Extending the Semantic Layer of the CompL-it Italian Lexicon: Traits, Semantic Types, and Definitions

Emiliano Giovannetti, Andrea Bellandi, Simone Marchi, Mafalda Papini

Cnr-Istituto di Linguistica Computazionale "A. Zampolli"
Via Moruzzi 1, 56124 Pisa, Italy
{emiliano.giovannetti, andrea.bellandi, simone.marchi, mafalda.papini}@ilc.cnr.it

Abstract

The growing impact of Large Language Models has highlighted the need for explicit, interpretable linguistic knowledge. Lexical resources respond to this need by offering structured representations that complement and constrain the implicit semantics of neural models. This paper presents an extension of CompL-it, currently the most comprehensive open computational lexicon of Italian. Building on the semantic layer inherited from LexicO—itself derived from the PAROLE-SIMPLE-CLIPS resource—the work enriches CompL-it with semantic traits and references to semantic types. Moreover, an experiment was conducted to generate missing definitions through an automatic process supported by LLMs. The resulting resource thus combines human-curated and machine-extended knowledge, ensuring both linguistic precision and scalability. This enriched semantic layer enhances CompL-it's interoperability within the Linguistic Linked Data framework and strengthens its usability for NLP tasks such as word sense disambiguation, semantic role labelling, and knowledge grounding.

Keywords: computational lexicon, Linguistic Linked Open Data, OntoLex-Lemon, Large Language Models, Word sense definitions and semantic enrichment

1. Introduction

In the era of Large Language Models (LLMs), lexical resources are progressively regaining a central role in linguistic and computational research. While for a certain period the availability of "end-to-end" distributional models had relegated lexicons to a secondary role, it is now increasingly evident that structured linguistic knowledge represents an indispensable complement for ensuring interpretability, semantic coherence, and controllability in artificial intelligence systems. Lexical resources, especially those enriched with explicit semantic information, can in fact support a wide range of tasks—from word sense disambiguation to semantic role labelling, from semantic parsing to information extraction, and even knowledge grounding in generative models (Navigli, 2018; Palmer et al., 2005; Bender et al., 2021).

Numerous studies have demonstrated the usefulness of resources such as WordNet, FrameNet, or VerbNet in improving the quality of neural models' semantic representations, both as knowledge bases for fine-tuning and as tools for validation and explainability (Luo et al., 2018; Peters et al., 2019). In these contexts, the availability of rich, structured, and interoperable semantic data is crucial not only for system training but also for integration with linguistic analysis pipelines, terminological tools, and ontological models within the Linguistic Linked Data framework (McCrae et al., 2017).

On this basis, the present work proposes an extension of the CompL-it resource (Sciolette et al., 2024b), currently the most comprehensive and richly annotated open computational lexicon of Italian. In particular, the semantic layer inherited from LexicO (Sciolette et al., 2023)—itself a computational lexicon derived from the PAROLE-SIMPLE-CLIPS (PSC) resource (Bel et al., 2000; Ruimy et al., 2002; AA.VV., 2016)—has been enriched through the formalization of semantic features and references to semantic types (see Section 3.1 for further details). Furthermore, a technique was tested for integrating missing definitions through a semi-automatic process supported by Large Language Models.

The article is structured as follows. Section 2 describes the main freely available lexical and computational resources for Italian, with references to the formal models used to represent linguistic information. Section 3 presents CompL-it, its development, and its distinctive semantic structure. Section 4 discusses the methodology adopted for lexical enrichment, while Section 5 reports the results and their evaluation. An example of resource application is presented in Section 6, followed by concluding remarks in the final section.

2. Related Work

A comprehensive overview of the main lexical and terminological resources for the Italian language is presented in (Sciolette et al., 2024a, Sec. 2), where the position of CompL-it within the landscape of existing computational lexicons is illustrated. This analysis identified numerous freely available resources for Italian, including EuroWordNet (Vossen, 1998), MultiWordNet (Pianta, Bentivogli and Girardi, 2002), ItalWordNet (Roventini et al., 2003), FrameNet

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 7857–7866
11–16 May 2026.

(Baker, Fillmore and Lowe, 1998), and VerbNet (Kipper Schuler, 2005).

Despite the richness and wide diffusion of these resources, only a subset of them explicitly conveys semantic information. WordNet-type resources focus primarily on conceptual relations (synonymy, hypernymy, meronymy), but do not provide a formal representation of the semantic traits or templates that describe the syntactic-semantic structure of lexical units. Frame-based or verb-class-based resources, such as FrameNet and VerbNet, offer more detailed descriptions of the semantic behaviour of predicates, but they do not reach the level of formal and structural granularity that characterizes the CompL-it model.

As for the reference models used in the literature, and not only within the Linguistic Linked Data community, it should be noted that—to our knowledge—no consolidated formal representation currently exists for the so-called "semantic trait" or "semantic type". In other words, current lexical models, such as OntoLex-Lemon (McCrae et al., 2017) and its extensions, provide well-defined structures for entries, senses, and conceptual mappings, but they do not offer a standard mechanism to directly represent traits and semantic types (e.g., descriptive properties, semantic attributes) as coherent entities within the model.

