# Linking Emotions: Affective and Lexical Resources for Italian in Linked Open Data

Eliana Di Palma[1,*,†], Valerio Basile[1,†], Agnese Vardanega[2,†], Giuliano Gabrieli[3,†] and Marco Vassallo[3,†]

[1]University of Turin - Computer Science Department, Corso Svizzera 185, 10149 Turin, Italy
[2]University of Teramo - Department of Political Sciences, Teramo, Italy
[3]CREA Research Centre for Agricultural Policies and Bio-economy, Rome (Italy)

**Abstract**

The growing interest in analysing emotions in Italian texts has led to the development of various affective resources, often independently constructed and lacking interoperability. To address this fragmentation, we adopt a Linked Open Data (LOD) approach. This paper presents three main contributions: (1) the release of Sentix 3.0, a revised and enriched polarity lexicon for Italian, together with two derivatives (MAL and WMAL) that address morphological and word frequency variation issues; (2) a new quartile-based methodology to discretize continuous polarity scores; and (3) the linking of Sentix 3. 0 and ELIta, an annotated emotion lexicon with categorical labels, to the LiITA Lemma Bank using standard ontologies (OntoLex-Lemon, MARL) and the newly introduced elita ontology for categorical emotion representation based on Plutchik's Wheel. At the heart of the linking is the word, the central node for aligning different lexical resources.

**Keywords**

Italian, Linguistic Linked Open Data, Emotions, Sentiment, Language Resources

## 1. Introduction

The increasing interest in analyzing emotions from Italian text has led to many specialized linguistic resources, such as lexicons and corpora. However, these resources are often independently built, with varying annotation methods, severely limiting their interoperability.

To address these challenges, we propose a Linked Open Data (LOD) approach. Inspired by projects such as LiLa [1], our work aims to integrate and formalize two Italian affective lexical resources, making them interoperable through linking to the LiITA project's lemma bank [2].

This article presents three original contributions:

- Resource linking and ontology development: we align ELIta, a lexicon annotated with categorical emotions, and the updated Sentix 3.0 with the LiITA project's lemma bank, using the MARL ontology to formally represent affective relations. Linking is achieved through the use of `ontolex:CanonicalForm`, and we introduce the new *elita ontology*, specifically designed to model categorical emotions based on Plutchik's Wheel, as applied to lexical entries.
- Release of Sentix 3.0: we introduce a revised and enriched version of the Sentix affective lexicon for Italian, along with two derivative resources, MAL and WMAL, designed to improve sentiment analysis by addressing morphological and frequency-based challenges.
- New polarity discretization methodology: we propose a novel approach to discretizing continuous polarity scores using quartile-based thresholds. This technique, initially introduced in Vassallo et al. [3], enables a more robust and interpretable classification of polarity levels.

The paper is structured as follows: Section 2 provides an overview of existing resources for Italian and introduces the Linked Open Data (LOD) paradigm, with a focus on its affective application in the LatinAffectus resource, developed as part of the LiLa (Linking Latin) project. Section 3 presents Sentix 3.0 along with its derivative resources, MAL and WMAL. Section 4 details the methodology adopted for polarity categorization and the ontological framework used. Section 5 outlines the LOD-based linking process, while Section 6 offers example queries that demonstrate how the linked data can be explored. Finally, Section 7 discusses future research directions and potential developments.

*CLiC-it 2025: Eleventh Italian Conference on Computational Linguistics, September 24 — 26, 2025, Cagliari, Italy*
*Corresponding author.
†These authors contributed equally.
eliana.dipalma@unito.it (E. D. Palma); valerio.basile@unito.it (V. Basile); avardanega@unite.it (A. Vardanega); giuliano.gabrieli@crea.gov.it (G. Gabrieli); marco.vassallo@crea.gov.it (M. Vassallo)
0000-0003-2154-2696 (E. D. Palma); 0000-0001-8110-6832 (V. Basile); 0000-0002-1419-9896 (A. Vardanega); 0009-0005-1153-5662 (G. Gabrieli); 0000-0001-7016-6549 (M. Vassallo)




CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2025-11-30

## 2. Related Works

In recent decades, the analysis of emotions has led to the development of numerous linguistic resources, particularly affective lexicons and annotated datasets. Despite the pervasive impact of extensive end-to-end language models (LLMs), affective and other annotated lexicons remain a dynamic area of research within computational linguistics, maintaining its vitality.They are particularly effective in the development of hybrid approaches [4], especially suitable for small or domain-specific corpora or low-resource languages [5, 6], where they ensure greater effectiveness and interpretability.

These lexical resources are designed to be FAIR (Findable, Accessible, Interoperable, Reusable) and transparent: as every system decision based on a lexicon can be traced back to specific entries, ensuring interpretability. Moreover, unlike computationally demanding large language models, lexicons are accessible to a broader research community, promoting the democratization of research.

For the Italian language, several lexicons [7, 8] and corpora [9, 10] annotated according to the sentiment [11] or emotion associated with the words or texts have been created and published, created by both automatic [12, 13] and manual methods [14]. However, much like the broader landscape of digital resources, there is a significant lack of clear, standardized guidelines for affective resources, resulting in poor interoperability. Each lexicon or corpus is typically developed independently, and although they often share similar objectives, such as sentiment analysis or emotion detection, they differ widely in terms of annotation categories, selected lemmas, and annotation methodologies.

In addition, the availability of these resources is highly fragmented. They are typically hosted on disparate platforms and, even when co-located within a shared repository infrastructure, interoperability between them remains limited or nonexistent.

A notable attempt to address these issues can be found in the LiLa project [15], which pioneered a new model of interoperability for Latin linguistic resources. This vision has since been adopted for Italian by the LiITA project [2]. Both initiatives share an ambitious goal: to build an interconnected system where the lemma acts as the central node linking multiple knowledge bases. The lemma becomes the foundation for connecting diverse databases through the use of shared vocabularies. At the heart of both projects lies the principle of Linked Open Data [16], enabling integration and reuse across resources.

Central to the LOD paradigm is the Semantic Web, which promotes the use of interoperable and interlinkable data schemas for online information. These schemas, commonly referred to as ontologies or vocabularies, support consistent data representation and semantic integration. To this end, the W3C has introduced foundational technologies such as RDF and OWL. Building upon these standards, the MARL ontology has been developed to formally describe opinions and to associate them with contextual information(such as opinion topic, features described in the opinion, etc.).

