Applications of Knowledge Graphs in the Portuguese Language: A Systematic Literature Review

Maurício Lima
Instituto de Informática
UFG
Goiânia, Brasil
k20x@outlook.com

Vinícius Oliveira
Instituto de Informática
UFG
Goiânia, Brasil
viniciustelesdeoliveira@gmail.com

Sávio Oliveira
Instituto de Informática
UFG
Goiânia, Brasil
savioteles@ufg.br

Elisângela Dias
Instituto de Informática
UFG
Goiânia, Brasil
elisangelasd@ufg.br

***Abstract*—This paper presents a Systematic Literature Review on applications of knowledge graphs in the Portuguese language. PRISMA 2020 and Parsifal were used for planning, execution, and documentation. Searches in databases such as ACM Digital Library, EI Compendex, IEEE, Web of Science, SciELO, ScienceDirect, Scopus, Sol SBC, and Springer Link retrieved 513 records; after deduplication, screening, and full-text eligibility assessment, 12 primary studies were included. For each study, we extracted the domain and objective, types of input data, graph size (when reported), construction and integration techniques, tools/frameworks, update and maintenance strategies, use of inference, evaluation metrics, and quantitative results. The findings reveal recurring benefits of publishing in open standards (RDF/OWL/SPARQL), practical utility in downstream tasks, and the feasibility of construction through extraction/clustering, rules/transformations, and, more recently, LLM-assisted workflows. Four usage archetypes were identified: broad-scope lexical resources, domain-oriented KGs, knowledge modules coupled with pipelines, and reference/evaluation resources.**

***Keywords*—Knowledge graphs; Portuguese language; Natural language processing.**

I. Introduction

Knowledge graphs (KGs) have been employed as semantic infrastructure in core Natural Language Processing (NLP) tasks, Information Retrieval, Question & Answering, Summarization, and Machine Translation, among others. By making entities, relations, and constraints available in queryable structures, KGs enable both symbolic reasoning and integration with statistical and neural methods. In the Portuguese language ecosystem, broad-scope lexical resources, such as Onto.PT/ECO, distributed in RDF/OWL and aimed at sense and relation inventories [1], [2], coexist with domain-oriented applications, such as visualization and debugging of narratives from annotations (Brat→DRS→KG) [3], mining of political-media positions with SPARQL publication [4], ontological induction in historical collections [5], and the use of common-sense knowledge in MT [6]. More recently, text↔graph interaction in multimodal and multilingual scenarios (e.g., Wikidata) has expanded the role of KGs as a source of structural truth for QG/QA and consistency evaluation [7].

Despite these advances, the literature remains fragmented regarding *how* Portuguese KGs are built, maintained, and integrated into NLP pipelines, and *which* benefits and limitations are recurrent in these scenarios. Recurring challenges remain in relation typing/normalization, standardization of metrics, and interoperability between Portuguese variants and other languages. In this context, we conducted a Systematic Literature Review (SLR) to map applications of KGs in Portuguese, synthesizing evidence on benefits, limitations, and integration strategies, as well as opportunities for interoperability and extrinsic evaluation.

This study contributes by presenting a synthesis of Portuguese KG applications, covering domains, objectives, and roles in the pipeline; a comparative mapping of input data, construction techniques, integration, tools, updating/maintenance, inference, and evaluation; and a critical discussion of gaps and research directions, with an emphasis on relation typing, interoperability of Portuguese with other languages, and task-based downstream evaluation.

Although consolidated resources exist in Portuguese (Onto.PT/ECO) [1], [2] and various sectoral applications [3]–[6], [8], [9], there is a lack of studies that integrate evidence on the benefits and limitations of using KGs in Portuguese; that make explicit the construction and integration patterns (RDF/OWL/SPARQL, property graphs, LLM-supervised pipelines); that compare updating/maintenance practices and metric reporting; and that identify barriers and opportunities for multilingual interoperability (e.g., via Wikidata QIDs, interlanguage links, or alignments to BabelNet/FrameNet) [4], [7], [10].

Furthermore, the rapid adoption of LLMs and GraphRAG approaches places KGs at the center of applications that require semantic governance, traceability, and resistance to hallucinations. Without a critical consolidation of Portuguese-language evidence, research and development teams tend to replicate known limitations (untyped edges, solely structural evaluation) and underexploit potential gains from standardized typing and integration. For these reasons, an SLR conducted with an explicit protocol (PRISMA 2020) and support tool (Parsifal) [11], [12] is justified, as it can produce a reproducible comparative framework that is useful to the community.

II. Objectives

The general objective of this study is to systematize the available evidence on applications of knowledge graphs in the Portuguese language, characterizing benefits, limitations, and interoperability aspects, as well as the ways in which KGs are employed and integrated into NLP pipelines.

The specific objectives are to: investigate which benefits, limitations, and multilingual interoperability aspects are reported in the use of KGs in Portuguese; analyze how KGs in Portuguese are employed and integrated into NLP pipelines (data, construction, updating, and evaluation) in the current literature (up to August 2025); collect, for each included study, the domain and objective, types of input data, graph size (when available), construction and integration techniques, tools and frameworks, update/maintenance strategies, use of inference, evaluation metrics/formats, and reported quantitative results; identify recurring gaps (relation typing, standardized reporting, extrinsic evaluation) and propose research directions and best practices aimed at reproducibility, interoperability, and integration in downstream tasks.

III. Methodology

The PRISMA 2020 methodology [11] was adopted, along with *Parsifal* for the planning, execution, management, and documentation of the research [12]. This is a Systematic Literature Review (SLR), characterized by a replicable methodology involving a comprehensive search to locate published and unpublished works on a given topic; a systematic integration of the results of that search; and a critical analysis of the scope, nature, and quality of the evidence in relation to a specific research question, synthesizing studies to extract broad theoretical conclusions about the meaning of a body of literature, connecting theory and evidence [13].

This SLR was guided by the following main research questions: **(RQ1)** What benefits, limitations, and multilingual interoperability aspects are reported in the use of knowledge graphs in Portuguese? **(RQ2)** How are knowledge graphs in Portuguese employed and integrated into NLP pipelines (data, construction, updating, and evaluation) in the identified applications?

