# ELIta: A New Italian Language Resource for Emotion Analysis

Eliana Di Palma[1,2]

[1]*Sapienza, University of Rome*
[2]*Roma Tre University*

**Abstract**
Emotions and language are strongly associated. In recent years, many resources have been created to investigate this association and automatically detect emotions from texts. Presenting ELIta (Emotion Lexicon for Italian), this study provides a new language resource for the analysis and detection of emotions in Italian texts. It describes the process of lexicon creation, including lexicon selection and annotation methodologies, and compares the collected data with existing resources. By offering a non-aggregated lexicon, ELIta fills a crucial gap and is applicable to various research and practical applications. Furthermore, the work utilises the lexicon by analysing the relationships between emotions and gender.

**Keywords**
emotions, language resource, italian, emotion lexicon, word-emotion association

## 1. Introduction and Related Works

Emotions and language are deeply interrelated human characteristics. Language serves as a tool to communicate our feelings, while affective studies have shown that emotion permeates all aspects of language [1, 2], such as morphology [3, 4, 5], phonology [6, 7], and semantics [8, 9]. This intricate relationship has recently attracted significant attention in fields such as computational linguistics, natural language processing (NLP), and affective computing. Research focusing on the identification of emotions from texts has produced various language resources, particularly emotion lexicons developed using diverse annotation methodologies, ranging from manual [10, 11] to automatic [12, 13], and from expert judgment [14, 15] to crowdsourcing [16, 17].

Most studies follow the dimensional approach to emotions [18, 19]. According to this perspective, the PAD (Pleasure, Arousal, Dominance) [20] or VAD (Valence, Arousal, Dominance) [19] model posits that the fundamental dimensions of VALENCE (the intrinsic attractiveness (positive VALENCE) or aversion (negative VALENCE) of an event, object or situation), AROUSAL (the level of physiological activation, ranging from sleep to excitement) and DOMINANCE (the degree of control a person feels over a situation) explain the majority of the emotional meaning of words. This approach has been highly productive for research on emotional language and the creation of language resources, exemplified by the ANEW (Affective Norms for English Words) [21, 22], NRC VAD [23], and the EmoBank corpus [11].

Alternatively, some researchers argue for the existence of a limited number of discrete primary emotion categories that have evolved to serve various adaptive functions through specific neural signatures, facial expressions, cognitive evaluations, and behavioral action tendencies [24, 25]. These basic emotions typically include JOY, SADNESS, DISGUST, ANGER, FEAR, SURPRISE, whereas Plutchik also considers TRUST and ANTICIPATION. Despite objections to the basic emotions model [27], it has inspired the creation of resources such as the NRC Lexicon (EmoLex) [17] (translated into over 100 languages, it's the most widely used lexicon in emotion detection), and the datasets Feel It [28] and Multiemotion It [29].

More recently, the field of computational linguistics and NLP has recognized the need for resources specifically created for languages other than English. Critics argue against relying solely on translations, advocating for lexicons created from texts in the target language and manually annotated [30, 31, 32]. This approach has led to the development of lexicons like the Portuguese emotional lexicon [30], which embodies the principle of "each language for itself."

For the Italian language several language resources with emotional annotations have been produced over the years. The initial ItEM lexicon [33] began by collecting seed words through an association task linking words to labels (Plutchik's basic emotions), then employed cosine similarity to expand the lexicon, assuming that neighboring words in semantic space share similar emotional connotations. The results, validated through crowdsourcing, showed low reliability for the emotions TRUST, ANTICIPATION (translated as 'attese') and SURPRISE. The more recent Depeche Mood ++ [34] was automatically created from judgements given by readers of articles on the 'Corriere della Sera' newspaper website and uses a

unique scale of emotions not directly comparable to others, such as ANNOYED, AFRAID, SAD, AMUSED, and HAPPY. [34].

In the case of Affective Norms [21], the tendency to create resources by adapting the English model with annotations in L1 languages other than English has resulted in Affective Norms for several languages, including Spanish [35] and Dutch [36]. For Italian, there has been a specific adaptation of the ANEW collected by [37].

Despite the existing resources in the literature, a notable gap persists. There is a lack of manually annotated Italian language resources that combine both discrete emotion annotations and dimensional evaluations. Furthermore, no available resource provides a non-aggregated version of the data.

This paper presents **ELIta** (**E**motion **L**exicon for **It**alian), an innovative resource designed for the analysis of emotions in the Italian language and emotion detection from text. **ELIta** aims to bridge this gap by providing a lexicon annotated using both categorical and dimensional approaches, and by offering a non-aggregated version of the data. This aligns with the perspectivist viewpoint, which values disagreement as valuable information [38, 39, 40, 41]. The development process of **ELIta**, including lexicon selection, annotation methodologies, and a comparative analysis with existing Italian sentiment lexicons, is thoroughly described. Finally, analyses of the relationships between emotions and between dimensions and gender are presented.