An important reference context is the work on PSC and its conversions into RDF/Linked Data. Del Gratta, Frontini, Khan, and Monachini (2015) describe the conversion of the Italian PSC lexicon into the lemon model, highlighting the challenges of reconciling PSC's "rich" semantic level—which incorporates semantic relations such as qualia roles, synonymy, and meronymy—with the more "lightweight" design of the lemon model (later evolved into OntoLex-Lemon), where LexicalSenses do not directly participate in semantic relations.

As described in Section 4, our approach is to define specific classes and properties for representing and linking semantic traits and types that extend the core of the OntoLex-Lemon model. These extensions are designed to integrate seamlessly with existing structures (e.g. Entry, LexicalSense) and to remain compatible with the semantic relations already adopted in CompL-it, enabling a coherent formalization of traits and semantic types within the computational lexicographic model.

Regarding the issue of missing lexical sense definitions, the advent of LLMs has profoundly reshaped the landscape of automatic definition generation. A first group of studies has explored the intrinsic capacity of these models to produce coherent, dictionary-like glosses, showing a qualitative leap compared to previous definition modeling approaches. Pham, Wong, Kim, Yin, and Skiena (2025) demonstrate that LLMs are capable of generating definitions comparable to those found in major English dictionaries in terms of semantic accuracy and lexical coverage, including rare lemmas. Meconi, Stirpe, Martelli, Lavalle, and Navigli (2025) further this line of research by evaluating the semantic competence of LLMs with respect to the notion of "lexical sense". Their results show that advanced models can not only effectively distinguish word senses but also produce contextually appropriate explanations and definitions, achieving performance comparable to specialized word sense disambiguation systems. In parallel, Almeman, Schockaert, and Espinosa Anke (2024) critically assess the quality of definitions and examples generated by LLMs in comparison with those from WordNet, noting that while semantic coherence is maintained, the models tend to deviate from canonical lexicographic style—highlighting the need for editorial control strategies.

A second line of research, closer to what is being done on CompL-it's missing definitions, focuses instead on constrained definition generation, i.e., definitions produced within structured contexts such as ontologies or terminological databases. Bischof, Filtz, Parreira, and Steyskal (2025) propose a guided generation approach in which LLMs are instructed through prompts and ontological constraints to draft definitions formally consistent with the conceptual structure of existing ontologies, showing promising results in industrial domains. Similarly, Benson, Sculley, Liebers, and Beverley (2025) explore the use of models grounded in Basic Formal Ontology (BFO) to support the automatic drafting of terminological definitions, demonstrating the potential of LLMs to maintain coherence with predefined ontological schemes, while also pointing to the need for mechanisms of stylistic verification and control.

3. CompL-it

CompL-it is a computational lexicon for contemporary Italian, developed as an open resource in accordance with the Linguistic Linked Open Data standards.

3.1 The genesis of CompL-it: focus on PSC

At the core of CompL-it—particularly with regard to its semantic component—lies the aforementioned computational lexicon for the Italian language, PSC, developed between 1996 and 2003 at the Institute of Computational Linguistics "A. Zampolli" of the Italian National Research Council (CNR-ILC), within the framework of national and European projects. PSC represents one of the most structured and information-rich lexical resources available for

Italian. It is organised into four interconnected linguistic layers—semantic, syntactic, morphological, and phonological—each composed of dedicated units linked by explicit relations (Sciolette et al., 2023, Fig. 4).

Word senses are described through multiple types of data, including semantic traits, templates, definitions, and examples, and are interconnected through a dense network of semantic relations (such as synonymy, hyponymy, meronymy, holonymy, etc.).

Semantic traits in PSC represent a fine-grained level of lexical meaning description, designed to formalize semantic information attributed to each semantic unit. Each trait consists of a name–value pair specifying a relevant semantic property of a word sense, contributing to the description of its conceptual behaviour and position within the lexicon's ontological structure. In essence, traits describe internal semantic properties of senses, such as polarity (Positive/Negative), experience (Subjective/Objective), or modal and temporal aspects. Through traits, for instance, it is possible to state that adjectives such as *bello* ("beautiful"), *amabile* ("lovable"), or *carino* ("nice") share the trait *PsychologicalProperty@Positive*, whereas terms like *freddo* ("cold") or *meschino* ("mean") convey traits of opposite polarity.

The semantic layer of PSC is also characterized by so-called templates—predefined semantic schemas that describe recurring types of meaning shared by groups of semantic units. Each template represents a semantic type, such as ANIMAL, INSTRUMENT, EVENT, or PSYCHOLOGICAL_PROPERTY, which captures a coherent set of conceptual properties and relations. These schemas act as semantic prototypes to which word senses can be associated in order to inherit typical behaviours, selectional restrictions, and relations of the represented concept type. Each template corresponds to a node in the SIMPLE Ontology (Bel et al., 2000), built according to the principles of Pustejovsky's Generative Lexicon (1995). Within this framework, templates serve to model the internal conceptual structure of lexical senses, specifying their ontological role (e.g., concrete entity, event, property) and the qualia relations that define them (formal, constitutive, agentive, telic).[1]

Following an in-depth analysis of PSC's four linguistic layers, the idea emerged to create LexicO. The goal was to identify and correct redundant, incorrect, or missing data that limited the usability of the resource. Through automatic procedures and manual checks, duplicates were removed, asymmetric relations (such as synonymy) were restored, and missing information was added. The revised data were then migrated to a new database, LexicO, designed to be interoperable and compliant with the OntoLex-Lemon model, in line with the principles of the Linguistic Linked Open Data (LLOD) framework.