Search string: The search string used Portuguese and English keywords directly related to knowledge graphs and the Portuguese language, as follows:

(“grafo de conhecimento” OR “grafos de conhecimento” OR “grafo semântico” OR “grafos semânticos” OR “knowledge graph” OR “semantic graph” OR “semantic network”) AND (“português” OR “língua portuguesa” OR “PT-BR” OR “PT-PT” OR “portuguese”)

Sources: *ACM Digital Library, EI Compendex, IEEE Digital Library, ISI Web of Science, SciELO, ScienceDirect, Scopus, Sol SBC, and Springer Link.*

Only primary studies that were published and peer-reviewed (conference papers and journal articles), with full text available, and that described, implemented, or evaluated applications using knowledge graphs in Portuguese were included in the review. This encompassed lexical resources, ontologies, domain-specific KGs, pipelines that consume or produce KGs, and multimodal text↔KG arrangements with an explicit Portuguese component, with no restriction on publication date or language, provided that the Portuguese dimension was central to the study.

All grey literature (white papers, blogs, slides, preprints), inaccessible studies, duplicates, out-of-scope works (those only in other languages without a Portuguese component; purely conceptual discussions without implementation or application; graphs without the nature of a knowledge graph), and articles whose methodology did not allow extraction of the data required for the research questions (domain, objective, graph size, technique, integration, updating, inference, and metrics) were excluded.

In the end, 12 studies met these criteria and were included in the synthesis, following the PRISMA 2020 flow, as shown in Figure 1.

A. Execution

The identification process retrieved 513 records, distributed across the following databases: ACM Digital Library (1), EI Compendex (30), IEEE Digital Library (5), ISI Web of Science (116), SciELO (1), ScienceDirect (3), Scopus (42), Sol SBC (1), and Springer Link (314). After deduplication (49 records), title and abstract screening, as well as full-text eligibility assessment, were conducted independently by two researchers; any disagreements were resolved by a third researcher. A total of 18 studies were excluded for being classified as grey literature, 351 for presenting applications not based on the Portuguese language, 32 for being secondary studies (systematic reviews and mappings), and 51 for not addressing applications of knowledge graphs in Portuguese.

In the end, 12 studies fully met the inclusion criteria and constitute the synthesis of this review. The complete process flow, following PRISMA 2020, is presented in Figure 1.

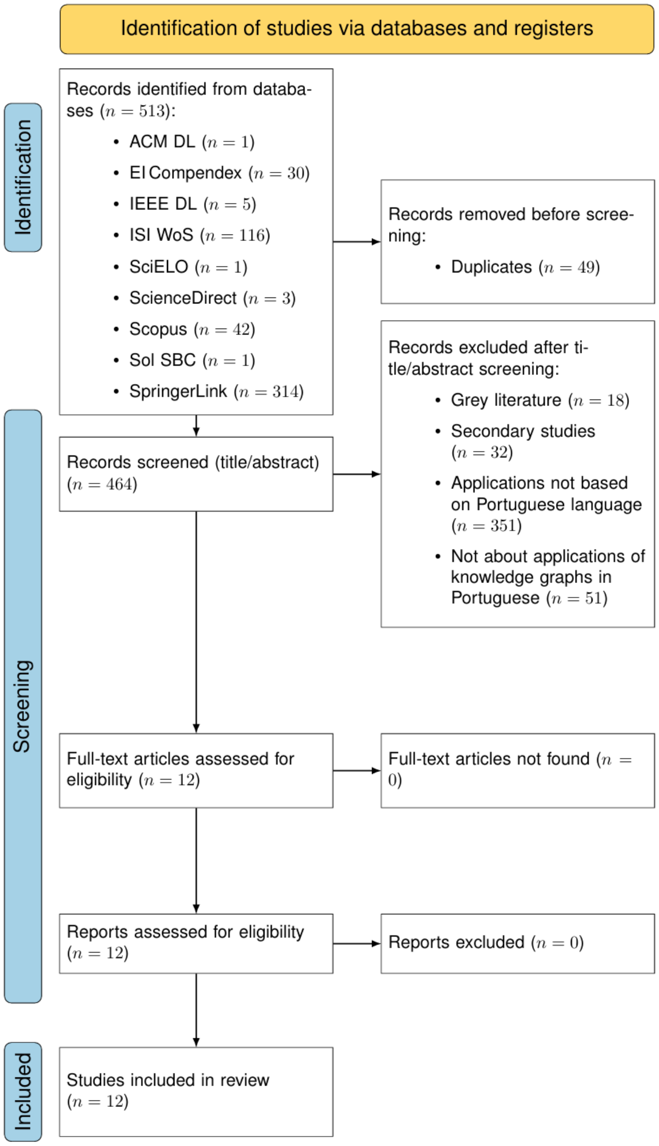

Figure 1. Study selection process following PRISMA 2020 [11].

The selected studies were classified according to their distribution by year, application domain, objectives, type of input data, graph size, construction and integration techniques, tools and frameworks used, updating and maintenance techniques, evaluation formats, and corresponding results.

IV. Results

Figure 2 presents the distribution of studies by year of publication. The analysis suggests a sparse start in the late 2000s and early 2010s, followed by an initial peak in 2014–2015, a slowdown between 2016–2020, and a recent resurgence from 2021 onward, with renewed concentration in 2021 and 2023. The record in 2025 indicates the continuity of this research agenda. It is worth noting that these values represent counts from the analyzed set and reflect the search and selection strategy; they do not, in themselves, imply quality, impact, or thematic saturation.

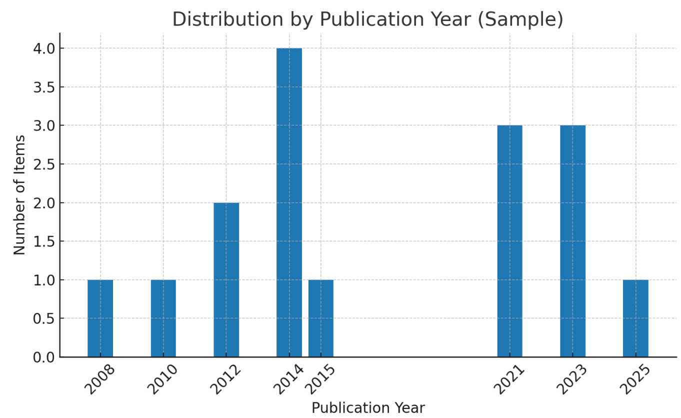

Figure 2. Distribution by year of the articles included in the systematic review

Table I synthesizes, for each of the 12 included studies, the application domain and the main objective of the respective knowledge graph–based computational application, as identified by the bibliographic reference.