## 2. Emotion Lexicon Creation

**Lexicon Selection** The lexicon for this study is constructed from existing resources in the literature. The major pool from which it draws is De Mauro's 'Nuovo Vocabolario di Base' (NVdB) [42]. This selection is made by reason of its representativeness of contemporary Italian language usage in different types of text. In line with EmoItaly [43], 186 emoji have been added to the lexicon so that it can also be used for texts from Social Networks.

Furthermore, as a foundational layer, the seed-words of ItEM [33] were incorporated. To ensure broad coverage, high-frequency words (recurring more than 200 times) from the Depeche Mood ++ [34] lexicon by Araque et al. were included.

The selection process favoured content words (verbs, nouns, adjectives and adverbs) over function words (determiners, conjunctions).

The final lexicon comprises 6905 items, including both words and emojis. The data set contains 21 % adjectives, 50 % of nouns, 21 % verbs and 8 % of words that can be considered both as adjective and noun. In addition, a smaller number of adverbs, expressions (e.g. 'restare a bocca aperta' *be looking open-mouthed*) and interjections

(e.g. 'beh', 'boh') have been included.

Consistent with the research of Montefinese et al. and previous studies [44], participants were not explicitly instructed to disambiguate words with multiple grammatical meanings.

**Annotation Schema** In order to collect a versatile dataset adaptable to different research approaches, the data collection involved an annotation process that included both the association of words with basic emotions [26] and the evaluation of the items according to VAD dimensions [20]. In the case of basic emotions, it was decided to use the translation 'aspettativa' for ANTICI-PATION instead of 'attese' as in ItEM in order to avoid misunderstandings and associations of the "attese-treno" type. Furthermore, to provide additional context for the analysis, participants were asked to share their demographic information.

**Data collection** Data were collected from two primary sources from April 2023 to May 2024. An online questionnaire in the form of a website [1] created from scratch was used to rate the words. The website was shared for annotation via mailing lists (such as LinguistList and AILC) and social networks. The participation was on a voluntary basis and without payment. In this system, the words to be rated in each questionnaire were randomly chosen. That is, each time the questionnaire was accessed, the system randomly chose the words from the entire list of 6, 905 words. Thus, each participant rated a different set of words.

When accessing the website, participants first agreed to an informed consent. Then, they were given the guidelines for both categorical and dimensional annotation. On the third screen, they had to provide the demographic information concerning age and gender, and select the time slot to spend on the annotation process (from 3 to 10 words, with the possibility of extending the annotation process at the annotator's discretion).

Participants were asked to rate the extent to which each word is associated to a list of emotions, using a scale from "non associated" (0), "weakly associated" (0.25), "moderately associated" (0.75) to "strongly associated" (1) [17]. Next, participants were shown the dimensions using the Self-Assessment-Manikin from the ANEW [21] (see Fig. 1) to assess the extent to which each word convey VALENCE, AROUSAL, and DOMINANCE using a 1 to 9 scale. The guidelines in the latter case were mutuated from Montefinese et al.(2014).

Additionally, the Prolific platform was used to recruit native Italian speakers as participants from March 2024 [2].

---

[1] https://emotionlexicon.com/
[2] In this case, the annotators were paid according to the rules established by the platform.
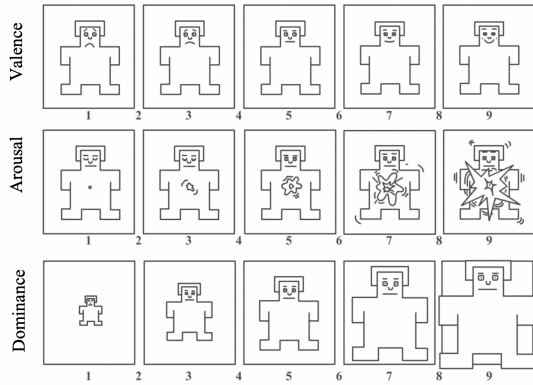
**Figure 1:** *Self-Assessment Manikin* da Bradley and Lang (1994)

A total of 100 different questionnaires were created and completed on the platform, each containing 65 words/emojis. Words and emojis were selected based on existing annotations to ensure a minimum annotation threshold of five per word (such as in the NRC lexicon [17]).

**Description of ELIta** The collected data underwent a rigorous filtering process to ensure quality and accuracy. Participants with exceptionally fast completion times were excluded. Additionally, despite the subjective nature of the task, annotations with clear anomalies were removed, such as associations deemed illogical (e.g., 'worsening' *peggioramento* strongly associated with JOY).

The total number of annotations gathered is 35, 412. For each of the 6905 words/emojis in the lexicon, from a minimum of 5 annotations to a maximum of 10 annotations were collected (on average 5.13 annotations per word).

From the demographic metadata, it can be observed that the majority of annotations come from women and the most frequent age group is 25-34 years old (see table 1).

