Expert Systems With Applications 303 (2026) 130434

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Cross-lingual transfer learning for knowledge graph acquisition: Paradigms, resources and challenges

Wei-Lin Chen [a], Kai-Qing Zhou [a,*], Arezoo Sarkheyli-Hägele [b], Feng Qin [c], Khairunnisa Hasikin [d,e], Di-Wen Kang [a,d], Azlan Mohd Zain [c]

[a] School of Communication and Electronic Engineering, Jishou University, Jishou, 416000, Hunan, China
[b] Department of Computer Science and Media Technology, Sustainable Digitalisation Research Centre (SDRC), Malmö University, Malmö, 20506, Scania, Sweden
[c] Faculty of Computing, Universiti Teknologi Malaysia, Skudai, 81310, Johor, Malaysia
[d] Department of Biomedical Engineering, Faculty of Engineering, Universiti Malaya, Kuala Lumpur, 50603, Malaysia
[e] Center of Intelligent Systems for Emerging Technology (CISET), Faculty of Engineering, Universiti Malaya, Kuala Lumpur, 50603, Malaysia

ARTICLE INFO

Keywords:
Knowledge graph
Cross-lingual transfer learning
Knowledge acquisition
Low-resource languages
Information extraction

ABSTRACT

Knowledge graphs play a pivotal role in structuring human knowledge within artificial intelligence systems. Nonetheless, knowledge distribution is markedly uneven across languages, and linguistic community activity can hinder the performance and scale. Cross-lingual transfer learning emerges as a predominant effective strategy to surmount linguistic barriers, facilitating knowledge transfer across natural languages. This paper reviews cross-lingual knowledge acquisition for knowledge graphs, offering the first systematic integration of cross-lingual transfer paradigms and resources in this field. It critically examines the state of research across subtasks (including named entity recognition, relation extraction, coreference resolution and entity linking). Despite the advancements facilitated by multilingual word embeddings, pre-trained language models and large language models, persistent challenges such as language bias-induced alignment difficulties and low transfer efficiency continue to impede progress. Enhancing model effectiveness through both paradigms and resources will benefit the future construction of multilingual or minor-language knowledge graphs.

1. Introduction

The knowledge graph (KG) (Singhal, 2012) is a semantic network that depicts innumerable entities and their interrelationships in the real world. It effectively carries a vast range of human knowledge in a structured format and has become an essential branch of the artificial intelligence field. KG breaks the traditional mode of data storage and usage, adopting a more intuitive graphical structure to present knowledge in different domains, making knowledge easier for humans to understand with excellent interpretability and computability. Therefore, KG is widely applied in intelligence reasoning applications, including semantic search (Singhal, 2012), question answering (Huang et al., 2019), and recommendation systems (Zhang et al., 2024).

Since the dawn of the information age, there has been a continuous pursuit to encode human cognition and thought processes into computer systems, the main development and some core formalisms summarised in Fig. 1.

A retrospective analysis of the historical trajectory identifies two distinct strands in the development of KGs. On the one hand, KGs represent an extension of earlier standards in graphical knowledge representation. Semantic networks, (Quillian, 1968), which use interconnected nodes and edges to encapsulate knowledge, are considered progenitors of KGs. Frames (Minsky, 1979) and conceptual graphs (Sowa, 1979) endeavored to imbue graphical representations with formal semantics; description logic (Brachman & Schmolze, 1985) provided a method for precise and well-defined semantics. Description Logic later evolved into the Web Ontology Language, which, alongside the World Wide Web (Berners-Lee, 1989), was integrated with other technical standards, such as the Uniform Resource Identifier and the Resource Description Framework, to form the foundations of the Semantic Web (Berners-Lee, 1998) and Linked Data (Berners-Lee et al., 2006). While KG inherits these technical standards and Semantic Web databases, it does not prioritise data openness but strictly controls content and data quality (Hitzler, 2021). On the other hand, KGs reveal a contemporary iteration of traditional knowledge-based systems. Early rule-based (Liao, 2005) or fuzzy logic-based expert systems(Qin et al., 2025; Ye et al., 2023), knowledge engineering (Studer et al., 1998), and modern KGs share the fundamental concept of knowledge-driven, AI-based exploration for representing

* Corresponding author.
E-mail addresses: chen0495@stu.jsu.edu.cn (W.-L. Chen), kqzhou@jsu.edu.cn (K.-Q. Zhou), arezoo.sarkheyli-haegele@mau.se (A. Sarkheyli-Hägele), qinfeng@graduate.utm.my (F. Qin), khairunnisa@um.edu.my (K. Hasikin), kangdiwen@jsu.edu.cn (D.-W. Kang), azlanmz@utm.my (A.M. Zain).

https://doi.org/10.1016/j.eswa.2025.130434
Received 7 May 2025; Received in revised form 5 September 2025; Accepted 12 November 2025
Available online 15 November 2025

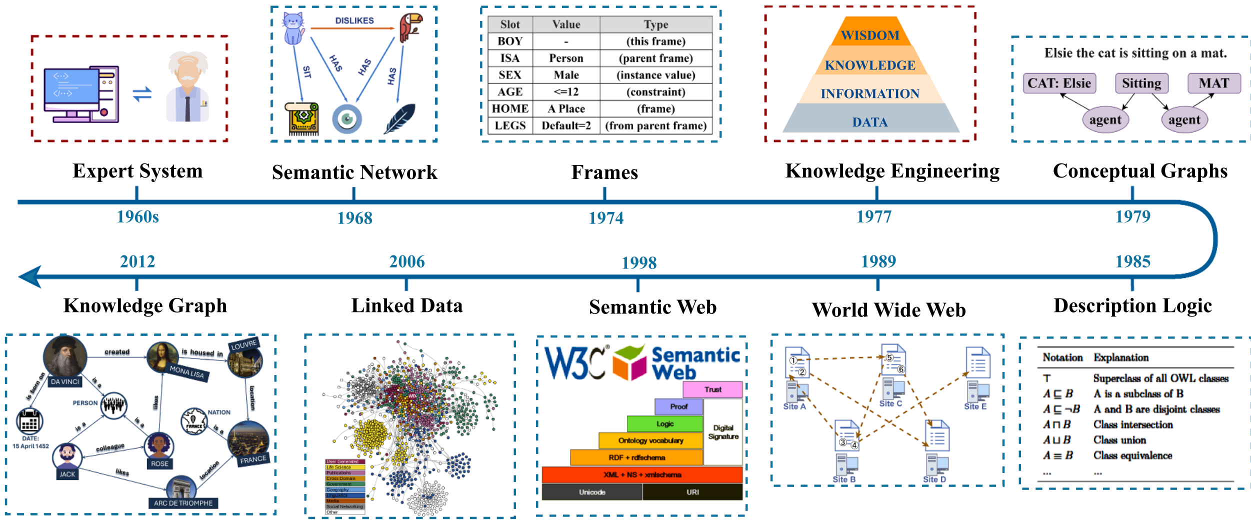


Fig. 1. Evolutionary history of KG.

knowledge and solving problems with computers. However, the distinction lies in the fact that early systems relied on experts to acquire knowledge manually, encountering challenges in knowledge acquisition, the inability to automate construction, and limited data scalability. In contrast, KGs leverage natural language processing (NLP) and machine learning technologies to extract knowledge from large-scale complex data, possessing more robust automated construction and reasoning capabilities.

The emergence of KGs has transformed the traditional model of knowledge acquisition, shifting away from relying on expert input to directly extracting knowledge from industry-specific sources, social media, and open knowledge bases. However, these diverse data sources are often global and multilingual, posing a significant challenge because much of the KG research focuses on major languages, such as English, often neglecting other languages. This disparity results in an unequal distribution of knowledge across languages (Klein et al., 2017; Seganti et al., 2021b). Major languages like English and German hold more extensive and higher-quality data, potentially leading to language bias in models and affecting their credibility and usability. Moreover, the structural and semantic variations between languages introduce additional complexities for text processing (Pikuliak et al., 2021). Conversely, as the global community increasingly prioritises marginalised linguistic regions, the demand for knowledge acquisition from low-resource languages is rapidly growing. This has spurred researchers to explore Cross-lingual Learning (CLL), a method of transferring knowledge from one natural language to another (Huang et al., 2022; Jin et al., 2019). CLL has emerged as a powerful tool for developing KG systems for low-resource languages, significantly reducing the need for model retraining in the target language and conserving time and computational resources.

Although CLL has shown potentials on the construction of KGs for low-resource languages, a systematic review is needed to identify the progress, challenges and directions of the field. Existing literature mainly focuses on specific techniques or application scenarios, and there has yet to be a comprehensive analysis that covers all aspects of cross-lingual knowledge acquisition. Therefore, it is necessary to conduct an in-depth investigation of this field to identify the current research progress, challenges, and future research directions.

To ensure the comprehensiveness and systematic nature of this analysis, we have formulated 11 research questions that address key aspects of cross-lingual knowledge acquisition from multiple perspectives. These questions aim to define the scope and focus of the research, refine the definition of subtasks, explore the practical application and challenges of cross-lingual approaches, and summarize the strengths and weaknesses of existing studies to guide future research. Table 1 lists these core questions, serving as the research framework for this literature investigation.

In response to the research questions in Table 1, the remainder of this paper is structured as follows: Section 2 details our research methodology and the strategy for literature collection. Section 3 then delves into the key concepts, providing a foundational understanding for the reader. Section 4 presents an in-depth analysis of the existing subtasks within cross-lingual knowledge acquisition and analyses the literature in detail. Section 5 synthesises the findings of this review and highlights future directions for KG.

2. Research methodology

Through comprehensive analysis of knowledge graph construction technologies, prior surveys are synthesized to establish thematic boundaries and elucidate pertinent research subtasks. Subsequently, a systematic literature acquisition strategy is implemented to isolate seminal publications for critical examination. The present systematic review conforms to PRISMA guidelines (Page et al., 2021) to ensure methodological transparency and analytical rigor.

2.1. Related surveys

KG is a rapidly evolving research field that merges its unique techniques with those developed in the NLP domain, sparking various discussions about constructing KGs. However, a clear and unified classification of the various subtasks involved in KG remains lacking, complicating the accurate identification and definition of application points for CLL. To systematically organize and summarize relevant research while minimizing redundant discussions, we reviewed secondary literature published between 2021 and 2024. This analysis identified potential KG knowledge acquisition subtasks relevant to CLL, as outlined in Table 2. The findings aim to clarify directions for integrating KG with CLL and establish a foundation for this investigation.

From the Table 2, some interesting findings could be obtained as follows. Hogan et al. (2021) provide an overview of essential KG technologies from ontology to schema construction. Similarly, Tamašauskaitė and Groth (2023) outline the development process of KG through a systematic review. As KG evolves, research has delved into advanced forms, dividing into two main directions. On the one hand, due to the static representation of knowledge in traditional KG, emerging variants such as event KG and temporal KG have focused on incorporating dynamic information, which was previously not included. For event KG,

Table 1
Research questions for the review.

No.	Description	Index
Q1.	How should the research topic and scope for this review be determined?	Section 2.1
Q2.	How can the subtasks within cross-lingual knowledge acquisition be delineated?	Section 2.1, 2.2
Q3.	What strategy should be employed to collect the literature?	Section 2.2
Q4.	What are the fundamental concepts associated with KG and CLL?	Section 3.1, 3.2
Q5.	What taxonomy has been established for the collected literature?	Section 4.1
Q6.	What are the advantages and disadvantages of various cross-lingual approaches and resources?	Section 4.2
Q7.	What are the specific concepts and operations involved in the subtasks identified in Q2?	Section 4.3
Q8.	How can the subtasks utilise the cross-lingual approaches and resources?	Section 4.3
Q9.	What challenges does cross-lingual knowledge acquisition face in practice?	Section 4.4
Q10.	What key findings could be obtained from this review?	Section 4.4
Q11.	What are the conclusions of this review and the open questions for future research?	Section 5

Table 2
A Comparison between Existing Surveys on Knowledge Graph Construction.

Publication	Year	Named Entity Recognition	Relation Extraction	Event Extraction	Attribute Extraction	Coreference Resolution	Entity Linking	Entity Alignment	Attribute Alignment	Knowledge Graph Completion	Domain
Hogan et al. (Hogan et al., 2021)	2021	✓	✓				✓			✓	KG.
Guan et al. (Guan et al., 2021)	2021		✓	✓		✓				✓	Event KG.
Ji et al. (Ji et al., 2022)	2022	✓	✓				✓	✓		✓	KG, Temporal KG.
Tamašauskaitė and Groth (Tamašauskaitė & Groth, 2023)	2023	✓	✓		✓			✓		✓	KG.
Zhong et al. (Zhong et al., 2023)	2023	✓	✓			✓	✓	✓	✓	✓	KG, Condition KG, Temporal KG.
Knez and Žitnik (Knez & Žitnik, 2023)	2023	✓	✓	✓	✓						Event-centric temporal KG.
Zhu et al. (Zhu et al., 2024b)	2024	✓	✓	✓			✓	✓		✓	Multi-modal KG.
Tang et al. (Tang et al., 2024)	2024	✓	✓	✓				✓		✓	KG.
Chen et al. (Chen et al., 2024a)	2024						✓	✓	✓	✓	KG.
Pikuliak et al. (Pikuliak et al., 2021)	2021										CLL.
Our	2024	✓	✓			✓	✓				KG, CLL.

(Guan et al., 2021) explore the emergence and evolution of event KGs from a perspective that diverges from the traditional entity-centric approach. Further insights into event-centric temporal KGs are presented by Knez and Žitnik (2023). For temporal KGs, Ji et al. (2022) provide insights into temporal knowledge embedding and reasoning, while Zhong et al. (2023) extend this discourse to a broader conceptualisation of conditional KGs. On the other hand, some surveys focus on interdisciplinary areas such as multi-modal and machine learning. The study by Zhu et al. (2024b) delves into KGs from the perspective of multi-modal data, encompassing voice and image. Tang et al. (2024) explore the integration of reinforcement learning-based technologies within KG frameworks. Furthermore, the work by Chen et al. (2024a) addresses privacy preservation challenges associated with openly sharing multiple KG datasets.

However, many of these studies (Guan et al., 2021; Hogan et al., 2021; Ji et al., 2022; Knez & Žitnik, 2023; Zhong et al., 2023) acknowledge the existence of CLL but fail to engage in-depth exploration or provide a comprehensive analysis. For instance, Knez and Žitnik (2023) limited their discussion to the task of cross-lingual event extraction. Similarly, Zhong et al. (2023) highlighted that constructing cross-lingual KGs as opposed to simply executing cross-lingual transfer remains an ongoing challenge, but they did not further investigate this topic. Therefore, despite the maturity of cross-lingual phenomena discussed in other reviews and literature, a notable research gap still exists in the domain of CLL and knowledge acquisition. Furthermore, current reviews on cross-lingual technologies do not include KG tasks in their analyses, as noted by Pikuliak et al. (2021). This observation underscores the importance and necessity of our review.

This phase clearly defines this paper's investigation scope. Based on previous classification, the construction of KG typically involves knowledge acquisition and refinement. The acquisition comprises named entity recognition (NER), relation extraction (RE), attribute extraction (AE), coreference resolution (CR), and entity linking (EL). Refinement entails alignment and completion. The scope excludes Event KG events as well as downstream processes like entity alignment, attribute alignment, and knowledge graph completion. AE was excluded from the literature screening process as there were fewer than 10 independent cross-lingual studies. Consequently, this review will investigate NER, RE, CR, and EL.

2.2. *Data collection*

A systematic literature review was conducted to collect this field's research comprehensively. The literature search parameters were defined as Table 3.

The search in Scopus yielded 911 results, while the search in Web of Science produced 492 results. Additionally, 68 preprint articles from arXiv published between 2024 and 2025 were identified. This resulted in a total of 1471 articles included in the preliminary screening. Subsequently, 5 inclusion criteria were then applied to initially filter these results:

Table 3
The literature search parameters.

Setting	Description
Scope	Analysis of secondary literature from 2021 to 2024, focusing on 5 subtasks (NER, RE, AE, CR, and EL), initially identified in Section 2.1.
Sources	The surveyed literature was gathered from the bibliographic databases Scopus and Web of Science.
Search	The search encompassed the article title, abstract, and keywords across two datasets.
Keyword	The keywords were *("cross-lingual" OR "multilingual") AND ("named entity recognition" OR "relation extraction" OR "attribute extraction" OR "coreference resolution" OR "entity linking")*.
Date	The cutoff date for literature retrieval was March 4, 2025.

(1) Duplicates: Duplicated articles within each subtask were excluded, allowing one piece of literature to appear in different subtasks since this study aims to conduct independent investigations within each subtask.
(2) Domain scope limitation: Sub-tasks containing fewer than 10 articles were excluded after de-duplication, as not every subtask holds practical research value.
(3) Non-English: All non-English results were excluded.
(4) Abstract and citation: This paper filtered the articles based on citations to select high-quality papers from various periods. Considering the impact of publication time, the following screening criteria were set for each subfield to ensure fairness and rigor. **1.** To objectively reflect paradigmatic shifts across developmental periods of KG and equitably evaluate both pioneering and contemporary contributions, temporal stratification was employed for literature screening. The first period (2011 and earlier) encompasses foundational literature predating KGs as an independent research domain. The second period (2012–2018), marked by Google Knowledge Graph's release, features rapid development driven by statistical machine learning and early neural approaches. The third period (2019–2020) witnessed a paradigm shift with widespread adoption of pre-trained models, transforming the field's methodology. The fourth period (2021–2022) represents the deepened application of these models. For each period, the five most highly cited papers were selected to identify influential contributions. **2.** For publications from 2023 to 2025, given citation limitations, we employed manual screening. Author h-index, venue h5-index, and impact factor data were provided to three authors who collaboratively evaluated papers' titles, abstracts, and publication metrics to identify promising studies, subsequently verifying that each paper: (1) addresses NER, RE, CR, or EL; (2) proposes novel methodologies; and (3) provides benchmark-comparable results.
(5) Paper Content: Full-text evaluation excluded reviews, commentaries, dataset analyses, and peripheral literature. Excluded papers were replaced with the next highest-cited article or one approved by the reviewing authors, until all criteria were satisfied.

To ensure comprehensiveness, paired authors conducted dual reviews. Any disagreements encountered during both initial screening and full-text evaluation stages were resolved through consensus discussion, consulting another research team member when necessary. This systematic approach enhanced selection reliability and precision.

Fig. 2 illustrates the literature selection process. Initially, 1471 papers were identified from two databases, with 373 duplicates removed. Then, 5 papers related to AE were excluded due to a lack of relevant literature, and 9 non-English publications were also eliminated. Finally, 913 results were excluded based on citation ranking and abstract screening, leaving 171 key papers. Although this approach may overlook high-quality papers published later in the selected timeframe, it effectively filters out most weakly relevant and low-quality results while retaining the most relevant and high-quality studies.

The next step involved a full-text evaluation of the 171 selected articles, from which 7 were excluded due to inaccessible full-text resources. No duplicates were found among the sub-tasks. Of the remaining papers, 99 were excluded for content-related reasons, such as lack of innovation, irrelevance of methodology to cross-linguistic transfer, or unrelated research topics. Ultimately, 65 papers were selected for the final review.

The literature collected using the aforementioned methods-totaling 1016 articles (excluding 68 arXiv preprints) without filtering by citation, reveals trends in cross-lingual knowledge acquisition research.

As illustrated in Fig. 3, research on NER has significantly increased since 2017, coinciding with the introduction of the Transformer architecture and pre-trained language models (PLMs), peaking in 2023. This indicates a sharp rise in research interest in recent years. At the same time, large language model (LLMs) began to emerge in late 2022 and early 2023, spreading into the NER field. In contrast, the growth trends in RE and CR have primarily emerged in the last few years, suggesting these fields are gradually gaining more attention. Our subsequent detailed analysis revealed that community competitions and the development of datasets have positively influenced the advancement of these areas. Research on EL has shown slower but steady growth overall.

As a more upstream task, NER has been more significantly impacted by PLMs, leading to a rapid expansion in research scale. Although other downstream tasks have also been affected, their progress is slower due to their nature and limitations such as dataset availability. This suggests that while downstream tasks tend to lag in the application of emerging technologies, they still possess considerable potential for development.

3. Preliminaries

This section introduces the core concepts in this paper, encompassing the elucidation of KGs, cross-lingual transfer definition, and the monolingual and multilingual word embeddings used in modern approaches. These discussions directly pertain to **Q4**. To ensure clarity and provide easy reference, a glossary of mathematical symbols, a list of common acronyms, and a key for language codes are provided in Table 4, Table 5, and Table 6, respectively.

3.1. Knowledge graph

KGs are logically composed of two basic units: entity and relation. Entities can refer to specific individuals, locations, organizations, events, or abstract concepts, each identifiable by a globally unique ID. Relations represent the external connections or associations between entities. In addition, there are attributes and attribute values. Attributes provide information about entities, describing their characteristics or parameters, while attribute values are specific instances or parameter values of attributes, representing the particular values of the attributes for a given entity.

In formal terms, KGs are represented as a triples $G = \{E, R, F\}$, where $E = \{e_1, e_2, \ldots, e_{|E|}\}$ denotes the collection of all entities, $R = \{r_1, r_2, \ldots, r_{|R|}\}$ denotes the collection of all relations, and $F = \{f_1, f_2, \ldots, f_{|F|}\}$ denotes the collection of all facts. Each fact f is represented as a directed triple $f = (h, r, t) \in F$, where h, r, and t represent the head entity, relation, and tail entity, respectively. The basic forms of facts include $f = \{entity, relation, entity\}$ and $f = \{entity, attribute, attribute\ value\}$. Attributes are entity-centric, comprehensively characterizing entities through the extraction of nuanced properties. Nevertheless, cross-lingual AE research remains insufficiently developed (as evidenced in Section 2). Given the scope of

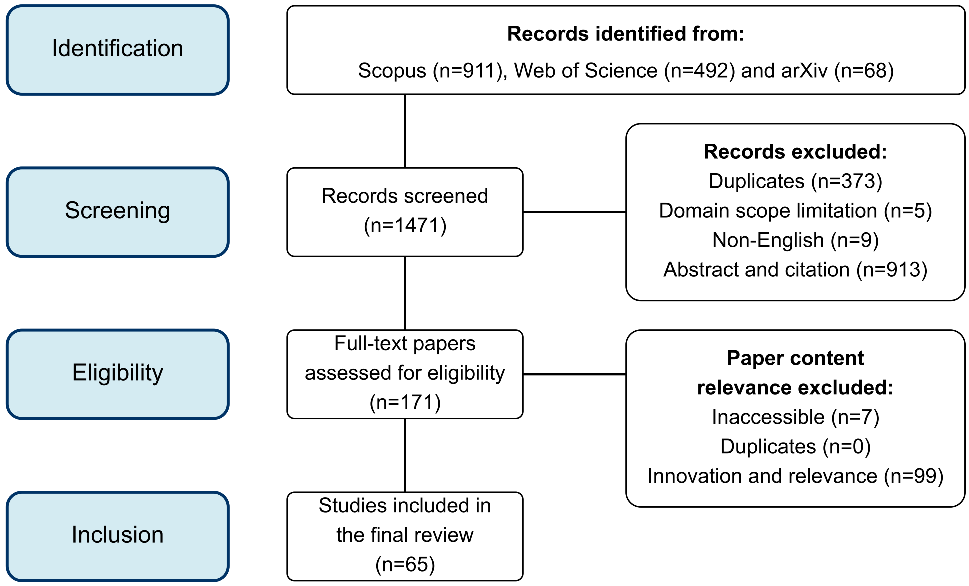


Fig. 2. Flowchart of the screening process.

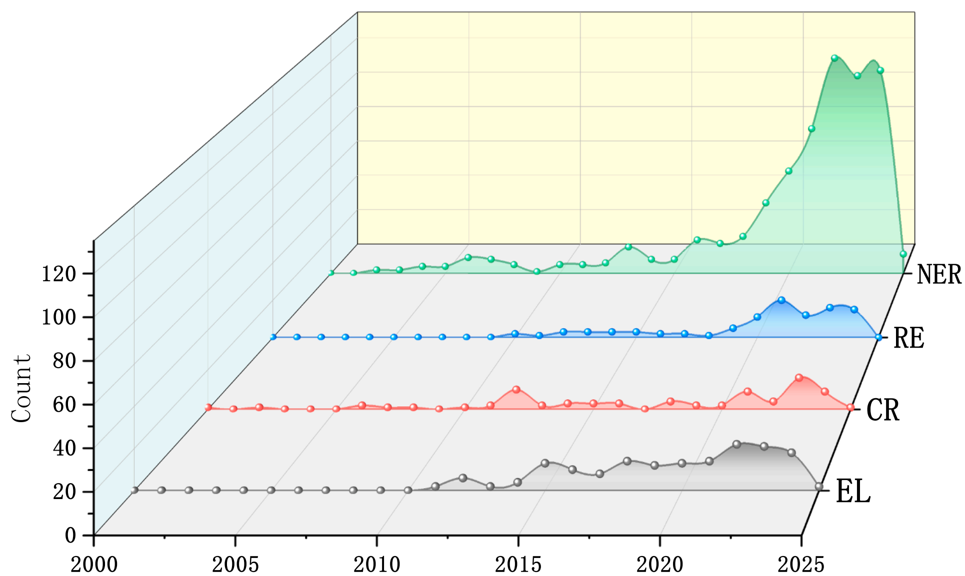


Fig. 3. Trends in Cross-Lingual knowledge acquisition Research.

Table 4
Symbols Reference Table.

Sym.	Term	Sym.	Term	Sym.	Term	Sym.	Term	Sym.	Term
G	Knowledge Graph	r	Relation	$P(\cdot)$	Prob. Distribution	$\mathscr{L}$	Language Set	t^e	Entity Type
E	Entity Set	h	Head Entity	T	Task	$\mathscr{P}$	Power Set	s	Coreference Chain
R	Relation Set	t	Tail Entity	C	Label Space	y	Label Space		
F	Fact Set	D	Domain	ℓ_t	Target Language	w	token		
e	Entity	$\mathscr{X}$	Feature Space	ℓ_s	Source Language	I	Index Position		

the present investigation, cross-lingual AE is not addressed in this manuscript.

Acquisition, refinement, evolution, and storage are the lifecycle stages of a KG from construction to deployment. The acquisition is the initial step in creating a KG, which primarily involves extracting essential information, such as entities and relations, from various data sources. These sources can be structured, semi-structured, or unstructured. The refinement process aims to improve the accuracy and completeness of the KG. Knowledge evolution analyses changes in the KG over time and introduces advanced semantic features. Storage refers to organising and preserving the information in the KG to facilitate retrieval and querying. Our review focuses explicitly on the acquisition process.

3.2. Cross-lingual learning

CLL is essentially a remedial strategy designed to address data insufficiency in language models. It is a flexible concept that has been widely

Table 5
Term Abbreviations Reference Table.

Abbr.	Term	Abbr.	Term	Abbr.	Term
AE	Attribute Extraction	EM	Expectation-Maximization	P@1	Precision@1
AUC	Area Under Curve	GAT	Graph Attention Network	PLMs	Pre-trained Language Models
BiLSTM	Bidirectional Long Short-Term Memory	GCN	Graph Convolutional Neural Network	POS	Part-of-Speech
BLEU	Bilingual Evaluation Understudy	HMM	Hidden Markov Model	RAG	Retrieval Augmented Generation
CCA	Canonical Correlation Analysis	IPA	International Phonetic Alphabet	RE	Relation Extraction
CLL	Cross-lingual Learning	KG	Knowledge Graph	RNN	Recurrent Neural Network
CNN	Convolutional Neural Network	LLMs	Large Language Models	seq2seq	Sequence-to-Sequence
CR	Coreference Resolution	LSTM	Long Short-Term Memory	SVD	Singular Value Decomposition
CRF	Conditional Random Field	mBERT	Multilingual BERT	TF-IDF	Term Frequency-Inverse Document Frequency
DFT	Discrete Fourier Transform	mT5	Multilingual T5	UD	Universal Dependencies
DQN	Deep Q-Network	NER	Named Entity Recognition	UMLS	Unified Medical Language System
EL	Entity Linking	NLP	Natural Language Processing		

Table 6
Language Codes Reference Table.

