

DOI <https://doi.org/10.30525/978-9934-26-277-7-21>

**LIWC-22 TO DEFINE THE CINEMATIC CHARACTERS'
VERBAL REPRESENTATION**

**ВИКОРИСТАННЯ LIWC-22 ДЛЯ ВИЗНАЧЕННЯ
ВЕРБАЛЬНОЇ РЕПРЕЗЕНТАЦІЇ
КІНЕМАТОГРАФІЧНИХ ПЕРСОНАЖІВ**

Berezhna M. V.

*Candidate of Philological Sciences,
Associate Professor at the Department
of English Translation Theory
and Practice
Zaporizhzhia National University
Zaporizhzhia, Ukraine*

Бережна М. В.

*кандидатка філологічних наук,
доцентка кафедри теорії та практики
перекладу з англійської мови
Запорізький національний університет
м. Запоріжжя, Україна*

It is implicit to current psychological language analysis research that several characteristics of someone's personality are embedded in their unique patterns of language use [7, p. 65]. The words that people use in everyday life tell us about their psychological states: their beliefs, emotions, thinking habits, lived experiences, social relationships, and personalities [5, p. 2]. Recent research has demonstrated that a powerful reflection of personality can be gleaned from the words people use in everyday life. As an increasing number of studies demonstrate, the ways in which people use words is reliable over time, internally consistent, predictive of a wide range of behaviors and even biological activity, and varies considerably from person to person. Language, then, is yet another fundamental dimension of personality [7, p. 64]. Throughout history, scholars and laypeople alike have believed that our words contain subtle clues about what we are like as people, psychologically speaking [2, p. 41]. However, the ways in which language has been used to infer psychological processes has seen dramatic shifts over time. Currently we find ourselves in the midst of a technological revolution whereby, for the first time, researchers can link daily word use to a broad array of real-world behaviors [1, p. 24].

Although promising, the early computer methods floundered because of the sheer complexity of the task. In order to provide a better method for studying verbal and written speech samples, a text analysis application called *Linguistic Inquiry and Word Count*, or *LIWC* was developed. The first *LIWC* was a part of an exploratory study of language and

disclosure. The second (*LIWC2001*), third (*LIWC2007*), fourth (*LIWC2015*), and now fifth (*LIWC-22*) versions updated the original application with increasingly expanded dictionaries and sophisticated software design. The most recent evolution, *LIWC-22*, has significantly altered both the dictionary and the software options to reflect new directions in text analysis [5, p. 2].

All previous versions of *LIWC* were written for researchers who typically analyzed large numbers of text files. Many users, however, have wanted to be able to simply dive into a single text to understand it more deeply through a close analysis. The *Case Studies* module essentially brings the other modules into a single location to allow users to focus on and explore a single text [5, p. 4]. The developers of *LIWC-22* present the researcher with a possibility to ‘bask in the cold, sterile glow of your computer screen, obsessively analyzing the narrative structure of a coworker’s email or *LIWC*ing the dialogue from your favorite episode of *Rick and Morty*’ [5, p. 4].

The *Test Kitchen* corpus was constructed from randomly selected subsets of text from across 15 different types of English language sets, including transcribed movie dialogs. Boyd et al. (2022) selected 1,000 movie samples, each averaging 6,600 words [4] to provide averaged statistics for the language used in different media.

The present paper examines the potential of the *LIWC-22* application to establish the verbal representation of the cinematic characters in mainstream movies. To increase the validity of data, the paper examines only the turns exceeding 150 words in total. The results are contrasted with the mean numbers and standard deviations of *LIWC* categories in movies gathered and presented by Boyd et al. [4]. The hypothesis is that deviations from mean figures demonstrate individual traits of the examined characters. To simplify the visual perception of the data, numbers exceeding the standard deviation are marked by different shades of brown; numbers lower than the standard deviation are marked with diverse shades of blue (Table 1).

LIWC-22 presents the researcher with 117 linguistic items, demonstrating different aspects of the personality. Among others, there are complex categories of ‘Analytical thinking’, ‘Clout’, ‘Authentic’ and ‘Emotional tone’.