Table 1: Number of annotations by gender and age. The highest number of annotations for each age and gender are highlighted.

|       | Women | Men  | Non binary | Not specified |
|-------|-------|------|------------|---------------|
| 18-24 | 4201  | 2318 | 108        | 73            |
| 25-34 | 9052  | 6797 | 654        | 18            |
| 35-60 | 6568  | 4766 | 8          | 11            |
| 60    | 267   | 550  | 13         | 8             |

**Versions**

The lexicon is provided in several versions [3]:

_____

[3]https://github.com/elianadipalma/ELIta

**RAW** Version including all annotations and demographic information with an inter-annotator agreement (calculated with Krippendorff [45]) of 0.67, which can be explained by the subjective nature of the task (associating words with emotions in isolation). Various factors such as gender, age and socio-cultural background can influence the IAA in such subjective tasks.

**GOLDEN** A second, non-aggregated version was also released, in which the five most similar annotations were selected for words with more than five annotations, thereby excluding the outliers. Additionally, an automatically generated 'golden standard' annotation was added for each word, calculated based on the majority vote from the five annotations for each emotion. This approach emphasizes the majority vote while retaining all individual entries. This 'ELIta-golden' version achieves an Inter-Annotator Agreement (IAA) of 0.874. The annotations are categorized by origin into 'ELIta,' 'ELIta-selected' for selections made from more than five annotations, and 'golden.' In this case, demographic information is absent, but association intensity is preserved.

**INTENSITY** One of the aggregate versions created from the golden version retains the intensity values of the original annotations, with the single value calculated as the average of the six annotations (five original + one golden). The decision to use the golden version is to balance the few annotations with one representative of the majority. In this case, the labels of LOVE automatically calculated from the values of JOY and TRUST and 'neutral' were also added.

**BINARY** The second aggregated version, converts the aggregated float values to integers, providing a binary representation of the basic emotions: 0 for values below 0.50 and 1 for values above 0.50.

## 3. Analyses and Discussion

### 3.1. Comparative Analyses

To evaluate the similarities and differences of the newly developed **ELIta** lexicon, it was conducted a comparative analyses with other language resources for Italian: EmoLex (NRC-AIL) [46], ItEM [33], and ANEW [37]. The **Intensity** version of **ELIta** was used for all analyses.

Correlations were calculated for each basic emotion and VAD dimension against the italian translation of EmoLex (NRC - AIL Affective Intensity Lexicon [46]), the cosine values of ItEM [33], and the dimensions of the Italian Affective Norms [37].

**ELIta vs. EmoLex**  Comparing the 2,388 shared items, the results showed a moderate correlation of 0.51. JOY exhibited the highest correlation ($r = 0.65$), while ANTICIPATION ($r = 0.38$) and SURPRISE ($r = 0.35$) showed the lowest. The results show even more the need to use lexicons specifically created for the target language.

**ELIta vs. ItEM**  With 3,299 shared items, Pearson correlations were calculated between the degree of association of ELIta for each basic emotion and the cosine similarities between the words and emotion-labels of the basic emotions. Correlations were generally low, with the highest for ANGER ($r = 0.29$). The lower correlations are in line with previous observations on the difficulty of annotating emotions such as TRUST ($r = 0.18$), ANTICIPATION ($r = 0.14$) and SURPRISE ($r = 0.13$).

**ELIta vs. ANEW**  The two resources share 762 items. The analysis revealed a strong correlation ($r = 0.88$) for VALENCE, while AROUSAL ($r = 0.48$) and DOMINANCE ($r = 0.61$) showed lower correlations. The observed outcomes are consistent with research showing AROUSAL and DOMINANCE as the dimensions most variable [35, 37].

To identify the words for which the two annotator groups provided significantly different ratings, a linear regression was used. This statistical model allows to estimate the extent to which ELIta ratings can be predicted by Affective Norms ratings and to identify the words for which this relationship is weaker. [4].

The results show a more negative connotation of words linked to the religious sphere, for example, 'church' *chiesa* and 'god' *dio* have shifted from positive to negative. Similarly, 'fur' *pelliccia* 'circus' *circo* and 'justice' *giustizia* have also transitioned from positive to negative. Conversely, the terms 'lesbian' *lesbica* and 'mad' *folle* have shifted from negative to positive.

Examining the associations of these words with basic emotions, it can be noted, for example, that the predominant emotion associated with the word 'church' is ANGER with a mean intensity of 0.54, followed by SADNESS and DISGUST. Analogously, 'fur' is associated most strongly with 0.75 to SADNESS and with 0.70 to ANGER, and 'circus' is more associated with DISGUST and SADNESS. The word 'god' presents an interesting contrast. Although it has a negative VALENCE ($M = 4.6$) compared to the ANEW result ($M = 8.3$), the primary emotions associated with it are TRUST (0.67) and ANTICIPATION(0.46). The word 'lesbian' does not appear to be associated with any emotion, except very weakly with JOY ($M = 0.20$), while 'mad' results associated more with JOY and SURPRISE ($M = 0.42$).