The creation of the CompL-it resource in its current form originated from three main sources: (i) the aforementioned LexicO, (ii) a lemmatized list of inflected forms obtained through a morphological analyser (M-GLF), and (iii) a set of annotated treebanks (Sciolette et al., 2024a, Sec. 3). In order to harmonize the representations coming from these different sources, a standardization process was carried out in accordance with LLOD community requirements, followed by conversion into the OntoLex-Lemon model.

3.2 CompL-it in its Current Form

In its current version, CompL-it contains approximately 100,000 lexical entries, 790,000 inflected forms, 56,000 lexical senses, and 86,000 instances of semantic relations. These figures place CompL-it among the Italian resources with the broadest lexical and semantic coverage. The structure of CompL-it is organised across multiple linguistic levels (morphology, lexicon, semantics), but the semantic layer, due to its richness, plays a central role in defining the relevance of the resource. In particular, CompL-it makes explicit 137 different types of semantic relations (e.g., hypernymy, meronymy, synonymy) inherited from LexicO. Among the semantic data not yet extracted from the original resources, the semantic traits and templates stand out. In this phase of CompL-it's enrichment, however, while all traits will be added, the treatment of templates will be limited to integrating the semantic types to be associated with individual senses.

In addition to semantic relations, traits, and templates, each lexical sense in CompL-it can also be described through a definition and a usage example, which clarify its meaning in an explicit and contextualized way. These elements make it possible to complement structured data with natural language descriptions, useful both for human consultation and for automatic applications. However, not all senses currently available in the resource include a definition or an example. For this reason, one of the activities described in this work concerns the automatic generation of missing definitions through the use

[1] For an overview of the theory and the relationship between qualia roles and relations, see (Sciolette et al., 2023). Following the terminology in the PSC documentation, entities refer to the concept expressed by the sense, conveyed by a specific entry. These entities can be connected to each other through semantic relations. Semantic relations are also classified according to qualia roles.

of LLMs, with the goal of completing and harmonizing the descriptive level of senses within CompL-it's semantic layer.

4. Methodology

This section illustrates the methodology adopted for enriching the semantic layer of CompL-it, describing the phases of extraction and integration of traits and semantic types from LexicO, as well as the experiments conducted for the automatic generation of definitions.

4.1 The addition of traits and semantic types

The activity of extracting and converting the semantic traits from LexicO aimed to transform the semantic information contained in the original relational database into a representation compliant with Linked Open Data standards, using the OntoLex-Lemon model. The procedure consisted of two main phases: (1) data extraction from the LexicO MySQL database and (2) conversion into RDF triples.

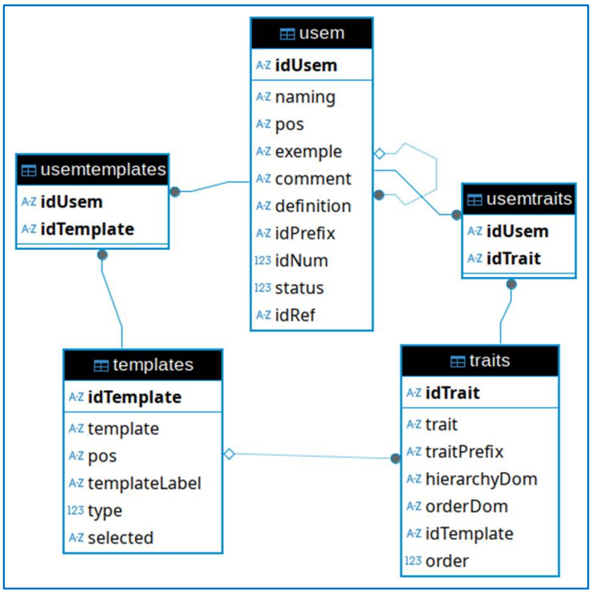

Figure 1: Part of LexicO's database schema involving semantic units, traits and semantic types.

For the first phase, it was necessary to analyze the internal structure of the database (Sciolette et al., 2022), which is composed of several interconnected tables (Figure 1). The semantic traits were stored in the *traits* table, which contains, for each trait, a unique identifier (*idTrait*), the name of the trait (*trait*), and its association with a semantic template (*idTemplate*). These data are linked to semantic units (*Semantic Units* or *SemU*), contained in the *usem* table, through an intermediate table named *usemtraits*, which specifies the relations between senses and traits. Finally, the *templates* table (connected to *usem* through the *usemtemplates* table) describes general semantic types (e.g., *PsychologicalProperty*). The extraction was carried out through an SQL query joining the four tables, resulting in 2,702 relations between semantic units and traits.