Code	Language	Code	Language	Code	Language	Code	Language	Code	Language
af	Afrikaans	ewe	Ewe	jv	Javanese	nya	Chichewa	th	Thai
amh	Amharic	fa	Persian	ka	Georgian	om	Oromo	ti	Tigrinya
ar	Arabic	fi	Finnish	kan	Kannada	pa	Punjabi	tk	Turkmen
asm	Assamese	fil	Filipino	kin	Kinyarwanda	pcm	Nigerian Pidgin	tl	Tagalog
awa	Awadhi	fon	Fon	ko	Korean	pl	Polish	tr	Turkish
bam	Bambara	fr	French	ky	Kyrgyz	pt	Portuguese	tsn	Tswana
bbj	Ghomala	gn	Guarani	kz	Kazakh	qu	Quechua	twi	Twi
ben	Bengali	grc	Ancient Greek	lo	Lao	ro	Romanian	ug	Uyghur
bg	Bulgarian	hau	Hausa	lt	Lithuanian	ru	Russian	uk	Ukrainian
bho	Bhojpuri	hbo	Ancient Hebrew	lug	Luganda	rw	Kinyarwanda	ur	Urdu
bn	Bengali	he	Hebrew	luo	Luo	si	Sinhala	uz	Uzbek
ca	Catalan	hi	Hindi	mhr	Meadow Mari	sna	Shona	uzn	Uzbek
cdo	Min Dong	hr	Croatian	mi	Maori	sq	Albanian	vi	Vietnamese
cs	Czech	hu	Hungarian	mk	Macedonian	sr	Serbian	wol	Wolof
cu	Church Slavonic	hy	Armenian	mos	Mossi	sv	Swedish	xho	Xhosa
da	Danish	ibo	Igbo	mr	Marathi	sw	Swahili	xmf	Mingrelian
de	German	id	Indonesian	ms	Malay	swa	Swahili	yor	Yoruba
el	Greek	ilo	Ilocano	my	Myanmar	swh	Swahili	zh	Chinese
en	English	is	Icelandic	nl	Dutch	ta	Tamil	zul	Zulu
es	Spanish	it	Italian	no	Norwegian	tam	Tamil		
eu	Basque	ja	Japanese	npi	Nepali	te	Telugu		

applied in NLP tasks such as machine translation (Siddhant et al., 2020), sentiment analysis (Xu et al., 2022), and question answering (Lewis et al., 2019), and has permeated various aspects of KG.

According to the definitions of transfer learning (Pan & Yang, 2010) and the formal definition of CLL (Pikuliak et al., 2021), CLL consists of two parts: domain $D = \{\mathscr{X}, P(\mathscr{X})\}$ and task $T = \{y, f : \mathscr{X} \to y\}$, where $\mathscr{X}$ is the feature space of a certain language ℓ, $P(\mathscr{X})$ is the marginal probability distribution over $\mathscr{X}$, y is the label space, and f is the target prediction function trained from $\mathscr{X}$ mapping y. Let D_s, T_s denote the domains and tasks created from source data, and D_t, T_t denote those created from target data. Generally, the source and target domains use the same language, and the tasks are identical, $\left(D_s = D_t = \ell \in \mathscr{L}\right) \wedge \left(T_s = T_t\right)$, where $\mathscr{L}$ is the set of all natural languages. However, in CLL, the source and target domains can both be non-empty language sets, $D_t, D_s \subseteq \mathscr{P}(\mathscr{L})$, where $\mathscr{P}(\mathscr{L})$ is the power set of $\mathscr{L}$. CLL is thus defined as:

$$\exists \ell_A \in D_t \quad \exists \ell_B \in D_s \quad \ell_A \neq \ell_B$$

This implies that knowledge transfer occurs between at least one pair of different languages in the task. Furthermore, if no target language exists in D_s, it indicates a zero-shot CLL task. Additionally, when the source domain contains more than one language $|D_s| > 1$, the training and testing language sets might be identical $D_t = D_s$, meaning the transfer happens within the language set.

4. Cross-lingual knowledge acquisition

Through the examination of transfer mechanisms and supportive resources in cross-lingual knowledge acquisition, researchers can attain a more precise comprehension of the most effective techniques and discern potential challenges or limitations. This survey is organised around two fundamental perspectives: the paradigms and resources that enable knowledge transfer. Each element was assessed and analysed independently according to its operational mechanisms and objectives. The resultant taxonomy is illustrated in Fig. 4.

This section reviews key papers and their subsequent citations. It begins by establishing a broad taxonomy of the literature. Building upon this foundation, we examine the core developments in cross-lingual knowledge acquisition by identifying and analysing various paradigms and resources (Sections 4.1 and 4.2). The investigation then proceeds with an in-depth exploration of specific methodologies, models, and their performance characteristics (Section 4.3). Finally, the applicability of existing research is evaluated, and current challenges are discussed (Section 4.4).

4.1. Paradigms

In cross-lingual learning, developing high-performance models for the target language is difficult due to scarce data. CLL is a framework that utilises knowledge from other languages to improve performance in the target language. CLL paradigms can be classified into three categories, each associated with specific knowledge transfer pathways and sharing mechanisms.

Label transfer entails the mapping of the label space from the source language to the target language through alignment methodologies, including dictionaries or machine translation. Labelled data from the source language is utilised to facilitate task learning in the target

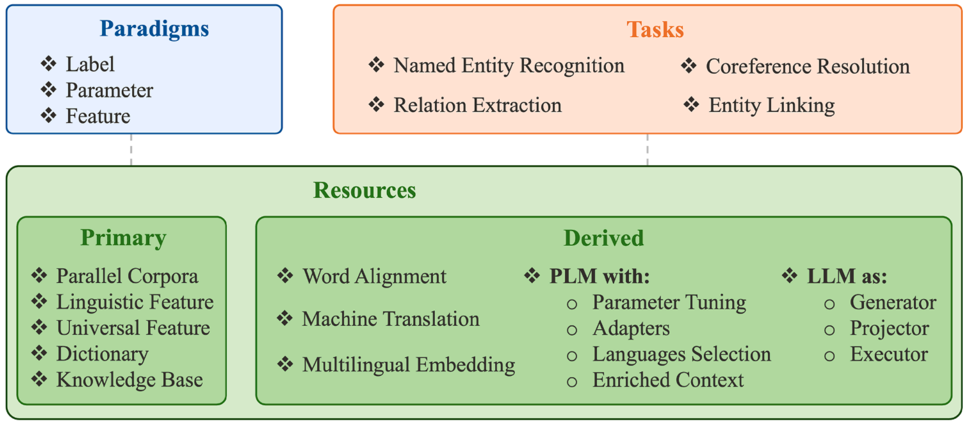


Fig. 4. Taxonomy of Cross-lingual Knowledge Acquisition in KG .

language. This paradigm assumes that the tasks in both languages share the same output label space, and that the content of samples across languages is comparable, ensuring effective transfer (Ji et al., 2015; Rahman & Ng, 2012). Label transfer is particularly advantageous for classification tasks and can markedly diminish the reliance of the target language on labelled data when the output label space is distinctly delineated.

Label transfer involves three main stages. First, correspondence mappings are established between the source and target languages using bilingual dictionaries, word alignment tools, or machine translation systems. Second, labels from the source language text are transferred to corresponding words or phrases in the target language based on these mappings. Finally, the target language model is trained using the projected annotations.

Despite its simplicity and straightforward implementation, making it a valuable baseline or component in more complex cross-lingual transfer frameworks, label transfer is susceptible to inaccuracies in the cross-lingual mapping process. Any error in mapping a source language vocabulary item to its target language counterpart can lead to erroneous label transfer. Errors can accumulate and propagate throughout the label transfer pipeline. Initial errors in the source language model, coupled with inaccuracies in cross-lingual mapping, and the inherent inapplicability during the label transfer process, collectively contribute to a decline in the quality of transferred labels in the target language (Faruqui & Kumar, 2015; Liu et al., 2021b). Issues such as lexical errors, syntactic and grammatical inaccuracies, and disfluency remain persistent challenges that hinder the effectiveness of these transfer tools (Mohamed et al., 2024). Consequently, training classifiers on noisy data can limit the performance of the learned model. Current solutions primarily focus on several aspects: enhancing word alignment quality, for instance, by employing more sophisticated alignment algorithms or incorporating syntactic information (Fei et al., 2023); and leveraging corpus-level information for re-alignment to improve contextual consistency of alignments (Ehrmann et al., 2011). These approaches offer partial mitigation of noise introduced by translation and alignment. Recent research has begun to explore the utilisation of LLMs and their natural language understanding capabilities (Chen et al., 2024b; Mishra et al., 2024). These methods connect label information from the source domain and unlabeled data from the target domain in natural language form, and leverage the reasoning capabilities of LLMs to generate corresponding labels for the target domain data. Nevertheless, while current LLMs excel in transferring general domain knowledge, they may struggle to accurately comprehend cross-lingual mappings when confronted with more specialized domain transfer tasks.

Parameter transfer emphasises the exchange of internal model parameters, including network weights and biases, to identify shared features and patterns across languages (Limkonchotiwat et al., 2023; Subburathinam et al., 2019). Many languages display grammatical, structural, or semantic similarities, and certain neural network parameters can effectively capture these commonalities. This paradigm lowers the need for large amounts of annotated target data by moving parameters from the source language model to it. Parameter transfer assumes that cross-lingual knowledge is shared to some extent and tends to be more effective when the source and target languages exhibit structural or semantic similarities (Nigatu et al., 2023).

To realize parameter transfer, a common strategy is to construct cross-lingual shared vector space representations (Lan et al., 2020; Peters et al., 2018a), often leveraging word embeddings or pre-trained language models. Within this shared space, semantically similar words or phrases from different languages are mapped to proximal locations in the vector space, thereby achieving cross-lingual semantic alignment. In practice, a typical approach involves initially training a model on a resource-rich source language, enabling it to learn language-agnostic semantic representations and task-relevant knowledge. Subsequently, the trained model can be directly applied to the target language, requiring minimal or no fine-tuning with target language data, as inputs from the target language are also mapped into the pre-constructed shared vector space. Subburathinam et al. (2019) encoded dependency tree structures of sentences in different languages and achieved structural transfer through parameter sharing. Pražák et al. (2021) proposed a multilingual CR model based on mBERT and emphasized the crucial role of consistent dataset annotation specifications for cross-lingual alignment. It is noteworthy, however, that parameter transfer methods may not perfectly align deep semantic spaces; nevertheless, they remain an effective means to enhance cross-lingual transfer efficiency. To further augment semantic alignment, researchers are continuously exploring novel methodologies. Con2GEN (Zhu et al., 2023) employs a contrastive learning strategy, constructing positive and negative sample pairs to encourage the model to discern semantically similar entities, thereby improving its ability to recognize complex or ambiguous entities. Furthermore, incorporating domain knowledge into the parameter transfer process has proven beneficial for achieving more effective cross-lingual alignment within specific domains, as exemplified by the teaBERT model (Chen et al., 2023). On the other hand, while generative models, exemplified by LLMs, have demonstrated significant potential in few-shot learning, they also encounter challenges such as relatively weaker output controllability, high training and fine-tuning costs, and potentially suboptimal performance on specific tasks.

Feature transfer attempts to align feature spaces between source and target languages to construct a language-independent feature space where instances from different languages are comparable (Szarvas et al., 2006). This approach generally necessitates domain expertise to create efficient universal features that can encapsulate task-relevant, critical information across various languages. Nonetheless, identifying effective features can be difficult when transitioning between various language pairs or tasks, thereby constraining the method's generalisability and applicability.

Feature transfer employs domain knowledge and linguistic proficiency to create universal features (Fernandes et al., 2014), encompassing lexical, syntactic, semantic, and positional information, to identify commonalities among languages at abstract levels. Models acquire the ability to utilise these universal features for prediction in the source language. Owing to their universality, trained models can be directly applied to target languages without the need for retraining. Nonetheless, these characteristics are frequently discrete, sparse, and necessitate domain expertise. Since PLMs came along in the era of deep learning, parameter transfer methods have mostly replaced traditional feature transfer methods. As a result, the role and methodology of feature engineering have evolved and become more integrated into the design and training of neural models.

4.2. Resources

This section examines key resources supporting cross-lingual learning, categorized by their roles within the research ecosystem into primary and derived resources. Primary resources constitute foundational data and knowledge sources-parallel corpora, dictionaries, external knowledge bases, and meta-linguistic features-serving as raw materials for analysis and training. Derived resources encompass computational tools and representations built upon these foundations, including word alignment systems, machine translation models, and learned representations such as multilingual embeddings and language models. Comprehensive understanding of these resources proves essential for researchers and developers constructing efficient multilingual and low-resource language applications.

4.2.1. Primary resources

Primary resources constitute the foundation of cross-lingual research, comprising raw data and structured knowledge that encode linguistic equivalences and real-world facts. Curated through extensive human effort or collected as direct linguistic evidence, these resources provide essential ground truth and foundational knowledge for computational models to bridge language barriers.

Parallel corpora. Parallel corpora are linguistic resources consisting of text pairs that contain identical content across multiple languages. These pairs typically contain source language text and its corresponding target language translation, aligned at the sentence, paragraph, or document level. As the primary source of cross-lingual knowledge, parallel corpora usually contain large amounts of textual data. They assist in the training of statistical models, establish an essential basis for analysing linguistic phenomena, and enable models to learn interconnections between languages.

Current studies frequently use accessible parallel corpora. Researchers use this as baseline data for feature extraction and model training, saving resources on data collection and alignment. A key advantage of using shared public corpora is the enhanced comparability and reproducibility of results across studies. Gui and Xiao (2024) used CCMatrix-v1[1], a large-scale multilingual resource with 576 language pairs and 4.5 billion parallel sentence pairs. Pfeiffer et al. (2020) trained the XLM-R model on the Common Crawl corpus, which covers 100 languages. Ehrmann et al. (2011) integrated multiple small-scale parallel corpora. However, general-purpose parallel corpora can differ significantly from specific task domains, potentially leading to suboptimal model performance in practice. Additionally, some corpora have infrequent updates, making it difficult to capture recent linguistic changes, especially in professional domains and emerging topics.

Acknowledging the constraints of directly employing existing corpora, certain researchers concentrate on comprehensive mining and application of available resources, generating new corpora via systematic preprocessing and reprocessing. These customised corpora are optimised for specific tasks or domains, effectively addressing particular application requirements. Corpus construction methodologies are becoming progressively varied and encompass enlisting multilingual participants for English text translation through crowdsourcing platforms (Harabagiu & Maiorano, 2000); utilising web crawlers to amass online multilingual textual resources (Klementiev & Roth, 2006); producing target language text via machine translation systems (Faruqui & Kumar, 2015); and aggregating and structuring content from knowledge bases with naturally multilingual versions, such as Wikipedia (Nothman et al., 2013).

Regardless of whether existing or new corpora are used, ensuring corpus quality remains crucial. For parallel corpora requiring quality improvement, researchers need to establish more precise correspondences between materials in different languages, with alignment at word, phrase, sentence, and entity levels. Nothman et al. (2013) innovatively used different hierarchical structures in Wikipedia as alignment signals, building a comprehensive parallel corpus with phrase-level, sentence-level, and paragraph-level alignment. Rijhwani et al. (2019) focused on low-resource language research, using parallel Wikipedia titles between low-resource languages and English as training data.

Linguistic features. Linguistic features are characteristics or attributes within a language that are used to describe and analyze its structure and usage. These features encompass various aspects of language, including Phonetics (e.g., phonemes, stress), syntax (e.g., word order, subject-verb-object structure), and morphology (e.g., noun gender, pluralisation) are important aspects of language study. Cross-linguistic studies rely heavily on these features because their identification and analysis allow researchers to gain a better understanding of how languages are built, function, and evolve.

Despite significant differences in acoustic realization among languages, phonological research shows that words often exhibit phonological correspondences across languages, particularly in proper nouns like personal and place names. This phenomenon largely arises from phonetic borrowing through language contact. Based on this, researchers have developed various phonological comparison methods. For example, discriminative transliteration models (Klementiev & Roth, 2006) identify potential named entity correspondences by comparing phonemic compositions across languages. Consonant Signatures (Ehrmann et al., 2011) capture word forms' core features by removing vowels and normalising specific consonants.

Morphological features are also critical, especially in languages with rich inflectional systems. Existing research employs diverse strategies, including surface form matching by calculating character-level similarity between source and target language entities (Jain et al., 2019). Other approaches focus on the systematicity of morphological changes, using machine translation systems to generate diverse variant forms to understand entity morphological paradigms (Ehrmann et al., 2011). The introduction of character-level BiLSTM has been particularly impactful, enabling models to directly process raw character sequences without relying on predefined vocabularies or word vectors (Rijhwani et al., 2019).

Overall, linguistic features that combine phonological, syntactic, and morphological knowledge are most effective in similar language groups like German, Dutch, and English. However, the performance of these features may suffer when applied to languages with significant differences.

Universal features. Universal features are applicable across different languages. These features can be independent of language-specific expressions, such as punctuation marks and word frequency statistics, or be abstract representations of language features, such as the International Phonetic Alphabet (IPA), part-of-speech (POS) tagging, and universal

[1] https://opus.nlpl.eu/CCMatrix.php

dependencies (UD). Universal features allow models to maintain stable performance in cross-lingual tasks without relying on specific vocabulary or grammar. POS tagging, for example, provides important syntactic clues for NER by identifying the grammatical functions of words (e.g., nouns, verbs, adjectives) and helps determine potential coreference relationships.

Klementiev and Roth (2006) proposed a cross-lingual entity alignment method based on temporal distribution features. This method analyses entity co-occurrence patterns over time and uses Discrete Fourier Transform (DFT) to quantify time series similarity, aiding the main model in identifying corresponding entities across languages. Building on this, (Jain et al., 2019) expanded the use of distribution features, using statistical distributions of entities in datasets to measure word importance for identifying keywords and significant entities. Beyond distribution statistics, distance features between entities also provide crucial evidence for CR (Fernandes et al., 2014; Kundu et al., 2018). These distance features capture the relative positions of entities, improving CR accuracy.

Other research directly uses cross-lingual unified annotation tools. IPA, for example, is a standard for phonological comparison across languages that lets you directly compare the sounds of words from different languages (Jain et al., 2019; Rijhwani et al., 2019). However, IPA notation's granularity might not capture all subtle phonetic differences. In this context, tools like PanPhon (Bharadwaj et al., 2016) offer more precise phonetic feature representations by detailing the position and state of articulatory organs, enhancing models' ability to recognise subtle phonological differences and improving cross-lingual transfer accuracy (Rijhwani et al., 2019). Furthermore, POS tagging, another significant universal feature, identifies words' grammatical roles in sentences (Szarvas et al., 2006; Zheng et al., 2023a) and provides syntactic-level information for NER. For example, consecutive proper noun tags often indicate a potential named entity, while pronouns may suggest coreference relationships. This POS tag-based syntactic analysis offers reliable language-independent features for cross-lingual NLP tasks.

Universal features have broad applicability across diverse linguistic environments. The multilingual POS tagging tool RDRPOSTagger [2] supports POS tagging for over 40 languages, and lexical distribution features are generally applicable to most current language writing systems. However, selecting appropriate features for different NLP tasks, such as dependency parsing and information extraction, is crucial. Researchers must also carefully consider how to ensure the accuracy of these mappings in research and application.

Dictionaries. Dictionaries are essential for cross-lingual knowledge alignment, providing one-to-one or many-to-one lexical mappings between languages. Their broad availability and language coverage make them crucial for cross-lingual transfer learning. Dictionaries come in various forms, including traditional printed versions, electronic versions, and structured databases. Bilingual dictionaries are most common, establishing direct vocabulary translation correspondences between two languages. These can be obtained online or compiled manually by linguistic experts (Harabagiu & Maiorano, 2000; Klementiev & Roth, 2006; Lan et al., 2020). Beyond these, structured lexical knowledge bases are also important dictionary resources. For example, Wikipedia's interlingual links act as generalised dictionaries, providing corresponding representations for entities or concepts across languages (Rijhwani et al., 2019). The Unified Medical Language System (UMLS), a large-scale biomedical knowledge base, contains multilingual biomedical terminology dictionaries, offering professional support for cross-lingual transfer in medical knowledge acquisition (Liu et al., 2021a).

Dictionary resources are crucial in practical cross-lingual transfer learning. In NER tasks, dictionaries expand candidate translation sets and, combined with language and universal features, enable effective disambiguation (Klementiev & Roth, 2006). Dictionaries are also valuable for correcting model mapping errors, especially during the fine-tuning of PLMs, enhancing their performance in cross-lingual knowledge acquisition (Lan et al., 2020). For language pairs lacking direct dictionary connections, researchers have proposed solutions based on dictionary transitivity (Rijhwani et al., 2019). This involves utilising the genealogical relationships of language systems when direct mappings between low-resource languages and source languages (like English) are unavailable. A two-step bridging mechanism is created using dictionaries between the low-resource language and high-resource languages within the same family, and then between those high-resource languages and the source language. This effectively links low-resource language entities to source language knowledge bases. This transitivity-based strategy significantly broadens the application of dictionaries in cross-lingual transfer.

However, dictionary resources present two main challenges. First, some dictionaries offer only primary translation correspondences, making it difficult to cover semantic variations in different contexts, potentially leading to ambiguous understanding. Researchers typically integrate contextual information and other language resources for semantic disambiguation to address this. Second, poor dictionary quality can introduce noise and errors, affecting downstream task performance. Common in current knowledge acquisition research, this issue usually requires researchers to develop robust quality control and error correction mechanisms tailored to specific tasks.

External knowledge bases. External knowledge bases contain vast structured or unstructured information, including entities, concepts, relations, and their multilingual associations. In cross-lingual transfer learning, they make cross-lingual text associations by giving multilingual versions of the data and useful knowledge supplements for languages that don't have a lot of resources, which makes the model work better. Common examples include Wikipedia, which offers cross-lingual knowledge via its multilingual versions and rich entity links; Wikidata, which stores structured cross-lingual entities and factual information; and UMLS, which focuses on multilingual knowledge in the biomedical domain. These knowledge bases provide valuable resources for cross-lingual transfer learning.

Weak supervision is a key mechanism for using external knowledge bases in cross-lingual transfer learning. Researchers primarily use cross-lingual entity links in knowledge bases for the automatic annotation of multilingual entities. Specifically, these links bridge the transfer of entity type information from English Wikipedia articles to other language versions, automatically creating entity type annotations and eliminating the need for extensive manual annotation (Nothman et al., 2013; Richman & Schone, 2008). Cross-lingual parallel corpora or dictionaries created through weak supervision significantly support subsequent model training (Lai & Ji, 2023). For example, extracting text at various levels from multilingual Wikipedia can build phrase-, sentence-, and paragraph-level parallel corpora, allowing models to optimise for cross-lingual tasks at different granularities (Liu et al., 2023a). A new technology trend called knowledge-enhanced pre-training improves models' ability to understand semantics and transfer knowledge across languages by adding information from a knowledge base to language models that have already been trained. For instance, during pre-training, information from other languages related to masked words is used as prompts, improving model generalisation by increasing the quantity and diversity of training data (Chen et al., 2023; Liu et al., 2021a).

However, external knowledge bases also have challenges. Wikipedia data requires processing to become usable knowledge, increasing complexity. Many knowledge bases lack sufficient coverage for low-resource languages, limiting their applicability. Errors, noise, and inconsistencies can also affect model learning. Future research should focus on building more comprehensive, high-quality multilingual knowledge bases; optimising data preprocessing; and improving knowledge supplementation and quality for low-resource languages.

[2] https://rdrpostagger.sourceforge.net/

4.2.2. Derived resources

Derived resources build upon primary resources, comprising computational artifacts and learned systems that automate and scale cross-lingual understanding. Unlike raw primary data, these resources emerge from algorithms, models, and training processes. They operationalize foundational knowledge into practical tools and abstract representations, enabling sophisticated transfer learning applications.

Word alignment. Word alignment tools map words in bilingual or multilingual sentences to identify equivalent terms across languages. This process can be visualised as a bipartite graph, with words as nodes connected by edges representing correspondences. Word alignment primarily uses statistical models or neural networks. FastAlign (Dyer et al., 2013) is an unsupervised tool trained with the Expectation-Maximisation (EM) algorithm. It is computationally efficient, requires no annotated data, and is suitable for quickly aligning large corpora. GIZA++ (Och & Ney, 2000), a widely used open-source statistical machine translation toolkit, offers comprehensive features and supports various models and fusion strategies, although it demands significant computational resources. MUSE (Søgaard et al., 2018) uses existing word alignment information to learn a linear mapping, projecting the target language's word vector space onto the source language's. Awesome-align[3], based on PLMs, computes similarities between word vectors to achieve alignment. It effectively uses contextual information, significantly improving accuracy but with substantial computational cost.

Research by Mulcaire et al. (2019) shows that word vector alignment effectively improves cross-lingual transfer learning. They obtain monolingual word vectors on each language's training set and then use MUSE to map target language word vectors into the English word vector space. This alignment allows models to connect semantically related vocabulary across languages, facilitating knowledge sharing and cross-lingual transfer. Kim et al. (2010) used GIZA++ to perform bidirectional word alignment on parallel corpora and projected source language entity and relation annotations onto target languages based on the alignment information. In cross-lingual dependency parsing, Fei et al. (2023) used Awesome-align to calculate alignment confidence scores between words in source and target language sentences, determining corresponding relationships between dependency tree nodes in different languages. These modified dependency trees significantly aid cross-lingual RE.

Selecting word alignment tools requires careful consideration of specific application scenarios, data resources, and computational constraints. Current methods still struggle with out-of-vocabulary words or those lacking translations in the target language, potentially hindering transfer learning effectiveness. Improvements could involve using contextual information to infer semantic representations of unaligned words or incorporating auxiliary information, like semantic role labelling, to enhance alignment and better utilise source language information. Moreover, cross-lingual processing faces a notable challenge in aligning discontinuous words, where other components separate word groups.

Machine translation. As a prominent example of a derived resource, machine translation systems use algorithms to translate text from a source language to a target language automatically. These systems are typically trained on large-scale foundational resources, most notably parallel corpora, to learn translation patterns. They are also helpful for a key task in cross-lingual learning: generating pseudo-parallel corpora. This process allows researchers to create synthetic training data for low-resource languages, thereby bridging data scarcity gaps. Currently, multiple advanced machine translation systems exist based on different technical architectures and algorithms. Among these, Google Translate[4] has gained widespread adoption due to its extensive language coverage. Moses (Koehn et al., 2007), as an open-source phrase-based statistical machine translation system, enables users to train translation models using their own data, showing outstanding performance in domain-specific translation tasks. In recent years, with the evolution of deep learning technology, neural machine translation models represented by Transformer (such as MarianMT Junczys-Dowmunt et al., 2018) have demonstrated excellent translation performance, particularly showing significant advantages in handling long-distance dependencies and complex syntactic structures. These systems provide a solid technical foundation for cross-lingual transfer learning.