‘Analytical thinking’. A high number reflects formal, logical, and hierarchical thinking; lower numbers reflect more informal, personal, here-and-now, and narrative thinking [3, p. 22]. The number is high in the turns of the Narrator (*Maleficent*, 2014), who recites the story of Maleficent and the Sleeping Beauty; Gramma Tala (*Moana*, 2016), who starts the movie with the legends of the tribe; and Denarian Dey (*Guardians of the Galaxy Vol. 1*, 2014), presenting the life stories of the main characters.

‘Clout’. A high number suggests that the author is speaking from the perspective of high expertise and is confident; low ‘Clout’ numbers suggest a more tentative, humble, even anxious style [3, p. 22]. Numbers surpassing the standard deviation, define the turns of Neytiri and Grace Augustine, having to respectively teach Jake Sully how to survive on the Pandora planet and how to operate the avatar body (*Avatar*, 2009). High numbers are also observed in the turns of the perfect soldier Captain America and the technical genius Tony Stark, responsible for the mission success or failure (*Avengers: Infinity War*, 2018). Combined with low numbers in the ‘I’ and ‘question_mark’ categories, ‘clout’ demonstrates the high social status, confidence, or leadership skills of the character [6].

‘Authentic’. Higher numbers are associated with a more honest, personal, and disclosing text; lower numbers suggest a more guarded, distanced form of discourse [3, p. 22]. Authenticity reflects the degree to which a person is self-monitoring and filtering what they say. The number is low in the turns of traitorous villains, pretending to be good guys: Bellwether (*Zootopia*, 2016), Supreme Intelligence and Yon-Rogg (*Captain Marvel*, 2019), Parker Selfridge (*Avatar*, 2009), and Loki (*Thor*, 2011).

‘Emotional tone’. A high number is associated with a more positive, upbeat style; a low number reveals greater anxiety, sadness, or hostility. A number around 50 suggests either a lack of emotionality or different levels of ambivalence [3, p. 22]. The number is often high in the turns of energetic, enthusiastic and optimistic characters: Aurora (*Maleficent*, 2014), Joy and Riley (*Inside Out*, 2015), Olaf (*Frozen*, 2013), and Maui (*Moana*, 2016). The number is frequently low in the turns of antagonists or villains: Nebula (*Guardians of the Galaxy Vol. 1*, 2014), Wanda (*Avengers: Age of Ultron*, 2015), Dreykov (*Black Widow*, 2021), Duke (*Frozen*, 2013), and Lady Ascot (*Alice in Wonderland*, 2010).

Function words, such as personal pronouns, reflect attentional allocation [6]. Thus, depressed and traumatized characters are focused on themselves, which is revealed via high numbers in the ‘I’ category: Maleficent (*Maleficent*, 2014), Elsa (*Frozen*, 2013), Red Queen (*Alice in Wonderland*,

2010), Valkyrie (*Thor: Ragnarok*, 2017), Loki (*The Avengers*, 2012), Ultron (*Avengers: Age of Ultron*, 2015), and Vanko (*Iron Man 2*, 2010).

Further research aims to create a comprehensive classification of verbal representation of characters frequenting English-language mainstream movies in the XXI century. *LIWC-22* proves to be a reliable way to automatically identify psychological characteristics of the researched characters. The combination of *LIWC-22* with qualitative (narrative) analysis appears promising for future complex investigations.

References:

1. Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>
2. Boyd, R. L., & Schwartz, H. A. (2021). Natural language analysis and the psychology of verbal behavior: The past, present, and future states of the field. *Journal of Language and Social Psychology*, 40(1), 21–41. <https://doi.org/10.1177/0261927X20967028>
3. Pennebaker, J. W., Booth, R. J., Boyd, R. L., & Francis, M. E. (2015). Linguistic Inquiry and Word Count: LIWC2015. Pennebaker Conglomerates. Retrieved November 19, 2022, from <https://www.liwc.net>
4. Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *LIWC-22: Descriptive Statistics and Norms*. Retrieved November 19, 2022, from <https://www.liwc.app>
5. Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). *The development and psychometric properties of LIWC-22*. Austin, TX: University of Texas at Austin. Retrieved November 19, 2022, from <https://www.liwc.app>
6. Kacewicz, E., Pennebaker, J. W., Davis, M., Jeon, M., & Graesser, A. C. (2014). Pronoun use reflects standings in social hierarchies. *Journal of Language and Social Psychology*, 33(2), 125–143. <https://doi.org/10.1177/0261927X13502654>
7. Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, 18, 63–68. <https://doi.org/10.1016/j.cobeha.2017.07.017>