In the second phase, the extracted data were exported in CSV format and subsequently converted into RDF triples, through which the following elements were defined:

- the property `complit:trait`, defined as an `owl:ObjectProperty` whose domain is `ontolex:LexicalSense`;
- the classes of traits, defined as `owl:Class` and labeled with textual identifiers (`rdfs:label`);
- the individuals (`owl:NamedIndividual`) representing specific traits of the aforementioned classes, labeled with semantic values (e.g., `positive`);
- the relations between senses and traits, established through the property `complit:trait`.

Overall, the procedure generated 3,790 RDF triples, which were added to the CompL-it resource.

As anticipated, the semantic layer of the original resource (PSC) is also described through so-called *templates*—predefined schemas that convey semantic-conceptual information and group together sets of semantic units associated with them. For the moment, we have focused on the integration of the semantic types, which are at the core of the templates.

Let us take the example of the word *albero* ('tree'). As shown in Figure 2, the word can refer either to a plant or to a part of a ship. The 254 different semantic types available were represented as a hierarchy of OWL classes within an ontology external to CompL-it.[2] Each lexical sense of a word refers to one of these classes via the property `hasSemanticType`, defined ad hoc.

The sense of *albero* as, in English "*the set of branches and leaves capable of growing in height thanks to a woody stem, called trunk, which usually starts branching a few meters above the ground*" refers to the semantic type "Plant", whereas the sense "*the wooden or steel shaft used on ships to support the sail or hang flags*", as shown in Figure 2, refers to the constitutive characteristic of being *part of something* (at the lexico-semantic level, the semantic relation

[2] https://github.com/klab-ilc-cnr/complit-extension

isPartOf links this sense with the ship sense of the word *nave*).

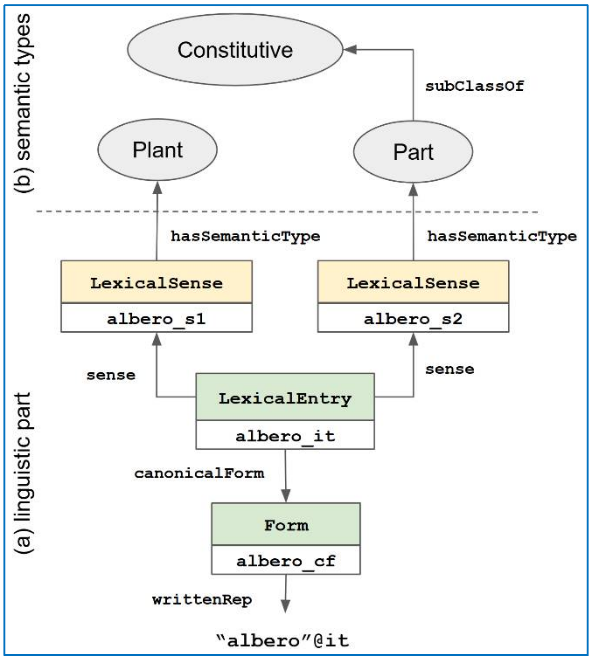

Figure 2: Representation of the polysemous lexical entry albero ('tree'). (a) The OntoLex-Lemon model defines the linguistic component – (b) The referenced semantic types are represented separately as ontological classes.

It is important to note that this kind of extra-linguistic information adds significant value to the resource. It enables not only conceptual searches within the resource itself but also advanced searches across texts annotated with it (see Section 6).

Overall, 54,487 instances of relations between lexical senses and semantic types were formalized.

4.2 Towards the creation of missing definitions

The task of adding missing definitions to the senses in CompL-it required a different approach: in this case, the goal was not to migrate data that had not yet been transferred from LexicO to CompL-it, but rather to add information entirely absent from the original resources. Although as many as 31,734 senses lack an example, in this phase we decided to focus on definitions, which are missing for 9,158 out of the total 55,713 senses.

More specifically, for this experiment we selected senses that lack a definition but include an example, semantic types (added as described above), and semantic relations—a total of 235 senses corresponding to 229 lexical entries. For this first approach, we selected the senses with the highest number of features to test which of these features impact the final output most significantly (for the ablation study, see Section 5). At this stage, we chose not to consider semantic traits, whose inclusion in an LLM prompt—given their formal and rather opaque nature—will require a careful mapping into natural language.

For the definition generation task, three multilingual models (including Italian) of medium-small size were selected, belonging to the *Gemma 3*, *Phi-4*, and *Mistral-Small* families.[3] The choice of using models of this scale is twofold: on the one hand, medium-small models allow for a more sustainable use of computational resources, enabling accessible and reproducible experiments even in research contexts with limited infrastructure; on the other hand, they make it possible to operate in controlled and localized environments, ensuring more secure and privacy-preserving data handling—often not guaranteed when using large-scale cloud-based models.

Once the prompt[4] was prepared, a Python program was developed to query the three models, which were asked to generate an Italian definition for each of the 235 selected senses based on the available data (i.e., example, semantic type, and semantic relations). All prompts, source code, and model outputs are freely available on GitHub.[5]

5. Results and evaluation

The activities described so far have enriched the CompL-it lexicon with 2,702 semantic traits and 54,487 references to semantic types. In addition, a first experiment for the automatic addition of missing definitions has been documented. Table 1 shows the top five semantic traits and semantic types, ordered by cardinality.

The experiment on adding missing definitions required adopting a specific evaluation process to assess the quality of the generated definitional texts.