This infrastructure underpins the design of LatinAffectus [17], a polarity lexicon originated within the LiLa project that adopts a multi-ontology framework for formal representation. Specifically, three standards are used: Lemon and Ontolex [18] to describe lexical data, and MARL [19] to encode sentiment information. In this schema, the lexicon itself is modeled as an instance of class E31 Document13 from the CIDOC Conceptual Reference Model (CRM), an ontology designed to represent entities and relationships in the cultural heritage domain. In parallel, LatinAffectus is also declared as a lexicon-type object following the LInguistic MEtadata (lime) module of Ontolex.

Lexical entries in the resource are connected to the lexicon through the `lime:entry` property and are instantiated as objects of the class `ontolex:LexicalEntry`. Each lexical entry is associated with a label, an `ontolex:canonicalForm` (linking it to the lemma in the LiLa Knowledge Base), and an `ontolex:sense`, which captures its meaning. Since LatinAffectus is concerned with prior polarity, each lexical entry has only one sense, modeled as an instance of the class `ontolex:LexicalSense`. This sense is further described by a label, the relation `marl:hasPolarity`, and the property `marl:polarityValue`. The `marl:hasPolarity` relation links the sense to a category within `marl:Polarity` (namely positive, negative, or neutral) while `marl:polarityValue` assigns a numerical sentiment score from the predefined set: 1.0, 0.5, 0.0, -0.5, or -1.0.

The integration of emotions with Linked Open Data is not a new challenge. Relevant contributions in this area include Iglesias et al. [20] and Sanchez-Rada et al. [21]. These works introduced Onyx, that is an RDF ontology designed to represent emotions in textual content within the framework of the Semantic Web and Linked Open Data. It builds upon and aligns with standards such as EmotionML [22], the NLP Interchange Format (NIF), and lemon (Lexicon Model for Ontologies) [18], providing a flexible yet structured approach to annotating emotional content. Central to Onyx is the concept of an `EmotionSet`, a container for one or more `Emotion` instances, each representing a specific emotional state. These emotions are linked to standardized `EmotionCategory` resources, which can reflect different psychological models, and may include numerical values for emotion intensity and affective dimensions such as arousal and valence. Onyx also enables the asso-

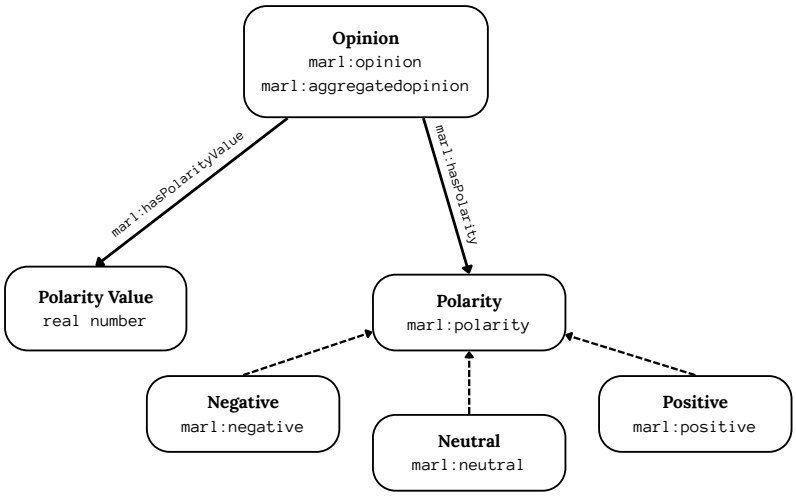


**Figure 1:** Marl Ontology [19]

ciation of emotional annotations with external resources or entities.

While Onyx is conceptually rich and expressive, its structure is relatively complex, as it is designed to represent the emotional content of texts.

**LiITA** LiITA (Linking Italian) is a Knowledge Base (KB) designed to foster interoperability among various Italian linguistic resources by leveraging the principles of Linked Open Data (LOD).

The core component of the LiITA KB is its Lemma Bank, a comprehensive collection of Italian lemmas. These lemmas, which are conventional lexical citation forms used across linguistic resources, serve as the central connection point for interlinking both lexical and textual data. The architecture of the LiITA KB mirrors that of the LiLa KB for Latin[1], operating under the assumption that all interoperable (meta)data sources within the KB are word-related.

Following the Linked Data paradigm, LiITA achieves conceptual interoperability among its distributed resources by applying vocabularies commonly used within the Linguistic Linked Open Data (LLOD). For the Lemma Bank specifically, this means adopting the vocabulary defined by OntoLex-Lemon [18], one of the most widely used models for representing and publishing lexical resources as Linked Data. The Lemma Bank of LiITA is a collection of canonical forms as intended by the Ontolex-Lemon ontology (`ontolex:canonicalForm` property, the conventionally chosen representation for the entire set of inflected forms belonging to a particular lexical entry), modeled as individuals of the Class `lila:Lemma`[2], which is a subclass of `ontolex:Form`, originally created for the LiLa project, and adopted in the LiITA Lemma Bank accordingly.

The Lemma Bank was initially populated with approximately $94,000$ lemmas derived from the online version of the Nuovo De Mauro dictionary, after excluding about $13,000$ multi-word expressions.

**ELIta** ELIta [23], a recently introduced linguistic resource, comprises a lexicon annotated via crowdsourcing. This annotation scheme incorporates both basic emotions, based on Plutchik's model [24] with corresponding association degrees, and the VAD (Valence, Arousal, Dominance) emotional dimensions, thus including sentiment (valence) [25]. The lexical items, primarily sourced from the De Mauro dictionary [26], were annotated in isolation. To date, four distinct versions of this lexicon have been released [8]:

- **RAW**: Full annotations with demographic data.
- **GOLDEN**: Selection of 5 consistent annotations + majority-vote golden label.
- **INTENSITY**: Aggregated intensities from GOLDEN; includes auto-generated "love" and "neutral".
- **BINARY**: Binary version of INTENSITY using 0.50 threshold.

[1] https://lila-erc.eu/data-page/

[2] http://lila-erc.eu/ontologies/lila/Lemma

## 3. Sentix 3.0 and two derived resources

Sentix is an affective lexicon for the Italian language created in 2013 by aligning several lexical resources, namely SentiWordNet [27] and MultiWordNet [28] through WordNet [29, 30].

The first version [11] was built by transferring the *synset* annotations from SentiWordNet to the respective Italian *synsets* of MultiWordNet, using an automatic mapping [cf. 31] to resolve the partial alignment of SentiWordNet's indices (based on WordNet 3.0), and those of MultiWordNet (based on WordNet 1.6) [cf. 32]. The performance of the lexicon was evaluated with data manually annotated by independent human judges.