Table II organizes, for the 12 included studies, *how* the knowledge graphs are constructed (techniques, heuristics, and algorithms) and *how* they are integrated into the target systems (publication/querying, visualization, graph databases, or use in pipelines). The range covers approaches from automatic extraction via hyperlinks and dictionaries with *clustering*/ontologization, to conversions from annotations (Brat→DRS→KG), LLM-based learning, NER+*entity linking*, association rules, and BFS-based propagation with rules.

Table III summarizes, for each included study, the type of input data used in the construction/consumption of the knowledge graph and, when available, the graph size (nodes/edges/synsets).

Table IV consolidates, for each included study, the technological ecosystem adopted (tools and frameworks) and the strategy for updating/maintaining the graph or the associated pipeline. Three recurring technological blocks are observed: the Semantic Web stack (RDF/OWL/SPARQL) for interoperable publication and querying; property graph databases (Neo4j) and NLP libraries (spaCy, NLTK, YAKE, Polyglot) for extraction and practical exploitation; and neural models/LLMs and modeling artifacts (mT5-small, Llama 3.2 3B, KELM) when the focus is on generation/training. In parallel, maintenance strategies are distributed among batch regeneration, incremental/modular evolution, data-driven re-execution/retraining, and alignment/enrichment with external datasets.

Table V consolidates, for each included study, three complementary dimensions: the use of inference (formal or procedural) associated with the knowledge graph, the evaluation strategy adopted (metrics, protocols, and focus), and the main quantitative results when reported.

Ref	Domain	Objective
[8]	Encyclopedic knowledge (Tourism and Philosophy)	Construction of domain KG for navigation/expansion
[1]	General (NLP in PT)	Sense inventory; support for IR, WSD, QA
[3]	News (PT-PT)	Visualization/debugging of narratives
[2]	General (lexical-semantic in PT)	Automatic construction of wordnet/KG
hline [14]	Legal/Legislative	Query/visualization; IR; summarization
[5]	History/Heritage	Domain modeling; queries/calculations
[4]	Politics/media	Position analysis
[9]	Education (Mathematics)	Search/retrieval; diagnosis
[7]	QA and QG (multi-modal)	QA/QG training/evaluation
[10]	Linguistics/Lexical resources	Preposition disambiguation
[15]	Sentiment/Opinion	Polarity lexicon induction
[6]	Machine translation	SMT post-editing/correction

Table I
CLASSIFICATION OF SELECTED STUDIES BY DOMAIN AND MAIN OBJECTIVE

There is a balanced distribution between broad-use lexical resources ([1], [2], [10], [15]) and task-oriented applications in specific domains, such as legal/legislative [14], politics/media [4], education [9], news/event records [3], machine translation [6], multimodal QA/QG [7], and extraction via Wikipedia and ontology in historical collections [5], [8]. The objectives are similarly concentrated in two main groups: the construction/curation of resources (sense inventories, wordnets, ontologies) and applied uses for search, analysis, summarization, debugging, lexicon induction, MT post-editing, and text↔graph consistency evaluation, underscoring the breadth of scenarios in which Portuguese KGs have been adding value.

Two main construction families can be observed: automatic/statistical approaches, such as extraction and clustering with ontologization in lexical resources [1], [2], NER+linking [4], and LLM-based NER/RE [14]; and rule/transformation-based approaches, such as Brat→DRS→KG [3], co-occurrence [9], and BFS with polarity rules [15].

Ref	Construction technique	System integration
[8]	Automatic extraction of concepts and links	Not reported
[1]	Extraction + *clustering*	RDF/OWL; query service
[3]	Brat→DRS conversion; KG	Web visualization
[2]	Extraction + *clustering*	RDF/OWL; query interface
[14]	LLM-based extraction	Graph database
[5]	YAKE; POS; association rules	Not reported
[4]	NER + entity linking	RDF/SPARQL
[9]	Intra-title co-occurrence	Not reported
[7]	"Silver" data + filters	Not applicable
[10]	Organization into synsets; layered structure	Lexical resource
[15]	BFS propagation + rules	Lexicon generation
[6]	KG integration in pipeline	Post-processing

Table II
CLASSIFICATION OF SELECTED STUDIES ACCORDING TO CONSTRUCTION AND INTEGRATION TECHNIQUES USED

Ref	Input Type	Graph Size
[8]	Portuguese Wikipedia	269–449 nodes; 573–6,322 edges
[1]	Lexicographic resources	≈109k synsets; ≈173k relations
[3]	Annotated text	Not reported
[2]	Dictionaries/thesauri	≈109k synsets; ≈173k relations
[14]	Legislative texts	Not reported
[5]	Handwritten text	Not reported
[4]	News headlines + Wikidata	Not reported
[9]	Article titles	Not reported
[7]	Wikidata + texts	Not applicable
[10]	Phrasal examples	13 PT synsets; 14 EN synsets
[15]	Lexical-semantic triples	PAPEL 2.0: ≈97k items
[6]	Parallel sentences	Not reported

Table III
CLASSIFICATION OF SELECTED STUDIES BY INPUT TYPE AND GRAPH SIZE

In terms of integration, Semantic Web patterns (RDF/OWL/SPARQL) stand out for interoperability [1], [2], [4], along with the use of property graphs for interactive exploration (graph database in [14]) and application-specific consumption artifacts (web visualization in [3], MT post-processing in [6], lexicon generation in [15]). Some studies do not detail the integration layer ("not reported") or do not apply it, as in the case of modeling protocol/dataset work [7], indicating an opportunity to standardize the reporting of this stage to enhance reuse and reproducibility.