For the evaluation of the 705 total definitions produced (corresponding to the 235 senses without definitions analysed by the three

3 The exact models used for the experiment were: gemma-3-12b-it-GGUF:Q8_0, phi-4-GGUF:Q6_K, and Mistral-Small-3.2-24B-Instruct-2506-GGUF:Q4_K_M.

4 The prompt was written entirely in Italian to prevent some models—despite being explicitly instructed to generate definitions in Italian—from responding exclusively in English, which would have compromised the outputs.

5 https://github.com/klab-ilc-cnr/definition-generation

generative models), an LLM-as-a-Judge approach was chosen. This technique employs large language models as automatic evaluators of the quality of texts generated by other systems. Instead of relying on lexical similarity metrics, the “judge” model receives as input the responses to be compared, along with an evaluation framework—including criteria, examples, and a scoring scale—and produces either a score or an argued assessment.

Semantic trait	***example***	**#**
BehaviourNegative	“arrogante”	131
StyleNeutral	“orgoglioso”	127
BehaviourPositive	“brillante”	121
NationalityNeutral	“scozzese”	109
BehaviourNeutral	“testardo”	107
Semantic type	***example***	**#**
Profession	“astronomo”	3,510
Human	“stratega”	3,317
Instrument	“pianoforte”	2,440
Agent of temp. activity	“narratore”	2,260
Purpose act	“assaltare”	2,205

Table 1: The most frequent semantic traits and semantic types within CompL-it

This methodology has proven particularly effective in reproducing human judgments, offering greater scalability and consistency compared to manual annotation. Recent studies, such as Huang et al. (2025), which conducted an empirical analysis of judge model behaviour across different benchmarks, and Desmond et al. (2025), who proposed strategies to simplify and standardize the use of LLMs as evaluators, confirm the robustness and reliability of this approach.

Three different medium-sized LLMs, not belonging to the families of the generative models used earlier, were chosen to act as judges in order to avoid bias. Specifically, models from the *llama-3.3*, *llama-4-maverick*, and *Qwen2.5* series were used.[6] Although this required greater computational resources and the use of cloud services, it was considered appropriate to rely on larger models for the evaluation phase to ensure a level of reliability that, while not equivalent to that of three human lexicographers, could be deemed sufficiently good.

The prompt submitted to the judge models did not differ substantially from the one used for generation. The judges were asked to assign a score from 0 to 10 evaluating the quality of each provided definition, based both on their internal linguistic competence and on the accompanying data (examples, semantic types, and relations). A precautionary exclusion criterion was applied: if even one of the judges assigned a score lower than 6 for a given sense, that sense was discarded. However, across the 235 sets of generated definitions, it never occurred that all three models produced failing definitions, and therefore no definition was excluded from the final output.

Among the selected definitions—chosen based on the highest average score—Gemma’s outputs were preferred in 59 cases, Phi’s in 147 cases, and Mistral’s in only 29 cases, showing that the latter model produced lower-quality definitions compared to the other two.

To assess the impact of different data types on definition quality, a feature ablation technique was also applied. Three additional generation runs were conducted, selectively removing from the inputs provided to the generative models (but not from those provided to the judges) the semantic relations, examples, and semantic types. The following tables report the results, presented “by judge” and “by generator”. Table 2 lists, for each row, the average scores assigned by each judge to the 705 total generations produced under the four configurations: all features included (*allFts*), without semantic relations (*noRel*), without examples (*noEx*), and without semantic types (*noTpe*). The last row of the table shows the arithmetic mean of these averages, highlighting the lowest and highest scores. From these data, two main observations emerge: removing examples significantly worsens definition quality, whereas removing semantic types surprisingly improves it.

judge	*allFts*	*noRel*	*noEx*	*noTpe*
llama-3.3	7.83	7.81	7.77	7.98
llama-4	7.79	7.83	7.46	7.89
Qwen2.5	7.88	7.86	7.68	8.08
AVG	7.83	7.83	**7.63**	**7.98**

Table 2: Average scores assigned by each judge to the generations produced by the three generative models in the various runs.

[6] The exact models used as judges were: llama-3.3-70b-versatile, llama-4-maverick-17b-128e-instruct, and Qwen2.5-72B-Instruct.

Table 3, instead, focuses on the performance of the individual generative models. Specifically, for each model, the table reports the average score obtained across all generated definitions and the number of times its definition was selected as the best one, both in the "all features" case and in the three feature ablation scenarios.

Model	*Feats*	*avg*	*#win*
Gemma	allFts	7.99	59
	noRel	7.87	43
	noEx	7.72	163
	noTpe	**8.61**	193
Phi	allFts	8.20	147
	noRel	8.21	151
	noEx	8.04	31
	noTpe	7.40	29
Mistral	allFts	7.30	29
	noRel	7.42	41
	noEx	**7.15**	39
	noTpe	7.93	12

Table 3: Average scores assigned by the three judges to the definitions generated by each model (avg) and the number of times each model's definition was selected as the best one (#win).

The most relevant data are also highlighted in this table: the lowest score was recorded for the Mistral model without examples, while the highest score was obtained by the Gemma model without semantic types.[7] These results confirm the previous findings regarding the positive impact of examples and the negative impact of semantic types.