The subsequent version, Sentix 2.0, aggregated the polarity scores of the different senses of a lemma into a single score (-1 to 1), using a weighted average with the sense frequencies calculated on the annotated SemCor corpus [33, 34]. This version, which includes 41,800 different lemmas, has been available on GitHub in the R package *sentixR* since 2019 [35], and has been used in various research projects over these years.

Other lexical resources have been developed from Sentix. The first derived lexicon was MAL [36], created by expanding Sentix 2.0 with inflected forms derived from *Morph-it!*, a morphological lexicon for the Italian language [37]. This was intended to address the inherent difficulties of lemmatization in Italian sentiment analysis - stemming from the language's morphological complexity and the limitations of available NLP tools - which are particularly exacerbated when analyzing user-generated content from social networks (often containing spelling errors, jargon, irregularities, and non-standard syntactic structures). MAL - which inherits Sentix 2.0's scores - has shown to achieve an improvement in overall sentiment analysis performance.

A second derived resource is WMAL [34], a dictionary of inflected forms like MAL, where MAL's scores were recalculated by weighting the original scores inversely with respect to their words frequencies in the TWITA corpus [38] by using the inversed version of the Zipf scale measure [39] that consists in a logarithmic scale based on the well-known Zipf law of word frequency distribution [40]. The two main WMAL lexical and methodological speculations were respectively to give more weight to low frequent terms and to reduce the polarity imbalance when using parametric threshold values to assign polarity classes: even small variations in these values in fact showed to have an opposite impact on the ability to correctly predict negative versus positive polarity [4]. WMAL has achieved better results in polarity classification, especially for negative messages.

On the occasion of WMAL's update, driven by the enrichment and update of the TWITA corpus to 2022, it was decided to harmonize all three resources, which were developed years apart; this involved not only updating WMAL's weights but also rectifying the interconnections among the three to ensure overall consistency. In particular, during the transition from Sentix to MAL, i.e., from lemmas to inflected forms, new identical forms with different scores and different senses are inevitably created. It was decided to manage these in a coordinated manner, primarily by revising Sentix.

A key part of this harmonization involved a deeper re-examination of the foundational Sentix lexicon. This specific effort, working backwards from the original Sentix version, led to an expansion in the number of linkable synsets between SentiWordNet (SWN) and MultiWordNet (MWN), which will be made available[3].

The revision subsequently involved external resources and supervised phases to identify forms present in Sentix that could generate unexpected duplicate entries in MAL (i.e., entries that could be either base forms or inflected forms), and to expand Sentix itself by back-linking lemmas present in *Morph-it!* traceable to pre-existing entries (717 entries). Finally, neutral terms from SentiWordNet not already present in Sentix were added (22, 117 entries). The new lexicon ultimately contains 63, 660 lemmas [41].

The resources used for the update were: SentiWordNet and MultiWordNet, using the *Open Multilingual Wordnet* (https://omwn.org/) [42, 43], the TreeTagger library [44], and, of course, *Morph-it!* itself.

## 4. Methodology

Following the methodology established for the affective lexicon of Latin within the LiLa project, LatinAffectus, the same approach was adopted for the ELIta and Sentix 3.0 lexicons by using the MARL ontology [19] to represent polarity properties.

In this context, neutrality thresholds for polarity labels were defined within the range between the first and third quartiles of the lexicon. The weighting methodology and polarity calculation based on the new updated version of WMAL are key to Sentix 3.0's polarity categorical classification. In this respect, the first and the third quartile-based interval $[Q1; Q3]$ was calculated on the WMAL scores to better individualize the neutral thresholds and consequently the positive and negative polarity values outside. The quartile-based strategy to detect neutral scores has already provided promising results

[3] Discrepancies between numerical IDs and verbal labels of numerous Italian *synsets*, due to changes between subsequent versions of Wordnet [cfr. 32], often prevents their retrieval when attempting to map from SWN to MWN through standard tools, such as the OMW Python package

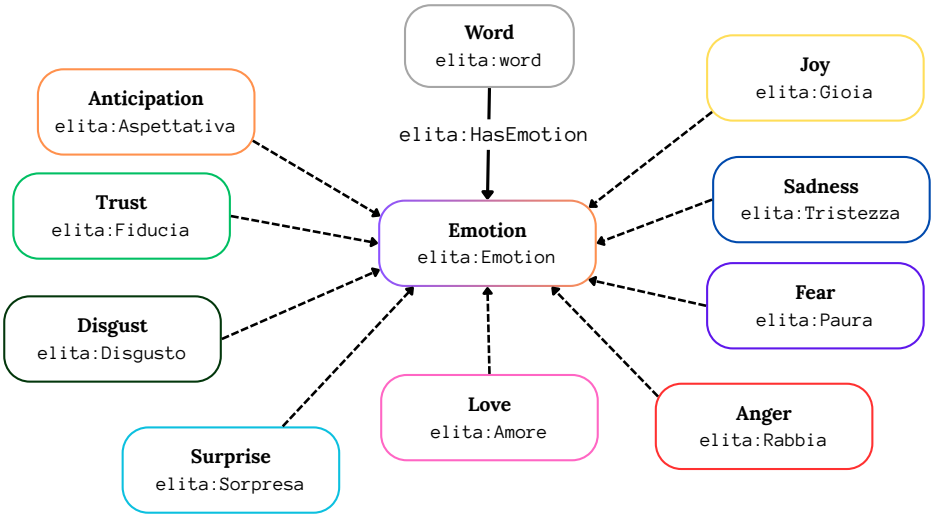


**Figure 2:** *elita ontology*

across annotated corpora[3]. The obtained polarity classification (Table 1) was used to assign categories within `marl:polarity` for bot h Sentix 3.0 and ELIta.

**Table 1**
Polarity classification thresholds based on numerical polarity values.

| Polarity Label | Polarity Value Range |
|---|---|
| Negative | $x < -0.1646$ |
| Neutral | $-0.1646 \leq x \leq 0.1250$ |
| Positive | $x > 0.1250$ |

The other relationship used to describe Sentix 3.0 data is `marl:hasPolaritValue` in which values are continuous from -1 to 1.

For example, the lemma in ELIta "abbandonare" was annotated as `marl:hasPolarity` "Negative", and `marl:hasPolarityValue` "$-0.833$".