Ref	Tools/Frameworks	Updating and Maintenance
[8]	Not reported	Batch regeneration
[1]	RDF/OWL (OntoBusca)	Regenerable; OWN.PT integration
[3]	Brat; NLTK/DRT; vis.js	Modular
[2]	RDF/OWL	Regenerable; gloss corrections
[14]	Llama 3.2 3B; Neo4j	Merging/post-processing
[5]	NLTK; YAKE; Polyglot	(Semi)automatic with supervision
[4]	spaCy; Elasticsearch; Jena	Adjustments/enrichment
[9]	Not reported	Batch reconstruction
[7]	mT5-small; KELM	Regeneration/retraining
[10]	FrameNet integration	Incremental
[15]	PAPEL 2.0; SentiLex-PT01	Re-execution
[6]	Moses; CSKT	Alignment/enrichment

Table IV
CLASSIFICATION OF SELECTED STUDIES BY TOOLS/FRAMEWORKS AND UPDATING/MAINTENANCE STRATEGIES USED

The largest reported structures come from the Onto.PT/ECO approaches [1], [2], with approximately 109k synsets and 173k relations, while the method based on Portuguese Wikipedia [8] generates smaller networks (269–449 nodes; 573–6,322 edges) targeted to specific domains. PAPEL 2.0 [15] reports approximately 97k items, and PrepNet.Br [10] remains at proof-of-concept stage (13 PT synsets aligned to 14 EN). In several cases, size is not reported (e.g., Brat2Viz [3], legislative [14], historical [5], political [4], co-occurrence in journals [9], ConceptNet/MT [6]), either due to a functional/visual focus, reliance on external databases (Wikidata), or inapplicability to the experimental design (multimodal QG/QA [7]). This heterogeneity highlights the usefulness of standardizing the reporting of structural metrics (nodes, edges, classes) to facilitate future comparisons.

In terms of tools, W3C standards stand out in widely used lexical resources [1], [2] and in SPARQL/Jena-based publication for the political graph [4], while legislative exploration combines an LLM (Llama 3.2 3B) with Neo4j [14]. The Brat→DRS→KG pipeline employs Brat, NLTK/DRT, and *vis.js* for inspection [3], whereas historical ontological induction uses NLTK, YAKE, and Polyglot [5]. For multimodal QG/QA, mT5-small and KELM structure the "silver" data cycle [7]; polarity propagation relies on PAPEL 2.0 and SentiLex-PT01 [15]; and MT post-editing integrates Moses and CSKT [6].

Regarding maintenance policies, batch regeneration is prevalent in graphs extracted through rules or co-occurrence [8], [9] and in PT wordnets [1], [2]; incremental/modular approaches are used for expanding resources [3], [10]; re-execution/retraining is adopted when quality depends on the generated dataset [7], [15]; and entity/relation alignment and enrichment with external databases is employed in [4], [6], alongside merging and post-processing for disambiguation in LLM-based scenarios [14].

Ref	Inference	Evaluation	Results
[8]	No	Sample validation	Philosophy: 64/67; Tourism: 91/153
[1]	Not reported	Accuracy/coverage	Hypernymy 78–79%; others 88–92%
[3]	Potential via DRS	Qualitative	Not applicable
[2]	Not reported	Sampling/coverage	78–92%
[14]	No	Coherence/EDA	Not reported
[5]	Calculations	Rule-based metrics	Lift > 2
[4]	No	P/R/F1	NER $F1 \approx 0.94$; type $F1 \approx 0.72$
[9]	No	C, L, degree distribution	$C \approx 0.76$–0.84; $L \approx 2.5$–2.7
[7]	No	BERTScore; F1/EM	+22.9 pts BERTScore (PT-BR)
[10]	Conceptual	Conceptual coherence	Not reported
[15]	Graph-based rules	Accuracy	+1.5 p.p. (3 seeds)
[6]	No	Perceived quality	Not reported

Table V
CLASSIFICATION OF SELECTED STUDIES ACCORDING TO INFERENCE TECHNIQUES, EVALUATION METHODS, AND EVALUATION RESULTS

Explicit inference is rare: there is the use of graph-based rules for polarity induction in [15] and a conceptual layer in PrepNet.Br [10], while most studies do not employ RDFS/OWL or formal rules. In terms of evaluation, lexical resources such as Onto.PT/ECO [1], [2] emphasize sample accuracy and coverage; the Wikipedia-based method [8] uses validation of sampled links; applied pipelines report component-level metrics, such as P/R/F1 for NER/RE and relation direction [4], or complex network metrics (C, L, degree distribution) in co-occurrence analysis [9]; multimodal scenarios adopt BERTScore, EM/F1, and reference-free evaluation [7]; and legislative and MT cases prioritize coherence/perceived quality [6], [14].

Among the results, notable findings include accuracy of 78–92% for non-hypernymy relations [2], 78–79% for hypernymy [1], 64/67 and 91/153 correct links in specific domains [8], gains of +22.9 BERTScore points in PT-BR [7], and +1.5 p.p. accuracy with lexicographic records [15]; in addition, selection by lift > 2 underpins the robustness of relations in historical texts [5].

A. Answering RQ1

1) Benefits: In [8], the authors show that semantic networks can be extracted directly from Portuguese Wikipedia using categories and internal links, avoiding dependencies like WordNet. This benefits low-resource languages and enables machine-readable semantic data in Portuguese. The method also supports domain-specific graphs (e.g., Tourism, Philosophy) and shows that coverage-oriented strategies (using "weak links") produce broader networks.

[1] describes the automatic creation of Onto.PT, freely available in RDF/OWL (via OntoBusca), facilitating reuse, standardization, and pipeline integration. It overcomes prior Portuguese resources' limits in size, coverage, creation mode, and availability, reaching scale and richness comparable to Princeton WordNet. Applications include synonym expansion and "cloze" QA.

In [3], Brat2Viz converts Brat annotations into an intermediate DRS, then into Knowledge Graphs and Message Sequence Charts. The DRS supports visualization, rewriting, evaluation, and inference without returning to the source text. The modular design allows new schemas and visualizations, merges redundant references, and is demonstrated on European Portuguese news.