Machine translation systems promote cross-lingual transfer learning through multiple key mechanisms. Among these, data augmentation is a core mechanism, constructing "synthetic" datasets by automatically translating annotated data from resource-rich languages (such as English) to resource-scarce languages (Chen et al., 2023; Ehrmann et al., 2011; Fei et al., 2023; Jain et al., 2019; Liu et al., 2021b). This process effectively increases the amount of training data that is available for target languages. This solves the issue of not having enough data in low-resource language model development. Meanwhile, machine translation plays a crucial role in mapping annotation tools from source language to target language (Faruqui & Kumar, 2015; Rahman & Ng, 2012). Specific implementation paths include translating target text to source language, applying annotation tools to the translated text, and then projecting annotations back to the original target language through alignment information. Furthermore, machine translation systems can serve as alternatives to dictionaries and word alignment tools (McNamee et al., 2011), providing valuable cross-lingual alignment information for transfer learning.

Although machine translation provides powerful support for cross-lingual transfer learning, existing machine translation systems still have some deficiencies that limit their effectiveness in this field. The most significant deficiency is that machine translation quality remains imperfect, and translation errors propagate during the transfer learning process, leading to performance degradation. Particularly for some specialised domains or target languages that differ significantly from the source, machine translation accuracy is often difficult to guarantee. Additionally, cross-lingual transfer learning methods overly rely on existing machine translation systems, using them as black-box tools while ignoring the internal information and mechanisms of machine translation systems. This results in poor model interpretability and also limits the space for further improvements.

Multilingual embeddings. Embedding maps words, phrases, and sentences from natural language into low-dimensional dense vector spaces, where linguistically similar units have similar vector representations. The concept of "word embedding" originated with Bengio et al. (2003), who pioneered the use of neural networks to simultaneously train language models and word vectors. Subsequently, word2vec (Mikolov et al., 2013a) captured word semantics through contextual learning, and GloVe (Pennington et al., 2014) expanded on word2vec by integrating global statistical information with context-based learning. These traditional methods create static representations by learning word contexts, which limit their adaptability to new contexts. CoVe (McCann et al., 2017) introduced methods for generating dynamic word representations based on different contextual content. ELMo (Peters et al., 2018a) then used a pre-trained deep bidirectional language model to obtain context-dependent word representations. Transformer-based methods, such as BERT (Devlin et al., 2019) and GPT (Brown et al., 2020), represent the latest advancements in dynamic representation.

Multilingual embedding extends this concept by mapping linguistic units from different languages into a shared vector space. This cross-lingual alignment ensures that semantically similar words have similar vector representations regardless of their language. For instance, the Malay word $Perancis$ corresponds to $France$ in English, and a good cross-lingual word embedding model should ensure $\vec{Perancis} \approx \vec{France}$. The

[3] https://github.com/neulab/awesome-align
[4] https://translate.google.com/

development of multilingual embeddings provides crucial technical support for cross-lingual transfer learning. In multilingual model architectures, different languages share parameters of key network components (such as word embedding layers and encoders). This allows models to learn universal language representation patterns from high-resource languages and transfer this knowledge to other languages through parameter sharing. Early research bridging language gaps often used statistical methods to construct multilingual representations. For example, the CCA-based cross-lingual model by Ammar et al. (2016) projects vocabulary from different languages into a shared space by identifying linear transformations that maximise correlations between word vectors. Another type of bilingual word embedding uses small seed dictionaries and Procrustes analysis to learn transformation matrices between two languages (Mikolov et al., 2013b). However, the limitations of linear transformations restrict the model's ability to capture complex non-linear relationships between languages. The rise of deep learning has led to significant breakthroughs in neural network-based multilingual representation learning. These methods leverage the powerful feature extraction capabilities of neural networks to automatically learn semantically rich, context-aware, cross-lingual representations from large monolingual or multilingual corpora. A common approach is to train shared encoders on multilingual annotated corpora, minimising alignment loss between embedding representations of different languages. For instance, Convolutional Neural Networks (CNNs) use convolutional kernels to slide over input text, effectively capturing local semantic features (Lin et al., 2017). Graph Convolutional Networks (GCNs) represent words or entities from different languages as graph structures, learning node representations that integrate structured information (Subburathinam et al., 2019). Long Short-Term Memory networks (LSTMs) effectively model long-distance dependencies in text. In multilingual embedding, LSTMs can capture sequential information and contextual dependencies of words in sentences, crucial for understanding sentence semantics (Rijhwani et al., 2019). The emergence of PLMs marks a new paradigm shift in NLP, further discussed in Paragraph 4.2.2.

The performance of multilingual embeddings directly affects the effectiveness of downstream cross-lingual tasks. Inaccurate embedding alignments or insufficient modelling of complex semantic relationships can significantly limit cross-lingual transfer. Current multilingual embedding models still show limitations in handling domain differences between source and target languages, constraining their generalisation ability in specific domains. Additionally, creating high-quality embeddings for low-resource languages remains a challenge when parallel corpora are scarce. Future research may focus on improving embedded quality, integrating multimodal information, and optimising low-resource languages to overcome current technical limitations.

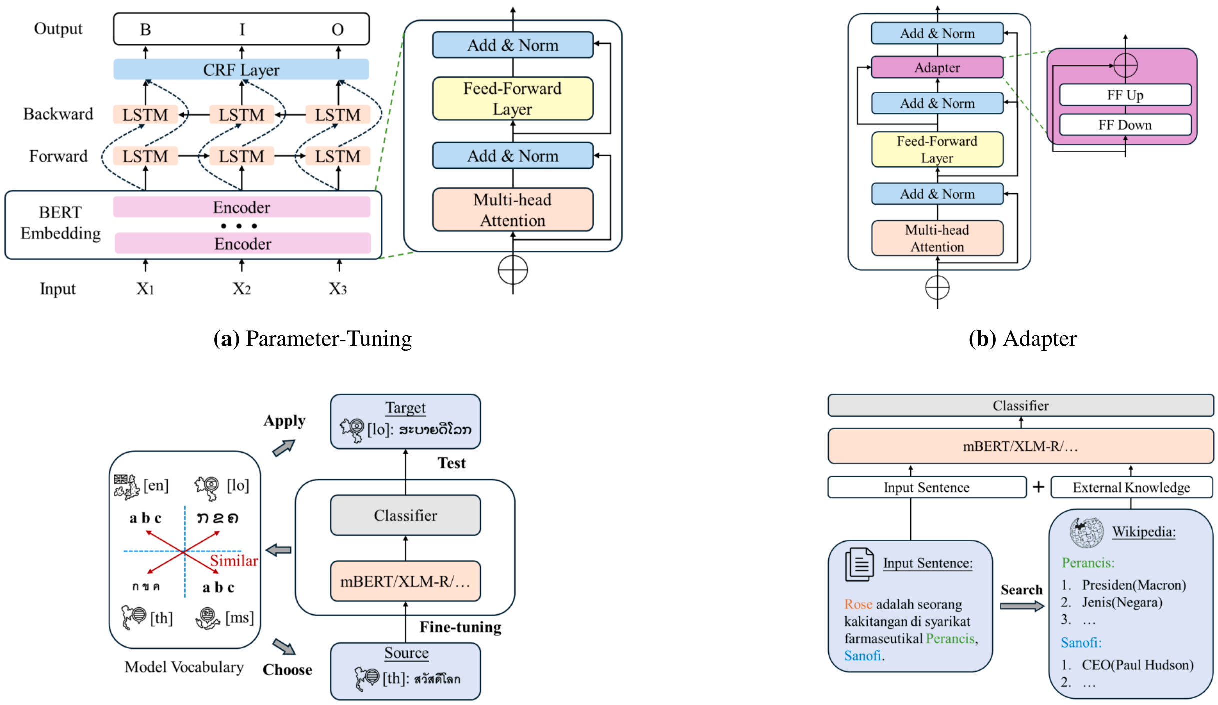
Pre-trained models. PLMs are a significant advancement in NLP. These models, with hundreds of millions of parameters, are trained using self-supervised learning on vast corpora and exhibit remarkable abilities in understanding and generating human language. Their importance extends to cross-lingual transfer learning. A key challenge here is bridging linguistic structural, vocabulary, and training data gaps between languages. PLMs address this by learning universal language representations that capture shared cross-lingual semantic and syntactic information. This enables knowledge transfer from high-resource languages like English to languages with limited annotated data. Multilingual BERT (mBERT) (Devlin et al., 2019), a highly influential model, is trained on concatenated Wikipedia data from 104 languages. Despite lacking explicit cross-lingual training objectives, mBERT demonstrates extraordinary zero-shot cross-lingual transfer capabilities. It can be fine-tuned on a downstream task in one language and applied directly to another without further training. This success is due to shared vocabulary and the model's ability to learn language-agnostic representations during pre-training. Architecturally, mBERT follows the original BERT model, using only the Transformer's encoder (Vaswani et al., 2017). This means BERT focuses on capturing relationships between words rather than text generation. XLM-R (Conneau et al., 2019), with a similar architecture but improved training objectives and larger scale, outperforms mBERT in some cross-lingual tasks. Another common model, mT5 (Xue et al., 2021), employs an encoder-decoder architecture, enabling both understanding and generating text. This is crucial for research that views knowledge acquisition as a sequence-to-sequence task. For fine-tuning on downstream tasks like knowledge acquisition, mBERT and XLM-R use similar procedures. They add a linear classification layer or a conditional random field (CRF) layer above the PLM's output layer to predict labels for each token and then fine-tune the model on annotated data using optimisers like AdamW (Ahmad et al., 2021; Tan et al., 2023). mT5 requires converting knowledge acquisition tasks into text generation (Bohnet et al., 2023).

To further enhance these multilingual models' performance on specific downstream tasks, researchers have developed various strategies. As depicted in Fig. 5, these strategies primarily fall into four categories: parameter tuning, adapters, language selection, and enriched context.

Parameter tuning, as shown in Fig. 5a, is an optimisation method for PLMs that systematically adjusts internal weights and parameters to better suit specific downstream tasks (Lange et al., 2020; Liu et al., 2021b). This process helps models accurately record grammatical structures, lexical features, and semantic connections across languages while keeping multilingual representations consistent. Parameter tuning involves two main stages: pre-training and fine-tuning.

Pre-training occurs on large unannotated corpora to learn universal language representations (Ogueji et al., 2021). Model development teams typically handle this stage and release the pre-trained weights. For cross-lingual models, the pre-training often uses mixed multilingual corpora. Researchers may re-execute pre-training if the model's performance is insufficient or if encoder architecture adjustments are necessary. Fine-tuning then involves further training the PLM on annotated datasets for specific tasks (Zhou et al., 2023). During this stage, researchers can choose to update all or some of the model's parameters for optimal task adaptation. Fig. 5a illustrates mBERT and BiLSTM-CRF models used for knowledge acquisition (from Lange et al., 2020). The mBERT layer, a PLM, extracts context-dependent representations of the input text. The BiLSTM network further captures long-distance dependencies in sequences. The CRF considers dependencies between labels to compute the optimal label sequence. In downstream tasks, the pre-trained BERT model is loaded as part of the entire model. During training, mBERT's parameters can be fine-tuned or fixed, depending on computational resources and dataset size. BiLSTM and CRF layer parameters require training from scratch. But fine-tuning on certain tasks can cause catastrophic forgetting (Luo et al., 2023), where the model loses the general language knowledge it learnt during pre-training. This makes it less good at generalisation in other tasks or languages. Keeping separate fine-tuned models for each downstream task and target language also takes up a lot of space, especially in programs that support more than one language and task.

Adapters are modular components designed to reduce the computational complexity of fine-tuning PLMs. As shown in Fig. 5b, adapter modules insert down-sampling and up-sampling layers between Transformer encoder layers. This design allows for local optimization at each layer without adjusting the overall model parameters. By optimizing only the adapter parameters during fine-tuning, while keeping the original PLM parameters unchanged, the computational resources and time needed for parameter updates are significantly reduced. This enables rapid model adaptation to new tasks or languages (Liu et al., 2023a). For example, applying the mBERT directly to low-resource language tasks often yields suboptimal performance. Introducing language-specific adapters can optimize model performance in such cases. By updating only the language adapter parameters while fixing mBERT's base parameters during training, efficient adaptation to low-resource language tasks is achievable (Ansell et al., 2021; Pfeiffer et al., 2020). This method preserves the model's basic language understanding capabilities while significantly improving processing effectiveness for specific



(a) Parameter-Tuning

(b) Adapter

(c) Language selection

(d) Enriched context

Fig. 5. The Method of Knowledge Acquisition utilising Multilingual Pre-trained Models.

low-resource languages. While adapters perform well on many tasks, their performance is generally slightly inferior to full parameter fine-tuning. Future research should explore more powerful adapter architectures, such as dynamic, composite, and transferable adapters. Additionally, the architectural design of adapter modules (e.g., insertion position, internal structure, dimension size) significantly impacts performance and requires careful design and tuning. Future research could explore automatically learning adapter structures and parameters, perhaps using neural architecture search technology.

Language selection is a strategy that uses annotated data from languages related to the target language to aid model training. Its central concept is leveraging similarities and transferability between languages. When there isn't enough or any annotated data for the target language (the few-shot or zero-shot), one or more high-resource languages that are related to the target language can be chosen to fine-tune the model. The fine-tuned model can then be applied to the target language (Eronen et al., 2023; Nigatu et al., 2023). This method effectively uses knowledge from high-resource languages to improve performance in low-resource languages. As shown in Fig. 5c, Malay and English share similarities due to cultural exchange and historical borrowing, making English a strong source language for Malay. However, Lao differs significantly from English in grammar, syntax, and writing systems, resulting in poor transfer performance. In such cases, Thai might be a more suitable source language due to its greater orthographic similarity with Lao. Determining the optimal source language, or combination of sources, for a given target language remains an open question. Current methods for measuring language similarity primarily rely on lexical or grammatical features and may not capture similarities at the semantic level.

Enriched context involves providing PLMs with more comprehensive information beyond the input text's literal content. This helps models better understand input semantics, especially for low-resource languages or those with unique cultural and linguistic features, leading to more accurate judgments. As shown in Fig. 5d, this approach's core idea is to build an external knowledge base or document collection. When the model reads input, it pulls relevant knowledge or information from the knowledge base based on what it reads and adds it as extra context (Elkaref et al., 2023; Tan et al., 2023). Different strategies for integrating knowledge into the model can affect how well it works, so it's important to think about how to do it in a way that doesn't introduce noise or interference. The knowledge base requires continuous updates to maintain its timeliness and accuracy, posing a maintenance challenge.

Large language models. The latest generation of Large Language Models (LLMs), including GPT-3 (Brown et al., 2020), GPT-4 (Achiam et al., 2023), PaLM (Chowdhery et al., 2024), and LLaMA (Touvron et al., 2023), The latest generation of LLMs, like GPT-3 (Brown et al., 2020), GPT-4 (Achiam et al., 2023), PaLM (Chowdhery et al., 2024), and LLaMA (Touvron et al., 2023), demonstrate significant scaling effects and are applicable to diverse scenarios like dialogue systems Wu et al. (2025), medical diagnosis Liu et al. (2025), and Internet of Things Yang et al. (2025) security under fine-tuning, few-shot, and even zero-shot settings. Similar to PLMs, LLMs learn more abstract and essential knowledge representations by mapping cross-lingual synonymous concept vocabularies to proximate semantic spaces. However, through training on large-scale multilingual corpora, LLMs are able to capture logical relationships from a multilingual perspective and deepen their understanding of semantics and context, thereby enhancing their logical reasoning and contextual understanding abilities. The parameter scale of LLMs has significantly increased from the hundred-million level of traditional pre-trained models to the hundred-billion level, consequently triggering a series of important emergent abilities (Zhao et al., 2023), including but not limited to In-context Learning, Instruction Following, and Step-by-step Reasoning. The development of LLMs provides unprecedented opportunities for addressing the numerous challenges in cross-lingual knowledge graph transfer learning. Given the excellent language understanding and generation abilities that LLMs have acquired through large-scale multilingual pre-training, and their demonstrated strong zero-shot and few-shot cross-lingual transfer capabilities, we have reason to believe that LLMs will effectively improve the efficiency of cross-lingual transfer learning and facilitate the direct or rapid transfer of knowledge between different languages.

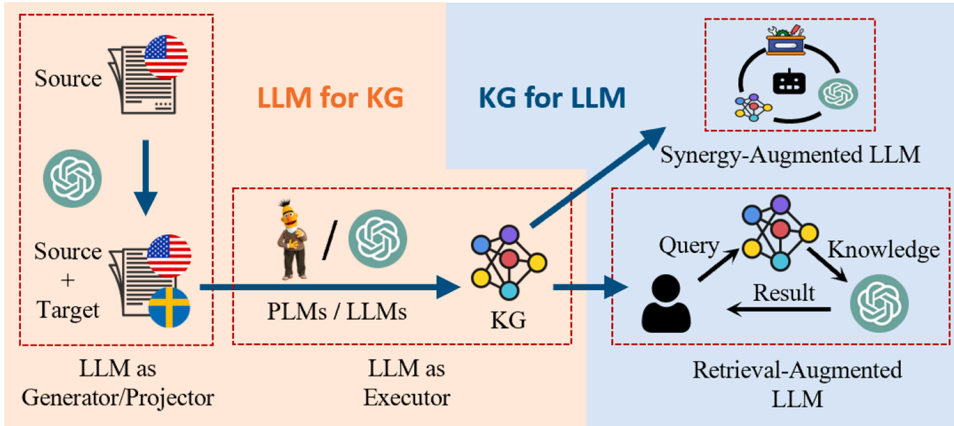


Fig. 6. KG and LLM Integration.

LLM for KG. In practical applications, LLMs currently enhance knowledge acquisition primarily through two approaches, as illustrated in Fig. 6. First, LLMs can directly perform knowledge acquisition tasks, typically by transforming diverse tasks into a generative format Xu et al. (2024). The advantage of LLMs lies in their ability to enhance or unlock specific capabilities, such as instruction following Zhao et al. (2023), by formatting task descriptions and input-output examples into flexible natural language instructions. Furthermore, models can leverage the vast corpus of knowledge accumulated during pre-training to exhibit exceptional generalization capabilities in zero-shot or few-shot learning scenarios, achieving performance comparable to that of smaller, specifically fine-tuned models. mRAKL (Nigatu et al., 2025) reformulates knowledge graph construction as a question-answering task, employing vector-based retrieval to extract bilingual contexts from Wikipedia for completing knowledge graphs in low-resource languages. M-NTA (Conia et al., 2023) generates more accurate and comprehensive non-English knowledge graph facts than single-source methods by synthesizing and cross-validating information from multiple noisy sources (machine translation, web search, and LLMs). On multilingual entity linking, M-NTA improved the baseline F1 score from 62.2 to 63.3. LS-CLDARE (Wang et al., 2025) combines small and large models for cross-lingual data augmentation: the small model identifies entities in low-resource texts, which the large model uses to generate high-quality cross-lingual data under constraints. However, LS-CLDARE's reliance on manual filtering prevents full automation and limits scalability. However, due to limitations in model size and computational resources, LLMs are less amenable to the fine-grained and in-depth adjustments possible with smaller models. Indeed, a single LLM often exhibits lower performance on specific tasks, particularly in specialized domains requiring fine-grained semantic understanding and reasoning (Lai et al., 2023; Tan et al., 2023). Second, LLMs can achieve dynamic projection mechanisms through enhanced contextual understanding. CLaP (Parekh et al., 2024) utilises instruction fine-tuning to adapt source language labels to target sentences, generating pseudo-labeled data that is not only highly accurate but also consistent and coherent with the target language sentences. Such research integrates the general knowledge representation capabilities of LLMs with the domain specificity and fine-tuning advantages of smaller, task-specific models. By generating high-quality pseudo-data, the performance of SOTA models can be further improved.

KG for LLM. To address the limitations of LLMs regarding knowledge accuracy, particularly in cross-lingual contexts, knowledge graph based retrieval-augmented generation (RAG) (Edge et al., 2025; Sharma et al., 2024) has emerged as an effective solution. The core principle is to use a knowledge graph as an external knowledge base for the LLM. Before the LLM generates text, relevant knowledge is retrieved from the knowledge graph and incorporated into the input prompt as contextual information, thereby guiding the LLM towards more accurate and reliable outputs. This approach mirrors the human practice of consulting references and acquiring background knowledge when tackling complex tasks, effectively compensating for the inherent knowledge limitations of LLMs and enhancing their ability to handle knowledge-intensive cross-lingual tasks. KG-MT (Conia et al., 2024) implements a RAG system for translation that dynamically retrieves entity translations from multilingual KG and integrates them into neural machine translation models end-to-end, achieving performance gains of 17.9 % over NLLB-200 and 25.3 % over GPT-4. However, when encountering rare, emerging, or absent entities, the system defaults to the base model's parametric knowledge. Building upon the retrieval-augmented paradigm, some research explores collaborative enhancement patterns between knowledge graphs and LLMs to further exploit the potential of knowledge graphs and achieve a deeper level of cross-lingual capability enhancement (Jiang et al., 2024a). Compared to simple retrieval augmentation, collaborative enhancement aims to facilitate deeper interaction between LLMs and knowledge graphs, completing complex tasks through multiple rounds of interaction. In this mode, the LLM can be viewed as an agent, capable of formulating action plans based on the current task objectives and utilising an external toolbox (e.g., web search, code execution, KG retrieval, and KG updates) to execute the plan. This collaborative approach empowers LLMs with stronger reasoning capabilities and more flexible interaction methods, enabling them to handle more complex cross-lingual tasks. MDKG-RAG (Zhang et al., 2025) employs an LLM-driven pipeline to construct domain-specific, multi-label, multilingual knowledge graphs from raw documents, which subsequently guide retrieval and generation. The retrieved context is processed by multiple collaborating LLM agents using multilingual prompts to produce accurate responses. This approach suits domains like ship design, where knowledge spans multiple languages and regions. Beyond natural language, knowledge graphs facilitate cross-programming language support for LLMs. CodeRCSG (Jiang et al., 2024b) generates semantic graphs by pairing code snippets with their cross-language counterparts, then retrieves both code text and semantic graphs as references during code generation.

LLM Evaluation. The cross-lingual capabilities of LLMs are typically evaluated along two dimensions: intrinsic and extrinsic evaluation. Intrinsic evaluation aims to directly measure the cross-lingual capabilities of LLMs, focusing on the quality of the model's language transfer process. This evaluation approach generally does not rely on specific downstream tasks but instead examines the quality of the model's translation output or language understanding. Notably, a tokenizer that supports multilingual tokens is crucial for enhancing the performance of LLM. The "fertility" of a tokenizer (Rust et al., 2021), defined as the average number of subtokens generated per tokenized word, is an important measure of its quality. A lower fertility score generally indicates better tokenizer quality. The Parity metric is used to measure the equality of tokenization processing across different languages (Petrov et al., 2023). For example, a Japanese Kanji character might be split into three tokens in GPT-2. Furthermore, accuracy and faithfulness are key metrics that directly reflect an LLM's cross-lingual understanding and its ability to generate multilingual pseudo-data (Parekh et al., 2024). Accuracy primarily concerns the quality of the model's translated labels, while faithfulness focuses on whether the translated labels remain semantically relevant to the translated sentences. Even if the model can accurately translate labels, its cross-lingual capabilities are still deficient if the translated labels are inconsistent with the content of the translated sentences. In addition to these metrics, ethical considerations (Agarwal et al., 2024), toxicity (de Wynter et al., 2024), and bias (Kaneko et al., 2022) are equally important, assessing whether the LLM adheres to the same human values across different languages (Zhu et al., 2024a).

Extrinsic evaluation, on the other hand, focuses on applying LLMs to practical downstream cross-lingual knowledge extraction tasks and evaluating their cross-lingual capabilities based on their performance in these tasks (Chen et al., 2024b; Parekh et al., 2024). This approach is closer to real-world application scenarios and can realistically reflect the model's ability to solve practical problems. Commonly used evaluation metrics include F1-score, precision, and recall. By comparing the performance of LLMs on tasks across different languages and domains, their extrinsic cross-lingual capabilities can be comprehensively assessed.

Larger parameter models generally demonstrate superior downstream task performance. For instance, GPT-4 achieved approximately 8 % better translation performance than GPT-3.5 in KG-MT (Conia et al., 2024) tests. In cross-lingual code generation, CodeRCSG (Jiang et al., 2024b) found GPT2-s underperformed GPT2-m by approximately 2.5 % on BLEU scores. However, increased parameters incur substantial computational costs. MDKG-RAG (Zhang et al., 2025) achieved competitive performance using Qwen2.5-7B, reducing inference time by over 50 % with minimal performance loss (2 %) compared to DeepSeek-R1-671b.

Model specialization often outweighs parameter count, particularly in low-resource domains. KG-MT (Conia et al., 2024) tests revealed mT5's poor Tigrinya performance, attributable to the language's absence from pre-training data. This underscores the importance of pre-training language coverage. While large models possess broader knowledge bases, task-specific success depends on selecting models with appropriate language coverage and fine-tuning suitability.

Well-designed frameworks can compensate for model limitations. Using GPT-3.5, M-NTA (Conia et al., 2023) surpassed single SOTA models including GPT-4. CodeRCSG (Jiang et al., 2024b) improved all models' BLEU scores by over 40 %, dwarfing inter-model performance differences. MKG-Rank (Li et al., 2025) demonstrated that graph-based RAG systems enhance most LLMs regardless of size or architecture, with improvements correlating to models' target language proficiency. However, for models already proficient in target languages (e.g., Qwen-2.5 72B for Chinese), additional frameworks may introduce interference.

In summary, generative LLMs are still in the early stages of exploration in the field of cross-lingual knowledge acquisition and knowledge graphs, with significant opportunities for improvement, especially in general-purpose, open-domain, and low-resource settings. Existing benchmark models are often tailored to specific domains or tasks, which, while potentially achieving higher performance on those specific tasks, limits their generality and applicability. LLM selection requires balancing parameter scale, model specialization including language coverage, and potential framework-induced performance gains, rather than defaulting to maximum model size. Furthermore, some generative tasks using LLMs also have limitations. For example, prompt design and compatibility issues with inputs of varying lengths and structures can lead to serious hallucination problems in other domains. Therefore, further development of general frameworks that can flexibly adapt to different domains and tasks is a promising research direction. Particularly noteworthy is how to collaboratively enhance LLM agents with KGs to provide suitable examples for different knowledge extraction scenarios, even if the KG itself is monolingual, enabling multilingual question answering through the LLM. Additionally, for tasks with predefined entity and relation types, such as NER and RE, the usability of LLMs may be questionable when the required granularity of the task does not match the granularity of the LLM's internal knowledge, especially given the inherent differences in knowledge granularity across languages. Therefore, future work could extend cross-lingual knowledge acquisition tasks to the open information extraction domain, guiding LLMs through prompting or fine-tuning to extract elements of undefined types at a specific granularity, which will place higher demands on the LLM's language understanding capabilities. Finally, generative knowledge extraction systems using LLMs still face challenges in few-shot and zero-shot scenarios, requiring further exploration of the in-context learning and step-by-step reasoning capabilities of multilingual LLMs to improve their performance, particularly in low-resource settings.

4.3. Task-specific adaptation

This section focuses on research progress in cross-language transfer learning for KG construction tasks. The article will thoroughly analyse representative models and methods applied in various subtasks, with particular emphasis on how these models and methods ingeniously utilise different transfer learning paradigms and available cross-language resources, aiming to provide a systematic overview and valuable reference for research in this field.

4.3.1. Named entity recognition

Monolingual NER is a foundational task in KG construction. Its primary goal is to accurately identify entity boundaries within unstructured text in a specific linguistic context. The accuracy of an NER system significantly affects the performance of downstream tasks and the quality of KG construction. Monolingual systems capitalise on the unique linguistic features, syntactic structures, and contextual nuances of the target language to enable high-precision entity recognition.