However, since the MARL ontology does not provide specific properties for representing categorical emotions, and the structure of Onyx is relatively complex, the *elita ontology* is introduced to fill this gap with a simpler and more transparent model, specifically designed for annotating individual words with emotion categories.

The *elita ontology* has been developed to represent categorical emotions in a structured and interoperable manner, with a particular focus on applications in linguistic and sentiment analysis. In contrast to the MARL ontology, which primarily addresses sentiment polarity (positive, neutral, negative), and Onyx, which incorporates multiple emotion models for text descriptions, *elita ontology* introduces explicit classes and properties for representing discrete emotional categories. These categories are based on Plutchik's Wheel of Emotions [24], and also include the dyad "Love," formed by the combination of "Joy" and "Trust".

At the core of the ontology is the `owl:class` defined `elita:Emotion`, which serves as the general category for all emotion instances. Specific emotions, such as *Gioia* (Joy), are modeled as individuals (instances) of this class.

To associate a resource, such as a lexical item, sentence, or document, with an emotion, the ontology defines the object property `elita:HasEmotion`. This `owl:ObjectProperty` links a subject (e.g., a word or expression) to an instance of `elita:Emotion`, thereby expressing the emotional content attributed to that element (Fig. 2).

Despite its simplicity, the ontology maintains conceptual continuity with Onyx through the use of the `elita:HasEmotion` property, which functionally corresponds to `onyx:hasEmotionCategory`. Moreover, the emotional categories defined in *elita* reflect the annotations present in the ELIta lexical resource, ensuring alignment with existing linguistic data.

This design enables the annotation and querying of resources using fine-grained emotional categories, effectively complementing polarity-based approaches in the representation of lexical entries. At the same time, it ensures interoperability with existing emotion ontologies while providing a lightweight, application-oriented model specifically tailored to lexical annotation.

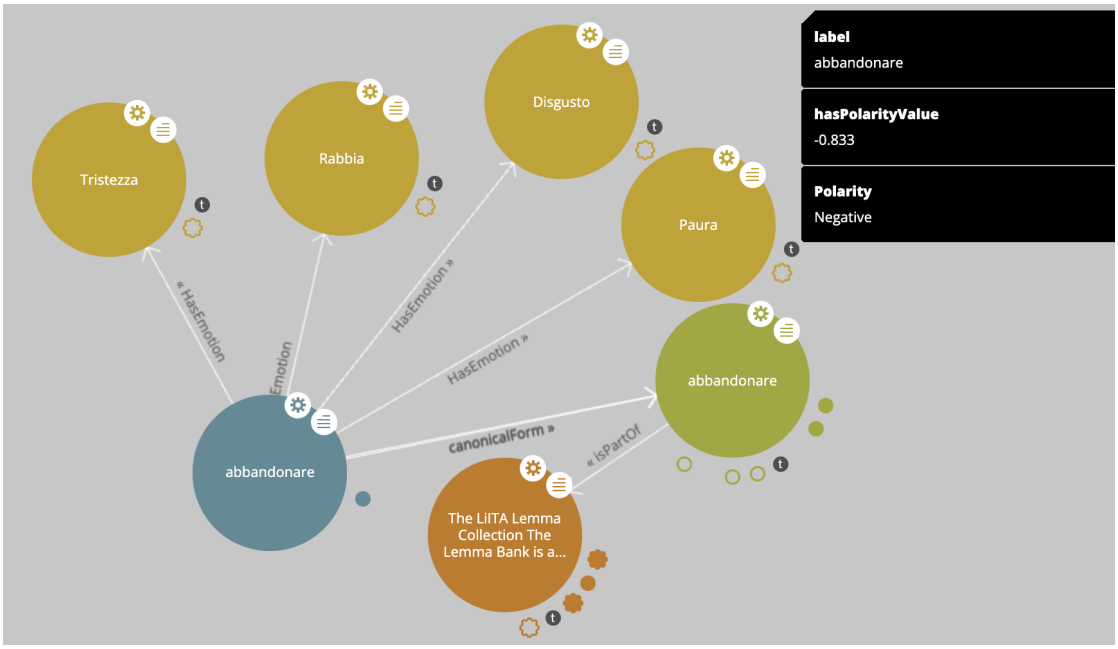


**Figure 3:** Linking visualization via LodLive of the lemma "abbandonare" (to abandon)

## 5. Linking

After selecting the ontologies for the Linked Open Data representation of the lexical resources, the lexicons were converted into RDF format. The linking procedure involved mapping lexical entries, each associated with a unique URI, to their corresponding lemmas within the LiITA's Lemma Bank. The results of this linking process are presented in the following sections (an example of a linked entry is shown in Figure 3).

### 5.1. Linking ELIta

The ELIta lexical resource comprises 6,905 entries, encompassing lemmas (as defined by LiITA), emojis, and multi-word expressions like "a malincuore" (reluctantly). Notably, ELIta's lexical entries include in some cases both masculine and feminine forms of adjectives and nouns. This inclusion aimed to facilitate the assessment of gender-based perceptual differences, particularly when morphological gender is the sole distinguishing factor.

To align ELIta with LiITA's Lemma Bank, emojis were initially removed. The remaining lexical entries were then compared and linked to LiITA's lemma URI where feasible. This process yielded 4,705 ELIta words that exhibited a one-to-one match with lemmas or hypolemmas in LiITA (as shown in Table 2). The remaining approximately 2,000 entries, however, matched multiple lemmas within the lemma bank, or none, necessitating further processing.

The lexical entries that presented a one-to-one match were associated with the lemmas in the Lemma Bank using the Ontolex ontology and the relation `ontolex:CanonicalForm`.

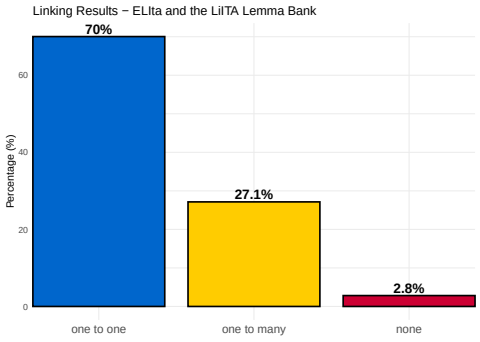


**Figure 4:** Percentages of linking between ELIta and LiITA Lemma Bank.

The high percentage of matches between ELIta and LiITA (shown in Fig. 4) is mainly due to the use of the same lexical source as the backbone of both resources. In particular, both rely heavily on the Nuovo De Mauro dictionary: about 70 percent of ELIta's entries come from the Nuovo Vocabolario di Base [26], while LiITA's Lemma Bank is based on the lexical base of an online version of

**Table 2**
Number of ELIta lexical entries (no emoji) with one link, multiple links, or no representation in the LiITA Lemma Bank.

| One to one | One to many | No one |
|---|---|---|
| 4705 | 1823 | 190 |

**Table 3**
Number of Sentix 3.0 lexical entries with one link, multiple links, or no representation in the LiITA Lemma Bank.

| One to one | One to many | No one |
|---|---|---|
| 22946 | 6244 | 34497 |

the dictionary Nuovo De Mauro[4]. The 190 lexical entries in ELIta that do not have a match in LiITA are mainly multi-word expressions or idiomatic phrases, which were not included in the construction of the LiITA lemma bank.