The ECO approach in [2] automatically builds a Portuguese wordnet/lexical ontology by extracting relations, clustering "synsets", and ontologizing them. The resulting Onto.PT, in RDF/OWL, integrates multiple Portuguese sources (PAPEL, TeP/OpenThesaurus.PT, Wiktionary.PT), offers broad coverage, and has been used for query expansion, WSD, and QA. It is easily updated and adaptable to other domains and languages.

[14] presents a complete, reproducible pipeline for Brazilian Portuguese KGs, from text cleaning to LLM-based extraction, post-processing, and Neo4j storage. Results show coherent graphs for legislative texts, with benefits for legal transparency and access. The work releases curated datasets, code, exploratory analyses, and readability metrics, suggesting applications in proposition similarity, IR, summarization, and GraphRAG.

[5] implements a (semi)automatic ontology learning process for historical Portuguese texts, combining NLP and ML to extract concepts, relations, and instances. The ontology supports queries and domain calculations (e.g., area, cereal production, workforce). The pipeline uses YAKE, NLTK, Polyglot, and Apriori rules, enabling domain modeling with specific classes and attributes.

In [4], a Portuguese KG enriched with Wikidata captures support/opposition relations from political news headlines, linking them to QIDs. Relations are published as RDF triples with SPARQL access. The NLP pipeline (NER, entity linking, relation type/direction classification) is reusable in other contexts, with graph, endpoint, and annotated set provided.

[9] uses Portuguese titles to build semantic networks that improve knowledge diffusion and search in journals via co-occurrence graphs. Such networks are presented as effective knowledge representation systems, with *small-world* properties useful for content curation and navigation.

[7] extends Brazilian Portuguese to multimodal QA tasks over both Wikidata and text. The method increases question coverage, improves intra/intermodal consistency, and enhances answerability. Code, data, and models are released to support reproducibility and reuse.

In [10], PrepNet.Br is proposed: a semantic network of Brazilian Portuguese prepositions organized in *synsets* with formal syntactic-semantic specifications. It models roles like location, instrumentality, and source-destination, aiming to reduce parsing/generation errors. The conceptual model includes image schemas, spatial axes, and FrameNet frames, with a proof of concept aligning bilingual *synsets*.

[15] treats the PAPEL lexical-semantic network as a graph to propagate polarity from seed terms, generating a Portuguese polarity lexicon. Incorporating metadata like "ironic" or "pejorative" improves labeling accuracy, highlighting their usefulness.

Finally, [6] uses Portuguese and English ConceptNet to filter and refine machine translation outputs, handling slang and culture-specific meanings. The approach integrates graph knowledge before, during, and after translation, demonstrated in a bilingual chat and adaptable to other domains.

2) Limitations: In [8], relations extracted from Portuguese Wikipedia are untyped (*related_to*), limiting semantic inference and interoperable reuse. Edge quality shows notable noise, with many incorrect links in the Tourism sample. The pilot study covers only two domains and a 2011 dump, calling for broader evaluations and future ontological typing.

In [1], there is a trade-off between scale and reliability: full automation yields lower accuracy than manual resources. Evaluation shows incomplete precision (78–79% for hypernymy, $\kappa \approx 0.47$; 88–92% for other relations, $\kappa \approx 0.48$), dropping further with incorrect synsets. The resource requires ongoing updates, adding glosses, integrating new sources (OWN.PT, Wikipedia), assigning confidence levels, and exploring new formats.

In [3], the main drawback is reliance on prior manual narrative annotation, costly and complex. The tool focuses on visualization, not full-scale KG building. Converting event→actor into $actor \leftrightarrow actor$ edges may flatten event structures, and portability needs label adjustments per annotation scheme. Current visualizations are limited to MSC and KG.

[2] shares the reliability–scale trade-off and identifies ontologization (synset assignment) as the weakest step, with 60–64% accuracy and moderate inter-annotator agreement. Hypernymy and part-of links are particularly problematic, and the taxonomy may contain cycles. Missing glosses hinder disambiguation and reuse.

In [14], issues include entity disambiguation, incomplete relation coverage from LLMs, and simplistic deduplication (Levenshtein distance ¡5), insufficient for complex semantic duplicates. Threats to validity arise from a single legislative source, possible LLM bias/hallucination, and readability metrics' limits for legal nuance. No extensive human annotation or fine-tuning is included; KG quality depends heavily on prompting and post-editing.

In [5], the process remains (semi)automatic, requiring human input. Seventeenth-century Portuguese challenges stopword removal, synonymy, and POS-tagging, demanding historical lexicon checks. Concept coverage is limited, axiomatization minimal, and relation inference from association rules risks missing rare but valid links. The corpus, one manuscript, limits generalization.

In [4], relying solely on headlines is restrictive; idioms and insufficient information hinder relation extraction. Disambiguation is imperfect, as some entities lack Wikidata QIDs. Corpus filtering may introduce bias; annotation relied on one annotator, and relation direction rules favored one class, limiting generalization.

In [9], evidence for semantic networks as knowledge representation remains preliminary. Title/co-occurrence-based construction leaves relations untyped and misses full-text semantics. Excluding one Portuguese journal (RPM) for comparability limits generalization, and initial networks have disconnected components despite a large giant component.

In [7], the pipeline depends on silver data and MT, introducing generator bias, limited verbalization coverage, and translation/alignment noise. No large gold-standard $text \leftrightarrow$ KG dataset exists in PT-BR; validation relies on heuristics and internal checks. Complex text-derived questions are excluded from KG QA training, and public KGs like Wikidata retain Anglocentric labels/values.

In [10], PrepNet.Br is a prototype covering only spatial prepositions, without multiword expressions or metaphorical senses. Validation is mainly qualitative, lacking broader quantitative evaluation. Cross-lingual alignment is non-trivial due to usage differences (e.g., between/among vs. entre).

In [15], register metadata improves labeling but is sparse for critical cases; markers like ironic and pejorative are rare, limiting overall gain. Evaluation is confined to adjectives via SentiLex-PT01, covering only part of the vocabulary. Performance depends on seed number/quality and underlying resource; only synonymy is explored.