While the positive contribution of examples was expected, the performance degradation associated with the presence of semantic types prompted further investigation. In fact, whereas some semantic types (such as "instrument", "substance", or "disease") can help refine the understanding of a given sense, several others—such as "Agent (temporary activity)" and "Patient (event)"—when incorporated too literally into the definition, introduced distortions. For example, for the word *abbagliatore* ("dazzler"), associated with the semantic type "Agent (temporary activity)", the Gemma model generated a definition (translated in English: "*A person who captures attention with talent and charisma, often in a temporary and spectacular way, like an artist or performer*") that introduced an element of temporariness which was not well evaluated by the judges.

A final observation concerns the poor performance of the Mistral model. By analysing the "usage metadata" provided in the model outputs, a high value of *input_tokens* was observed—significantly higher than those of Gemma and Phi. This finding aligns with reports from researchers regarding the high token fertility of Mistral models for the Italian language (Moroni et al., 2025), suggesting a comparatively lower lexical competence of the Mistral model used in this study with respect to Italian word meanings.

6. An Example of Application

Beyond the work presented so far—which we believe is already valuable in itself, as it continues to enrich the lexicon by transferring into it all the linguistic data derived from one of its original sources (PSC)—the enhancement of CompL-it can make the resource even more effective for various tasks.

One of the applications in which we intend to use it is full-text search, in processes involving query expansion supported by lexical resources. A prototype application was presented in (Giovannetti et al., 2022), where it was shown that using a rich and structured linguistic resource can provide an advantage over query expansion techniques based on WordNet-like resources. Even in its preliminary conversion phase and with limited coverage, the integration of lexical and morphosyntactic information improved the relevance of the search results.

More specifically, a text search technique supported by a lexical resource is considered to be more effective the more the resource itself contains and interconnects the largest possible number of words present in the text.

For this reason, it is important that a supporting resource also include specialized terminology when conducting searches on domain-specific texts. In such cases, a linguistic resource like CompL-it, however rich, will not provide sufficient lexical coverage. To overcome this limitation, we plan to formally link CompL-it to specialized or terminological lexical resources, leveraging the LLOD paradigm to create an extensive resource in which CompL-it serves as a "lexical superstructure" through its rich network of semantic relations. An example of this type of linking has already been presented in (Litta et al., 2025a), where CompL-it was connected to the LiITA Knowledge Base (Litta et al., 2025b). The linking, which involved over 100,000 lexical entries, improved the coherence and descriptive

[7] Extending this approach to senses that do not have semantic types, we could generate definitions of comparable quality for a total of 1,420 senses, compared to the 235 initially considered.

richness of both resources, while introducing new lemma variants and linking strategies. The experiment demonstrated the validity of the Linked Open Data–based approach and the usefulness of shared ontologies for integrating heterogeneous resources.

However, whereas in the cited work the linking was performed at the level of lexical entries, future developments aim to establish a linking process that connects words through their respective senses. To achieve this, we consider it crucial to link senses to other senses (e.g., to their hypernyms), also exploiting the additional information provided by semantic types, traits, and textual descriptions of meanings.

7. Conclusions and future work

This article has presented work on the expansion of CompL-it, a computational lexicon for the Italian language. Compared with the previously available open version, the resource has been enriched in its semantic layer through the addition, to its lexical senses, of 2,702 semantic traits and 54,487 references to 254 different semantic types. The updated resource is available on GitHub.[8]

A first experiment was also conducted to add missing definitions with the support of LLMs. Using an LLM-as-a-Judge technique, it was possible to compare the performance of three medium-small models, revealing interesting differences: Gemma achieved the best results when semantic types were not provided, suggesting that certain conceptual categories may introduce noise into the generation process; Phi stood out for overall consistency, while Mistral showed less stable performance, possibly due to weaker training in Italian compared to the other two. A sample-based human review of the judge's outputs is planned to assess the reliability of the automatic evaluations and to detect potential biases or edge cases. This validation will also help refine the judging criteria and guide iterative improvements to the overall workflow.

As for future work, the next step will involve integrating the actual templates, building on previous work (Toral and Monachini, 2007), to be appropriately merged with the resource modeled in OntoLex-Lemon. Once the methodology for definition generation is consolidated, the resulting definitions will be added to the resource, and the same technique—after suitable adaptation—will be applied to the generation of missing examples. For both definitions and examples, we also plan to experiment with the inclusion of semantic traits, once the necessary mapping has been developed to express them in natural language within a prompt. In addition, we will explore the expansion of existing definitions and examples when they are deemed too concise (for instance, in CompL-it there are 3,619 senses with one-word definitions).

The potential applications of CompL-it in its extended form are numerous and relevant to various areas of computational linguistics, particularly in contexts where the resource is combined with neural models. Beyond traditional tasks such as word sense disambiguation, semantic role labelling, and knowledge grounding for Italian, the resource can significantly support processes of query expansion and information retrieval, as demonstrated in the application described in Section 6. In this context, CompL-it enables the extension of textual searches based on shared semantic relations and traits, improving both result relevance and lexical coverage. Furthermore, integration with specialized terminological resources will make it possible to perform semantic-based searches within specialized corpora.