Regarding AROUSAL, terms such as 'optimism' *ottimismo*, 'erotic' *erotico*, 'success' *successo*, 'food' *cibo*, and 'in love' *innamorato* have shown increased activation.

---

[4]Plots are available in the appendix, see Fig. .6.

In contrast, terms like 'unpleasant' *spiacevole*, 'discouraged' *scoraggiato*, 'boredom' *noia*, 'cold' *freddo*, and 'rain' *pioggia* are associated with less activation in the ELIta lexicon.

For DOMINANCE, there is an increased sense of dominance associated with terms such as 'hatred' *odio*, 'optimism' *ottimismo*, 'triumph' *trionfo*, 'triumphant' *trionfante*, 'to sleep' *dormire* and 'to travel' *viaggiare*. Conversely, the sense of submission is associated with terms like 'earth' *terra*, 'nature' *natura*, 'circus' *circo*, and states of illness such as 'fever' *febbre*.

These differences may reflect the different sensibilities of the annotators. The affective norms of [37] were published in 2014, while the majority of the annotators of the proposed ELIta lexicon belong to the age range of 25-34 years. ELIta can thus be seen as a limited update to the norms proposed by Montefinese et al. (2014).

Although generational characteristics may influence the results, it is important to consider that the comparison was based on the means of responses from approximately 20 persons for the Norms and 5 persons per word for ELIta. The lower number of annotators for ELIta could imply that the individuality and socio-cultural background of each participant have a greater impact on the results. Therefore, further analyses should be conducted.

## 3.2. Correlations and Gender Variation

Once the data as a whole had been analysed in comparison with other lexicons, the annotations were analysed to examine the relationship between the different emotions and dimensions, and whether there were differences between genders in the association between words and emotions.

**Correlations**  Firstly, Pearson correlations between categories and dimensions were calculated. (see Fig. 2).

The results show a moderate correlation between AROUSAL and negative emotions, particularly FEAR ($r = 0.45$) and ANGER ($r = 0.40$). Consequently, the correlation between AROUSAL and VALENCE turns out to be weakly negative ($r = -0, 17$).

Furthermore, it can be noticed that negative emotions tend to co-occur, suggesting that words associated with SADNESS may also be linked to ANGER, DISGUST, or FEAR [47]. Conversely, JOY shows a moderate to strong correlation with TRUST ($r = 0.66$), ANTICIPATION ($r = 0.62$) and SURPRISE ($r = 0.49$).

Interestingly, there is a moderate correlation between DOMINANCE and JOY ($r = 0.53$), indicating that words with positive VALENCE are also associated with a greater sense of control ($r = 0.7$), while negative ones are associated to a sense of submission ($r = -0.40$ to $r = -0.53$) [48]. An exception is given by words such as 'nature' which, as
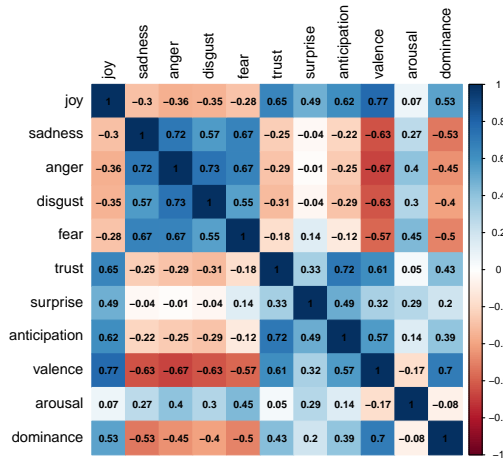
**Figure 2:** Correlations between basic emotions and VAD dimensions.



**Figure 3:** Scatterplots of the distributions of ELIta according to DOMINANCE and AROUSAL dimensions, and VALENCE and AROUSAL dimensions. The lines represent the linear regression according to the values before the VALENCE median (in purple) or the DOMINANCE median (in red), and after the VALENCE median (in green) or the DOMINANCE median (in teal).

we have seen, has a low rating (M = 3.5) for DOMINANCE but is strongly associated with JOY (M = 1).

SURPRISE shows positive correlations both modestly with JOY and ANTICIPATION, and weakly with FEAR, and TRUST, although it is a more neutral emotion than the others, it is generally more prone to have a positive VALENCE ($r = 0.32$).

**ELIta**'s findings for Italian corroborate patterns previously identified by Ferré et al. (2016) for Spanish and Sarli and Justel (2021) for Argentinian Spanish.

The correlations and regression analyses revealed patterns consistent with the other resources: a U-shaped relationship between VALENCE and AROUSAL, DOMINANCE and AROUSAL (see Fig. 4, and a linear relationship between DOMINANCE and VALENCE (see Fig. 3). These results suggest that highly negative or positive items, as well as words associated with low or high control, tend to elicit greater emotional and physiological activation. Meanwhile, greater positivity corresponds to a greater sense of control.