Definition: Given a sequence of tokens $C = w_1, w_2, \ldots, w_n$ derived from a text, the task of monolingual NER is to identify the named entities mentioned, denoted as $e_i = (I_1, I_2, t^e)$. Here, I_1 and I_2 indicate the position of the entity e_i within the sequence C, representing the start and end indices, respectively. The term t^e refers to a predefined set of entity categories.

There are various approaches to monolingual NER. Early rule-based systems depend on clearly defined linguistic patterns (Kim & Woodland, 2000) and specialised vocabularies (Hanisch et al., 2005). Expert-crafted rule sets identify and categorise entities when text conforms to them. Although efficient in specific, narrow domains, such systems frequently experience low recall and restricted adaptability to novel contexts. Unsupervised methods bypass the need for labelled data by leveraging inherent linguistic patterns. These methods may cluster words based on distributional semantics (Etzioni et al., 2005; Liu et al., 2019a) or use language models to estimate the likelihood of a word being a named entity based on its context (Veena et al., 2023), identifying entities through patterns in usage. A more widely used class of techniques is supervised machine learning, which frames NER as a classification or sequence labelling problem. Traditional models such as Hidden Markov Models (HMMs) (Bikel et al., 1998) predict entity labels by analysing a limited window of preceding words, whereas CRFs (McCallum & Li, 2003) consider the entire sequence for more accurate predictions. Other supervised approaches, such as Maximum Entropy (Grishman & Borthwick, 1999) and Conditional Markov Models (Downey et al., 2007), provide alternative frameworks for modelling relationships between words and their entity labels. However, the performance of these methods often depends on the quality of hand-engineered features. To overcome this limitation, supervised deep learning methods have gained prominence, enabling models to automatically learn complex feature representations from data. CNNs (Collobert et al., 2011) extract local features, while Iterated Dilated Convolutional Neural Networks (ID-CNNs)

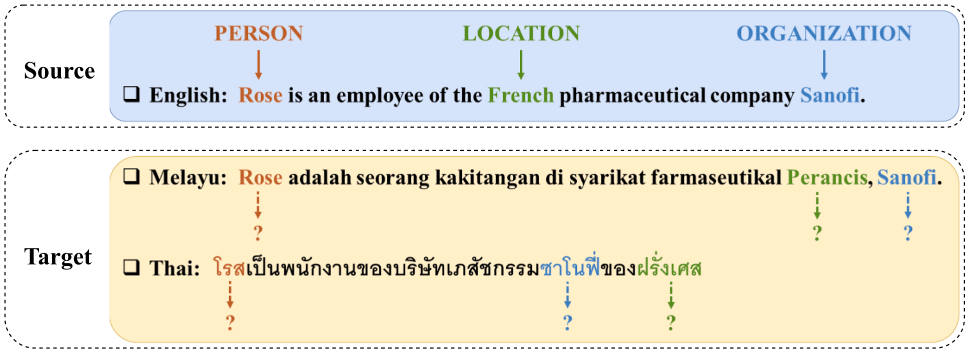


Fig. 7. The cross-lingual NER process.

(Strubell et al., 2017) expand the receptive field to capture broader contextual information. A popular approach, the BiLSTM-CRF (Huang et al., 2015), integrates both past and future context for more accurate predictions. Deep learning methods have been further enhanced by dynamic word embeddings like ELMo, which create contextualised word representations based on the surrounding text, capturing nuanced meanings (Peters et al., 2018b). Advancing even further, BERT-based models leverage pre-trained language representations to produce highly contextualised embeddings, achieving state-of-the-art results in NER tasks (Liu et al., 2023b). These models excel at capturing complex relationships within text and are often fine-tuned for NER or employed in a reading comprehension framework to directly predict entity spans.

Cross-lingual NER extends the traditional task by aiming to create a unified framework for entity recognition across multiple languages. Non-English systems, however, face significant challenges, primarily due to limited training data. Researchers often use cross-lingual transfer learning to solve this problem. In this method, a model that was trained on labelled data in the source language is adapted so that it can perform well on test sets in the target language. The main challenge lies in establishing semantic bridges between languages to enable effective knowledge transfer while preserving the unique linguistic features of each language.

Definition: Cross-lingual NER considers a sentence $C = w_1, w_2, \dots, w_n$ in a target language ℓ_t, along with an annotated sentence in a source language ℓ_s. The objective is to identify the position of each named entity mention $e_{j,k} = \{w_j, w_{j+1}, \dots, w_k\}$ within the sentence and accurately categorize them into predefined entity types y. Under a few-shot setting, label acquisition in ℓ_s is limited; in a zero-shot setting, labels can only be obtained from ℓ_s, typically English.

Example: Fig. 7 illustrates a toy example, aiming to identify and correctly classify entities in a Malay or Thai sentence using English data.

Label transfer considers mappings between words of different languages; if the English word “French” is labeled as “Location”, and its Malay equivalent is known to be “Perancis” and the Thai equivalent is ฝรั่งเศส , then both “Perancis” and ฝรั่งเศส are also considered as “Location”, similarly for other entity words. Parameter transfer establishes a shared vector space for words of different languages that supports the transfer of model parameters. If English word vectors $[V_1, V_2, V_3] = [\vec{Rose}, \vec{France}, \vec{Sanofi}]$ and their Malay equivalents $[V_1', V_2', V_3'] = [\vec{Rose}, \vec{Perancis}, \vec{Sanofi}]$ (with their Thai equivalents being โรส, ฝรั่งเศส, and ซาโนฟี respectively) are known to approximate $V_1 \simeq V_1', V_2 \simeq V_2', V_3 \simeq V_3'$, then the NER model trained on English can be applied to Malay or Thai to some extent, since the vector spaces show no significant linguistic discrepancies. Feature transfer constructs the same features for both languages, and feature-driven machine learning methods can be transferred to other languages. If “Rose” in both English and Malay appears at the beginning of sentences and shares string similarities, then from positional and linguistic features, “Rose” can be categorized as “Person” in both languages.

Table 7[5,6] synthesises representative cross-lingual NER research, outlining their paradigms, resources, supported languages, and datasets.

(1) Machine learning: Early cross-lingual NER research primarily used traditional machine learning. Klementiev and Roth proposed a weakly supervised learning method based on discriminative transliteration models (Klementiev & Roth, 2006). As seeds, they used a small group of known transliteration pairs along with weak temporal alignment data from multilingual corpora. This method determined if a Russian mention was an English entity’s transliteration to facilitate label transfer. However, accuracy was limited (Russian: 65 %), and it was highly sensitive to corpus temporal alignment quality. Similarly, Szarvas et al. (2006) used classical ensemble learning algorithms-AdaBoostM1 and C4.5 decision trees-for business news. They used boosting to combine decision tree models, which made them more accurate and reliable by adding general and language-specific features through feature engineering. Later, with resources like Wikipedia, researchers used these to enhance training. Richman and Schone (2008) developed an HMM-based system that redirected cross-lingual NER decisions to English when possible. It generated classification information using English Wikipedia’s resources and structure and then returned results to non-English mentions via Wikipedia’s cross-language links. To better utilise Wikipedia’s structure, the C&C tagger (Nothman et al., 2013), a maximum entropy model-based NER annotator, also generated training data through article classification and link annotation. The maximum entropy model, a flexible statistical model, incorporated multiple features, improving NER performance. However, such external resources may not be available in other domains, and cross-language link coverage with English varies, causing performance disparities across domains and languages. Regarding existing NER models, Ehrmann et al. (2011) obtained a list of all possible translations for named entities and then determining if a target language word is present in this list based on string matching. For words that do not match, consonant signature matching is performed. The NER annotations are then transferred to the target language for successfully matched entities. The effectiveness of this approach is dependent on both the quality of the initial NER system and the accuracy of the matching process.

(2) Deep learning: Neural network models have gained significant traction in cross-lingual NER research due to their robust representation learning capabilities. Xie et al. (2018) proposed a model based on BiLSTM-CRF. Instead of directly using output embeddings from other models, they employed bilingual word embedding spaces to translate source language training data into the target language before training. They also implemented self-attention mechanisms to address word order differences. Inspired by Xie et al. (2018) and label transfer, Jain et al. (2019) proposed projecting annotation results to target lan-

[5] http://www.appenbutlerhill.com/.
[6] http://au-kbc.org/nlp/NER-FIRE2013/.

Table 7
Multilingual NER Papers.

Work	Paradigms	Resources	Languages\Performance	Databases
Klementiev and Roth (2006)	Label	- Parallel corpora - Universal feature - Linguistic feature - Dictionary	A :en(−),ru(.65)	Self-generated
Szarvas et al. (2006)	Feature	- Universal feature	F :en(.89),hu(.95)	CONLL-2003 (Sang & De Meulder, 2003) Szeged Treebank (Csendes et al., 2004)
Richman and Schone (2008)	Label	- Wikipedia - Dictionary	F :en(−),es(.84),fr(.85),pl(.86), pt(.80), ru(.80),uk(.78)	ACE-2007 (Chen et al., 2014) Self-generated
Ehrmann et al. (2011)	Label	- Parallel corpora - Machine translation - Dictionary - Linguistic feature	R :cs(.92),de(.95),en(−),es(.94), fr(.95),ru(.87)	Self-generated
Nothman et al. (2013)	Label	- Wikipedia	F :de(−),en(−),es(.90),fr(−), it(−),nl(−),pl(−),pt(−),ru(.88)	CoNLL-2002 (Sang, 2002) CoNLL-2003 (Sang & De Meulder, 2003) ABH BBN (Ralph Weischedel, 2005) Europarl (Faruqui et al., 2010) Wikigold (Balasuriya et al., 2009)
Fang et al. (2017)	Parameter	- Multilingual embedding	F :de(.63),en(−),es(.56),nl(.56)	CoNLL-2002 (Sang, 2002) CoNLL-2003 (Sang & De Meulder, 2003)
Xie et al. (2018)	Label	- Multilingual embedding	F :de(.58),en(−),es(.72),nl(.70), ug(.32)	CoNLL-2002 (Sang, 2002) CoNLL-2003 (Sang & De Meulder, 2003) DARPA LORELEI (Mayhew et al., 2017)
Jain et al. (2019)	Label	- Parallel corpora - Machine translation - Linguistic feature - Universal feature - Word alignments	F :de(.62),en(−),es(.74),hi(.42),hy(.62), nl(.70),ta(.34),zh(.50)	CoNLL 2002 (Sang, 2002) CoNLL 2003 (Sang & De Meulder, 2003) FIRE 2013 OntoNotes 4.0 (Weischedel et al., 2011) Self-generated
Mulcaire et al. (2019)	Parameter	- Multilingual embedding - Word alignments	F :ar(.78),en(−),zh(.76)	CoNLL-2012 (Pradhan et al., 2012)
Pfeiffer et al. (2020)	Parameter	- Adapter	$\widetilde{F}$:ar(.59),cdo(.27),en(.56),gn(.26), ilo(.31),is(.53),ja(.53),jv(.40), mhr(.28),mi(.24),my(.25),qu(.34), sw(.51),tk(.31),xmf(.33),zh(.56)	WikiANN (Pan et al., 2017)
Liu et al. (2021b)	Parameter	- Machine translation - Parameter tuning	F :ar(.54),en(.79),fi(.75),it(.76), ja(.37),tr(.75),zh(.41)	CoNLL-2002 (Sang, 2002) CoNLL-2003 (Sang & De Meulder, 2003) WikiANN (Pan et al., 2017)
Ogueji et al. (2021)	Parameter	- Parameter tuning	F :amh(.74),hau(.90),ibo(.87),kin(.74), lug(.79),luo(.70),pcm(.86),swa(.88), wol(.62),yor(.81)	MasakhaNER (Adelani et al., 2021)
Ansell et al. (2021)	Parameter	- Adapter	F :en(−),hau(.77),ibo(.70),kin(.66), lug(.54),luo(.33),pcm(.73),swa(.73), wol(.32),yor(.69)	MasakhaNER (Adelani et al., 2021)
(Zhou et al., 2023)	Parameter	- Parameter tuning	F :ar(.72),de(.76),en(−),es(.85), hi(.83),nl(.84),zh(.61)	CoNLL-2002 (Sang, 2002) CoNLL-2003 (Sang & De Meulder, 2003) WikiANN (Pan et al., 2017)
Tan et al. (2023)	Parameter	- Enriched context	F :bn(.95),de(.90),en(.91),es(.94), fa(.92),fr(.93),hi(.95),it(.95), pt(.93),sv(.95),uk(.92),zh(.90)	MultiCoNER v2 (Fetahu et al., 2023a)
Gui and Xiao (2024)	Parameter	- Parameter tuning	F :ar(.55),de(.76),en(.84),es(.75), fr(.79),ru(.66),vi(.68)	PANX (Pan et al., 2017)
Chen et al. (2024b)	Label Parameter	- LLM as Executor - Parameter tuning	F :bam(.60),bbj(.52),ewe(.72),fon(.54), hau(.72),ibo(.38),kin(.56),lug(.68) luo(.59),mos(.49),nya(.76),pcm(.76) sna(.51),swh(.73),tsn(.71),twi(.76) wol(.59),xho(.63),yor(.52),zul(.44)	MasakhaNER2 (Adelani et al., 2022)
Parekh et al. (2024)	Label Parameter	- LLM as Projector - Parameter tuning	F :af(.74),ar(.49),bg(.81),el(.76), es(.75),eu(.69),fa(.59),fr(.79) id(.56),it(.80),ja(.45),jv(.65) ka(.71),ko(.60),ms(.74),my(.62) ru(.68),th(.13),vi(.76),zh(.50)	WikiANN (Pan et al., 2017)
Wang et al. (2024b)	Parameter	- Parameter tuning	F :de(.64),en(.71),es(.72),fr(.65)	Self-generated
Golde et al. (2024)	Parameter	- Enriched context	F :ar(.15),en(.35),zh(.27)	Self-generated
Kumar et al. (2025)	Parameter	- Parameter tuning	F :ar(.65),en(.98)	Self-generated
Bouabdallaoui et al. (2025)	Parameter	- Enriched context	F :de(.65),es(.67),fr(.67),it(.67)	Self-generated

Note: A represents accuracy, F denotes the F Score, and R indicates recall.
Note: $\widetilde{F}$ signifies the F Score when the current language is treated as the source language, with other languages being considered as targets.

guages via Google Translate. To mitigate the impact of machine translation on linguistic features, they used orthographic and phonetic alignment methods, and realigned missed entities at the corpus level using Term Frequency-Inverse Document Frequency (TF-IDF). When applied to Xie et al. (2018)'s base NER model, this system improved F-scores compared to the original embedding-based translation method (Spanish: 72 %→74 %, German: 58 %→62 %), with substantial gains in languages significantly different from English (Hindi: 27 %→42 %, Chinese: 3.5 %→50 %). To further explore embedding representation impact, Mulcaire et al. (2019) introduced Rosita, an ELMo-based multilingual embedding model. Trained on mixed multilingual corpora, its context-sensitive word representations served as input for BiLSTM-CRF NER models. Experiments showed Rosita's dynamic word representations significantly outperformed static character-level representations, offering valuable insights for future cross-lingual word representation research. Differently, Fang et al. (2017) focused on active learning, proposing a Deep Q-Network (DQN) and linear-chain CRF-based method. This approach learned active learning strategies through DQN, using current states (sentence information, model-predicted marginal probability distributions, and confidence levels) as input to output Q-values for each action (annotating the current sentence). For cross-lingual transfer, the method used multilingual embeddings trained with Canonical Correlation Analysis (CCA), enabling strategy reuse from source to target languages, achieving modest F-score results (German: 63 %, Spanish: 56 %, Dutch: 56 %).

(3) PLMs: PLMs have become powerful tools for cross-lingual NER tasks, and many studies focus on improving their cross-lingual transfer capabilities and efficiency. A key limitation is vocabulary mismatch with target languages, requiring parameter tuning. One solution reserves space for each language in multilingual vocabularies for better representation, followed by retraining. XLM-V (Liang et al., 2023) creates vocabulary fingerprints from the chances of words appearing and arranges vocabulary representation vectors so that each language has enough space. XLM-V enhances language clustering algorithms and optimises vocabulary distribution for more semantically coherent partitioning and minimal resegmentation, especially for low-resource languages. Other research focuses on specialised vocabularies for these languages. AfriBERTa (Ogueji et al., 2021) pre-trains on sentences from eleven African languages, developing subword vocabularies and using BERT-like masked language modeling. This achieved good F-scores across African languages (Hausa: 90 %, Igbo: 87 %, Swahili: 88 %). However, larger vocabularies can increase computational cost, and performance may plateau or decline after a point (Liang et al., 2023). While language-specific models avoid this, they can generalise poorly due to less training data (Mhaske et al., 2023; Ogueji et al., 2021). Similarly, IndicNER (Mhaske et al., 2023), a multilingual NER model for Indian languages, sometimes outperforms mBERT. Because IndicNER's main corpus includes English transliterations, transliteration quality can impact performance. These studies suggest joint training on similar low-resource languages improves performance; however, incorporating high-resource languages needs careful consideration due to model capacity and transfer quality. KDCT (Kumar et al., 2025) leverages a pre-trained XLM-RoBERTa model as a teacher to effectively distill knowledge into a more compact DistilBERT student model, further enhancing the model's robustness on a low-resource target language (Arabic) through consistency training. The results suggest that using a small amount of labeled data (rather than a completely zero-shot approach) combined with semi-supervised consistency training is a more judicious choice, achieving better performance under resource-constrained conditions.

Besides retraining, aligning cross-lingual corpus features (structural and semantic) during fine-tuning improves cross-lingual understanding and generation. PhoneXL (Nguyen et al., 2023) combines phonetic transcription with orthographic representations to address writing system differences in PLMs. The ContProto model (Zhou et al., 2023) uses contrastive self-training to build positive sample pairs, which may include same-category entities, possibly appearing within the same language or across languages. By training a teacher model on annotated source language data and applying it to unannotated target language data to generate pseudo-labels, ContProto achieves better language alignment. However, this works best when training and testing domains align, limiting its generalisation in cross-domain tasks. The Mulda framework (Liu et al., 2021b) translates annotated sequences, converting source language data into multiple target languages, and then fine-tunes using these translated datasets. Results show that combining source language data with translated multilingual data significantly improves recognition accuracy in target languages, even with limited source language data. LITSET (Golde et al., 2024) posits that existing few-shot NER methods are constrained in the label interpretation learning phase by limitations in the number of entity types and the scarcity of semantic information within training datasets. To address this, LITSET leverages XLM-RoBERTa for efficient processing of large-scale entity type spaces. It learns to associate entity types described in natural language with corresponding NER annotations. Consequently, the model can generalize to recognizing previously unseen entity types by requiring only descriptions of new entity types and, optionally, a few supporting examples. 2M-NER (Wang et al., 2024b) is presented as a language-multimodal NER model. Its core innovation lies in employing contrastive learning to align textual and visual representations. Furthermore, it effectively integrates multimodal information through fine-tuning a Transformer-based multimodal collaboration module. Research findings corroborate the beneficial role of multilingual and multimodal information in enhancing NER performance.

Adapters are essential for efficiently fine-tuning PLMs as training costs increase. The MAD-X framework (Pfeiffer et al., 2020) introduces three adapter types: language-specific adapters that capture linguistic features and knowledge for each language; task adapters that learn task-specific knowledge; and Invertible Adapters, designed to accommodate both input and output representations, better addressing vocabulary mismatches between multilingual and target language vocabularies. However, performance still relies on the underlying multilingual model quality and pre-training data. Invertible Adapter effectiveness decreases when source languages are not high-resource. Building on MAD-X (Pfeiffer et al., 2020), Deb et al. (2023) developed a method using Singular Value Decomposition (SVD) to construct affine subspaces for each Transformer layer per target language, capturing language-specific information. During task-specific fine-tuning with source language labeled data, source language representations are selectively projected onto target language subspaces, enhancing cross-lingual transfer. The FewTopNER framework (Bouabdallaoui et al., 2025) deeply integrates topic modeling with the NER task, leveraging topic information to enhance the contextual awareness of entity recognition. Built upon the XLM-RoBERTa pre-trained model, this framework incorporates cross-task bridging and language-specific adapters to facilitate information exchange between the entity recognition and topic modeling tasks. This enables the model to effectively handle nuanced variations across different languages. However, training independent adapters for each language is time-consuming and computationally expensive, posing challenges for low-resource languages. To address this, MAD-G (Ansell et al., 2021) uses typological feature vectors as input to generate corresponding language adapter module parameters. This allows generated adapters to support multiple typologically similar languages simultaneously, significantly improving cross-lingual transfer efficiency. However, MAD-G's (Ansell et al., 2021) generated language adapter performance heavily relies on the availability of typologically related languages during training, limiting its effectiveness for unseen languages with significant typological differences. Additionally, MAD-G's (Ansell et al., 2021) NER performance, when trained on English datasets, may underperform specialized PLMs like AfriBERTa for certain African languages (Luo: 33 %→70 %, Wolof: 32 %→62 %).

In few-shot and zero-shot cross-lingual approaches (e.g., MAD-X (Pfeiffer et al., 2020), MAD-G (Ansell et al., 2021), and multilevel alignment frameworks (Gui & Xiao, 2024)), source language selection criteria are understudied. These tasks involve fine-tuning with data from one or more languages and testing on unseen target languages. While these models perform reasonably well on unseen languages by identifying linguistic commonalities, most research intuitively selects resource-rich languages like English or German as source languages due to their large datasets, tools, and technical support. Recent research shows that source language selection significantly impacts model performance, especially for low-resource languages (Eronen et al., 2023). Eronen et al. (2023) introduced a new language similarity metric based on nearly 200 linguistic features from the World Atlas of Language Structures. Their experiments show a strong correlation between language similarity and cross-lingual transfer performance. Similarly, Nigatu et al. (2023)'s work indicates that models often rely on writing system similarities for effectively representing unseen languages. However, semantic relationships between languages in PLMs are under-explored, and it's unclear which shared features best contribute to transferability for tasks like NER. Moreover, shared features beneficial for NER may not be useful for all NLP tasks.

In the recent SemEval-2023 Task 2 competition (Fetahu et al., 2023b), transformer models using rich contextual information made significant progress in multilingual tracks. While traditional multilingual NER systems primarily improved performance by augmenting data during pre-training or fine-tuning, these methods often struggled with complex, fine-grained named entities due to insufficient knowledge and underuse of contextual information. The DAMO-NLP team (Tan et al., 2023) improved model input comprehension by acquiring multilingual knowledge from Wikipedia and Wikidata. They retrieved information relevant to the input sentences and integrated it into the XLM-R model. This approach used a CRF layer to decode fusion vectors, outputting entity label sequences. The model achieved F-scores above 90 % in multiple languages, showing the potential of context information fusion. Similarly, the NLPeople team (El-karef et al., 2023) enhanced entity boundary recognition through external context retrieval. They concatenated retrieved contextual information with target sentences and entity forms to strengthen entity representations and improve classification accuracy, although F-scores for some languages could still be improved (Persian: 83 %, Ukrainian: 89 %). Although these context-enhanced systems were not specifically for cross-lingual transfer, they significantly improved performance in specific applications. They were particularly effective in addressing contextual information scarcity in low-resource languages, significantly contributing to multilingual model generalisation.

(4) LLMs: Recent research has explored the use of generative LLMs in NLP tasks. Wang et al. (2023) combined NER task information, including task descriptions, few-shot demonstrations, and input sentences, into prompts for the GPT-3 model. While GPT-3 achieved F-scores of 91 % and 82 % on the CoNLL2003 and OntoNotes5.0 datasets, respectively, this was notably lower than existing top PLMs (95 % and 92 %, respectively). LLM performance limitations were more evident in multilingual settings. DAMO-NLP (Tan et al., 2023) systematically evaluated LLMs using single-round prompting, multi-round prompting, and multi-ICL strategies. Their research showed that even GPT-3.5 models optimised through prompt engineering significantly underperformed XLM-R models with external retrieval, achieving F-scores of only 15 % and 16 % for German and Spanish (compared to 90 % and 94 % for XLM-R). Lai et al. (2023)'s research further supported this, with experiments showing ChatGPT significantly underperforming the DAMO model across 11 languages in the MultiCoNER dataset evaluation (F1 scores below 40 %). In-depth analysis revealed ChatGPT consistently produced verbose outputs and exhibited entity category confusion in multilingual NER tasks.

To address the issue of hallucinations in LLMs when applied to low-resource languages, some research endeavors have explored leveraging high-resource languages to guide and constrain these models. CLaP (Parekh et al., 2024) utilises Google Translate to generate parallel data. It then employs instructions to compel the LLM to consider pre-existing labels within the target language sentence during label projection. This ensures that the projected labels effectively correspond to the target language sentence semantically and contextually, ultimately achieving an absolute improvement of 0.7 % F1 score compared to the Awesome-align approach. The TransFusion framework (Chen et al., 2024b) proposes a three-step "translate-annotate-fuse" process based on LLMs. This involves using machine translation to convert low-resource languages into high-resource languages, followed by annotation using information extraction Large Models. Finally, it fuses the annotation information from the translated text with the original low-resource language text, learning this fusion-reasoning paradigm during instruction fine-tuning. LLM with TransFusion, by using English annotations as context, moves beyond passively accepting translated data. Instead, it actively learns to utilise translation as a tool to enhance its capabilities in low-resource languages. The results demonstrate an average F1 score improvement of 6.6 % for baseline models in African languages.

Future Directions: The core task within NER is the accurate identification of entity boundaries. However, the inherent complexity of natural language, exemplified by phenomena such as nested entities and the emergence of fine-grained entity types, presents significant challenges to precise entity boundary recognition. Linguistic diversity further exacerbates this issue. While PLMs have demonstrated remarkable performance in recognizing common entity types, their efficacy can be improved when confronted with complex scenarios characterized by ambiguous entity boundaries or heightened contextual dependency. Future models necessitate enhanced contextual understanding capabilities to effectively delineate entity boundaries, particularly in contexts exhibiting high ambiguity and nuanced entity types.

Recent research in multilingual NER utilising LLMs heavily relies on the intrinsic multilingual capabilities of LLMs and English-centric instruction guidance. However, this approach does not directly address the fundamental problem of data scarcity in low-resource languages. Future research should proactively explore the construction of customized instructions and contextual examples specifically tailored for low-resource languages. While this may necessitate a degree of manual annotation effort, even a limited quantity of high-quality, low-resource language instructions and examples can provide LLMs with more direct and target language-specific guidance.

In the field of Named Entity Recognition, research directions encompassing multimodality and multilingualism are currently in nascent stages. Real-world textual data frequently co-occurs with multimodal information, harboring substantial developmental potential. Integrating visual, auditory, and other modalities into NER models, alongside expanding NER technology to a broader spectrum of languages, represents a crucial direction for future research.

4.3.2. Relation extraction

Monolingual RE aims to identify and categorise potential semantic relationships between entities within texts. This process not only reveals complex associations among entities but also enriches and refines semantic links within KGs, thereby enhancing their overall performance and utility.

Definition: Given a natural language text or document C, the objective of the RE task is to predict one or more factual triples (h, r, t). Here, the head entity h and the tail entity t are words, phrases, or other syntactic units derived from C, while the relation r belongs to a predefined set of relation types R. The relation r describes the semantic connection between h and t within the context of C.