8. Bibliographical References

Fatemah Yousef Almeman, Steven Schockaert, and Luis Espinosa Anke. 2024. WordNet under Scrutiny: Dictionary Examples in the Era of Large Language Models. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17683–17695, Torino, Italia. ELRA and ICCL URL: https://aclanthology.org/2024.lrec-main.1538/.

Collin F. Baker, Charles J. Fillmore, and John B. Lowe. 1998. The Berkeley FrameNet Project. In *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Volume 1*, pages 86–90, Montreal, Quebec, Canada. Association for Computational Linguistics URL: https://aclanthology.org/P98-1013/.

Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Alessandro Lenci, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas, and Antonio Zampolli. 2000. SIMPLE: A General Framework for the Development of Multilingual Lexicons. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, Athens, Greece. European Language Resources Association (ELRA) URL: http://www.lrec-conf.org/proceedings/lrec2000/pdf/61.pdf.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell.

[8] https://github.com/klab-ilc-cnr/complit-extension

2021. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? . In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 610–623, Virtual Event Canada. ACM URL: https://dl.acm.org/doi/10.1145/3442188.3445922.

Carter Benson, Alec Sculley, Austin Liebers, and John Beverley. 2025. *My Ontologist: Evaluating BFO-Based AI for Definition Support*. arXiv. arXiv:2407.17657 URL: http://arxiv.org/abs/2407.17657.

Stefan Bischof, Erwin Filtz, Josiane Xavier Parreira, and Simon Steyskal. 2025. LLM-Based Guided Generation of Ontology Term Definitions. In *The Semantic Web: ESWC 2024 Satellite Events*, pages 133–137, Cham. Springer Nature Switzerland DOI: https://doi.org/10.1007/978-3-031-78952-6_13.

Riccardo Del Gratta, Francesca Frontini, Anas Fahad Khan, and Monica Monachini. 2015. Converting the PAROLE SIMPLE CLIPS lexicon into RDF with lemon. *Semantic Web*, 6(4):387–392 DOI: https://doi.org/10.3233/SW-140168.

Michael Desmond, Zahra Ashktorab, Werner Geyer, Elizabeth M. Daly, Martín Santillán Cooper, Qian Pan, Rahul Nair, Nico Wagner, and Tejaswini Pedapati. 2025. EvalAssist: LLM-as-a-Judge Simplified. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(28):29637–29639 DOI: https://doi.org/10.1609/aaai.v39i28.35351.

Emiliano Giovannetti, Davide Albanesi, Andrea Bellandi, Simone Marchi, Mafalda Papini, and Flavia Sciolette. 2022. The Role of a Computational Lexicon for Query Expansion in Full-Text Search. In *Proceedings of the Eighth Italian Conference on Computational Linguistics CliC-it 2021*, pages 162–168. Accademia University Press URL: http://books.openedition.org/aaccademia/10638.

Hui Huang, Xingyuan Bu, Hongli Zhou, Yingqi Qu, Jing Liu, Muyun Yang, Bing Xu, and Tiejun Zhao. 2025. An Empirical Study of LLM-as-a-Judge for LLM Evaluation: Fine-tuned Judge Model is not a General Substitute for GPT-4. In *Findings of the Association for Computational Linguistics: ACL 2025*, pages 5880–5895, Vienna, Austria. Association for Computational Linguistics URL: https://aclanthology.org/2025.findings-acl.306/.

Karin Kipper Schuler, and Martha S. Palmer. 2005. Verbnet*: a broad-coverage, comprehensive verb lexicon*. University of Pennsylvania, USA. ISBN 0-542-20049-X.

Eleonora Litta, Marco Carlo Passarotti, Valerio Basile, Cristina Bosco, Andrea Di Fabio, and Paolo Brasolin. 2025a. LiITA: a Knowledge Base of Interoperable Resources for Italian. In *Proceedings of the 5th Conference on Language, Data and Knowledge*, pages 130–135, Naples, Italy. Unior Press URL: https://aclanthology.org/2025.ldk-1.14/.

Eleonora Litta, Marco Passarotti, Giovanni Moretti, Paolo Brasolin, Francesco Mambrini, Valerio Basile, Andrea Di Fabio, Eliana Di Palma, Emiliano Giovannetti, Simone Marchi, Andrea Bellandi, and Flavia Sciolette. 2025b. Linking CompL-it to the LiITA Knowledge Base. In *Proceedings of the Eleventh Italian Conference on Computational Linguistics (CLiC-it 2025)*, pages 595–602, Cagliari, Italy. CEUR Workshop Proceedings URL: https://aclanthology.org/2025.clicit-1.57/.

Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang, and Zhifang Sui. 2018. Incorporating Glosses into Neural Word Sense Disambiguation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2473–2482, Melbourne, Australia. Association for Computational Linguistics URL: https://aclanthology.org/P18-1230/.

John McCrae, Julia Bosque-Gil, Jorge Gracia, Paul Buitelaar, and Philipp Cimiano. 2017. The OntoLex-Lemon Model: Development and Applications. In *Electronic lexicography in the 21st century: Lexicography from scratch. Proceedings of eLex 2017*, pages 587–597, Brno. Lexical Computing CZ s.r.o. URL: https://elex.link/elex2017/wp-content/uploads/2017/09/paper36.pdf.