These analyses have positioned **ELIta** as a valuable resource for emotional language research. Despite variations in sample size, the data mirror the trends and distributions observed in existing emotion analysis literature [35, 21, 48, 22, 49]. Consequently, **ELIta** can be considered a psychologically valid resource for emotion research.

**Gender variation** Gender is a significant factor influencing the annotation of subjective constructs such as emotions. Previous research has shown that men and women often respond differently to the same stimuli [37, 21, 22].

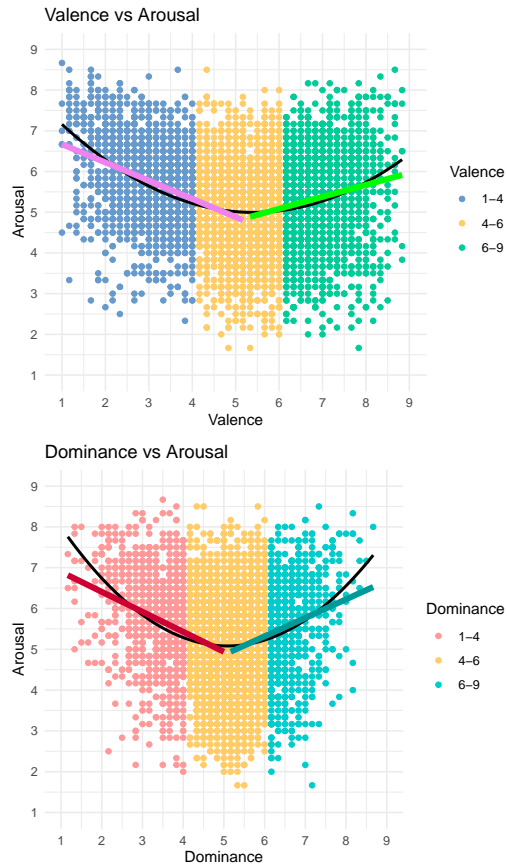To investigate the impact of gender on emotion anno-

tation, a subgroup of words/emojis annotated by both men and women (n=6, 219) was considered. For each word, the mean emotional ratings provided by the different gender groups were calculated. Subsequently, the correlation between the mean ratings was assessed, and statistical tests were conducted to identify any significant differences between the groups.

The most significant differences were found in annotations of AROUSAL, with a correlation of 0.20 and a statistically significant difference calculated using a t-test with a $p-value < 0.005$ (M = 5.39 for women and M = 5.13 for men). As also reported in the literature, women tend to annotate words not only as more arousing, but also with more extreme values on the valence scale, i.e. rating unpleasant words as more negative and pleasant words as more positive.
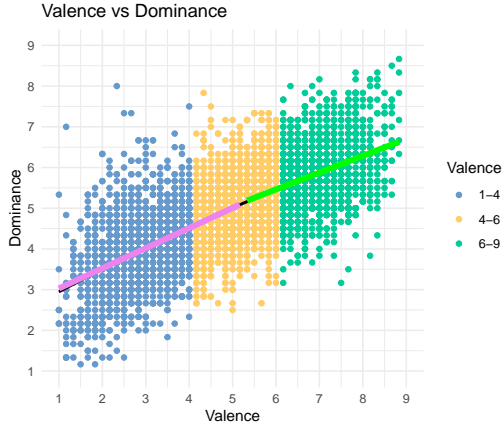
**Figure 4:** Scatterplot showing the relationship between DOM-INANCE and VALENCE in **ELIta**. The lines represent the linear regression according to the values before the VALENCE median (in purple), and after the VALENCE median (in green).

VALENCE also showed a significant difference ($p - value = 0.017$), with women assigning higher AROUSAL and lower VALENCE ratings compared to men (M = 5.08 for women and M = 5.15 for men), although it showed a stronger correlation ($r = 0.64$) than the other dimensions. These results confirm previous findings [37].

Unlike previous studies [37], the results did not show significant differences in DOMINANCE ($p - value > 0.05$, $r = 0.30$).

Regarding basic emotions, women reported significantly higher levels of FEAR ($p < 0.001$) and lower levels of TRUST and SURPRISE (both $p < 0.01$) compared to men, according to the t-test. For example, female participants expressed significantly lower levels of TRUST towards relationship-related words than male participants, with mean scores for 'partner' *partner*, 'spouse' *sposo*, and 'wedding' *nozze* ranging from 0.5 to 0.87 compared to mean male rating of 1.

These findings indicate that gender significantly influences emotion annotation, particularly for AROUSAL and VALENCE (see Fig. 5). The outcomes again corroborate trends observed in the literature for other languages [49], underlining the importance of offering non-aggregated resources to better represent the differences between speakers.