Monolingual RE methods have evolved from early non-neural approaches to more sophisticated neural network architectures. Non-neural methods include rule-based systems, supervised machine learning, and distant supervision. Rule-based methods rely on manually defined patterns (as described in Appelt et al., 1993; Hearst, 1992), which have limitations in terms of adaptability and scalability. Supervised machine learning utilises feature engineering (as described in

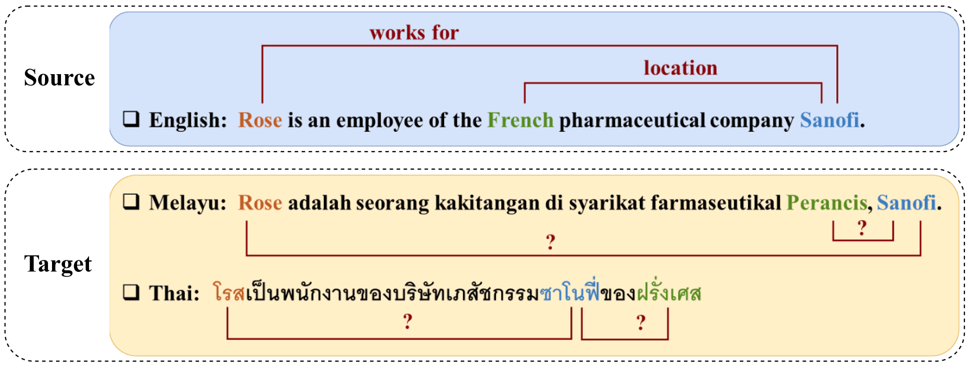


Fig. 8. The cross-lingual RE process.

Kambhatla, 2004; Zhou et al., 2005) or kernel methods (as described in Culotta & Sorensen, 2004; Zelenko et al., 2002; Zhang et al., 2008) to train classifiers, but these methods struggle with capturing long-distance dependencies. Distant supervision (Mintz et al., 2009) uses knowledge bases to automatically label data, but it is prone to propagating erroneous labels.

Neural network methods primarily fall into three strategies: pipeline-based approaches, joint training, and distant supervision. Pipeline-based methods handle entity detection and RE separately. MV-RNN (Socher et al., 2012) uses syntactic trees for semantic composition. GCN-based methods (as described in Guo et al., 2019; Zhang et al., 2018) transform syntactic trees into graph structures to better capture long-distance dependencies. Joint approaches (as described in Dai et al., 2019; Markus & Adrian, 2020; Miwa & Bansal, 2016; Sui et al., 2023; Zheng et al., 2017) perform both tasks simultaneously to reduce error propagation. These methods involve using shared parameters and complex labelling schemes or redefining tasks as set prediction problems (Sui et al., 2023). Distant supervision has also been applied to neural network methods to address data scarcity (as described in Feng et al., 2018; Ye & Ling, 2019; Zeng et al., 2015). These studies primarily mitigate noise issues and improve extraction accuracy through techniques such as multi-instance learning, reinforcement learning, and attention mechanisms.

Cross-lingual RE aims to construct a unified framework capable of understanding relation expressions across multiple languages. The expression of relations varies significantly across languages, such as word order variations and diversity in relation markers, making it difficult to directly transfer monolingual RE models to other languages. However, through alignment and language-independent semantic representations, cross-language RE is breaking through language barriers, offering new possibilities for multilingual knowledge acquisition.

Definition:In cross-lingual RE, given a sentence $C_t = w_1, w_2, \ldots, w_n$ in a target language ℓ_t and a corresponding sample C_s in a source language ℓ_s, the primary objective is to identify the specific positions of entities within the sentence, denoted as $e_{j,k} = \{w_j, w_{j+1}, \ldots, w_k\}$. The subsequent task is to infer the relationship r between a given head entity h_e and a tail entity t_e, where r belongs to a predefined set of relationship types R. In a zero-shot learning setting, the relevant label information is only available in the source language ℓ_s.

Fig. 8 illustrates a toy example of cross-lingual RE task from English to Malay and Thai.

In the label transfer paradigm, annotations in another language can be obtained through cross-lingual mapping of the head and tail entity words. Assuming the English fact $f = \{Sanofi, Location, French\}$ is known. The mappings of the head and tail entity words in the same Malay sentence are “Sanofi” and “Perancis”, respectively; it can be inferred that the Malay fact is $f = \{Sanofi, Lokasi(Location), Perancis\}$. Similarly, the case can be transferred to Thai. Regarding the parameter transfer paradigm, similar to the NER task, model transfer can be achieved through the shared space of English and other language words.

(1) Machine learning: To more clearly trace the research trajectory of cross-language RE, Table 7 summarises representative works in this field. Early research primarily focused on machine learning methods, where Kim et al. (2010) proposed a model based on statistical machine translation, comprising three components: annotation, projection, and evaluation. The annotation component utilises existing entity recognisers and relation detectors for entity recognition and relation annotation in the source language, then transfers the annotation information to the target language through word alignment-based projection. However, this model performed poorly on Korean with an F1 score of only 37 %, possibly due to incorrect word alignments and source language annotation errors. Faruqui and Kumar (2015) utilised the Google Translate API to translate source language sentences into English and obtain word alignment information, then used the OLLIE system for open-domain RE on the translated English sentences, and finally mapped the extraction results back to the source language based on BLEU score phrase similarity metrics. Although this paper proposed a BLEU score-based projection method, cross-language phrase alignment remains a challenging problem, and inherent linguistic features and structural information of the source language are easily lost during translation.

(2) Deep learning: Among mainstream neural network-based methods, Verga et al. (2016) used BiLSTM to encode Universal Schema (USchema) components of text to capture combined semantic information of sentence entities and external knowledge base relations. The model provides cross-language capability by binding parameters of different language sentence encoders and achieved performance exceeding the best TAC 2013 system on English tasks (41 % F1 score), but performance on Spanish tasks still had room for improvement (20 % F1 score), indicating limited effectiveness of the parameter binding strategy in cross-language transfer. MNRE (Lin et al., 2017) utilises CNN for sentence encoding and introduces monolingual attention and cross-lingual attention mechanisms, used respectively for selecting information-rich sentences and capturing consistency in cross-language relation patterns, thereby improving extraction accuracy. (Subburathinam et al., 2019) used GCN to encode sentence dependency tree structures and combined word distributional representations for cross-language structure transfer. Its language-independent tree structure representation and node representation are key, while utilising both distributed semantic information from word embeddings and symbolic information such as POS tags and dependency relations, enabling the model to better understand sentence structure and semantics. The model, trained on English and tested on Chinese and Arabic test sets, achieved F1 scores of 43 % and 59 %, respectively, slightly lower than supervised models trained on these languages (69 % and 67 %, respectively).

Projection-based methods perform shallow label-specific transfer, generating noisy yet target-specific training data. While straightforward, this approach remains fragile, highly susceptible to cascading errors and syntactic divergence. Conversely, embedding-based methods attempt deeper transfer of abstract semantic representations, enabling clean source models to process imperfectly represented target

Table 8
Multilingual RE Papers.

Work	Paradigms	Resources	Languages	Databases
Kim et al. (2010)	Label	- Parallel corpora - Word alignments - Bilingual dictionary	F :en(−),ko(.37)	ACE-2003 (Mitchell et al., 2004) self-generated
Faruqui and Kumar (2015)	Label	- Machine translation	B :en(−),fr(.47),hi(.38),ru(.62)	self-generated
Verga et al. (2016)	Parameter	- Multilingual embedding	F :en(.41),es(.20)	TACKBP 2012 TACKBP 13–14 (Ellis et al., 2014)
Lin et al. (2017)	Parameter	- Multilingual embedding	$P@1$:en(−),zh(−)	self-generated
Subburathinam et al. (2019)	Parameter	- Universal feature - Multilingual embedding	F :ar(.59),en(.68),zh(.42)	ACE-2005 (Walker et al., 2006)
Lan et al. (2020)	Parameter	- Parameter tuning	F :ar(.84),en(.79)	ACE-2005 (Walker et al., 2006)
Ahmad et al. (2021)	Parameter	- Parameter tuning	F :ar(.67),en(−),zh(.55)	ACE-2005 (Walker et al., 2006)
Nguyen et al. (2021)	Parameter	- Parameter tuning	F :ar(.68),en(−),zh(.58)	ACE-2005 (Walker et al., 2006)
Rathore et al. (2022)	Parameter	- Parameter tuning	AUC :de(.82),en(.83),es(.85),fr(.87)	DiS-ReX (Bhartiya et al., 2022)
Liu et al. (2023a)	Parameter	- Adapter - Enriched context	F :de(−),en(.64),es(−),fr(−),tr(−)	RELX (Köksal & Özgür, 2020)
Hsu et al. (2023)	Parameter	- Parameter tuning	F :ar(68),en(−),zh(.77)	ACE-2005 (Walker et al., 2006) self-generated
Fei et al. (2023)	Parameter	- Parameter tuning	F :ar(.71),en(−),zh(.64)	ACE-2005 (Walker et al., 2006)
Jinensibieke et al. (2024)	Label Parameter	- LLM as Projector - Language selection	F :fa(.07),fil(.13),he(.09) id(.08),ky(.14),kz(.11),ug(.03) uz(.13)	Self-generated
Santini (2024)	Label Parameter	- LLM as Executor	-	Self-generated
Jiang et al. (2024c)	Parameter	- Enriched context - LLM as Executor	-	DocRED (Yao et al., 2019)
Wei et al. (2025)	Parameter	- Parameter tuning	F :en(.79),zh(.79)	Re-DocRED (Tan et al., 2022) HacRED (Zheng et al., 2023b)

Note: $P@1$ (Precision@1) is a precision metric that only considers whether the relation predicted by the model with the highest confidence score is correct. AUC (Area Under Curve) is a metric for measuring the overall classification performance of a model. B (BLEU metric) quantifies the similarity between machine translation outputs and one or more human reference translations.

data. Though more robust to sentence-level structural variations, these methods suffer from imperfect semantic alignment and limited contextual comprehension. Method selection typically depends on available resources (high-quality machine translation systems versus bilingual lexicons) and typological distance between languages. Notably, the most successful early approaches integrated language-agnostic symbolic features-part-of-speech tags and universal dependency paths-presaging the centrality of capturing shared linguistic structures that would define the subsequent pre-training era.

(3) PLMs: Recent research has increasingly utilised PLMs. For English-to-Arabic cross-lingual transfer learning, Lan et al. (2020) proposed a customised bilingual BERT model called GigaBERT. GigaBERT employs a specialised vocabulary designed specifically for English and Arabic to better handle subword units in both languages. The paper also suggested a way to improve cross-lingual transfer by adding more data by combining code-switching data from English-Arabic dictionaries (such as PanLex, MUSE, and Wikipedia parallel titles). By replacing portions of vocabulary in sentences to simulate cross-lingual scenarios, this method effectively improved the model's cross-lingual capabilities. Experimental results show that GigaBERT performs excellently on Arabic RE tasks, achieving an F1 score of 84 % and even maintaining 48 % in zero-shot transfer settings.

The GATE model (Ahmad et al., 2021) utilises multilingual pre-trained encoders to generate word representations and combines UD parsers to construct syntactic tree structures. The model also has graph attention mechanisms that change the weights of attention based on syntactic distance. These mechanisms capture long-distance dependencies and bridge word order differences across languages, making it easier to use across languages. Finally, it obtains contextualised entity representations through maximum pooling and feeds them into a linear classifier for prediction. Building upon this, Nguyen et al. (2021) proposed an innovative cross-lingual alignment method aimed at addressing monolingual bias in RE tasks. This method similarly uses mBERT encoders to generate source and target language representations but introduces a category-based alignment mechanism to align vectors in the same category across languages. Additionally, by aligning language-independent features such as POS tags and dependency relations, the model enhances cross-lingual representation consistency. To further optimise word class representations, the method employs gradient reversal layers learnt from adversarial training to filter out irrelevant contextual information, thereby improving word class alignment accuracy and achieving certain improvements over the GATE model. The PARE model (Rathore et al., 2022) concatenates multiple sentences into paragraphs for mBERT encoding, enabling the model to capture inter-sentence interactions. This approach to directly utilising PLMs for encoding and downstream tasks has become one of the mainstream methods in the NLP field. However, when paragraphs exceed BERT's input limits, truncation may lead to loss of critical information. Moreover, PARE's insensitivity to sentence order might cause the model to overlook logical relationships or contextual connections, potentially affecting accuracy in certain cases. Furthermore, Fei et al. (2023) proposed a method based on code-mixed UD trees to address syntactic structure bias caused by language differences. This method is based on aligning the vocabulary and grammatical nodes of two languages and then merging them into code-mixed UD forests. This makes the syntactic distance between entity pairs smaller. Similar to previous work, this method uses mBERT to generate text representations, then uses Graph Attention Networks (GAT) to encode these representations to extract syntactic features from source and target languages. Finally, it predicts semantic relations between entity pairs through a biaffine decoder. This method achieved F1 scores of 71 % and 64 % in the Arabic and Chinese tracks, respectively.

In recent years, prompt-tuning technology has achieved significant results in NLP. Inspired by this, Prompt-XRE (Hsu et al., 2023) first introduced the concept of prompt-tuning into cross-lingual RE tasks, utilising prompts to activate cross-lingual knowledge in PLMs. This method converts input sentences and entity pairs into formats suitable for the mBART model and fine-tunes it using constructed data, enabling the model to more accurately predict relation types between entity pairs under prompt guidance. Experimental results show that Prompt-XRE achieved state-of-the-art performance, reaching a 77 % F1 score in English-Chinese transfer, significantly higher than the GATE model's 55 %, highlighting the immense potential of prompt tuning in cross-lingual tasks.

GranCATs (Liu et al., 2023a) proposed an efficient fine-tuning method that enhances cross-lingual alignment by incorporating contrastive adapters at different text granularities (phrase, sentence, and paragraph levels). GranCAT uses entity alignment, sentence summaries, and paragraph description data from multilingual sources (such as Wikipedia) for training and adopts contrastive learning as a pre-training task to capture global semantic information. By reducing and expanding dimensions, these adapters are built into transformer layers. This lets the model learn and share cross-lingual alignment information while keeping its original structure.

PLM-based cross-lingual RE has evolved from relying solely on intrinsic model capabilities to integrating syntactic information, refined alignment strategies, and parameter-efficient fine-tuning techniques including prompt tuning and adapters. These advances address fundamental challenges of structural linguistic divergence and monolingual bias, enabling robust cross-lingual knowledge transfer.

(4) LLMs: Additionally, the application prospects of LLMs in RE tasks are beginning to emerge. Using English prompts, ChatGPT achieved an average F1 score of 69.4 % on the SMiLER (Seganti et al., 2021a) dataset. Language-specific prompts appeared to yield better or comparable performance for ChatGPT in RE tasks, averaging 70.2 % (Lai et al., 2023). Notably, ChatGPT performed the worst on SMiLER's English tasks (only 61.8 %), which could be attributed to the English test data being larger and more complex. This indicates that ChatGPT still needs further development to effectively handle more difficult and nuanced tasks. Santini (2024) utilised instruction-tuned ChatGPT-4 to transform unstructured Italian text into semi-structured English triples. Subsequently, a seq2seq Transformer model was employed to map natural language relations to predefined knowledge ontology properties within Wikidata. The study outlined a comprehensive workflow for low-resource KG construction, however, it did not address the potential risks of hallucinations and error propagation inherent in such approaches.

Jinensibieke et al. (2024) translated a high-quality English RE dataset into languages spoken in regions such as Central Asia and Southeast Asia, employing both direct translation and pivot language translation methodologies. Data quality was subsequently enhanced through a language model perplexity filtering mechanism. The resulting low-resource language RE datasets were utilised to evaluate the performance of LLMs in low-resource environments. Experiments revealed that the F1 scores of several open-source LLMs were generally low. Another finding was that the strategy of first translating English into Chinese and then into Central Asian languages might improve translation quality to some extent. Knowledge-LLM (Jiang et al., 2024c) is introduced as a two-stage knowledge-enhanced multilingual relation extraction framework. It employs entity linking techniques to connect entities identified in the text to Wikidata, a large-scale multilingual knowledge base. Subsequently, the LLM is instructed to validate if the identified relation is substantiated within Wikidata and the given input context. Experimental results have demonstrated the crucial role of external world knowledge in enhancing the performance of LLMs in complex NLP tasks.

LLMs enable parameter-free cross-lingual RE through natural language prompting, which provides task instructions and exemplars demonstrating input-output mappings in few-shot settings. This approach offers remarkable flexibility, as adapting to new relation schemas requires only prompt modification. However, deploying LLMs for cross-lingual RE presents substantial challenges. Most critically, performance gains in English translate poorly to low-resource settings. Due to severe underrepresentation in pre-training corpora, LLMs often underperform smaller, fine-tuned PLMs on low-resource languages, indicating that current LLMs fail to bridge the performance gap across the world's languages.

A significant trend in LLM-based relation extraction involves integration with external KBs, paralleling developments in NER and broader NLP. This directly addresses LLMs' inherent limitations: susceptibility to hallucination, insensitivity to fine-grained patterns, and lack of domain-specific world knowledge. This evolution signals a shift from relying on monolithic LLMs as universal solutions toward a pragmatic approach positioning LLMs as powerful, specialized components within structured, robust information extraction pipelines.

Future Directions: While PLMs can capture contextual information, they may be more adept at identifying lexical co-occurrence patterns rather than profoundly comprehending deep semantic relations between entities. In relation extraction, determining "no relation" or "negative relation" between two entities is inherently more challenging. Especially in open-domain, cross-lingual relation extraction, a greater diversity of potential relation types exists, and it becomes necessary for models to discern during training and fine-tuning to differentiate between genuinely present relationships, non-existent relationships, and negative relationships. This necessitates that models possess enhanced relation discrimination capabilities, rather than merely relation identification capabilities. Contrastive learning methods partially mitigate this challenge; however, the effective construction of negative samples remains crucial.

Traditional remote supervision methods, while effective in reducing annotation costs, remain heavily dependent on knowledge bases. The performance of remote supervision is largely contingent upon the quality and coverage of knowledge bases and is vulnerable to the impact of erroneous information within them. LLMs, by virtue of their robust semantic understanding capabilities, empower them to autonomously recognize and extract entities and relations from unstructured text, obviating the reliance on predefined patterns or manual annotation. Future research can harness this attribute to bypass the strict dependence of traditional remote supervision methods on knowledge bases, expand the applicability to a wider range of less-explored domains, and concurrently further minimize human labor costs. However, to realize high-precision, cross-lingual, remote-supervised relation extraction, it remains imperative to investigate robust evaluation methodologies in order to assess the reliability of outputs generated by LLMs and to devise mechanisms for noise filtration and output quality enhancement.

Although certain progress has been made in cross-lingual RE from English to Arabic and Chinese, research in this field remains primarily focused on these few languages. Compared to other knowledge acquisition tasks, this limitation leaves the availability of KGs across a broader range of languages still unclear. Therefore, future research should encompass a wider linguistic environment and advocate for experimental studies on RE tasks for other low-resource languages, as well as the development of high-quality benchmark datasets to promote the widespread application.

4.3.3. Coreference resolution

Monolingual CR task aims to address the issue of redundant references to a particular entity within documents, offering a streamlined method to manage and unify these references. This reduces redundant information and ensures the consistency and accuracy of the references, thereby enhancing the accuracy and overall quality of downstream tasks such as EL and question-answering systems. Since MUC-6 (Sundheim, 1995), CR has been defined and evaluated as an independent task.

Definition: Given a sequence of text segments $C = w_1, w_2, \dots, w_n$, the objective of the CR task is to identify a set of coreference chains $S = \{s_1, s_2, \dots, s_{|S|}\}$, where each chain $s_i = \{s_{i,1}, s_{i,2}, \dots, s_{i,|s_i|}\}$ consists of words or phrases that refer to the same entity.

In the early stages, monolingual CR research relied on feature-based methods utilising statistical features from lexical, syntactic, and semantic levels to train a traditional machine learning model such as decision tree or CRF. For example, Soon et al. (2001) designed generic features and built a CR engine using C5 decision trees; Durrett and Klein

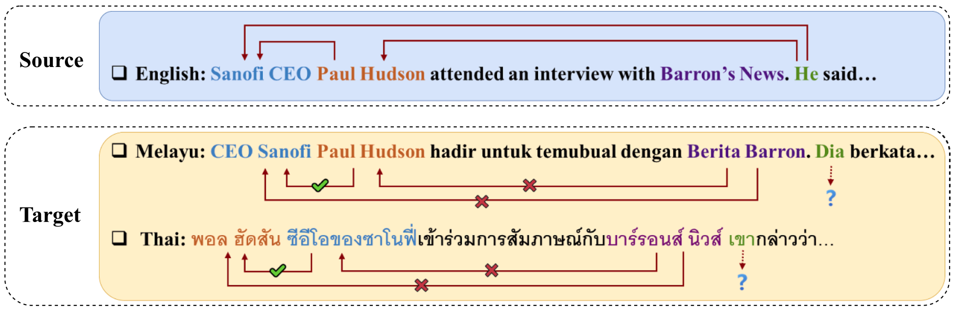


Fig. 9. The cross-lingual CR process.

(2014) proposed a structured CRF model that jointly modelled CR, NLP, and EL, capturing task interactions through factors. CNN were used to extract features from mention pairs, such as SNF, which considered both textual and numerical features. RNN were widely adopted for their ability to capture global features. Wiseman et al. (2016) used RNNs to directly learn latent representations of entity clusters; Lee et al. (2017) treated all text spans as potential mentions, using BiLSTM to learn representations and combining them with attention mechanisms for span embedding. Their subsequent work Lee et al. (2018) further incorporated ELMo dynamic word representations to enhance higher-order reasoning capabilities. PLMs significantly improved CR performance, with Joshi et al. (2019) replacing LSTM in higher-order reasoning models with SpanBERT, enhancing BERT-like models' ability to distinguish similar entities. Dobrovolskii (2021) used RoBERTa (Liu et al., 2019b) to establish coreference links at the word level, reconstructing word spans to reduce model complexity and facilitate integration. Bohnet et al. (2023) employed T5 (Raffel et al., 2020), treating CR as a sequence-to-sequence conversion task, eliminating the need for separate mention detection or higher-order decoders. Another trend is knowledge-based methods, which enhance training using external knowledge. Aralikatte et al. (2019) rewarded coreference resolvers through knowledge base consistency and combined RE tasks for multi-task reinforcement learning. Zhang et al. (2019) directly integrated KG triples into neural networks, using attention mechanisms to select relevant knowledge and improve pronoun CR.

Cross-lingual CR task involves detecting mentions that refer to the same entity in unstructured text. Although studies on cross-lingual CR began earlier (Harabagiu & Maiorano, 2000), this area did not attract widespread attention until the release of the multilingual CR shared tasks at SemEval-2010 (Recasens et al., 2010) and CoNLL-2012 (Pradhan et al., 2012). However, progress in the field has remained slow. Recently, the CRAC 2022 (Žabokrtský et al., 2022) and CRAC 2023 (Žabokrtský et al., 2023) workshops have drawn more scholarly attention to the cross-lingual challenges, injecting new vitality into this domain.

Definition: Cross-lingual CR considers sentences in the target language ℓ_t, denoted by $C_t = w_1, w_2, \dots, w_n$, and corresponding samples C_s in the source language ℓ_s. The primary objective is to identify the positions of entities within the sentences, $e_{j,k} = \{w_j, w_{j+1}, \dots, w_k\}$, and then determine whether two given entities should be merged as the same entity. The target language ℓ_t lacks necessary annotation information in cross-lingual challenges. Under a zero-shot setting, labels are only available in the source language ℓ_s.

Example: Fig. 9 illustrates a toy example of cross-lingual CR task from English to Malay and Thai.

Assuming the English mentions "Sanofi CEO" and "He" refer to the same person, "Paul Hudson," thus forming a coreference chain. In contrast, the mention of "Barron's News" neither points to other mentions nor is pointed at by them, making it an independent entity. Within the parameter or label transfer paradigm, if the direct mapping from English to other languages is known, or if mappings from English and other languages to a shared vector space are available, this correspondence can guide us to recognize the coreference of พอล ฮัดสัน, เขา, and ซีอีโอของซาโนฟี in a parallel Thai sentence. Similarly, this case can be extended to Malay and other languages that can be linked to the source language.

(1) Machine learning: Table 7 summarises representative related work. In early CR research, SWIZZLE (Harabagiu & Maiorano, 2000) explored rule-based approaches. This method automatically generated heuristic rules by analysing large amounts of annotated data for performing CR on both English and Romanian texts. Its core idea was to construct bilingual heuristic rules by comparing features of referring expressions and their antecedents in both languages. This process utilised POS tagging features and bilingual lexical resources, aiming to achieve better bilingual performance than monolingual CR. However, due to its reliance on manual translation and Romanian data annotation, its scalability was limited. Subsequently, researchers began exploring machine learning methods to address cross-lingual CR, focusing on cross-lingual transfer of labels and features. Rahman and Ng (2012) investigated the potential of using machine learning for label and feature projection in cross-lingual CR tasks. Their method involved translating target language sentences into English, applying mature English coreference parsers, and then using word alignment techniques for projection. To handle discontinuous entities, they adopted a strategy aiming to cover all entities while minimising text spans. Research results showed that automatic projection remained effective in cross-lingual CR scenarios with certain alignment resources but lacking target language coreference annotation data. Fernandes et al. (2014) proposed a cross-lingual CR model based on latent coreference trees and entropy-guided feature induction. This model used a set of universal basic features that had certain applicability across different languages. To capture differences between languages, the model automatically combined basic features into more complex feature templates through entropy-guided feature induction methods to obtain richer contextual information. Experimental results showed that this method achieved the then-best performance on Arabic, Chinese, and English (CoNLL scores of 54 %, 63 %, and 63 %, respectively).

(2) Deep learning: For a long time, despite significant progress in neural network methods for monolingual CR, effectively applying them to CR tasks remained challenging. Related research didn't begin to emerge until around 2018. Kundu et al. (2018) proposed a neural network-based cross-lingual CR model that addressed cross-lingual coreference issues by combining multilingual embeddings and language-independent features. This model converted mentions from different languages into vector representations using multilingual word embeddings and processed named entities and common noun mentions sequentially. To differentiate the importance of different types of mention pairs, the model employed feature embeddings and attention mechanisms for weighting and performed feature transformation and output through ReLU and Sigmoid layers to ultimately determine whether two mentions refer to the same entity. Other studies (Cruz et al., 2018; Urbizu et al., 2019) also explored different neural network methods, focusing on specific language pairs.

Table 9
Multilingual CR Papers.