In addition, some entries without correspondence correspond to feminine forms that were explicitly annotated in ELIta. These forms often carry different affective annotations than their male counterparts, which are usually used as the canonical form of the lemma in LiITA. Consequently, these female variants were not included in the linking process.

One-to-many correspondences, accounting for 27.1 percent of the total, are largely attributable to words that correspond to multiple parts of speech (PoS) in LiITA, and thus correspond to multiple lemmas. In ELIta, the annotation process, following the methodology described in [14], did not specify the part of speech for each entry. As a result, PoS-based disambiguation is not possible in the current version of the resource.

### 5.2. Linking Sentix 3.0

The same procedure was applied to Sentix 3.0, with each entry linked to a lemma from the lemma bank whenever possible. The results revealed that most entries were new to the lemma bank and therefore had no matching lemma. This is primarily because the lexicon contains a large number of multi-word expressions, such as "oggetti per la casa" (household items), "vedova nera" (black widow), "difficoltà di apprendimento" (learning difficulties). As in the case of ELIta, the one-to-many links are due to the presence of lemmas in Sentix that may belong to different parts of speech in LiITA Lemma Bank. To address the one-to-many mappings, a possible solution would be to disambiguate entries based on part of speech. However, the current version of Sentix consists of isolated lexical entries, and PoS tagging typically relies on contextual information. Since such context is not available in the lexicon, it is not currently possible to assign PoS labels reliably. While previous versions of Sentix included some PoS information, this is not present in the current release. As a result, any attempt at disambiguating part of speech would be arbitrary.

Nevertheless, successful linking was achieved in 36% (as shown in Fig. 5) of cases, with 22,946 lemmas matched on a one-to-one basis (see Table 3).

[4]https://dizionario.internazionale.it/

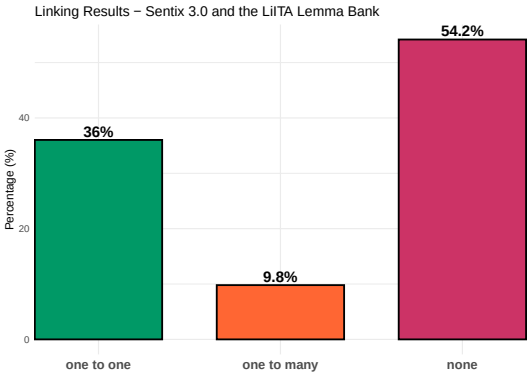


**Figure 5:** Percentages of linking between Sentix 3.0 and LiITA Lemma Bank.

## 6. Query

One of the cardinal principles of Linked Open Data as mentioned above is also the use of standards such as RDF and SPARQL to provide useful information on what is identified by a URI, for the purpose of (meta)data representation and retrieval. If RDF (Resource Description Framework) [45] is the data model underlying the Semantic Web, SPARQL[5] is a query language for (meta)data represented in RDF.

Integrating affective resources into LiITA significantly enhances its query capabilities, allowing for advanced SPARQL interrogations across LiITA's own data and its linked lexical and textual resources.

For instance, it is possible to[6]:

**Query 1. Retrieve the Distribution of Emotions in ELIta** It's possible to query the distribution of emotions as defined within the ELIta lexicon. the SPARQL query counts the number of lemmas associated with each emotion label in the ELIta resource. By linking lemmas from the LiITA Lemma Bank to their corresponding emotion annotations in ELIta, the query retrieves the textual label of each emotion and aggregates the lemmas accordingly. The results are grouped by emotion label and sorted in descending order based on lemma count.

[5]https://www.w3.org/TR/rdf-sparql-query/
[6]The SPARQL queries used to generate these examples are available in the appendix and can be executed via the LiITA endpoint.

Table 4 shows the result of such distribution, with "Aspettativa" (Anticipation) being the most frequent emotion.

**Table 4**
Distribution of lemmas across different emotions.

| Lemmas | Emotion |
|---|---|
| 1257 | "Aspettativa" |
| 1030 | "Gioia" |
| 787 | "Amore" |
| 765 | "Paura" |
| 745 | "Fiducia" |
| 710 | "Rabbia" |
| 645 | "Tristezza" |
| 597 | "Sorpresa" |
| 360 | "Disgusto" |

**Query 2. Retrieve Polarity Distribution in Sentix 3.0:** The SPARQL query counts the number of lemmas associated with each polarity label (e.g., "Positive", "Negative", "Neutral") in the Sentix 3.0 resource. By linking lemmas from the LiITA Lemma Bank to their corresponding polarity annotations in Sentix 3.0, the query retrieves the textual label of each polarity and aggregates the lemmas accordingly. The results are grouped by polarity and sorted in descending order based on lemma count.

This query type helps determine the count of Italian words marked as positive, negative, or neutral based on lemmas shared between Sentix 3.0 and LiITA. Table 5 provides the result, indicating a higher number of neutral lemmas.

**Table 5**
Distribution of lemmas across different polarities.

| Lemmas | polarityLabel |
|---|---|
| 15058 | "Neutral" |
| 3730 | "Negative" |
| 2780 | "Positive" |

**Query 3. Return the average Sentix polarity score for each emotion annotated in ELIta:** Another possible query can be used to identify the most negative emotion in ELIta based on Sentix 3.0 Polarity Value.

The query retrieves the average Sentix polarity value for each emotion label found in the ELIta resource. It does so by:

1. Linking lemmas from LiITA (via lila:lemma) to their associated emotions in ELIta.
2. Retrieving corresponding polarity values from Sentix.
3. Grouping the results by emotion label and calculating the average polarity for each.
4. Sorting the results in ascending order of polarity, so the most negative emotions appear first.