## 4. Conclusions

This research introduces a new lexicon for Italian that collects word-emotion associations. Notably, it is the first lexicon, to the authors' knowledge, to be annotated using both categorical and dimensional approaches. Furthermore, it offers an innovative non-aggregated version of
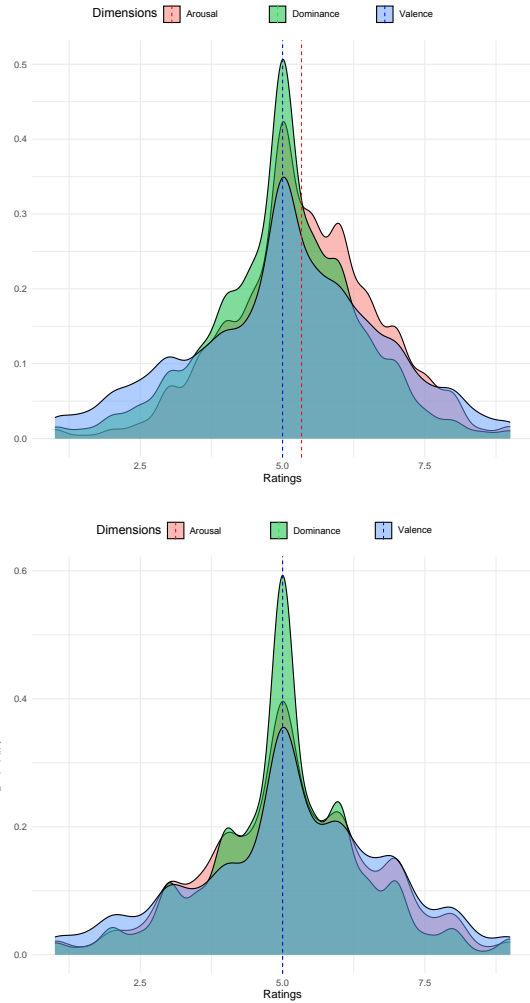


**Figure 5:** Dimensions distribution in the annotations of men (bottom) and women (top)

the data, reflecting a 'perspectivist' approach that values disagreement as valuable information, such as women showing a greater tendency towards negative VALENCE and higher AROUSAL ratings than man. Analyses using correlations between basic emotions and dimensions, along with comparisons to existing resources such as ANEW, underscore the lexicon's potential to deepen our understanding of the interplay between emotions and language. While **ELIta** represents a significant step forward in capturing the complexity of emotion-language interactions in Italian, continued development will be essential to addressing its current limitations and maximizing its utility as a comprehensive tool for emotional analysis.

# References

[1] F. M. Citron, M. A. Gray, H. D. Critchley, B. S. Weekes, E. C. Ferstl, Emotional valence and arousal affect reading in an interactive way: Neuroimaging evidence for an approach-withdrawal framework, Neuropsychologia 56 (2014) 79–89. URL: https://www.sciencedirect.com/science/article/pii/S0028393214000062. doi:https://doi.org/10.1016/j.neuropsychologia.2014.01.002.

[2] J. A. Hinojosa, E. M. Moreno, P. Ferré, Affective neurolinguistics: towards a framework for reconciling language and emotion, Language, Cognition and Neuroscience 35 (2020) 813–839. doi:10.1080/23273798.2019.1620957.

[3] J. A. Hinojosa, J. Haro, R. Calvillo-Torres, L. González-Arias, C. Poch, P. Ferré, I want it small or, rather, give me a bunch: the role of evaluative morphology on the assessment of the emotional properties of words, Cognition and Emotion 36 (2022) 1203–1210. doi:10.1080/02699931.2022.2093840.

[4] V. Kuperman, Accentuate the positive: Semantic access in english compounds, Frontiers in Psychology 4 (2013). URL: https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00203. doi:10.3389/fpsyg.2013.00203.

[5] G. Lapesa, S. Padó, T. Pross, A. Rossdeutscher, Are doggies really nicer than dogs? the impact of morphological derivation on emotional valence in German, in: C. Gardent, C. Retoré (Eds.), Proceedings of the 12th International Conference on Computational Semantics (IWCS) — Short papers, 2017. URL: https://aclanthology.org/W17-6922.

[6] J. S. Adelman, Z. Estes, M. Cossu, Emotional sound symbolism: Languages rapidly signal valence via phonemes, Cognition 175 (2018) 122–130. URL: https://www.sciencedirect.com/science/article/pii/S0010027718300374. doi:https://doi.org/10.1016/j.cognition.2018.02.007.

[7] M. Conrad, S. Ullrich, D. Schmidtke, S. Kotz, Erps reveal an iconic relation between sublexical phonology and affective meaning, Cognition 226 (2022) 105182. URL: https://www.sciencedirect.com/science/article/pii/S0010027722001706. doi:https://doi.org/10.1016/j.cognition.2022.105182.

[8] J. Haro, R. Calvillo, C. Poch, J. A. Hinojosa, P. Ferré, Your words went straight to my heart: the role of emotional prototypicality in the recognition of emotion-label words, Psychological Research 87 (2023) 1075–1084. URL: https://doi.org/10.1007/s00426-022-01723-6. doi:10.1007/s00426-022-01723-6.