Work	Paradigms	Resources	Languages	Databases
Harabagiu and Maiorano (2000)	Label	- Parallel corpora - Universal feature - Word Alignments - Bilingual dictionary	R :en(.77),ro(.70)	MUC-6 (Sundheim, 1995) MUC-7 (Chinchor, 1998) self-generated
Rahman and Ng (2012)	Label	- Machine translation - Bilingual dictionary	F :en(−),es(.61),fr(−),it(.59)	SemEval-2010 (Recasens et al., 2010)
Fernandes et al. (2014)	Feature	- Universal feature	C :ar(.54),en(.63),zh(.63)	CoNLL-2012 (Pradhan et al., 2012)
Kundu et al. (2018)	Parameter	- Universal feature - Multilingual embedding	C :es(.89),en(−),zh(.93)	TAC-2015 (Ji et al., 2015)
Pražák et al. (2021)	Parameter	- Parameter tuning	C :ca(.50),cs(.59),de(.46),es(.51), pl(.44),ru(.65)	CorefUD (Nedoluzhko et al., 2021)
Lai and Ji (2023)	Parameter	- Parameter tuning - Enriched context	C :ar(.67),en(−),es(.76),nl(.60)	CoNLL-2012 (Pradhan et al., 2012)
SemEval-2010 (Recasens et al., 2010)				
Porada and Cheung (2023)	Parameter	- Parameter tuning	C :ca(.72),cs(.68),de(.42),en(.67), es(.74),fr(.65),hu(.66),lt(.66), no(.74),pl(.76),ru(.77),tr(.45)	CorefUD 1.1 (Žabokrtský et al., 2023)
Straka (2023)	Parameter	- Parameter tuning	C :ca(.83),cs(.79),de(.72),en(.77), es(.83),fr(.70),hu(.69),lt(.76), no(.79),pl(.80),ru(.82),tr(.56)	CorefUD 1.1 (Žabokrtský et al., 2023)
Tang and Hardmeier (2023)	Parameter	- Adapter	C :ar(.75),ca(.75),de(.74),en(.73), es(.75),fr(.74), it(.75), nl(.75), ru(.75), zh(.75)	OntoNotes 5.0 (Weischedel et al., 2013) self-generated
Bohnet et al. (2023)	Parameter	- Parameter tuning	C :ar(.69),ca(.84),de(.86),en(.83), es(.84),it(.66),nl(.67),zh(.74)	CoNLL-2012 (Pradhan et al., 2012) SemEval-2010 (Recasens et al., 2010)
Mishra et al. (2024)	Label Parameter	- LLM as Projector - Parameter tuning	C :asm(.28),awa(.27),ben(.56),bho(.28), kan(.55),npi(.57),tam(.54),uzn(.54)	Self-generated
Liu et al. (2024)	Parameter	- Parameter tuning	-	CorefUD 1.1 (Žabokrtský et al., 2023)
(Straka, 2024)	Parameter	- Parameter tuning	C :ca(.82),cs(.75),cu(.62),de(.70) en(.76),es(.83),fr(.68),grc(.71) hbo(.72),hu(.63),lt(.76),no(.80) pl(.79),ru(.83),tr(.68)	CorefUD 1.2 (Novák et al., 2024)

Note: C (CoNLL average F1 score) is a comprehensive evaluation metric commonly used in CR tasks, calculated by taking the arithmetic mean of several standard F1 metrics to provide a more robust and thorough assessment of model performance.

(3) PLMs: With the rise of PLMs, subsequent research increasingly relied on PLMs to address challenges related to text representation and cross-lingual knowledge transfer. CR models typically consist of two key components: entity mention detection and coreference relation scoring. To reduce information loss, these modules are usually trained jointly in end-to-end models, with the encoder being fine-tuned during training. (Pražák et al., 2021) proposed an end-to-end CR model based on mBERT. This model first identifies mentions, then uses BERT-based embeddings as input, applying feed-forward neural networks to score coreference relations between mentions and their antecedents. Experimental results showed that joint training on datasets mixing multiple languages can effectively improve CR performance for low-resource languages, particularly prominent in CR, indicating that these languages benefit from knowledge of other languages and the unified annotation standards of the CorefUD dataset. MSCAW-coref (Liu et al., 2024) extends the existing efficient monolingual word-level CR model, CAW-coref, to multilingual environments, with a specific emphasis on the identification of singleton mentions. By introducing antecedent links, the approach ensures that the initial mention in each coreference chain possesses at least one antecedent link. Through this mechanism, the model is enabled to differentiate between words lacking valid antecedents and words functioning as singleton mentions. The study further validated that the explicit modeling of singleton mentions constitutes a key driving factor in performance enhancement.

At CRAC2023′s task, the McGill team (Porada & Cheung, 2023) proposed a multilingual CR model based on entity ranking. This model first identifies mentions, then generates candidate referents for each mention, and scores them using binary cross-entropy loss. The model uses language embeddings along with different joint training methods, like uniform weighting, proportional weighting, and maximum weighting strategies, to handle multilingual data well. Additionally, the model particularly focuses on mention head matching and uses dependency tree parsing to estimate mention head information. In contrast, CorPipe (Straka, 2023) generates contextual embeddings through the pre-trained mT5 model, followed by mention detection. It adopts an antecedent-maximisation strategy for coreference linking, jointly trains mention detection and CR, and uses dynamic decoding algorithms during inference to ensure correct label ordering. A crucial factor was the expanded context window (from 512 to 2560 tokens) during inference, which proved essential for long-distance coreference resolution. Optimal performance was achieved through ensemble averaging across multiple training checkpoints, validating this approach for performance enhancement and stabilization. The CorPipe model achieved a CoNLL score of 74.9 % on CRAC 2023 and obtained the best performance across all 17 corpora, demonstrating stronger competitiveness compared to the McGill model. CorPipe-2stage (Straka, 2024) employs a pre-trained language encoder model, serving as a baseline system, to predict empty nodes within text. These empty nodes are treated as potential instances of zero anaphora. Subsequently, the original CorPipe is utilised to perform the downstream CR task. CorPipe-2stage decouples the prediction of empty nodes from the conventional CR task, processing each with distinct pre-trained models. Evaluated at CRAC 2024, CorPipe-2stage achieved an average CoNLL score of 73.9 %.

Furthermore, Lai and Ji (2023) addressed the scarcity of annotated data in non-English languages by proposing to construct enhanced contextual information using natural coreference relations from anchor texts in external knowledge bases. They proposed an ensemble framework based on transfer learning techniques. A span-based architecture is at the heart of their model. It uses multilingual transformer encoders to create contextual text representations and feed-forward neural networks to guess whether candidate spans are entity mentions. To mitigate the scarcity of annotated data in non-English languages, the authors employed three transfer learning methods: continuous learning, joint learning, and distant supervision using Wikipedia anchor texts. During training, multiple models are trained using these methods, and during inference, their predictions are combined through unweighted averaging to generate final CR results. Tang and Hardmeier (2023) proposed methods for enhancing PLMs through parallel data and efficient fine-tuning. Their approach employs two channels: one for processing annotated source language data and another for processing unannotated target language data. To adapt to the target language, randomly initialised adapters are integrated into mention and coreference scorers. Model performance is improved by jointly optimising source language coreference loss and cross-lingual coreference loss, utilising potential coreference information in parallel data. Despite widespread adoption of joint training to address error propagation in two-stage mention detection and coreference prediction pipelines, inference remains inherently sequential, rendering mention detection errors irrecoverable. This fundamental constraint has driven architectural innovations. Link-Append (Bohnet et al., 2023) is a sequence-to-sequence model based on the pre-trained language model mT5. Their method combines text generation tasks with coreference link prediction, processing documents sentence by sentence and directly generating coreference links. Using “link” and “append” operations, this method predicts mention coreference relationships within a transformation system framework. This gets rid of the need for separate mention detection steps and makes the CR process easier. Compared to traditional mention ranking or span ranking-based methods, transformation-based systems are simpler, more efficient, and more aligned with how humans process coreference relations. However, the model also faces hallucination issues, occasionally generating words that don’t exist in the original text, although this occurs relatively infrequently.

(4) LLMs: Mishra et al. (2024) automatically translated an English CR dataset into South Asian languages using NLLB, an LLM specifically designed for translation tasks. To ensure translation quality and cross-lingual mention alignment, they employed awesome-align, a word alignment tool with high recall. Subsequently, a Transformer-based multilingual training model, combined with a dependency syntax tree-based end-to-end approach, achieved an average zero-shot CoNLL score of 37 % in South Asian languages. P et al. (2025) also adopted a label transfer strategy based on LLMs. However, in their subsequent CR model, they incorporated a cross-attention mechanism to more finely capture the relationships between pronouns and candidate referents, thereby enhancing CR performance.

Future Directions: Zero anaphora refers to instances in sentences where a semantically necessary constituent is not explicitly expressed in the syntactic structure, necessitating inference of its referent based on context. Variations in discourse organization across languages, habits of information ellipsis, and differences in cultural backgrounds all influence contextual inference for zero anaphora. For PLMs or LLMs, their strength lies in capturing explicit patterns and relations within text. However, zero anaphora, by its very nature, lacks explicit linguistic cues. Consequently, models may have limited exposure or ineffective learning during training regarding how to process coreference phenomena involving pronoun omission. Future research needs to augment models’ comprehension of deep semantics, discourse structure, and even world knowledge.

Current multilingual CR research primarily focuses on intra-document or cross-document coreference relations, with scant attention to spoken language data. Nevertheless, the demand for processing multilingual spoken language data is increasingly pertinent in scenarios such as international conferences, multilingual customer service, and multinational team collaboration. Compared to written language, the inherent spontaneity and informality of spoken information, including features like more frequent pauses, repetitions, and discourse markers, significantly elevates the complexity of linguistic analysis. Furthermore, information-rich features inherent in spoken language, such as intonation, emotion, and speaker information, while potentially beneficial for CR, also impose heightened demands on models’ capacity for effective utilization. Critically, the inherent error rate of automatic speech recognition systems further exacerbates the complexity of downstream tasks. In light of this, future research should proactively focus on multilingual, multimodal spoken language tasks, striving to enhance model robustness against speech recognition errors and spoken language spontaneity, to better address challenges in real-world scenarios.

4.3.4. Entity linking

Monolingual EL involves associating entity mentions within a given text with corresponding entities in a knowledge base or graph. This process requires identifying which words or phrases in the text are entity mentions and accurately mapping these mentions to existing entities in the database. EL enhances the understanding and processing of textual information, enabling more effective use of existing knowledge bases or graphs, thereby improving the accuracy and intelligence of information retrieval, question-answering systems, and text analysis applications.

Definition: Monolingual EL consists of two distinct phases. Initially, given a document D with a set of identified entity mentions $M = \{m_1, m_2, \dots, m_{|M|}\}$, the candidate entity generation is responsible for quickly locating in the knowledge base $E = \{e_1, e_2, \dots, e_{|E|}\}$ the candidate entities $C_i = \{c_{i,1}, c_{i,2}, \dots, c_{i,|C|}\}$ related to the mention m_i, where $C \subseteq M$. This is followed by the candidate entity ranking, which evaluates the probability of each mention m_i being linked to each candidate $c_{i,j}$ and determines the optimal mapping entity.

Candidate entity generation can utilise methods such as name dictionary lookups, document expansion, and search engine-based approaches. Name dictionaries map entity names to corresponding entities and may include variations such as abbreviations, expansions, aliases, and approximations (Chen et al., 2021; Shen et al., 2013). These dictionaries can be manually created by experts or generated through external knowledge sources like Wikipedia. Variants of entity names are likely to appear together within a single document, prompting methods that map these variants to names based on document content. For example, heuristic rules match contextual expansions and their acronyms (Chen et al., 2021; Shen et al., 2013), or supervised methods like SVM (Zhang et al., 2011) and CRF (Liu et al., 2017) can be employed to identify variations. Additionally, some studies (Dredze et al., 2010; Fang et al., 2020; Han & Zhao, 2009) generate candidate entities using search engines like Google and knowledge bases such as Wikipedia by submitting the mentioned texts and filtering the returned results to retain relevant entity pages for further processing.

Candidate entity ranking employs both traditional machine learning and modern embedding methods. Traditional supervised machine learning approaches frame the problem as a binary classification task (Lehmann et al., 2010; Pilz & Paaß, 2011), assessing whether a document’s entity mentions points to a specified knowledge base entity. When multiple mention-entity pairs are marked positive, the best match is selected using confidence level (Lehmann et al., 2010), learning-to-rank (Pilz & Paaß, 2011), or vector space models (Zhang et al., 2010). Unlike classification or regression, learning-to-rank focuses on the relationships and relative ordering among all samples, with standard methods including Ranking-SVM (Joachims, 2002) and ListNet (Cao et al., 2007). Vector space models, an unsupervised approach, compute similarity using various techniques and select the highest-scoring candidate as the mapping entity. Additionally, some linking methods (Han et al., 2011; Moro et al., 2014) model mention-entity pairs as graph models, using graph inference algorithms to deduce entities, ensuring highly consistent semantic interpretations. Modern methods generally employ neural models like CNNs, RNNs, attention mechanisms, or Transformers to embed words, mentions, and entities into real-valued vectors containing

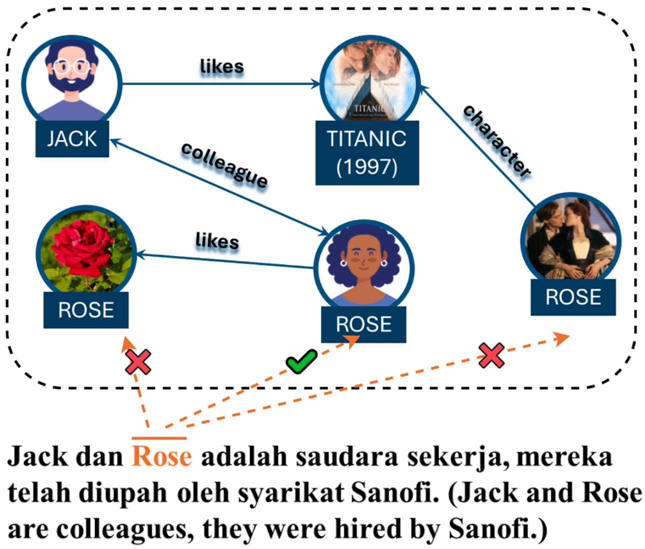


Fig. 10. The cross-lingual EL process.

natural language information, capturing the data's semantic, structural, and surface features. Different features indicate different mention-entity mappings; the most straightforward and compelling feature is prior popularity (Ganea & Hofmann, 2017), which considers the probability of an entity's occurrence given a known mention. For instance, when searching for the keyword Trump, former U.S. President Donald Trump is prioritized over his father, Fred Trump. Surface form similarity Chen et al. (2021) is another practical feature, with studies employing various distance metrics to determine the likelihood of a link to the mention. Additionally, entity type (Onoe & Durrett, 2020), context (Broscheit, 2020), and global topic features (Kolitsas et al., 2018) significantly enhance linking accuracy.

Cross-lingual EL is a key NLP task that aims to link mentions of the same real-world entity from texts in different languages to unique identifiers in a shared knowledge base. Its significance lies in its ability to bridge language barriers, enabling machines to understand and integrate information from different languages, thereby advancing applications such as cross-lingual information retrieval, machine translation, and KG construction, ultimately facilitating broader global knowledge sharing and interconnection.

Definition: Cross-lingual EL considers a sentence in the target language ℓ_t represented as $C_t = \{w_1, w_2, \ldots, w_n\}$, and a sample C_s in the source language ℓ_s. The objective of EL is to identify entity mentions $e_{j,k} = \{w_j, w_{j+1}, \ldots, w_k\}$ within the sentence and determine whether these mentions match an entity $e' \in E$ in the source language's KG or knowledge base. Resolving entity ambiguity is a critical issue in the EL task.

Example: Fig. 10 illustrates a toy example of linking sentence mentions to a knowledge base, where they may be in different languages.

As word *Rose* could refer to a type of flower, it might also denote the protagonist *Rose Calvert* from the movie Titanic or other individuals with the same name. Thus, EL begins by identifying all possible entities related to a mention in the knowledge base, forming a set of candidate entities. The most likely entity selected from these candidates is the predicted entity.

(1) Deep learning: Since 2011, the TAC-KBP (Ji et al., 2011) EL shared task has included Chinese/Spanish to English EL. Cross-lingual EL was almost simultaneously introduced as a new task (McNamee et al., 2011), featuring a two-stage Wikipedia linking approach. Initially, fast name-matching technology identifies potential entity candidates, followed by scoring and ranking these candidates using supervised learning. This method expanded from monolingual to bilingual by transcribing text into the knowledge-base language using statistical translation, allowing supervision information to be applied to the transcribed text for monolingual linking in the knowledge-base language. This is a viable approach for cross-lingual linking since monolingual systems can utilise more features, and some feature similarities are ineffective across languages. Most participants in the subsequent TAC KBP tracks (Ji et al., 2015) used one of two approaches: performing EL in a foreign language and then finding the corresponding English titles from the generated foreign language titles or translating the query document into English and then performing English EL.

Tsai and Roth (2016) utilised bilingual embeddings to represent vocabulary and Wikipedia page titles, relying on inter-language links without machine translation technology, making it applicable to all languages. However, compared to previous work McNamee et al. (2011), their method could not handle entities that exist only in the English Wikipedia, a limitation that reintroducing statistical translation models could address. In scenarios where no direct transfer resources (dictionaries, machine translation, etc.) are available between low-resource languages and English, linking from low-resource languages to a closely related hub language within the same language family and then to English appears viable, and their method achieved a P@1 value of over 80 % on all evaluated languages. Rijhwani et al. (2019) selected a closely related hub language for low-resource languages and then trained a character-level similarity encoder using bilingual dictionaries between the hub language and English, mapping entities from the hub and English languages into vector representation space. This method is suitable for some languages with very limited resources on Wikipedia, achieving 69 % accuracy in Tigrinya. Sil et al. (2018) introduced a neural model based on core citation chains from query mentions to extract relevant sentences as context. This model employs a CNN to encode the sentence context and a LSTM network to model the fine context around the mention, enhancing linking performance. Furthermore, it projects representations of each language into the vector space of pre-trained English word embeddings using multilingual CCA. The model achieved results higher than bilingual embeddings (Tsai & Roth, 2016). As the candidate generation system significantly impacts low-resource language models' performance, Zhou et al. (2019) proposed three methods to improve candidate generation and disambiguation, including combining two candidate generation systems, designing language-independent disambiguation features, and using recurrent neural networks for feature combination. The model achieved 68 % and 54 % accuracy on extremely low-resource languages such as Rwandan and Oromo, respectively.

(2) PLMs: Early studies utilising multilingual PLMs predominantly employed a dual encoder architecture (Botha et al., 2020). This method encodes the input mention and candidate entity independently via a mention encoder and an entity encoder, both initialised with PLMs. Upon parameter tuning, the mention encoder produces the vector representation of the input mention, whereas the entity encoder generates vector representations for all potential entities. The dot product of the outputs from the two encoders is calculated to determine the most suitable entity match. EL tasks generally presuppose that the entity mentions are recognised and directly associate mentions with entities in the knowledge base. Nevertheless, these studies frequently exclude the mention detection phase owing to the absence of high-quality multilingual training corpora. Consequently, a comprehensive linking system that encompasses both mention detection and entity disambiguation tasks is necessary. mReFinED (Limkonchotiwat et al., 2023) executes mention detection and encoding through two mBERT models, simultaneously implementing a bootstrapping mention detection framework to produce superior multilingual training data. The model utilises a multitask training methodology by assigning weights to the losses from mention detection, entity type prediction, entity description matching, and entity disambiguation tasks, thereby facilitating efficient end-to-end EL. In language-agnostic knowledge bases with different entity identifiers, the dual encoder can handle entities with no English descriptions but descriptions in other languages, providing a competitive advantage. The

Table 10
Multilingual EL Papers.

Work	Paradigms	Linguistic resources	Languages	Databases
McNamee et al. (2011)	Label	- Machine translation	$P@1$:ar(.93),bg(.89),cs(.86),da(.86), de(.88),el(.85),en(−),es(.89), fi(.94),fr(.88),hr(.92),it(.91), mk(.85),nl(.96),pt(.98),ro(.96), sq(.91),sr(.84),sv(.96),tr(.97), ur(.84),zh(.91)	TAC-KBP 2010 (Ji et al., 2010) self-generated
Tsai and Roth (2016)	Parameter	- Multilingual embedding - Wikipedia	$P@1$:ar(.86),de(.81),en(−),es(.81), fr(.80),he(.84),it(.80),ta(.84), th(.89),tl(.85),tr(.85),ur(.91), zh(.85)	TAC-KBP 2015 (Ji et al., 2015) self-generated
Sil et al. (2018)	Parameter	- Multilingual embedding	A :en(−),es(.82),zh(.84)	TAC-KBP 2015 (Ji et al., 2015)
Rijhwani et al. (2019)	Parameter	- Multilingual embedding - Bilingual dictionary - Linguistic feature	A :bn(.53),en(−),jv(.87),lo(.28), mr(.62),pa(.48),te(.46),ti(.69), uk(.56),ug(.40)	DARPA-LRL self-generated
Zhou et al. (2019)	Parameter	- Multilingual embedding - Wikipedia	A :en(−),om(.54),rw(.68),si(69), ti(.35)	TAC-KBP 2011 (Ji et al., 2011) TAC-KBP 2015 (Ji et al., 2015) DARPA-LRL
Botha et al. (2020)	Parameter	- Parameter tuning	$R@1$:ar(.92),de(.92),en(.87),es(.89), fa(.92),ja(.88),sr(.93),ta(.88), tr(.88)	TR2016hard (Tsai & Roth, 2016) WikiNews-2018 (Gillick et al., 2019) self-generated
Liu et al. (2021a)	Parameter	- Parameter tuning - Machine translation - External knowledge base	$P@1$:en(−),es(.59),de(.35),fi(.30), ru(.42),tr(.57),ko(.20),zh(.25), ja(.24),th(.05)	self-generated
De Cao et al. (2022)	Parameter	- Parameter tuning	A :ar(.95),en(.87),fa(.94),de(.92), ja(.91),sr(.95),es(.90),ta(.94), tr(.92),cs(.70),fr(.73),it(.57), pl(.69),pt(.76),ru(.66),zh(.52)	Mewsli-9 (Botha et al., 2020) TR2016hard (Tsai & Roth, 2016) TAC-KBP 2015 (Ji et al., 2015) self-generated
Limkonchotiwat et al. (2023)	Parameter	-Parameter tuning	R :ar(.62),fa(.55),de(.69),en(.64), es(.68),fr(.25),it(.26),ja(.44), sr(.85),ta(.34),tr(.50)	Mewsli-9 (Botha et al., 2020) TR2016hard (Tsai & Roth, 2016)
Zhu et al. (2023)	Parameter	- Parameter tuning	A :de(.31),en(.87),es(.53),fi(.24), fr(.57),ja(.26),ko(.17),nl(.51), ru(.40),th(.19),tr(.41),zh(.17)	XL-BEL (Liu et al., 2021a) Mantra GSC (Kors et al., 2015)
Chen et al. (2023)	Parameter	- Parameter tuning - Machine translation	A :en(−),zh(.76)	ICD10-CN CHPO self-generated
Remy et al. (2024)	Parameter	- Parameter tuning - LLM as Generator	A :en(.73),es(.53),de(.58)	XL-BEL (Liu et al., 2021a)
Vassileva et al. (2024)	Parameter	- Parameter tuning - LLM as Executor	A :en(.75),es(.61),fr(.73),it(.73) nl(.73)	SympTEMIST (Vassileva et al., 2024)
Wang et al. (2024a)	Parameter	- Parameter tuning	A :de(.65),es(.86),fr(.64),it(.64) zh(.89)	TAC-KBP 2015 (Ji et al., 2015) TR2016hard (Tsai & Roth, 2016)

limitation of this dual encoder architecture is that it establishes a singular language representation and omits the entity's linguistic information from the model parameters, hindering the complete exploitation of the interrelations among various languages. Wang et al. (2024a) proposed an adversarial learning-based framework. By employing a language discriminator to guide the model in learning language-invariant entity representations and leveraging the robust multilingual feature extraction capabilities of XLM-Roberta, their approach significantly enhanced generalization performance in cross-lingual zero-shot settings.

Other research has concentrated on tackling language differentiation tasks through generative models. The autoregressive sequence-to-sequence model mGENRE (De Cao et al., 2022), derived from an mBART model fine-tuned on Wikipedia hyperlink data, initially predicts a language identifier and subsequently generates the entity name in that language character by character utilising beam search. It achieved state-of-the-art performance at the time on the Mewsli-9, TAC-KBP2015, and TR2016hard datasets, surpassing the dual encoder architecture (Botha et al., 2020). The Con2GEN sequence-to-sequence generation model, founded on contrastive learning, generates positive and negative samples for each input during the training process. The objective of contrastive learning is to instruct the model to differentiate between analogous entities and enhance its capacity to identify ambiguous entities. Con2GEN (Zhu et al., 2023) employs predefined natural language templates to organise the output, facilitating the model's generation of entity names, entity types, and language information within a natural language sentence during the decoding process. In tasks requiring differentiation between entities in different languages, generative models are often preferred because traditional dual pre-trained encoder architecture models may not meet these requirements. The previously mentioned works typically presume that entity mentions are recognised and directly correlate mentions to entities within the knowledge base. Nevertheless, these studies frequently overlook the mention detection phase owing to the scarcity of high-quality multilingual training corpora. Consequently, a comprehensive linking system that encompasses both mention detection and entity disambiguation tasks is essential. mReFinED employs mBERT for mention detection and encoding, establishing a bootstrapping framework for mention detection to produce superior multilingual training data. The model utilises a multi-task training methodology by assigning weights to the losses from mention detection, entity type prediction, entity description matching, and entity disambiguation tasks to attain effective end-to-end EL.

Furthermore, recent efforts have concentrated on correlating mentions in diverse languages with the UMLS. Given that medical entities are associated with unique concept identifiers, distinguishing between output languages is unnecessary. Liu et al. (2021a) presented a language-agnostic self-alignment pre-training model that utilises phrase-level synonym sets derived from the UMLS to augment the cross-lingual proficiency of domain-specific representation models in the medical domain. Chen et al. (2023) aggregated all English medical entities from the medical graph and employed a translation engine to convert them into Chinese, thereby creating a translation-augmented training set. A new PLM named TeaBERT was subsequently introduced, which correlates Chinese medical entities with the UMLS. This study proposed a dimensionality reduction scheme for embedding vectors utilising principal component analysis to decrease the model's hardware requirements.

(3) LLMs: BioLORD-2023 (Remy et al., 2024) significantly augmented training data by leveraging GPT-3.5 to generate definitions for 400,000 biomedical concepts. Employing cross-lingual distillation, BioLORD-2023 transferred biomedical knowledge and EL proficiency from an English teacher model to a multilingual student model. Experimental findings indicate that the definitions generated by LLMs can significantly enhance the quality of the student models, achieving encouraging results in German and Spanish (53 % and 58 %), but performing worse than monolingual models in English tasks. Vassileva et al. (2024) leveraged the UMLS dictionary for efficient candidate entity retrieval via exact matching. For entities absent from the dictionary, they employed the cross-lingual SapBERT model to generate candidate sets, assessing the semantic relevance between mentions and candidates through cosine similarity. Subsequently, they utilised GPT-3.5 to rerank the top-5 candidates, capitalizing on the contextual understanding prowess of large language models to determine the most appropriate entity link. This hybrid strategy effectively integrates the precision of knowledge bases with the semantic comprehension capabilities of Transformer models, achieving an average accuracy of 73 % across five languages.

Future Directions: EL tasks encompass several interconnected but distinct subtasks, including NER, candidate generation, and ranking. Pipeline approaches typically address these subtasks sequentially, with each stage employing disparate objectives and methodologies. Particularly in multilingual settings, this staged processing paradigm is susceptible to error propagation. Errors introduced in earlier stages are inevitably carried forward, significantly impacting overall EL performance. To mitigate this issue, future research should increasingly focus on span-based end-to-end models and generative end-to-end models, especially for low-resource languages. For end-to-end, developing more effective multilingual PLM fine-tuning strategies is crucial, potentially incorporating techniques such as contrastive learning, replacement task adapters, and language adapters.

Multilingual EL often necessitates interaction with extensive knowledge bases, even multiple ones. Directly tasking LLMs with candidate entity generation within such vast search spaces can incur substantial computational costs and time overhead. The primary objective of candidate generation is to efficiently narrow the search scope while ensuring the inclusion of true entities within the candidate set. Consequently, LLMs are typically deployed in context augmentation and entity ranking phases.