The results indicate that Disgust, rather than Sadness, consistently emerges as the most negative emotion when analyzing average polarity scores for emotions based on Sentix 3.0 annotations. This is visualized in Figure 6, where the color gradient reflects polarity intensity.

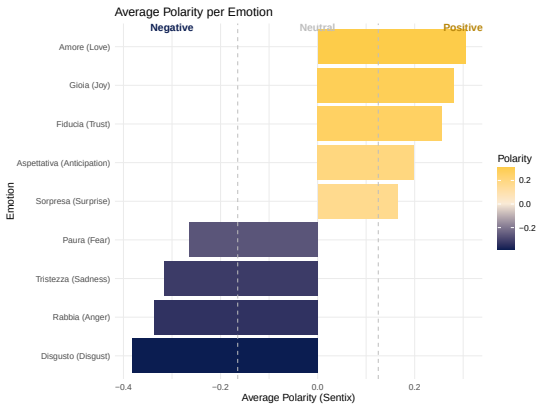


**Figure 6:** Average polarity scores for different emotions based on Sentix annotations. The color gradient represents polarity intensity, with dark blue indicating strong negative valence, dark gold indicating strong positive valence, and pale shades indicating neutrality. Dashed lines mark the thresholds separating negative, neutral, and positive polarity regions.

**Query 4. Determine the polarity of lemmas annotated with contrasting emotions:** Since words in ELIta can be associated with multiple emotions, we explored instances where Joy and Sadness, two emotions of opposing polarities, co-occurred in the annotation of the same lemma.

As a first step, we queried the number of words in ELIta that are simultaneously associated with both Joy and Sadness, grouped by their Sentix polarity label.

More specifically, the query:

1. Selects lemmas from the LiITA Lemma Bank that are linked to ELIta entries.
2. Filters for those lemmas tagged simultaneously with the emotions `elita:Gioia` (Joy) and `elita:Tristezza` (Sadness).
3. Matches these lemmas to their corresponding Sentix 3.0 polarity labels (Positive, Negative, or Neutral).
4. Counts how many lemmas fall into each polarity category.

5. Sorts the output by descending frequency.

We found that in most instances the simultaneous presence of Joy and Sadness corresponded to a neutral polarity. The second most common polarity observed was positive (as shown in Table 6).

**Table 6**
Distribution of lemmas in ELIta associated with both Joy and Sadness, by Sentix polarity.

| Polarity | Lemmas |
|---|---|
| "Neutral" | 24 |
| "Positive" | 18 |
| "Negative" | 2 |

Interestingly, only two words, "invocare" (to invoke) and "umore" (mood), identified through a dedicated query[7], consistently exhibited a negative polarity according to Sentix 3.0. Their respective polarity values are shown in Table 7.

The examples showcased a range of queries that extract information not only from individual resources but also by integrating data from both Sentix 3.0 and ELIta, highlighting how interoperability enables more comprehensive analysis of affective lexical information.

## 7. Conclusions and Future Works

This paper introduces, for the first time in Italian, two affective lexical resources, Sentix 3.0 and ELIta, published according to the Linked Open Data (LOD) paradigm. It also presents the new version of the Sentix 3.0 resource and its derivatives, MAL and WMAL, now available on GitHub. Additionally, the ontology developed for rendering the ELIta emotional lexicon within the Linguistic Linked Open Data (LLOD) framework is introduced.

Both resources have been linked to the LiITA Lemma Bank, thus contributing to and enriching the possibilities of investigation and promoting interoperability among LLOD resources.

Through the LOD paradigm, these resources also support interdisciplinary applications, particularly within the digital humanities (e.g., cultural heritage, social sciences), where linguistic knowledge graphs find practical applications (e.g., through frameworks like CIDOC CRM or LiLa).

Nonetheless, this work represents only the initial phase of fully aligning these affective resources with LiITA. Future efforts will focus on resolving one-to-many mappings and incorporating new lemmas into the LiITA Lemma Bank where applicable.

[7]The corresponding query is provided in the appendix.

**Table 7**
Lemmas associated with both Joy and Sadness in ELIta that exhibit negative polarity in Sentix 3.0.

| Lemmas label | Polarity Value |
|---|---|
| "invocare" | −0.25 |
| "umore" | −0.5 |

Further challenges, such as the emotion analysis of literary texts or interlingual evaluations between regional variants of Italian, can be addressed through interoperability in an ecosystem where the word is the basis of knowledge.

## Acknowledgments

This contribution is funded by the European Union — Next Generation EU, Mission 4 Component 1 CUP J53D2301727OOO1 — PRIN 2022 PNRR project *LiITA: Interlinking Linguistic Resources for Italian via Linked Data.*

## References

[1] F. Mambrini, M. Passarotti, E. Litta, G. Moretti, Interlinking valency frames and wordnet synsets in the lila knowledge base of linguistic resources for latin, in: Further with Knowledge Graphs, IOS Press, 2021, pp. 16–28.

[2] E. Litta, M. Passarotti, P. Brasolin, G. Moretti, V. Basile, A. Di Fabio, C. Bosco, The lemma bank of the LiITA knowledge base of interoperable resources for Italian, in: F. Dell'Orletta, A. Lenci, S. Montemagni, R. Sprugnoli (Eds.), Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024), CEUR Workshop Proceedings, Pisa, Italy, 2024, pp. 517–522. URL: https://aclanthology.org/2024.clicit-1.61/.

[3] M. Vassallo, G. Gabrieli, V. Basile, C. Bosco, Neutral score detection in lexicon-based sentiment analysis: The quartile-based approach, in: F. Dell'Orletta, A. Lenci, S. Montemagni, R. Sprugnoli (Eds.), Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024), CEUR Workshop Proceedings, Pisa, Italy, 2024, pp. 976–982. URL: https://aclanthology.org/2024.clicit-1.105/.

[4] M. Polignano, V. Basile, P. Basile, G. Gabrieli, M. Vassallo, C. Bosco, A hybrid lexicon-based and neural approach for explainable polarity detection, Information Processing & Management 59 (2022) 103058. doi:10.1016/j.ipm.2022.103058.

[5] M. Alfreihat, O. S. Almousa, Y. Tashtoush, A. Al-Sobeh, K. Mansour, H. Migdady, Emo-sl framework: Emoji sentiment lexicon using text-based

features and machine learning for sentiment analysis, IEEE Access 12 (2024) 81793–81812. doi:10.1109/ACCESS.2024.3382836.