Finally, [6] is ongoing work with untested prototypes, including ConceptNet alignment and enrichment. The approach assumes sufficient KG coverage and requires human post-editing, limiting scalability. The SMT was trained in a narrow domain (Pesquisa FAPESP articles) and gains are not measured with BLEU, hindering standard translation quality comparisons.

3) Interoperability: In [8], interoperability is mainly methodological portability: the extraction process is language-independent and applicable to any Wikipedia edition, leveraging each version's community-specific knowledge. However, no cross-lingual alignments (e.g., via interlanguage links or DBpedia/BabelNet mappings) are implemented, making it potential, not realized, interlingual integration.

For Onto.PT, [1] emphasizes built-in compatibility: following the WordNet model and publishing in RDF/OWL aligns it with the Semantic Web, enabling links to other resources. Comparisons with other wordnets and later integrations (e.g., OWN.PT) strengthen resource interoperability, but no ready-made cross-lingual mapping (e.g., to BabelNet) is provided, only suggestions via Wikipedia/Wiktionary.

[3] is language-agnostic by design: using DRS as an intermediate layer between any text and visualization supports pipeline portability. Adherence to annotation standards (ISO 24617-1/9, LIRICS) aids tool and resource interoperability. Still, the study remains monolingual (PT-PT) and lacks cross-lingual KG mappings or links to multilingual KBs.

In [2], interoperability is also strong technically: the WordNet model and RDF/OWL formats ensure compatibility with other ontologies and KGs. The authors discourage direct transfers from English, stressing language-specific wordnet construction. Like [1], multilingual potential is noted (Wikipedia/Wiktionary, SUMO/DOLCE) but without delivered cross-lingual alignments.

The legislative PT-BR study [14] treats interoperability mainly as process portability: the (sanitization + LLM + KG) workflow is adaptable to other Romance languages. Neo4j/Cypher persistence supports queries, but there are no RDF/OWL alignments or Wikidata integration; cross-lingual interoperability is limited to potential replication. The LLM's multilingual ability is noted as capacity, not validated KG alignment.

In [5], interoperability is likewise methodological: multilingual tools like Polyglot could enable adaptation, but the experiment is monolingual and lacks $PT \leftrightarrow EN/ES$ mappings or links to multilingual KBs (DBpedia/Wikidata). PT–EN bilingual tables aid readability, not semantic alignment.

The political KG linked to Wikidata [4] shows strong technical interoperability: RDF triples, SPARQL, and Wikidata QIDs allow integration with other KBs and tools, with QIDs enabling translingual enrichment. However, there is no dedicated cross-lingual reconciliation module; multilingualism derives from Wikidata's ecosystem, not an implemented KG alignment.

In [9], interoperability appears as method/metric portability: Portuguese and English networks share similar topological patterns, indicating applicability across languages. Still, no interlingual vocabulary alignment or graph merging is done, the analysis is comparative, not integrative.

[7] is explicitly cross-lingual: PT-BR QG/QA over often English-labeled Wikidata relies on multilingual labels and MT for names. Results show gains in cross-modal consistency and correlation with human judgments, but gaps in multilingual labels and name variation still require heuristics and filters.

In [10], PrepNet.Br is designed for $PT \leftrightarrow EN$ synset alignment, with frame links supporting structural connections to FrameNet. However, it is preliminary: the proof of concept aligns only $13 \leftrightarrow 14$ synsets, and ambiguous cases need specific alignment policies.

[15] does not implement interlingual alignment; interoperability is intralingual, using a Variant field to handle Portuguese varieties. The method is transferable to other languages in principle but not demonstrated.

Finally, [6] considers parallel PT-BR and EN ConceptNets, identifying direct concept/edge matches (e.g., book–learn livro–aprender) to enable cross-graph knowledge transfer. The authors propose aligning PT–EN concepts/relations (via GIZA++ and graph structure) and enriching one ConceptNet with the other's data and parallel texts, outlining an explicit cross-lingual path. No links to external standards like RDF/OWL are reported; the focus is on interoperability between ConceptNets for MT.

B. Answering RQ2

The twelve studies analyzed cover a wide range of NLP scenarios in Portuguese, from automatic network extraction from Wikipedia and dictionaries to LLM-supervised pipelines, annotation-oriented logical visualization, ontological induction in historical texts, mining of political-media relations, co-occurrence-based representation in journals, multimodal QG/QA aligned with Wikidata, lexical modeling of prepositions, and polarity propagation in a graph-centered lexicon. Despite this diversity, they can be systematically compared along the dimensions (data, construction, update, and evaluation) as well as by how each KG integrates into the pipeline.

1) Data: from web corpora and lexicons to highly specialized domains: Four main data source lines stand out. The first extracts directly from encyclopedic or lexical repositories: Portuguese Wikipedia in specific domains in [8], and a set of Portuguese dictionaries/thesauri (PAPEL, Dicionário Aberto, Wiktionary.PT, TeP, OpenThesaurus.PT) in [1], [2]. The second uses annotated or curated data to enable structural analysis: annotated PT-PT news for DRS generation and visualization in [3]. The third focuses on niche corpora: legislative texts from ALRN in [14] and a 17th-century manuscript in [5]. The fourth works with short texts and external bases for semantic anchoring: Portuguese political news headlines anchored in Wikidata in [4], titles of mathematics education journals in [9], graph–text pairs from WEBNLG/Wikidata translated into PT-BR in [7], phrasal examples and corpus evidence for prepositions in [10], lexical triples from PAPEL 2.0 in [15], and two parallel ConceptNets (PT-BR and EN) in [6]. In terms of linguistic coverage, PT-PT, PT-BR, or "neutral" Portuguese predominate depending on the source; studies with Wikidata and ConceptNet offer natural multilingual anchoring points [4], [6], [7].