While LLM-generated context leverages world knowledge, it can be prone to both information overload and insufficiency. Specifically, generated contexts may contain excessive irrelevant information, thereby hindering subsequent entity ranking. Alternatively, they might be overly generalized, lacking sufficient relevance to the entity disambiguation task even when factually accurate. This can prevent the provision of crucial information for distinguishing similar entities. Future research should explore guiding LLMs to retrieve and utilise relevant knowledge base information, ensuring the accuracy and informativeness of generated content. This can be achieved through integration with multilingual KG embeddings, Knowledge Retrieval-Augmented Generation (RAG), and related techniques.

Ranking with LLMs frequently involves iterative inference processes, such as generating descriptions for each candidate entity and computing context similarity. This significantly escalates computational demands, particularly in applications with numerous candidate entities or real-time response requirements. To fully exploit rich contextual information for ranking, longer text segments might be necessary as input for LLMs. However, LLMs possess limited context window lengths, and processing excessively long contexts can lead to information loss or performance degradation. Furthermore, context truncation can result in the exclusion of true entities from the candidate set. Future research should investigate expanding frameworks to accommodate this challenge, perhaps by generating or retrieving entities beyond the initial candidate pool.

4.4. Discussion

Task and language bias: In cross-lingual transfer learning research for KGs, NER as an upstream task has received widespread attention, while downstream tasks such as RE, CR, and EL have been relatively less studied. Through statistical analysis of language distribution in core literature (see Tables 7,8[7],9,10[8,9,10,11]), as shown in Fig. 11, English dominates as the primary source language across all tasks. However, there are notable differences in the focus on other languages across tasks: RE research primarily concentrates on a few test languages, with Chinese ranking second, while in the similarly understudied CR and EL tasks, the language variety is more diverse, with greater attention given to Spanish and German. These distribution characteristics not only reflect the significant impact of existing datasets on research directions but also highlight the urgent need for developing richer multilingual data resources in the RE field. Notably, current research primarily focuses on transfer between mainstream languages, with insufficient attention to truly low-resource languages, indicating that language bias exists not only in the training process of parameter transfer models but is more deeply rooted in the distribution of data resources itself.

Label transfer: Based on systematic analysis of core literature, several important findings have emerged. While current research primarily focuses on parameter transfer methods, label transfer remains uniquely valuable as an efficient semi-supervised data generation method in scenarios lacking quality annotated data. However, label transfer faces inherent annotation error and noise issues, mainly stemming from the oversimplified assumption that source and target language label spaces are completely consistent. In practice, label definitions and granularity often differ significantly between languages, particularly in cross-domain scenarios. Therefore, research must not only overcome cross-lingual transfer challenges but also address the additional complexities of cross-domain transfer. This highlights the importance of exploring effective methods for handling label space inconsistencies, including label mapping, multi-label learning, and open-set recognition techniques.

Parameter transfer: When source and target languages differ significantly, direct parameter transfer may lead to model performance degradation. This phenomenon primarily occurs because models may learn source language-specific features irrelevant to the target language task, subsequently causing negative transfer. Such language bias issues are common in multilingual PLMs. Researchers have proposed various mitigation strategies to address this challenge: reducing language bias by limiting training language scope (Ogueji et al., 2021), introducing language adapters for modulation (Ansell et al., 2021), and utilising mixed dependency syntactic forests for cross-lingual alignment (Fei et al., 2023). Future research should focus on identifying and transferring parameters with cross-lingual universality, which may require fine-grained alignment using cross-lingual syntactic or semantic information and incorporating external knowledge to guide parameter transfer training and inference processes.

Furthermore, cross-lingual end-to-end models have demonstrated efficient and holistic processing advantages in downstream tasks such as coreference resolution and entity linking. Compared to traditional pipeline approaches, end-to-end models jointly perform multiple tasks,

[7] https://tac.nist.gov/2012/KBP/data.html

[8] https://www.darpa.mil/program/low-resourcelanguages-for-emergent-incidents

[9] https://www.darpa.mil/program/low-resource-languages-for-emergent-incidents

[10] http://www.nhc.gov.cn/cms-search/xxgk/getManuscriptXxgk.htm?id=52905

[11] https://www.chinahpo.net

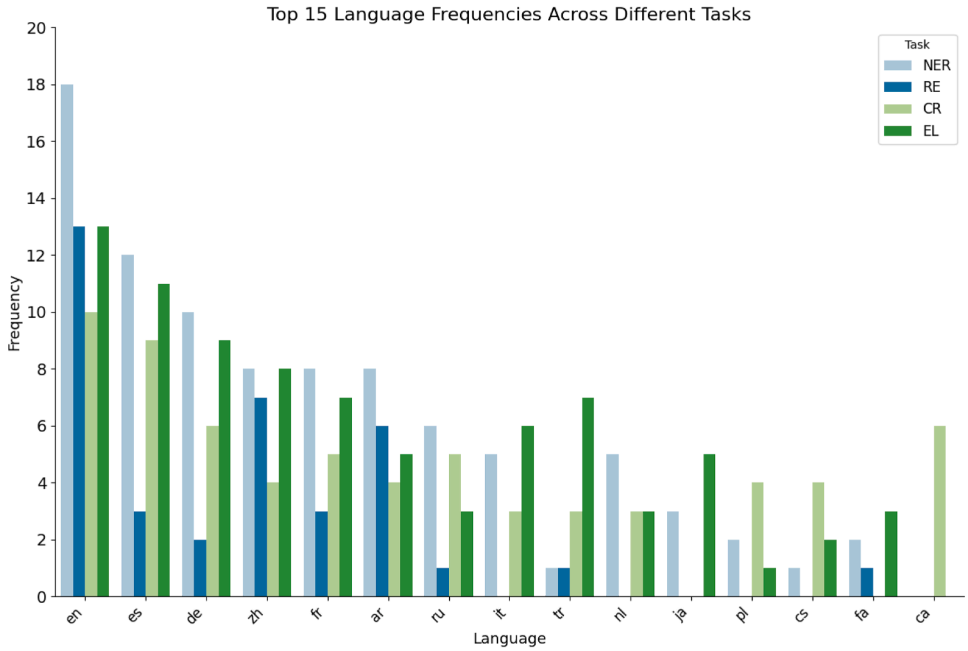


Fig. 11. Language frequencies.

mitigating the issue of error propagation between different steps. They can also optimize shared representations across tasks within a unified framework, thereby fully leveraging the interdependencies between them. Despite the significant progress end-to-end models have achieved in these tasks, they still face numerous challenges when confronted with more complex linguistic phenomena, such as zero anaphora and entity ambiguity. Future research could focus on developing more comprehensive evaluation benchmarks and metrics to systematically assess model performance in handling complex linguistic phenomena. Simultaneously, exploring more powerful model architectures and reasoning mechanisms is crucial to further enhance their performance in challenging cross-lingual tasks.

LLM: LLMs have been extensively applied in tasks such as cross-lingual translation, pseudo-data generation, and label projection, demonstrating significant potential for cross-lingual knowledge acquisition. However, this field also faces critical challenges. The performance of LLMs is highly dependent on the quality and diversity of their training data. Existing multilingual training datasets often exhibit biases, with data volumes for high-resource languages far exceeding those for low-resource languages. Furthermore, training data may be skewed towards common domains (e.g., news, web pages) while neglecting specialized fields like medicine and law. This implies that LLMs trained on general domains may struggle to handle translation and projection tasks in specialized domains or low-resource languages. Future efforts should intensify the collection and annotation of high-quality data for low-resource languages and various domains, and establish mechanisms for bias evaluation and detection specifically tailored for task-oriented LLMs. Additionally, LLMs are prone to hallucinations and inconsistencies when processing low-resource languages. Therefore, when LLMs are utilised as pseudo-generators or projectors, rigorous control over generation quality is necessary. This includes strategies such as instruction-tuning alignment, controllable generation techniques, and optimized evaluation metrics and validation mechanisms to enhance the credibility and robustness of the models.

Meanwhile, LLMs have showcased remarkable potential in tasks like NER and RE. Benefiting from pre-training on massive multilingual corpora, LLMs exhibit robust zero-shot and few-shot cross-lingual transfer capabilities. Although their performance may not consistently outperform specifically trained smaller models, the potential of LLMs in world knowledge understanding and few-shot learning is undeniable. Future research should focus on developing more effective language or task adaptation methods and explore techniques such as data augmentation and knowledge distillation to integrate the strengths of both LLMs and smaller models, thereby improving performance in specific tasks. Furthermore, Prompt Engineering, a crucial methodology in instruction tuning, deserves greater attention. Well-designed prompts can further unlock the capabilities of LLMs in specific tasks.

However, LLMs have precipitated an evaluation crisis. Their generative nature disrupts traditional exact-match evaluation paradigms-models may produce semantically correct relations using synonyms or abbreviations, which standard F1 metrics penalize as incorrect, necessitating costly human evaluation. The proposed automated alternative employing LLMs as evaluators has proven unreliable, exhibiting poor performance in specialized domains and inherent biases (Laskar et al., 2025). Without robust, scalable evaluation metrics, quantifying progress and comparing models becomes prohibitively difficult, creating a critical bottleneck in the field.

Beyond the applications of LLMs in KGs emphasized in this paper, it's crucial to consider the support that Knowledge Graphs can provide to LLMs. Given the high training time and cost of LLMs, and their potential limitations in processing the latest information or domain-specific knowledge, their capabilities can be constrained. Simultaneously, the generation process of LLMs is often a black box, making it difficult to trace the source and basis of generated results. This poses a significant drawback in scenarios requiring high credibility and explainability. Future research can focus on leveraging multilingual Knowledge Graphs as external knowledge bases for retrieval-augmented generation, enhancing the generation capabilities of LLMs by incorporating retrieved knowledge. Furthermore, Knowledge Graphs can serve as core components of LLM agents, collaboratively executing more complex tasks. For instance, LLMs can be used to convert natural language into KG queries, which are then executed, and the answers can be transformed back into natural language by LLMs. Additionally, toolkits can be employed to crawl web content in real-time, and LLMs can be used for knowledge extraction, dynamically incorporating updated results into the Knowledge Graph. This necessitates a deeper and more integrated collaboration between LLMs and KG, leveraging multi-turn interactions to achieve more complex and efficient task processing.

Multimodal multilingual KG: With advances in multimodal information processing technology, research focus has gradually expanded to exploring how to leverage cross-lingual associations in modalities such as images, videos, and audio for knowledge transfer.KGs hold significant promise as a tool for the effective integration of knowledge from diverse modalities. Looking ahead, cross-modal knowledge graphs will not only incorporate information spanning text, images, and video, but will also facilitate multi-dimensional knowledge sharing across languages. Such knowledge graphs are poised to offer more precise support for applications like intelligent search and question answering systems. Future research should explore more advanced methodologies for cross-modal alignment and fusion, such as GANs, GNNs and variational autoencoders. These innovative techniques can map data from different modalities and languages into a shared latent space, thereby enhancing information transfer and interaction efficiency.

5. Conclusion

Cross-lingual knowledge acquisition technologies enable KGs to overcome language barriers to a certain extent, to acquire the necessary supervision or knowledge from other languages for KG construction, and to create previously infeasible KG systems for resource-poor languages. This article comprehensively reviews the current research dynamics in cross-lingual knowledge acquisition for KGs, examining the methods and resources used from both horizontal and vertical perspectives. Additionally, combining technological research progress and key findings from this review, several open questions remain unresolved: (1) Can a generalised model be applied to all target languages? (2) Can the best source language range be identified for a given target language? (3) What benefits can existing work bring to the study of KGs? In response to these questions, we provide the following analysis and outlook:

(1): Grammar, semantics, and cultural differences are common across languages. The most advanced Transformer-based PLMs, limited by the current state of the art, need help to fully overcome these differences, resulting in suboptimal performance in some languages. No generalised model has been found that performs excellently across all languages. This has led some research to focus on a subset of languages, as reducing the number of languages in the model parameters improves the performance of each language within the model and reduces the training data required. This underscores the growing importance of research directions that prioritize models specialized for specific language families or architectures incorporating adaptable, language-specific adapters. Such strategies are poised to become central to the future trajectory of PLM-based research.

(2): Choosing the appropriate source language helps better capture the features of the target language, thereby improving the accuracy and reliability of knowledge acquisition. Some studies set multiple source languages to bridge the gap between the target and source languages, enhancing model performance across different languages by adding language-specific parameters to the PLM. Thus, finding the best source language for a given target language is highly feasible. However, researchers need to strike a balance between English, which has the most resources, and pivot languages, which have fewer resources but better transfer effects. For truly low-resource languages that cannot be linked directly to English, choosing a closely related pivot language with more resources and then linking it to English is a viable solution.

(3): For KGs derived from multilingual and geographically diverse datasets, cross-lingual knowledge acquisition enables developers to prioritize technological innovation rather than linguistic data processing, eliminating the prerequisite for specialized language expertise. Furthermore, for established multilingual KGs currently constrained to predominant languages, cross-lingual knowledge acquisition methodologies can facilitate expansion to less-represented linguistic domains. Specialized KGs serving minority language communities and comprehensive multilingual KGs addressing diverse global audiences will likely constitute the two predominant trajectories for future development. And KGs for low-resource languages remain significantly underdeveloped in both qualitative sophistication and quantitative scope.

LLMs and KGs demonstrate complementary capabilities. KGs provide LLMs with rigorously structured, verified, and contemporaneous factual information, thereby enhancing the precision and empirical foundation of LLM-based reasoning processes. Conversely, LLMs enhance KGs with sophisticated contextual interpretation, facilitating knowledge enrichment, nuanced user intent disambiguation, and enabling more intuitive natural language interfaces for human-system interaction. Significantly, the multilingual proficiency inherent in advanced LLMs expands the accessibility of KGs across linguistic boundaries, substantially reducing the technical and resource barriers to multilingual KG development.

Additionally, the convergence of multimodality and multilinguality enables computational systems to integrate and synthesize information across diverse representational domains-visual, auditory, and linguistic-culminating in more comprehensive and nuanced semantic understanding. This developmental trajectory suggests an imminent future characterized by increasingly sophisticated integration between multimodal and multilingual KGs, catalyzing advancements in practical applications including intelligent customer engagement systems, clinical decision support frameworks, and algorithmically optimized content recommendation architectures.

CRediT authorship contribution statement

Wei-Lin Chen: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization; **Kai-Qing Zhou:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding acquisition; **Arezoo Sarkheyli-Hägele:** Methodology, Writing – review & editing, Supervision; **Feng Qin:** Investigation, Data curation, Writing – review & editing, Visualization; **Khairunnisa Hasikin:** Methodology, Writing – review & editing, Supervision; **Di-Wen Kang:** Investigation, Data curation, Formal analysis, Writing – review & editing; **Azlan Mohd Zain:** Writing – review & editing.

Data availability

No data was used for the research described in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is supported by the National Natural Science Foundation of China under grant number 62066016, the Natural Science Foundation of Hunan Province of China under grant number 2024JJ7395, the Scientific Research Project of Education Department of Hunan Province of China under grant number 22B0549, the Liye Qin Bamboo Slips Research Special Project of Jishou University under grant number 25LYY03, the International and Regional Science and Technology Cooperation and Exchange Program of the Hunan Association for Science and Technology under grant number 025SKX-KJ-04, the Postgraduate Scientific Research Innovation Project of Hunan Province under grant number CX20251611.

References

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S. et al. (2023). Gpt-4 technical report. arXiv preprint arXiv:2303.08774.

Adelani, D. I., Abbott, J., Neubig, G., D'souza, D., Kreutzer, J., Lignos, C., Palen-Michel, C., Buzaaba, H., Rijhwani, S., Ruder, S. et al. (2021). MasakhaNER: Named entity recognition for african languages. *Transactions of the Association for Computational Linguistics*, *9*, 1116–1131.

Adelani, D. I., Neubig, G., Ruder, S., Rijhwani, S., Beukman, M., Palen-Michel, C., Lignos, C., Alabi, J. O., Muhammad, S. H., Nabende, P., Dione, C. M. B., Bukula, A., Mabuya, R., Dossou, B. F. P., Sibanda, B., Buzaaba, H., Mukiibi, J., Kalipe, G., Mbaye, D., …& Klakow, D. (2022). MasakhaNER 2.0: Africa-centric transfer learning for named entity recognition. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 4488–4508). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.298

Agarwal, U., Tanmay, K., Khandelwal, A., & Choudhury, M. (2024). Ethical reasoning and moral value alignment of LLMs depend on the language we prompt them in. In N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, & N. Xue (Eds.), *Proceedings of the 2024 joint international conference on computational linguistics, language resources and evaluation (LREC-COLING 2024)* (pp. 6330–6340). Torino, Italia: ELRA and ICCL. https://aclanthology.org/2024.lrec-main.560/.

Ahmad, W. U., Peng, N., & Chang, K.-W. (2021). Gate: Graph attention transformer encoder for cross-lingual relation and event extraction. In *Thirty-fifth AAAI conference on artificial intelligence, thirty-third conference on innovative applications of artificial intelligence and the eleventh symposium on educational advances in artificial intelligence* (pp. 12462–12470). Assoc Advancement Artificial Intelligence (*vol. 35*). AAAI Conference on Artificial Intelligence. Electr Network, FEB 02-09, 2021.

Ammar, W., Mulcaire, G., Tsvetkov, Y., Lample, G., Dyer, C., & Smith, N. A. (2016). Massively multilingual word embeddings. abs/1602.01925.

Ansell, A., Ponti, E. M., Pfeiffer, J., Ruder, S., Glavaš, G., Vulić, I., & Korhonen, A. (2021). MAD-G: Multilingual adapter generation for efficient cross-lingual transfer. In *Findings of the association for computational linguistics: EMNLP 2021* (pp. 4762–4781).

Appelt, D. E., Hobbs, J. R., Bear, J., Israel, D., & Tyson, M. (1993). FASTUS: A finite-state processor for information extraction from real-world text. In *IJCAI* (pp. 1172–1178). (*vol. 93*).

Aralikatte, R., Lent, H., Gonzalez, A. V., Hershcovich, D., Qiu, C., Sandholm, A., Ringaard, M., & Søgaard, A. (2019). Rewarding coreference resolvers for being consistent with world knowledge. arXiv preprint arXiv:1909.02392.

Balasuriya, D., Ringland, N., Nothman, J., Murphy, T., & Curran, J. R. (2009). Named entity recognition in wikipedia. In *Proceedings of the 2009 workshop on the people's web meets NLP: Collaboratively constructed semantic resources (People's web)* (pp. 10–18).

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*(6), 1137–1155. Workshop on Machine Learning Methods for Text and Images, Vancouver, Canada, 2001. https://doi.org/10.1162/153244303322533223

Berners-Lee, T. (1989). Berners-Lee's original proposal to CERN.

Berners-Lee, T. (1998). Semantic web road map. https://www.w3.org/DesignIssues/Semantic.html.

Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., & Sheets, D. (2006). Tabulator: Exploring and analyzing linked data on the semantic web. In *Proceedings of the 3rd international semantic web user interaction workshop* (p. 159). Athens, Georgia (*vol. 2006*).

Bharadwaj, A., Mortensen, D. R., Dyer, C., & Carbonell, J. G. (2016). Phonologically aware neural model for named entity recognition in low resource transfer settings. In *Proceedings of the 2016 conference on empirical methods in natural language processing* (pp. 1462–1472).

Bhartiya, A., Badola, K., & Mausam (2022). DiS-ReX: A multilingual dataset for distantly supervised relation extraction. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Proceedings of the 60th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 849–863). Dublin, Ireland: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-short.95

Bikel, D. M., Miller, S., Schwartz, R., & Weischedel, R. (1998). Nymble: A high-performance learning name-finder. arXiv preprint arXiv:cmp-lg/9803003.

Bohnet, B., Alberti, C., & Collins, M. (2023). Coreference resolution through a seq2seq transition-based system. *Transactions of the Association for Computational Linguistics*, *11*, 212–226. https://doi.org/10.1162/tacl_a_00543

Botha, J. A., Shan, Z., & Gillick, D. (2020). Entity linking in 100 languages. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 7833–7845). Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.630

Bouabdallaoui, I., Guerouate, F., Bouhaddour, S., Saadi, C., & Sbihi, M. (2025). Fewtop-NER: Integrating few-shot learning with topic modeling and named entity recognition in a multilingual framework. https://arxiv.org/abs/2502.02391.

Brachman, R. J., & Schmolze, J. G. (1985). An overview of the KL-ONE knowledge representation system. *Cognitive Science*, *9*(2), 171–216.

Broscheit, S. (2020). Investigating entity knowledge in BERT with simple neural end-to-end entity linking. *CoRR*. https://arxiv.org/abs/2003.05473.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A. et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901.

Cao, Z., Qin, T., Liu, T.-Y., Tsai, M.-F., & Li, H. (2007). Learning to rank: From pairwise approach to listwise approach. In *Proceedings of the 24th international conference on machine learning* (pp. 129–136).

Chen, C., Zheng, F., Cui, J., Cao, Y., Liu, G., Wu, J., & Zhou, J. (2024a). Survey and open problems in privacy-preserving knowledge graph: Merging, query, representation, completion, and applications. *International Journal of Machine Learning and Cybernetics*, (pp. 1–20).

Chen, L., Qi, Y., Wu, A., Deng, L., & Jiang, T. (2023). TeaBERT: An efficient knowledge infused cross-lingual language model for mapping chinese medical entities to the unified medical language system. *IEEE Journal of Biomedical and Health Informatics*, *27*(12), 6029–6038. https://doi.org/10.1109/JBHI.2023.3315143

Chen, L., Varoquaux, G., & Suchanek, F. M. (2021). A lightweight neural model for biomedical entity linking. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 12657–12665). (*vol. 35*).

Chen, S., Maeda, K., Walker, C., & Strassel, S. (2014). ACE 2007 multilingual training corpus LDC2014t18. *Web Download. Philadelphia: Linguistic Data Consortium*.

Chen, Y., Shah, V., & Ritter, A. (2024b). Translation and fusion improves zero-shot cross-lingual information extraction. https://arxiv.org/abs/2305.13582.

Chinchor, N. (1998). Overview of MUC-7. In *Seventh message understanding conference (MUC-7): Proceedings of a conference held in fairfax, Virginia, April 29-May 1, 1998*.

Chowdhery, A., Narang, S., Devlin, J., Bosma, M., Mishra, G., Roberts, A., Barham, P., Chung, H. W., Sutton, C., Gehrmann, S., Schuh, P., Shi, K., Tsvyashchenko, S., Maynez, J., Rao, A., Barnes, P., Tay, Y., Shazeer, N., Prabhakaran, V.,…Fiedel, N. (2024). PaLM: Scaling language modeling with pathways. *Journal of Machine Learning Research*, *24*(1), 1–113. https://doi.org/10.5555/3648699.3648939

Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (Almost) from scratch. *Journal of Machine Learning Research*, *12*, 2493–2537.

Conia, S., Lee, D., Li, M., Minhas, U. F., Potdar, S., & Li, Y. (2024). Towards cross-cultural machine translation with retrieval-augmented generation from multilingual knowledge graphs. In Y. Al-Onaizan, M. Bansal, & Y.-N. Chen (Eds.), *Proceedings of the 2024 conference on empirical methods in natural language processing* (pp. 16343–16360). Miami, Florida, USA: Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.emnlp-main.914

Conia, S., Li, M., Lee, D., Minhas, U., Ilyas, I., & Li, Y. (2023). Increasing coverage and precision of textual information in multilingual knowledge graphs. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 1612–1634). Singapore: Association for Computational Linguistics. https://aclanthology.org/2023.emnlp-main.100/. https://doi.org/10.18653/v1/2023.emnlp-main.100

Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). Unsupervised cross-lingual representation learning at scale. In *Annual meeting of the association for computational linguistics*.

Cruz, A. F., Rocha, G., & Cardoso, H. L. (2018). Exploring spanish corpora for portuguese coreference resolution. In *2018 Fifth international conference on social networks analysis, management and security (SNAMS)* (pp. 290–295). IEEE.

Csendes, D., Csirik, J., & Gyimóthy, T. (2004). The szeged corpus: A POS tagged and syntactically annotated hungarian natural language corpus. In *Text, speech and dialogue: 7th international conference, TSD 2004, Brno, Czech Republic, September 8–11, 2004. proceedings 7* (pp. 41–47). Springer.

Culotta, A., & Sorensen, J. (2004). Dependency tree kernels for relation extraction. In *Proceedings of the 42nd annual meeting of the association for computational linguistics (ACL-04)* (pp. 423–429). Barcelona, Spain. https://doi.org/10.3115/1218955.1219009

Dai, D., Xiao, X., Lyu, Y., Dou, S., She, Q., & Wang, H. (2019). Joint extraction of entities and overlapping relations using position-attentive sequence labeling. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 6300–6308). (*vol. 33*).

De Cao, N., Wu, L., Popat, K., Artetxe, M., Goyal, N., Plekhanov, M., Zettlemoyer, L., Cancedda, N., Riedel, S., & Petroni, F. (2022). Multilingual autoregressive entity linking. *Transactions of the Association for Computational Linguistics*, *10*, 274–290. https://doi.org/10.1162/tacl_a_00460

Deb, U., Parab, R., & Jyothi, P. (2023). Zero-shot cross-lingual transfer with learned projections using unlabeled target-language data. In *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 449–457).

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *2019 Conference of the North American chapter of the association for computational linguistics: Human 8language technologies (NAACL HLT 2019), Vol. 1* (pp. 4171–4186). Assoc Computat Linguist, N Amer Chapter; Comp Res Assoc, Comp Community Consortium; Natl Sci Fdn; Natl Rees Council Canada; Google. Minneapolis, MN, JUN 02-07, 2019.

Dobrovolskii, V. (2021). Word-level coreference resolution. In *Conference on empirical methods in natural language processing*.

Downey, D., Broadhead, M., & Etzioni, O. (2007). Locating complex named entities in web text. In *IJCAI* (pp. 2733–2739). (*vol. 7*).

Dredze, M., McNamee, P., Rao, D., Gerber, A., & Finin, T. W. (2010). Entity disambiguation for knowledge base population. In *International conference on computational linguistics*.

Durrett, G., & Klein, D. (2014). A joint model for entity analysis: Coreference, typing, and linking. *Transactions of the Association for Computational Linguistics*, *2*, 477–490. https://doi.org/10.1162/tacl_a_00197

Dyer, C., Chahuneau, V., & Smith, N. A. (2013). A simple, fast, and effective reparameterization of IBM model 2. In *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 644–648).

Edge, D., Trinh, H., Cheng, N., Bradley, J., Chao, A., Mody, A., Truitt, S., Metropolitansky, D., Ness, R. O., & Larson, J. (2025). From local to global: A graph RAG approach to query-focused summarization. https://arxiv.org/abs/2404.16130.

Ehrmann, M., Turchi, M., & Steinberger, R. (2011). Building a multilingual named entity-annotated corpus using annotation projection. In *Recent advances in natural language processing*. https://api.semanticscholar.org/CorpusID:18220955.

El-karef, M., Herr, N., Tanaka, S., & De Mel, G. (2023). NLPeople at SemEval-2023 task 2: A staged approach for multilingual named entity recognition. In *Proceedings of the 17th international workshop on semantic evaluation (SemEval-2023)* (pp. 1148–1153).

Ellis, J., Getman, J., & Strassel, S. (2014). TAC KBP english sentiment slot filling – comprehensive training and evaluation data 2013–2014 LDC2021t08. *Web Download. Philadelphia: Linguistic Data Consortium*,

Eronen, J., Ptaszynski, M., & Masui, F. (2023). Zero-shot cross-lingual transfer language selection using linguistic similarity. *Information Processing & Management*, *60*(3), 103250.

Etzioni, O., Cafarella, M. J., Downey, D., Popescu, A.-M., Shaked, T., Soderland, S., Weld, D. S., & Yates, A. (2005). Unsupervised named-entity extraction from the web: An experimental study. *Artificial Intelligence*, *165*, 91–134.