[6] F. Koto, T. Beck, Z. Talat, I. Gurevych, T. Baldwin, Zero-shot sentiment analysis in low-resource languages using a multilingual sentiment lexicon, in: Y. Graham, M. Purver (Eds.), Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, St. Julian's, Malta, 2024, pp. 298–320. URL: https://aclanthology.org/2024.eacl-long.18/. doi:10.18653/v1/2024.eacl-long.18.

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

[8] E. Di Palma, ELIta (emotion lexicon for italian), 2024. ILC-CNR for CLARIN-IT repository hosted at Institute for Computational Linguistics "A. Zampolli", National Research Council, in Pisa.

[9] 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.

[10] 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.

[11] V. Basile, M. Nissim, Sentiment analysis on Italian tweets, in: Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2013, pp. 100–107.

[12] 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.

[13] 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.

[14] 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.

[15] M. Passarotti, The project of the index thomisticus treebank, in: Digital Classical Philology, De Gruyter Saur, 2019, pp. 299–320.

[16] T. Berners-Lee, J. Hendler, O. Lassila, The semantic web, Scientific american 284 (2001) 34–43.

[17] F. Mambrini, M. Passarotti, Representing etymology in the lila knowledge base of linguistic resources for latin, in: Proceedings of the 2020 Globalex Workshop on Linked Lexicography, 2020, pp. 20–28.

[18] J. P. McCrae, J. Bosque-Gil, J. Gracia, P. Buitelaar, P. Cimiano, The ontolex-lemon model: development and applications, in: Proceedings of eLex 2017 conference, 2017, pp. 19–21.

[19] A. Westerski, C. A. Iglesias, F. T. Rico, Linked opinions: Describing sentiments on the structured web of data, in: SDoW@ISWC, 2011. URL: https://api.semanticscholar.org/CorpusID:10397646.

[20] C. Iglesias, J. Sánchez-Rada, G. Vulcu, P. Buitelaar, Linked data models for sentiment and emotion analysis in social networks, in: F. A. Pozzi, E. Fersini, E. Messina, B. Liu (Eds.), Sentiment Analysis in Social Networks, Morgan Kaufmann, Boston, 2017, pp. 49–69. URL: https://www.sciencedirect.com/science/article/pii/B9780128044124000048. doi:https://doi.org/10.1016/B978-0-12-804412-4.00004-8.

[21] J. F. Sánchez-Rada, C. A. Iglesias, Onyx: A linked data approach to emotion representation, Information Processing Management 52 (2016) 99–114. URL: https://www.sciencedirect.com/science/article/pii/S030645731500045X. doi:https://doi.org/10.1016/j.ipm.2015.03.007, emotion and Sentiment in Social and Expressive Media.

[22] F. Burkhardt, C. Pelachaud, B. W. Schuller, E. Zovato, EmotionML, Springer International Publishing, Cham, 2017, pp. 65–80. URL: https://doi.org/10.1007/978-3-319-42816-1_4. doi:10.1007/978-3-319-42816-1_4.

[23] E. Di Palma, ELIta: A new Italian language resource for emotion analysis, in: F. Dell'Orletta, A. Lenci, S. Montemagni, R. Sprugnoli (Eds.), Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024), CEUR Workshop Proceedings, Pisa, Italy, 2024, pp. 297–307.

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

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

[26] 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.

[27] S. Baccianella, A. Esuli, F. Sebastiani, Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining., in: Lrec, volume 10, Valletta, 2010, pp. 2200–2204.

[28] E. Pianta, L. Bentivogli, C. Girardi, MultiWordNet:

Developing an aligned multilingual database, in: First International Conference on Global WordNet, 2002, pp. 293–302.
[29] C. Fellbaum, Towards a representation of idioms in WordNet, in: Usage of WordNet in Natural Language Processing Systems, 1998.
[30] C. Fellbaum, WordNet: An Electronic Lexical Database, MIT press, Boston, 1998.
[31] J. Daude, L. Padro, G. Rigau, Mapping WordNets Using Structural Information, 2000. URL: https://arxiv.org/abs/cs/0007035. doi:10.48550/arXiv.cs/0007035.
[32] E. Kafe, How Stable are WordNet Synsets?, in: LDK Workshops, 2017, pp. 113–124.
[33] H. Langone, B. R. Haskell, G. A. Miller, Annotating WordNet, in: Proceedings of the Workshop Frontiers in Corpus Annotation at HLT-NAACL 2004, Association for Computational Linguistics, Boston, Massachusetts, USA, ????, pp. 63–69.
[34] M. Vassallo, G. Gabrieli, V. Basile, C. Bosco, Polarity Imbalance in Lexicon-based Sentiment Analysis, in: F. Dell'Orletta, J. Monti, F. Tamburini (Eds.), Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it 2020 : Bologna, Italy, March 1-3, 2021, Collana Dell'Associazione Italiana Di Linguistica Computazionale, Accademia University Press, Torino, 2020, pp. 457–463. doi:10.4000/books.aaccademia.8964.
[35] V. Basile, Valeriobasile/sentixR, 2019-2024.
[36] M. Vassallo, G. Gabrieli, V. Basile, C. Bosco, The tenuousness of lemmatization in lexicon-based sentiment analysis, in: Proceedings of the Sixth Italian Conference on Computational Linguistics, volume 2481, Ceur, 2019, pp. 1–6.
[37] E. Zanchetta, M. Baroni, Morph-it! A free corpus-based morphological resource for the Italian language, in: Proceedings of Corpus Linguistics Conference Series 2005 (ISSN 1747-9398), volume 1, University of Birmingham, 2005, pp. 1–12.
[38] V. Basile, M. Lai, M. Sanguinetti, Long-term social media data collection at the university of turin, in: Proceedings of the Fifth Italian Conference on Computational Linguistics (CLiC-it 2018), Torino, Italy, December 10-12, 2018., 2018. URL: http://ceur-ws.org/Vol-2253/paper48.pdf.
[39] W. J. B. van Heuven, P. Mandera, E. Keuleers, M. Brysbaert, Subtlex-uk: A new and improved word frequency database for british english, Quarterly Journal of Experimental Psychology 67 (2014) 1176–1190. URL: http://dx.doi.org/10.1080/17470218.2013.850521. doi:10.1080/17470218.2013.850521.
[40] G. K. Zipf, Human Behaviour and the Principle of Least Effort: an Introduction to Human Ecology, Addison-Wesley, 1949.
[41] V. Basile, M. Nissim, C. Bosco, M. Vassallo, G. Gabrieli, A. Vardanega, Sentix, 2025. doi:10.5281/zenodo.15609215.
[42] F. Bond, M. W. Goodman, E. Rudnicka, L. M. da Costa, A. Rademaker, J. P. McCrae, Documenting the Open Multilingual Wordnet, in: G. Rigau, F. Bond, A. Rademaker (Eds.), Proceedings of the 12th Global Wordnet Conference, Global Wordnet Association, University of the Basque Country, Donostia - San Sebastian, Basque Country, 2023, pp. 150–157.
[43] F. Bond, K. Paik, A survey of wordnets and their licenses, in: Proceedings of the 6th Global WordNet Conference (GWC 2012), Matsue, 2012, pp. 64–71.
[44] H. Schmid, M. Baroni, E. Zanchetta, A. Stein, Il sistema 'tree-tagger arricchito'–The enriched Tree-Tagger system, IA Contributi Scientifici 4 (2007) 22–23.
[45] E. J. Miller, An introduction to the resource description framework, Journal of library administration 34 (2001) 245–255.