Domenico Meconi, Simone Stirpe, Federico Martelli, Leonardo Lavalle, and Roberto Navigli. 2025. Do Large Language Models Understand Word Senses? In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*, pages 33897–33916, Suzhou, China. Association for Computational Linguistics URL: https://aclanthology.org/2025.emnlp-main.1720/.

Luca Moroni, Giovanni Puccetti, Pere-Lluís Huguet Cabot, Andrei Stefan Bejgu, Alessio Miaschi, Edoardo Barba, Felice Dell'Orletta, Andrea Esuli, and Roberto Navigli. 2025. Optimizing LLMs for Italian: Reducing Token Fertility and Enhancing Efficiency Through Vocabulary Adaptation. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 6661–6675, Albuquerque, New Mexico. Association for Computational Linguistics URL: https://aclanthology.org/2025.findings-naacl.371/.

Roberto Navigli. 2018. Natural Language Understanding: Instructions for (Present and Future) Use. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 5697–5702. International Joint Conferences on Artificial Intelligence Organization URL: https://doi.org/10.24963/ijcai.2018/812.

Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106 DOI: https://doi.org/10.1162/0891201053630264.

Matthew E. Peters, Mark Neumann, Robert Logan, Roy Schwartz, Vidur Joshi, Sameer Singh, and Noah A. Smith. 2019. Knowledge Enhanced Contextual Word Representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 43–54, Hong Kong, China. Association for Computational Linguistics URL: https://aclanthology.org/D19-1005/.

Bach Pham, JuiHsuan Wong, Samuel Kim, Yunting Yin, and Steven Skiena. 2025. Word Definitions from Large Language Models. In *2025 19th International Conference on Semantic Computing (ICSC)*, pages 158–162 DOI: https://doi.org/10.1109/ICSC64641.2025.00028.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*.

James Pustejovsky. 1995. *The Generative Lexicon*. The MIT Press. ISBN 978-0-262-28196-6 URL: https://direct.mit.edu/books/book/4726/The-Generative-Lexicon.

A. Roventini, A. Alonge, F. Bertagna, N. Calzolari, A. Zampolli, C. Girardi, B. Magnini, R. Marinelli, M. Speranza, and J. Cancila. 2003. “ItalWordNet”: Building a Large Semantic Database for the Automatic Treatment of Italian. *Linguistica computazionale : XVIII/XIX, 1998/1999*, :745–791 DOI: https://doi.org/10.1400/18178.

Nilda Ruimy, Monica Monachini, Raffaella Distante, Elisabetta Guazzini, Stefano Molino, Marisa Ulivieri, Nicoletta Calzolari, and Antonio Zampolli. 2002. CLIPS, a Multi-level Italian Computational Lexicon: a Glimpse to Data. In *Proceedings of the Third International Conference on Language Resources and Evaluation, LREC 2002, May 29-31, 2002, Las Palmas, Canary Islands, Spain*. European Language Resources Association URL: http://www.lrec-conf.org/proceedings/lrec2002/sumarios/197.htm.

Flavia Sciolette, Andrea Bellandi, Emiliano Giovannetti, and Simone Marchi. 2024a. CompL-it: a Computational Lexicon of Italian. *AIDAinformazioni*, 42(3–4):119–148 DOI: https://doi.org/10.57574/596545646.

Flavia Sciolette, Emiliano Giovannetti, and Simone Marchi. 2023. LexicO: an Italian Computational Lexicon derived from Parole-Simple-Clips. *Umanistica Digitale*, 7(15):169–193 DOI: https://doi.org/10.6092/issn.2532-8816/15176.

Antonio Toral, Monica Monachini, and others. 2007. SIMPLE-OWL: a Generative Lexicon Ontology for NLP and the Semantic Web. In *Workshop of Cooperative Construction of Linguistic Knowledge Bases, 10th Congress of Italian Association for Artificial Intelligence* URL: https://www.researchgate.net/profile/Monica-Monachini/publication/228339605_SIMPLE-OWL_a_generative_lexicon_ontology_for_nlp_and_the_semantic_web/links/0fcfd50a356005c27d000000/SIMPLE-OWL-a-generative-lexicon-ontology-for-nlp-and-the-semantic-web.pdf.

Piek Vossen, editor. 1998. *EuroWordNet: a multilingual database with lexical semantic networks*. Kluwer Academic Publishers, USA. ISBN 0-7923-5295-5.

9. Language Resource References

AA.VV. 2016. *PAROLE-SIMPLE-CLIPS*. ILC4CLARIN: ILC Data & Tools, CLARIN-IT URL: http://hdl.handle.net/20.500.11752/ILC-88.

Flavia Sciolette, Emiliano Giovannetti, and Simone Marchi. 2022. *LexicO*. ILC4CLARIN: ILC Data & Tools, CLARIN-IT URL: http://hdl.handle.net/20.500.11752/ILC-977.

Flavia Sciolette, Emiliano Giovannetti, Simone Marchi, and Andrea Bellandi. 2024b. *CompL-it*. CLARIN-IT URL: http://hdl.handle.net/20.500.11752/ILC-1007