-3614.

[15] Kailer, A., & Chung, C. K., “The russian liwc2007 dictionary”, *Austin, TX: LIWC. net*., 2011

[16] Bjekić, J., Lazarević, L. B., Živanović, M., & Knežević, G., “Psychometric evaluation of the Serbian dictionary for automatic text analysis-LIWCser”, Psihologija, 2014, 47(1), 5-32.

[17] Ramirez-Esparza, N., Chung, C., Kacewic, E., & Pennebaker, J. ,”The psychology of word use in depression forums in English and in Spanish: Testing two text analytic approaches”, In Proceedings of the international AAAI conference on web and social media (Vol. 2, No. 1, pp. 102-108), 2008

[18] Zasiekin, S., “Exploring Bohdan Lepky’s Translation Ethics Using Linguistic Inquiry and Word Count”, East European Journal of Psycholinguistics, 8(2)., 2021

[19] Popale, Lata, and Pushpak Bhattacharyya.,"Creating Marathi WordNet.", The WordNet in Indian Languages : 147-166., 2017

[20] Falotico, R., & Quatto, P., “Fleiss’ kappa statistic without paradoxes”, Quality & Quantity, 2015, 49, 463-470.