# Appendix

This appendix reports the SPARQL Queries illustrated in 6.

## Query 1

Retrieve the distribution of emotions:

```
PREFIX lila: <http://lila-erc.eu/
    ontologies/lila/>
PREFIX elita: <http://w3id.org/elita
    />
PREFIX ontolex: <http://www.w3.org/ns
    /lemon/ontolex#>
PREFIX rdfs: <http://www.w3.org
    /2000/01/rdf-schema#>

SELECT ?emotionLabel (COUNT(*) as ?
    count)
WHERE {
    ?lemma a lila:Lemma .
    ?elitaLemma ontolex:canonicalForm
        ?lemma .
    ?elitaLemma elita:HasEmotion ?
        emotion .
    ?emotion rdfs:label ?emotionLabel
}
GROUP BY ?emotionLabel
ORDER BY DESC (?count )
```

### Query 2

Return the distribution of polarity:

```
PREFIX lila: <http://lila-erc.eu/
    ontologies/lila/>
PREFIX marl: <http://www.gsi.upm.es/
    ontologies/marl/ns#>
PREFIX ontolex: <http://www.w3.org/ns
    /lemon/ontolex#>
PREFIX rdfs: <http://www.w3.org
    /2000/01/rdf-schema#>

SELECT ?polarityLabel (COUNT(*) as ?
    count)
WHERE {
    ?lemma a lila:Lemma .
    ?sentixLemma ontolex:
        canonicalForm ?lemma .
    ?sentixLemma marl:hasPolarity ?
        polarity .
    ?polarity rdfs:label ?
        polarityLabel
}
GROUP BY ?polarityLabel
ORDER BY DESC (?count )
```

### Query 3

Return the average Sentix polarity score for each emotion annotated in ELIta:

```
 PREFIX lila: <http://lila-erc.eu/
     ontologies/lila/>
PREFIX elita: <http://w3id.org/elita
    />
PREFIX ontolex: <http://www.w3.org/ns
    /lemon/ontolex#>
PREFIX rdfs: <http://www.w3.org
    /2000/01/rdf-schema#>
PREFIX marl: <http://www.gsi.upm.es/
    ontologies/marl/ns#>
PREFIX xsd: <http://www.w3.org/2001/
    XMLSchema#>

SELECT ?emotionLabel (AVG(?
    polarityValue) AS ?avgPolarity)
WHERE {
  ?lemma a lila:Lemma .
  ?elitaLemma ontolex:canonicalForm ?
      lemma .
  ?elitaLemma elita:HasEmotion ?
      emotion .
  ?sentixLemma ontolex:canonicalForm
      ?lemma .
  ?sentixLemma marl:hasPolarityValue
      ?polarityValue .
  ?emotion rdfs:label ?emotionLabel .
}
GROUP BY ?emotionLabel
ORDER BY ASC(?avgPolarity)
```

### Query 4

Determine the polarity of lemmas annotated with contrasting emotions (Joy and Sadness):

```
PREFIX lila: <http://lila-erc.eu/
    ontologies/lila/>
PREFIX elita: <http://w3id.org/elita
    />
PREFIX ontolex: <http://www.w3.org/ns
    /lemon/ontolex#>
PREFIX rdfs: <http://www.w3.org
    /2000/01/rdf-schema#>
PREFIX marl: <http://www.gsi.upm.es/
    ontologies/marl/ns#>

SELECT ?polarityLabel (COUNT(*) as ?
    count)
WHERE {
    ?lemma a lila:Lemma .
    ?elitaLemma ontolex:canonicalForm
        ?lemma .
    ?elitaLemma elita:HasEmotion
        elita:Gioia .
    ?elitaLemma elita:HasEmotion
        elita:Tristezza .
    ?sentixLemma ontolex:
        canonicalForm ?lemma .
    ?sentixLemma marl:hasPolarity ?
        polarity .
    ?polarity rdfs:label ?
        polarityLabel
}
GROUP BY ?polarityLabel
ORDER BY DESC (?count )
```

Retrieve the polarity value and label of lemmas that are identified in LiITA, annotated in ELIta with both Joy and Sadness, and are associated with a negative polarity according to Sentix 3.0:

```
PREFIX lila: <http://lila-erc.eu/
    ontologies/lila/>
PREFIX elita: <http://w3id.org/elita
    />
PREFIX ontolex: <http://www.w3.org/ns
    /lemon/ontolex#>
PREFIX rdfs: <http://www.w3.org
    /2000/01/rdf-schema#>
PREFIX marl: <http://www.gsi.upm.es/
    ontologies/marl/ns#>

SELECT ?label ?value
WHERE {
    ?lemma a lila:Lemma .
    ?elitaLemma ontolex:canonicalForm
        ?lemma .
    ?elitaLemma elita:HasEmotion
        elita:Gioia .
    ?elitaLemma elita:HasEmotion
        elita:Tristezza .
    ?sentixLemma ontolex:
        canonicalForm ?lemma .
    ?sentixLemma marl:hasPolarity ?
        polarity .
    ?sentixLemma marl:
        hasPolarityValue ?value .
    ?elitaLemma rdfs:label ?label .
    ?polarity rdfs:label "Negative"
        @en .
}
```

## Declaration on Generative AI

During the preparation of this work, the author(s) used ChatGPT (OpenAI), Grammarly, and DeepL Write / DeepL Translate in order to: Text translation, Paraphrase and reword, and Grammar and spelling check. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.