[9] C. Herbert, M. Junghofer, J. Kissler, Event related potentials to emotional adjectives during reading, Psychophysiology 45 (2008) 487–498. doi:https://doi.org/10.1111/j.1469-8986.2007.00638.x.

[10] P. Subasic, A. Huettner, Affect analysis of text using fuzzy semantic typing, IEEE Transactions on Fuzzy Systems 9 (2001) 483–496. doi:10.1109/91.940962.

[11] S. Buechel, U. Hahn, A flexible mapping scheme for discrete and dimensional emotion representations, Cognitive Science (2017).

[12] J. Staiano, M. Guerini, Depeche mood: a lexicon for emotion analysis from crowd annotated news, in: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), Association for Computational Linguistics, Baltimore, Maryland, 2014, pp. 427–433. URL: https://aclanthology.org/P14-2070. doi:10.3115/v1/P14-2070.

[13] M. Köper, S. Schulte im Walde, Automatically generated affective norms of abstractness, arousal, imageability and valence for 350 000 German lemmas, in: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), European Language Resources Association (ELRA), Portorož, Slovenia, 2016, pp. 2595–2598. URL: https://aclanthology.org/L16-1413.

[14] C. Strapparava, A. Valitutti, WordNet affect: an affective extension of WordNet, in: Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), European Language Resources Association (ELRA), Lisbon, Portugal, 2004. URL: http://www.lrec-conf.org/proceedings/lrec2004/pdf/369.pdf.

[15] H. Schuff, J. Barnes, J. Mohme, S. Padó, R. Klinger, Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus, in: Proceedings of the 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 13–23. URL: https://aclanthology.org/W17-5203. doi:10.18653/v1/W17-5203.

[16] M. Brysbaert, A. B. Warriner, V. Kuperman, Concreteness ratings for 40 thousand generally known english word lemmas, Behavior Research Methods 46 (2014) 904–911. URL: https://doi.org/10.3758/s13428-013-0403-5. doi:10.3758/s13428-013-0403-5.

[17] S. Mohammad, P. Turney, Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon, in: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text, Association for Computational Lin-

guistics, Los Angeles, CA, 2010, pp. 26–34. URL: https://aclanthology.org/W10-0204.

[18] J. A. Russell, A circumplex model of affect., Journal of Personality and Social Psychology 39 (1980) 1161–1178.

[19] J. A. Russell, A. Mehrabian, Evidence for a three-factor theory of emotions, Journal of Research in Personality 11 (1977) 273–294. URL: https://www.sciencedirect.com/science/article/pii/009265667790037X. doi:https://doi.org/10.1016/0092-6566(77)90037-X.

[20] A. Mehrabian, Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament, Current Psychology 14 (1996) 261–292.

[21] M. M. Bradley, P. J. Lang, M. M. Bradley, P. J. Lang, Affective Norms for English Words (ANEW): Instruction manual and affective ratings, Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

[22] A. Warriner, V. Kuperman, M. Brysbaert, Norms of valence, arousal, and dominance for 13,915 english lemmas, Behavior research methods 45 (2013). doi:10.3758/s13428-012-0314-x.

[23] S. M. Mohammad, Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words, in: Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL), Melbourne, Australia, 2018.

[24] P. Ekman, An argument for basic emotions, Cognition & emotion 6 (1992) 169–200.

[25] P. Ekman, Emotion in the human face, 2 ed., Cambridge University Press, New York, 1982.

[26] R. Plutchik, A general psychoevolutionary theory of emotion, in: Theories of emotion, Elsevier, 1980, pp. 3–33.

[27] L. F. Barrett, K. A. Lindquist, M. Gendron, Language as context for the perception of emotion, Trends in Cognitive Sciences 11 (2007) 327–332. URL: http://dx.doi.org/10.1016/j.tics.2007.06.003. doi:10.1016/j.tics.2007.06.003.

[28] F. Bianchi, D. Nozza, D. Hovy, FEEL-IT: Emotion and sentiment classification for the Italian language, in: Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, Association for Computational Linguistics, Online, 2021, pp. 76–83. URL: https://aclanthology.org/2021.wassa-1.8.

[29] R. Sprugnoli, Multiemotions-it: A new dataset for opinion polarity and emotion analysis for italian, in: Proceedings of the Seventh Italian Conference on Computational Linguistics (CLiC-it 2020), Accademia University Press, Torino, 2020, pp. 402–408. URL: http://hdl.handle.net/10807/165687.

[30] D. Santos, A. Simões, C. Mota, Broad

coverage emotion annotation, Lang. Resour. Eval. 56 (2022) 857–879. URL: https://doi.org/10.1007/s10579-021-09565-1. doi:10.1007/s10579-021-09565-1.