2) Construction: from link/co-occurrence graphs to ontologization and text $\leftrightarrow$ graph alignment: Construction strategies vary substantially. In [8], the network is induced from the Wikipedia link mesh by acquiring concepts and relations, with untyped edges (*related_to*). In [1], [2], relation extraction from lexical definitions is followed by *synset* induction (clustering) and "ontologization" mapping relations to *synset* pairs, resulting in a scalable wordnet/lexical ontology (Onto.PT). The pipeline in [3] transforms Brat annotations into DRS and then into KG for narrative inspection, while [14] uses a PT-BR LLM for NER, RE, and coreference, generating triples for Neo4j. In [5], construction combines YAKE, POS-tagging, and association rules (Apriori) to induce concepts, relations, and instances in a historical domain. Knowledge-supervised approaches appear in [4], which chains NER, entity linking to QIDs, relation type and direction classification, and publishes triples in SPARQL; in [9], the graph is derived from intra-title co-occurrence (each title becomes a clique). In the text $\leftrightarrow$ graph axis, [7] builds "silver" data for QG/QA conditioned simultaneously on graphs (Wikidata) and texts, ensuring dual answerability; [10] builds prepositional *synsets* with linguistic and conceptual layers (frames, *image schemas*); [15] models polarity propagation in a lexical graph (PAPEL) adjusting logic with register metadata; and [6] integrates the commonsense KG into the MT pipeline, focusing on post-editing for slang and culturally marked meanings.

3) Update: automatic regeneration, metadata-driven iterations, and "silver" cycles: Update mechanisms reflect the objectives. In [8], broader extraction strategies are favored, with future steps including relation typing and replication in other languages. Onto.PT is conceived as a non-static resource, regenerable with new sources, glosses, and structural corrections [1], [2]. In [3], modularity allows adding new schemes and visualizations. The legislative pipeline performs duplicate merging via Levenshtein and can be rerun as prompts and post-processing are refined [14]. The historical ontology evolves under human supervision, with refinements and expert validation [5]. The political graph adjusts disambiguation heuristics and foresees enrichment with full news content [4]. Co-occurrence networks tend to be rebuilt as new titles appear [9]. In multimodal QG/QA, the underlying KG is unchanged; updates occur in the renewal of the "silver" set and task rebalancing [7]. PrepNet.Br foresees incremental expansion (more senses, expressions, PT↔EN alignment) [10]. Polarity propagation in PAPEL is refined with systematic incorporation of register metadata [15]. In MT assisted by ConceptNet, there is an explicit plan for alignment and cross-enrichment between PT-BR and EN ConceptNets [6].

4) Evaluation: structural metrics, manual samples, pipeline components, and functional utility: Evaluation criteria also differ. In [8], node/edge counts per domain are combined with manual CA/RA checks. In Onto.PT, evaluation includes large-scale coverage and sample accuracy by relation type, with moderate agreement [1], [2]. In [3], evaluation is functional/visual (inspection and debugging via DRS/KG). In [14], verification is empirical (KG coherence with the text), supported by EDA and readability. The political pipeline measures performance per component (NER, linking, relation type/direction classification) before publishing the graph [4]. Co-occurrence networks are validated by complex network metrics and *small-world* properties [9]. In QG/QA, evaluation uses BERTScore, F1/EM, and text↔graph consistency [7]. PrepNet.Br validates coherence by interchangeability and conceptual compatibility [10]. Polarity induction is compared to SentiLex-PT01 and quantifies gains from handling metadata [15]. In MT with ConceptNet, emphasis is on perceived cultural/semantic quality in a chat prototype [6].

5) KG use in the pipeline: lexical resource, consultable final product, knowledge module, and evaluation tool: Regarding the KG's role, four recurring patterns emerge. As a lexical resource/ontology, Onto.PT provides sense and relation inventories for query expansion, WSD, and QA, integrable in RDF/OWL [1], [2]; PrepNet.Br complements wordnets with formalized prepositional senses, useful for parsing and generation [10]; PAPEL serves as the structural substrate for inducing polarity lexicons [15]. As a consultable final product, the political graph is materialized in RDF and published with a SPARQL endpoint [4]; the legislative graph is stored in Neo4j for queries/visualization and downstream tasks [14]. As a knowledge module, ConceptNet supports MT post-editing, disambiguating slang and cultural meanings [6]. As an evaluation/inspection tool, Brat2Viz's DRS→KG enables annotation debugging [3]; co-occurrence networks support "state of the art" diagnostics and navigation [9]; and, in QG/QA, the KG serves as structural ground truth to check cross-modal semantic consistency [7]. The link-based method in [8] positions its KG as a lexical-conceptual base still with limited inference due to untyped edges.

6) Interoperability patterns and differentiators: Technical interoperability is strongest when W3C standards are adopted: Onto.PT in RDF/OWL with compatible services [1], [2] and the political graph in RDF/SPARQL with Wikidata QIDs [4]. Conceptual and cross-modal interoperability is prominent in [7], which requires text↔graph alignment and use of Wikidata multilingual labels, and in [6], which plans to align PT-BR and EN ConceptNets. Resources like PrepNet.Br aim for PT↔EN *synset* alignment in its early phase [10]. Neo4j-based pipelines [14] prioritize agility and exploration but do not offer out-of-the-box semantic interoperability; DRS as an intermediate layer favors visualization portability [3].

7) Scalability, cost, and robustness: Automatic strategies quickly extract large volumes of relations and senses but with reliability trade-offs: the Wikipedia link method delivers broad coverage with noise and untyped edges [8]; large-scale ontologization in Onto.PT scales but requires structural corrections and glosses [1], [2]. LLM-supervised pipelines are agile and adaptable but sensitive to prompting, disambiguation, and entity merging [14]. Approaches with strong external anchoring (Wikidata/PRON) tend to be robust in consumption and analysis [4], while graph↔text methods for QG/QA improve consistency but depend on "silver" data and rigorous filtering [7]. In historical domains, human supervision remains essential for fine semantic decisions [5].

V. Conclusion

This systematic review mapped 12 primary studies on applications of knowledge graphs (KGs) in the Portuguese language, synthesizing benefits, limitations, interoperability patterns, and ways of integrating them into NLP pipelines. In response to RQ1, recurrent benefits include increased transparency and reusability when published using W3C standards (RDF/OWL/SPARQL), practical utility in downstream tasks (IR, WSD, QA, MT, visualization/inspection), and feasibility of construction under different regimes (extraction/"clustering," rules/transformations, and LLM-assisted workflows). Limitations are mainly related to poor relation typing (generic edges), heterogeneity of metrics and reporting, gaps in extrinsic evaluation, and challenges in multilingual alignment ($PT \leftrightarrow EN$/ES). Interoperability proved stronger when resources adopt global identifiers and open semantic formats, and more limited when persistence is done only as property graphs without semantic publication layers.