[31] K. Kajava, E. Öhman, H. Piao, J. Tiedemann, Emotion preservation in translation: Evaluating datasets for annotation projection, in: Proceedings of the 5th Conference of Digital Humanities in the Nordic Countries, 2020, pp. 38–50.

[32] B. Maia, D. Santos, Language, emotion, and the emotions: The multidisciplinary and linguistic background, Language and Linguistics Compass 12 (2018) e12280. doi:https://doi.org/10.1111/lnc3.12280, e12280 LNCO-0696.R3.

[33] L. Passaro, L. Pollacci, A. Lenci, Item: A vector space model to bootstrap an italian emotive lexicon, in: Second Italian Conference on Computational Linguistics CLiC-it 2015, Academia University Press, 2015, pp. 215–220.

[34] O. Araque, L. Gatti, J. Staiano, M. Guerini, Depechemood++: A bilingual emotion lexicon built through simple yet powerful techniques, IEEE Transactions on Affective Computing 13 (2022) 496–507. doi:10.1109/TAFFC.2019.2934444.

[35] J. Redondo, I. Fraga, I. Padrón, M. Comesaña, The spanish adaptation of ANEW (affective norms for english words), Behav. Res. Methods 39 (2007) 600–605.

[36] L. J. Speed, M. Brysbaert, Ratings of valence, arousal, happiness, anger, fear, sadness, disgust, and surprise for 24, 000 dutch words, Behavior Research Methods (2023). URL: http://dx.doi.org/10.3758/s13428-023-02239-6. doi:10.3758/s13428-023-02239-6.

[37] M. Montefinese, E. Ambrosini, B. Fairfield, N. Mammarella, The adaptation of the affective norms for english words (anew) for italian, Behavior Research Methods 46 (2014) 887–903.

[38] V. Basile, It's the end of the gold standard as we know it. on the impact of pre-aggregation on the evaluation of highly subjective tasks, in: DP@AI*IA, 2020. URL: https://api.semanticscholar.org/CorpusID:229344921.

[39] V. Basile, M. Fell, T. Fornaciari, D. Hovy, S. Paun, B. Plank, M. Poesio, A. Uma, We need to consider disagreement in evaluation, Proceedings of the 1st Workshop on Benchmarking: Past, Present and Future (2021). URL: https://api.semanticscholar.org/CorpusID:236486317.

[40] J. Baan, R. Fernández, B. Plank, W. Aziz, Interpreting predictive probabilities: Model confidence or human label variation?, 2024. URL: https://arxiv.org/abs/2402.16102. arXiv:2402.16102.

[41] A. M. Davani, M. Díaz, V. Prabhakaran, Dealing with disagreements: Looking beyond the majority

vote in subjective annotations, Transactions of the Association for Computational Linguistics 10 (2022) 92–110. URL: https://aclanthology.org/2022.tacl-1.6. doi:10.1162/tacl_a_00449.

[42] T. De Mauro, Il nuovo vocabolario di base della lingua italiana, Internazionale (2016). URL: https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana.

[43] F. Tamburini, Neural Models for the Automatic Processing of Italian, Pàtron, Bologna, 2022.

[44] A. Moors, K. R. Scherer, The role of appraisal in emotion, in: M. Robinson, E. Watkins, E. Harmon-Jones (Eds.), Handbook of cognition and emotion, Guilford Press, 2013, pp. 135–155.

[45] K. Krippendorff, Content analysis: An introduction to its methodology, Sage publications, 2018.

[46] S. M. Mohammad, Word affect intensities, in: Proceedings of the 11th Edition of the Language Resources and Evaluation Conference (LREC-2018), Miyazaki, Japan, 2018.

[47] P. Ferré, M. Guasch, N. Martínez-García, I. Fraga, J. A. Hinojosa, Moved by words: Affective ratings for a set of 2, 266 spanish words in five discrete emotion categories, Behavior Research Methods 49 (2016) 1082–1094. URL: http://dx.doi.org/10.3758/s13428-016-0768-3. doi:10.3758/s13428-016-0768-3.

[48] L. Sarli, N. Justel, Emotional words in spanish: Adaptation and cross-cultural differences for the affective norms for english words (anew) on a sample of argentinian adults, Behavior Research Methods 54 (2021) 1595–1610. URL: http://dx.doi.org/10.3758/s13428-021-01682-7. doi:10.3758/s13428-021-01682-7.

[49] A. P. Soares, M. Comesaña, A. P. Pinheiro, A. Simões, C. S. Frade, The adaptation of the affective norms for english words (anew) for european portuguese, Behavior Research Methods 44 (2011) 256–269. doi:10.3758/s13428-011-0131-7.

# Appendix

**Figure .6:** The following plots show the relationship between the **ELIta** lexicon and the Italian adaptation of the ANEW of Montefinese et al.. For each dimension, it is possible to see the regression line and the words that are furthest from the line, i.e. the words that were rated differently by the annotators between the two lexicons.

Dominance