Regarding RQ2, four usage archetypes were identified: broad-scope ontologies/lexical resources with semantic reuse;

domain-oriented KGs published as queryable triples or in graph databases; knowledge modules coupled to the pipeline (e.g., DRS→KG, MT post-editing); and reference/evaluation resources (multimodal QG/QA, polarity, co-occurrence). In construction, two dominant poles stand out (statistical extraction and rules/transformations), with a third emerging (LLMs for NER/RE/coreference). For updating, batch regeneration or incremental and modular evolution strategies prevail; for evaluation, structural metrics, manual sampling, and task-based evaluations coexist. The results reflect the search/selection strategy (exclusion of grey literature, focus on studies with an explicit Portuguese component) and the heterogeneous reporting of the works (graph sizes and metrics not always provided).

For future work, it is recommended to prioritize relation typing and normalization, adopting explicit semantic labels and standardized vocabularies to reduce generic edges (*related_to*) and enable inference; establish a standardized minimum reporting that includes counts of nodes, edges, and classes, inventory of relations, source/dump and license, as well as access methods (SPARQL endpoint, RDF/Neo4j dump) and schemas; invest in extrinsic evaluation that measures impact on downstream tasks (IR, WSD, QA, MT, summarization), with replicable baselines and partitions; promote multilingual interoperability through global identifiers (QIDs) and $PT \leftrightarrow EN$/ES mappings (via interlanguage links, BabelNet, or FrameNet) where relevant; explore LLMs with semantic governance, using KGs for grounding, reconciliation, and consistency checking, with provenance recording and traceability to mitigate hallucinations; release open artifacts (code, prompts/rules, annotated data, dumps, and example queries) together with the complete PRISMA workflow; and foster Portuguese benchmarks, including variants, with reference sets and tasks that reduce Anglocentric bias.

In conclusion, Portuguese KGs have evolved from sector-specific prototypes to reusable ecosystems; the next leap depends on relation typing, task-oriented evaluation, and systematic multilingual interoperability. Adoption of the above recommendations is likely to enhance reproducibility, practical impact, and integration between KGs and neural models in real-world scenarios.

References

[1] H. Gonçalo Oliveira, "The creation of onto. pt: A wordnet-like lexical ontology for portuguese," in *International Conference on Computational Processing of the Portuguese Language*. Springer, 2014, pp. 161–169.

[2] H. Gonçalo Oliveira and P. Gomes, "Eco and onto. pt: a flexible approach for creating a portuguese wordnet automatically," *Language resources and evaluation*, vol. 48, no. 2, pp. 373–393, 2014.

[3] E. Amorim, A. Ribeiro, I. Cantante, A. Jorge, B. Santana, S. Nunes, M. d. P. Silvano, A. Leal, and R. Campos, "Brat2viz: a tool and pipeline for visualizing narratives from annotated texts," in *Proceedings of Text2Story-Fourth Workshop on Narrative Extraction From Texts held in conjunction with the 43rd European Conference on Information Retrieval (ECIR 2021)*, 2021.

[4] D. S. Batista, "Extracção de relações de apoio e oposição em títulos de notícias de política em português," *Linguamática*, vol. 15, no. 1, pp. 91–101, 2023.

[5] J. P. Carvalho, O. Belo, and A. Barros, "Extracting automatically a domain ontology from the "book of properties" of the archbishop's table of braga," in *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, 2023, pp. 233–244.

[6] H. de Medeiros Caseli, B. A. Sugiyama, and J. C. Anacleto, "Using common sense to generate culturally contextualized machine translation," in *Proceedings of the NAACL HLT 2010 Young Investigators Workshop on Computational Approaches to Languages of the Americas*, 2010, pp. 24–31.

[7] K. Han and C. Gardent, "Multilingual generation and answering of questions from texts and knowledge graphs," in *The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)*. Association for Computational Linguistics, 2023, pp. 13 740–13 756.

[8] C. C. Xavier and V. L. S. de Lima, "A method for automatically extracting domain semantic networks from wikipedia," in *International Conference on Computational Processing of the Portuguese Language*. Springer, 2012, pp. 93–98.

[9] T. Henrique, I. d. S. Fadigas, M. G. Rosa, and H. B. d. B. Pereira, "Mathematics education semantic networks," *Social Network Analysis and Mining*, vol. 4, no. 1, p. 200, 2014.

[10] D. D. Garcia and B. C. D. da Silva, "Prepnet. br: a semantic network for prepositions," in *Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL)*. SBC, 2015, pp. 75–79.

[11] M. Page, J. McKenzie, P. Bossuyt, I. Boutron, T. Hoffmann, C. Mulrow, L. Shamseer, J. Tetzlaff, E. Akl, S. Brennan, R. Chou, J. M. Glanville, J. Grimshaw, A. Hrõbjartsson, M. Lalu, T. Li, E. Loder, E. Mayo-Wilson, S. McDonald, L. McGuinness, L. Stewart, J. Thomas, A. Tricco, V. Welch, P. Whiting, and D. Moher, "The prisma 2020 statement: An updated guideline for reporting systematic reviews." *Journal of clinical epidemiology*, 2021.

[12] A. Carrera-Rivera, W. Ochoa, F. Larrinaga, and G. Lasa, "How-to conduct a systematic literature review: A quick guide for computer science research," *MethodsX*, vol. 9, p. 101895, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2215016122002746

[13] A. P. Siddaway, A. Wood, and L. Hedges, "How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses." *Annual review of psychology*, vol. 70, pp. 747–770, 2019.

[14] G. L. A. d. Oliveira, B. S. Santos, M. Silva, and I. Silva, "Exploring legislative textual data in brazilian portuguese: Readability analysis and knowledge graph generation," *Data*, vol. 10, no. 7, p. 106, 2025.

[15] A. P. Santos, H. G. Oliveira, C. Ramos, and N. C. Marques, "The role of language registers in polarity propagation," in *International Conference on Computational Processing of the Portuguese Language*. Springer, 2012, pp. 235–240.