Fang, M., Li, Y., & Cohn, T. (2017). Learning how to active learn: A deep reinforcement learning approach. arXiv preprint arXiv:1708.02383.

Fang, Z., Cao, Y., Li, R., Zhang, Z., Liu, Y., & Wang, S. (2020). High quality candidate generation and sequential graph attention network for entity linking. In *Proceedings of the web conference 2020* (pp. 640–650).

Faruqui, M., & Kumar, S. (2015). Multilingual open relation extraction using cross-lingual projection. In R. Mihalcea, J. Chai, & A. Sarkar (Eds.), *Proceedings of the 2015 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 1351–1356). Denver, Colorado: Association for Computational Linguistics. https://doi.org/10.3115/v1/N15-1151

Faruqui, M., Padó, S., & Sprachverarbeitung, M. (2010). Training and evaluating a german named entity recognizer with semantic generalization. In *Konvens* (pp. 129–133).

Fei, H., Zhang, M., Zhang, M., & Chua, T.-S. (2023). Reinforcement learning for relation classification from noisy data. *Proceedings of the AAAI conference on artificial intelligence*, *32*(1), 5779–5786. https://doi.org/10.1609/aaai.v32i1.12063.

Feng, J., Huang, M., Zhao, L., Yang, Y., & Zhu, X. (2018). Reinforcement learning for relation classification from noisy data. In *AAAI Conference on artificial intelligence*. https://doi.org/10.1609/aaai.v32i1.12063

Fernandes, E. R., dos Santos, C. N., & Milidiú, R. L. (2014). Latent trees for coreference resolution. *Computational Linguistics*, *40*(4), 801–835. https://doi.org/10.1162/COLI_a_00200

Fetahu, B., Chen, Z., Kar, S., Rokhlenko, O., & Malmasi, S. (2023a). MulticoNER v2: A large multilingual dataset for fine-grained and noisy named entity recognition. abs/2310.13213.

Fetahu, B., Kar, S., Chen, Z., Rokhlenko, O., & Malmasi, S. (2023b). Semeval-2023 task 2: Fine-grained multilingual named entity recognition (multiconer 2). arXiv preprint arXiv:2305.06586.

Ganea, O., & Hofmann, T. (2017). Deep joint entity disambiguation with local neural attention. *CoRR*. http://arxiv.org/abs/1704.04920.

Gillick, D., Kulkarni, S., Lansing, L., Presta, A., Baldridge, J., Ie, E., & Garcia-Olano, D. (2019). Learning dense representations for entity retrieval. arXiv preprint arXiv:1909.10506.

Golde, J., Hamborg, F., & Akbik, A. (2024). Large-scale label interpretation learning for few-shot named entity recognition. https://arxiv.org/abs/2403.14222.

Grishman, R., & Borthwick, A. (1999). A maximum entropy approach to named entity recognition. https://api.semanticscholar.org/CorpusID:60779558.

Guan, S., Cheng, X., Bai, L., Zhang, F., Li, Z., Zeng, Y., Jin, X., & Guo, J. (2021). What is event knowledge graph: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *35*, 7569–7589.

Gui, A., & Xiao, H. (2024). Multi-level multilingual semantic alignment for zero-shot cross-lingual transfer learning. *Neural Networks*, *173*, 106217. https://doi.org/10.1016/j.neunet.2024.106217

Guo, Z., Zhang, Y., & Lu, W. (2019). Attention guided graph convolutional networks for relation extraction. arXiv preprint arXiv:1906.07510.

Han, X., Sun, L., & Zhao, J. (2011). Collective entity linking in web text: A graph-based method. *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*,

Han, X., & Zhao, J. (2009). NLPR_KBP in TAC 2009 KBP track: A two-stage method to entity linking. In *TAC*.

Hanisch, D., Fundel, K., Mevissen, H. T., Zimmer, R., & Fluck, J. (2005). Prominer: Rule-based protein and gene entity recognition. *BMC Bioinformatics*, *6*(1). https://doi.org/10.1186/1471-2105-6-S1-S14

Harabagiu, S., & Maiorano, S. J. (2000). Multilingual coreference resolution. In *Sixth applied natural language processing conference* (pp. 142–149).

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Coling 1992 volume 2: The 14th international conference on computational linguistics*.

Hitzler, P. (2021). A review of the semantic web field. *Communications of the ACM*, *64*(2), 76-83. https://doi.org/10.1145/3397512

Hogan, A., Blomqvist, E., Cochez, M., D'amato, C., Melo, G. D., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngomo, A.-C. N., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, *54*(4). https://doi.org/10.1145/3447772

Hsu, C., Zan, C., Ding, L., Wang, L., Wang, X., Liu, W., Lin, F., & Hu, W. (2023). Prompt-learning for cross-lingual relation extraction. In *2023 IEEE International joint conference on neural networks, IJCNN*. Int Neural Network Soc; IEEE Computat Intelligence Soc. Broadbeach, Australia. https://doi.org/10.1109/IJCNN54540.2023.10192002

Huang, X., Zhang, J., Li, D., & Li, P. (2019). Knowledge graph embedding based question answering. In *Proceedings of the twelfth ACM international conference on web search and data mining* WSDM '19 (p. 105–113). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3289600.3290956

Huang, Z., Li, Z., Jiang, H., Cao, T., Lu, H., Yin, B., Subbian, K., Sun, Y., & Wang, W. (2022). Multilingual knowledge graph completion with self-supervised adaptive graph alignment. abs/2203.14987.

Huang, Z., Xu, W., & Yu, K. (2015). Bidirectional LSTM-CRF models for sequence tagging. abs/1508.01991.

Jain, A., Paranjape, B., & Lipton, Z. C. (2019). Entity projection via machine-translation for cross-lingual NER. In *2019 Conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP 2019): Proceedings of the conference* (pp. 1083 +). Google; Facebook; Apple; ASAPP; Salesforce; Huawei; Baidu; Deepmind; Amazon; PolyAI; Naver; ByteDance; Megagon Labs; Zhuiyi; Verisk; MI. Hong Kong, NOV 03-07, 2019.

Ji, H., Grishman, R., & Dang, H. T. (2011). Overview of the TAC2011 Knowledge Base Population Track. *Proceedings of the Text Analysis Conference (TAC 2011) - KBP Track Overview Papers*, National Institute of Standards and Technology (NIST).

Ji, H., Grishman, R., Dang, H. T., Griffitt, K., & Ellis, J. (2010). Overview of the TAC 2010 knowledge base population track. In *Third text analysis conference (TAC 2010)* (pp. 3–3). National Institute of Standards and Technology (NIST). (*vol. 3*).

Ji, H., Nothman, J., Hachey, B., & Florian, R. (2015). Overview of TAC-KBP2015 tri-lingual entity discovery and linking. In *TAC*.

Ji, S., Pan, S., Cambria, E., Marttinen, P., & Yu, P. S. (2022). A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), 494–514. https://doi.org/10.1109/TNNLS.2021.3070843

Jiang, J., Zhou, K., Zhao, W. X., Song, Y., Zhu, C., Zhu, H., & Wen, J.-R. (2024a). KG-Agent: An efficient autonomous agent framework for complex reasoning over knowledge graph. https://arxiv.org/abs/2402.11163.

Jiang, Z., Shi, Z., Gao, X., & Xiong, Y. (2024b). Enhancing code generation through retrieval of cross-lingual semantic graphs. In *2024 31st Asia-pacific software engineering conference (APSEC)* (pp. 151–160). IEEE.

Jiang, Z., Su, J., Chen, T., Wang, Z., & Procheta, S. (2024c). Knowledge base-enhanced multilingual relation extraction with large language models. In *LKM 2024: The first international OpenKG workshop large knowledge-enhanced models@ IJCAI 2024*.

Jin, H., Li, C., Zhang, J., Hou, L., Li, J.-Z., & Zhang, P. (2019). XLORE2: Large-scale cross-lingual knowledge graph construction and application. *Data Intelligence*, *1*, 77–98.

Jinensibieke, D., Maimaiti, M., Xiao, W., Zheng, Y., & Wang, X. (2024). How good are LLMs at relation extraction under low-resource scenario? Comprehensive evaluation. https://arxiv.org/abs/2406.11162.

Joachims, T. (2002). Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 133–142).

Joshi, M., Levy, O., Weld, D. S., & Zettlemoyer, L. (2019). Bert for coreference resolution: Baselines and analysis. abs/1908.09091.

Junczys-Dowmunt, M., Grundkiewicz, R., Dwojak, T., Hoang, H., Heafield, K., Neckermann, T., Seide, F., Germann, U., Aji, A. F., Bogoychev, N. et al. (2018). Marian: Fast neural machine translation in c + +. arXiv preprint arXiv:1804.00344.

Kambhatla, N. (2004). Combining lexical, syntactic, and semantic features with maximum entropy models for information extraction. In *Proceedings of the ACL interactive poster and demonstration sessions* (pp. 178–181). Barcelona, Spain: Association for Computational Linguistics. https://aclanthology.org/P04-3022.

Kaneko, M., Imankulova, A., Bollegala, D., & Okazaki, N. (2022). Gender bias in masked language models for multiple languages. In M. Carpuat, M.-C. de Marneffe, & I. V. Meza Ruiz (Eds.), *Proceedings of the 2022 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 2740–2750). Seattle, United States: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.naacl-main.197

Kim, J.-H., & Woodland, P. C. (2000). A rule-based named entity recognition system for speech input. In *Interspeech*.

Kim, S., Jeong, M., Lee, J., & Lee, G. G. (2010). A cross-lingual annotation projection approach for relation detection. In C.-R. Huang, & D. Jurafsky (Eds.), *Proceedings of the 23rd international conference on computational linguistics (Coling 2010)* (pp. 564–571). Beijing, China: Coling 2010 Organizing Committee. https://aclanthology.org/C10-1064.

Klein, P., Ponzetto, S. P., & Glavaš, G. (2017). Improving neural knowledge base completion with cross-lingual projections. Association for Computational Linguistics.

Klementiev, A., & Roth, D. (2006). Weakly supervised named entity transliteration and discovery from multilingual comparable corpora. In *COLING/ACl 2006, vols 1 and 2, proceedings of the conference* (pp. 817–824). Assoc Computat Linguist. 21st International Conference on Computational Linguistics/44th Annual Meeting of the Association for Computational Linguistics, Sydney, AUSTRALIA, JUL 17–21, 2006.

Knez, T., & Žitnik, S. (2023). Event-centric temporal knowledge graph construction: A survey. *Mathematics*, *11*(23). https://doi.org/10.3390/math11234852

Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R. et al. (2007). Moses: Open source toolkit for statistical machine translation. In *Proceedings of the 45th annual meeting of the association for computational linguistics companion volume proceedings of the demo and poster sessions* (pp. 177–180). Association for Computational Linguistics.

Köksal, A., & Özgür, A. (2020). The RELX dataset and matching the multilingual blanks for cross-lingual relation classification. In T. Cohn, Y. He, & Y. Liu (Eds.), *Findings of the association for computational linguistics: EMNLP 2020* (pp. 340–350). Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.findings-emnlp.32

Kolitsas, N., Ganea, O.-E., & Hofmann, T. (2018). End-to-end neural entity linking. In *Conference on computational natural language learning*.

Kors, J. A., Clematide, S., Akhondi, S. A., Van Mulligen, E. M., & Rebholz-Schuhmann, D. (2015). A multilingual gold-standard corpus for biomedical concept recognition: The mantra GSC. *Journal of the American Medical Informatics Association*, *22*(5), 948–956.

Kumar, S., ElKholy, M., Liu, D., & Boulenger, A. (2025). Bridging the gap: Efficient cross-lingual NER in low-resource financial domain. In *Proceedings of the joint workshop of the 9th financial technology and natural language processing (finNLP), the 6th financial* *narrative processing (FNP), and the 1st workshop on large language models for finance and legal (LLMFinlegal)* (pp. 54–62).

Kundu, G., Sil, A., Florian, R., & Hamza, W. (2018). Neural cross-lingual coreference resolution and its application to entity linking. In I. Gurevych, & Y. Miyao (Eds.), *Proceedings of the 56th annual meeting of the association for computational linguistics (volume 2: Short papers)* (pp. 395–400). Melbourne, Australia: Association for Computational Linguistics. https://doi.org/10.18653/v1/P18-2063

Lai, T., & Ji, H. (2023). Ensemble transfer learning for multilingual coreference resolution. In M. Strube, C. Braud, C. Hardmeier, J. J. Li, S. Loaiciga, & A. Zeldes (Eds.), *Proceedings of the 4th workshop on computational approaches to discourse (CODI 2023)* (pp. 24–36). Toronto, Canada: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.codi-1.3

Lai, V. D., Ngo, N., Pouran Ben Veyseh, A., Man, H., Dernoncourt, F., Bui, T., & Nguyen, T. H. (2023). ChatGPT beyond English: Towards a comprehensive evaluation of large language models in multilingual learning. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the association for computational linguistics: EMNLP 2023* (pp. 13171–13189). Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.878

Lan, W., Chen, Y., Xu, W., & Ritter, A. (2020). An empirical study of pre-trained transformers for Arabic information extraction. In B. Webber, T. Cohn, Y. He, & Y. Liu (Eds.), *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 4727–4734). Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-main.382

Lange, L., Iurshina, A., Adel, H., & Strötgen, J. (2020). Adversarial alignment of multilingual models for extracting temporal expressions from text. In S. Gella, J. Welbl, M. Rei, F. Petroni, P. Lewis, E. Strubell, M. Seo, & H. Hajishirzi (Eds.), *Proceedings of the 5th workshop on representation learning for NLP* (pp. 103–109). Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.repl4nlp-1.14

Laskar, M. T. R., Jahan, I., Dolatabadi, E., Peng, C., Hoque, E., & Huang, J. (2025). Improving automatic evaluation of large language models (LLMs) in biomedical relation extraction via LLMs-as-the-judge. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds.), *Proceedings of the 63rd annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 25483–25497). Vienna, Austria: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.acl-long.1238

Lee, K., He, L., Lewis, M., & Zettlemoyer, L. (2017). End-to-end neural coreference resolution. arXiv preprint arXiv:1707.07045.

Lee, K., He, L., & Zettlemoyer, L. (2018). Higher-order coreference resolution with coarse-to-fine inference. In *North American chapter of the association for computational linguistics*.

Lehmann, J., Monahan, S., Nezda, L., Jung, A., & Shi, Y. (2010). LCC Approaches to knowledge base population at TAC 2010. In *TAC*.

Lewis, P., Oğuz, B., Rinott, R., Riedel, S., & Schwenk, H. (2019). MLQA: Evaluating cross-lingual extractive question answering. abs/1910.07475.

Li, F., Chen, Y., Liu, H., Yang, R., Yuan, H., Jiang, Y., Li, T., Taylor, E. M., Rouhizadeh, H., Iwasawa, Y. et al. (2025). MKG-rank: Enhancing large language models with knowledge graph for multilingual medical question answering. arXiv preprint arXiv:2503.16131.

Liang, D., Gonen, H., Mao, Y., Hou, R., Goyal, N., Ghazvininejad, M., Zettlemoyer, L., & Khabsa, M. (2023). XLM-V: Overcoming the vocabulary bottleneck in multilingual masked language models. arXiv preprint arXiv:2301.10472.

Liao, S.-H. (2005). Expert system methodologies and applications-a decade review from 1995 to 2004. *Expert Systems with Applications*, *28*(1), 93–103.

Limkonchotiwat, P., Cheng, W., Christodoulopoulos, C., Saffari, A., & Lehmann, J. (2023). mReFinED: An efficient end-to-end multilingual entity linking system. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Findings of the association for computational linguistics: EMNLP 2023* (pp. 15080–15089). Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.1007

Lin, Y., Liu, Z., & Sun, M. (2017). Neural relation extraction with multi-lingual attention. In R. Barzilay, & M. Y. Kan (Eds.), *Proceedings of the 55th annual meeting of the association for computational linguistics (ACL 2017), vol 1* (pp. 34–43). Alibaba Grp; Amazon; Apple; Baidu; Bloomberg; Facebook; Google; Samsung; Tencent; eBay; Elsevier; IBM Res; KPMG; Maluuba; Microsoft; Naver Line; NEC; Recruit Inst Technol; SAP; Adobe; Bosch; CVTE; Duolingo; Huawei; Nuance; Oracle; Sogou; Grammarly; Toutiao; Yandex. Vancouver, Canada, JUL 30-AUG 04, 2017. https://doi.org/10.18653/v1/P17-1004

Liu, A., Du, J., & Stoyanov, V. (2019a). Knowledge-augmented language model and its application to unsupervised named-entity recognition. In *North American chapter of the association for computational linguistics*.

Liu, F., Li, Z., Yin, Q., Huang, J., Luo, J., Thakur, A., Branson, K., Schwab, P., Yin, B., Wu, X. et al. (2025). A multimodal multidomain multilingual medical foundation model for zero shot clinical diagnosis. *npj Digital Medicine*, *8*(1), 86.

Liu, F., Vulic, I., Korhonen, A., & Collier, N. (2021a). Learning domain-specialised representations for cross-lingual biomedical entity linking. In *Annual meeting of the association for computational linguistics*.

Liu, H., Bauer, J., D'Oosterlinck, K., Potts, C., & Manning, C. D. (2024). MSCAW-coref: Multilingual, singleton and conjunction-aware word-level coreference resolution. In *Proceedings of the seventh workshop on computational models of reference, anaphora and coreference* (pp. 33–40).

Liu, J., Liu, C., & Huang, Y. (2017). Multi-granularity sequence labeling model for acronym expansion identification. *Information Sciences*, *378*, 462–474. https://doi.org/10.1016/j.ins.2016.06.045

Liu, L., Ding, B., Bing, L., Joty, S., Si, L., & Miao, C. (2021b). MulDA: A multilingual data augmentation framework for low-resource cross-lingual NER. In *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 5834–5846).

Liu, M., He, J., Guo, X., Chen, J., Hui, S. C., & Zhou, F. (2023a). GranCATs: Cross-lingual enhancement through granularity-specific contrastive adapters. In *Proceedings of the 32nd ACM international conference on information and knowledge management* CIKM '23 (p. 1461-1471). New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3583780.3614896

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019b). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Liu, Y., Wei, S., Huang, H., Lai, Q., Li, M., & Guan, L. (2023b). Naming entity recognition of citrus pests and diseases based on the BERT-biLSTM-CRF model. *Expert Systems with Applications*, *234*, 121103.

Luo, Y., Yang, Z., Meng, F., Li, Y., Zhou, J., & Zhang, Y. (2023). An empirical study of catastrophic forgetting in large language models during continual fine-tuning. arXiv preprint arXiv:2308.08747.

Markus, E., & Adrian, U. (2020). Span-based joint entity and relation extraction with transformer pre-training. In *ECAI 2020*, IOS Press. http://dx.doi.org/10.3233/FAIA200321. https://doi.org/10.3233/faia200321

Mayhew, S., Tsai, C.-T., & Roth, D. (2017). Cheap translation for cross-lingual named entity recognition. In *Proceedings of the 2017 conference on empirical methods in natural language processing* (pp. 2536–2545).

McCallum, A., & Li, W. (2003). Early results for named entity recognition with conditional random fields, feature induction and web-enhanced lexicons. In *Proceedings of the seventh conference on natural language learning at HLT-NAACL 2003*, (*vol. 4*), (pp. 188–191). Edmonton, Canada: Association for Computational Linguistics. https://doi.org/10.3115/1119176.1119206.

McCann, B., Bradbury, J., Xiong, C., & Socher, R. (2017). Learned in translation: Contextualized word vectors. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems* (*30*, 6297–6308). Curran Associates, Inc. https://doi.org/10.5555/3295222.3295377

McNamee, P., Mayfield, J., Lawrie, D., Oard, D., & Doermann, D. (2011). Cross-language entity linking. In H. Wang, & D. Yarowsky (Eds.), *Proceedings of 5th international joint conference on natural language processing* (pp. 255–263). Chiang Mai, Thailand: Asian Federation of Natural Language Processing.

Mhaske, A., Kedia, H., Doddapaneni, S., Khapra, M. M., Kumar, P., Murthy, R., & Kunchukuttan, A. (2023). Naamapadam: A large-scale named entity annotated data for Indic languages. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 10441–10456). Toronto, Canada: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.acl-long.582

Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013a). Efficient estimation of word representations in vector space. In *International conference on learning representations*.

Mikolov, T., Le, Q. V., & Sutskever, I. (2013b). Exploiting similarities among languages for machine translation. abs/1309.4168.

Minsky, M. (1979). A Framework for Representing Knowledge. In D. Metzing (Ed.), *Frame Conceptions and Text Understanding* (*vol. 5*), (pp. 1–25). Dieter Metzing (Ed.). Berlin, Boston: De Gruyter. https://doi.org/10.1515/9783110858778-003.

Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009). Distant supervision for relation extraction without labeled data. In *Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP* (pp. 1003–1011).

Mishra, R., Desur, P., Shah, R. R., & Kumaraguru, P. (2024). Multilingual coreference resolution in low-resource south asian languages. arXiv preprint arXiv:2402.13571.

Mitchell, A., Strassel, S., Przybocki, M., Davis, J. K., Doddington, G. R., Grishman, R., Meyers, A., Brunstein, A., Ferro, L., & Sundheim, B. (2004). Tides extraction (ACE) 2003 multilingual training data LDC2004t09. *Web Download. Philadelphia: Linguistic Data Consortium*. https://doi.org/10.35111/7xtm-ys65

Miwa, M., & Bansal, M. (2016). End-to-end relation extraction using lstms on sequences and tree structures. arXiv preprint arXiv:1601.00770.

Mohamed, Y. A., Khanan, A., Bashir, M., Mohamed, A. H. H. M., Adiel, M. A. E., & Elsadig, M. A. (2024). The impact of artificial intelligence on language translation: A review. *IEEE Access*, *12*, 25553–25579. https://doi.org/10.1109/ACCESS.2024.3366802

Moro, A., Raganato, A., & Navigli, R. (2014). Entity linking meets word sense disambiguation: A unified approach. *Transactions of the Association for Computational Linguistics*, *2*, 231–244.

Mulcaire, P., Kasai, J., & Smith, N. A. (2019). Polyglot contextual representations improve crosslingual transfer. In *2019 Conference of the North American chapter of the association for computational linguistics: human language technologies (NAACL HLT 2019), Vol. 1* (pp. 3912–3918). Assoc Computat Linguist, N Amer Chapter; Comp Res Assoc, Comp Community Consortium; Natl Sci Fdn; Natl Rees Council Canada; Google. Minneapolis, MN, JUN 02-07, 2019.

Nedoluzhko, A., Novák, M., Popel, M., Žabokrtskỳ, Z., & Zeman, D. (2021). Coreference meets universal dependencies–a pilot experiment on harmonizing coreference datasets for 11 languages. *ÚFAL MFF UK, Praha, Czechia,*

Nguyen, H. H., Zhang, C., Zhang, T., Rohrbaugh, E., & Yu, P. S. (2023). Enhancing cross-lingual transfer via phonemic transcription integration. arXiv preprint arXiv:2307.04361.

Nguyen, M. V., Nguyen, T. N., Min, B., & Nguyen, T. H. (2021). Crosslingual transfer learning for relation and event extraction via word category and class alignments. In *2021 Conference on empirical methods in natural language processing (EMNLP 2021)* (pp. 5414–5426). Conference on Empirical Methods in Natural Language Processing (EMNLP), Punta Cana, Dominican Rep, NOV 07–11, 2021.

Nigatu, H. H., Li, M., Ter Hoeve, M., Potdar, S., & Chasins, S. (2025). mRAKL: Multilingual retrieval-augmented knowledge graph construction for low-resourced languages. In W. Che, J. Nabende, E. Shutova, & M. T. Pilehvar (Eds.), *Findings of the association for computational linguistics: ACL 2025* (pp. 13072–13089). Vienna, Austria: Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.findings-acl.678

Nigatu, H. H., Tonja, A. L., & Kalita, J. (2023). The less the merrier? Investigating language representation in multilingual models. arXiv preprint arXiv:2310.13228.

Nothman, J., Ringland, N., Radford, W., Murphy, T., & Curran, J. R. (2013). Learning multilingual named entity recognition from wikipedia. *Artificial Intelligence*, *194*(SI), 151–175. https://doi.org/10.1016/j.artint.2012.03.006

Novák, M., Dohnalová, B., Konopík, M., Nedoluzhko, A., Popel, M., Pražák, O., Sido, J., Straka, M., Žabokrtskỳ, Z., & Zeman, D. (2024). Findings of the third shared task on multilingual coreference resolution. arXiv preprint arXiv:2410.15949.

Och, F. J., & Ney, H. (2000). Improved statistical alignment models. In *Proceedings of the 38th annual meeting of the association for computational linguistics* (pp. 440–447).

Ogueji, K., Zhu, Y., & Lin, J. (2021). Small data? No problem! exploring the viability of pretrained multilingual language models for low-resourced languages. In *Proceedings of the 1st workshop on multilingual representation learning* (pp. 116–126).

Onoe, Y., & Durrett, G. (2020). Fine-grained entity typing for domain independent entity linking. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 8576–8583). (*vol. 34*).

P, R., Rajeev, A. J., Anuranjana, K., & Mamidi, R. (2025). Bridge the GAP: Multi-lingual models for ambiguous pronominal coreference resolution in South Asian languages. In K. Sarveswaran, A. Vaidya, B. Krishna Bal, S. Shams, & S. Thapa (Eds.), *Proceedings of the first workshop on challenges in processing south asian languages (CHiPSAL 2025)* (pp. 104–114). Abu Dhabi, UAE: International Committee on Computational Linguistics. https://aclanthology.org/2025.chipsal-1.10/.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., …& Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71.

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*, 1345–1359.

Pan, X., Zhang, B., May, J., Nothman, J., Knight, K., & Ji, H. (2017). Cross-lingual name tagging and linking for 282 languages. In *Proceedings of the 55th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1946–1958).

Parekh, T., Hsu, I.-H., Huang, K.-H., Chang, K.-W., & Peng, N. (2024). Contextual label projection for cross-lingual structured prediction. https://arxiv.org/abs/2309.08943.

Pennington, J., Socher, R., & Manning, C. (2014). GloVe: Global vectors for word representation. In A. Moschitti, B. Pang, & W. Daelemans (Eds.), *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Doha, Qatar: Association for Computational Linguistics. https://doi.org/10.3115/v1/D14-1162

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018a). Deep contextualized word representations. In M. Walker, H. Ji, & A. Stent (Eds.), *Proceedings of the 2018 conference of the North American chapter of the association for computational linguistics: Human language technologies, volume 1 (long papers)* (pp. 2227–2237). New Orleans, Louisiana: Association for Computational Linguistics. https://doi.org/10.18653/v1/N18-1202

Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018b). Deep contextualized word representations. abs/1802.05365.

Petrov, A., La Malfa, E., Torr, P., & Bibi, A. (2023). Language model tokenizers introduce unfairness between languages. *Advances in Neural Information Processing Systems*, *36*, 36963–36990.

Pfeiffer, J., Vulic, I., Gurevych, I., & Ruder, S. (2020). MAD-X: An adapter-based framework for multi-task cross-lingual transfer. In *Proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP)* (pp. 7654–7673). Bloomberg Engn; Google Res; Apple; Amazon Sci; Baidu; Megagon Labs; Facebook; DeepMind; Grammarly; ByteDance; Zeta Alpha; Babelscape; Naver; Adobe; Hitachi; Salesforce; Univ So Calif, Viterbi Sch Engn, Informat Sci Inst. Electr network, NOV 16-20, 2020.

Pikuliak, M., Šimko, M., & Bieliková, M. (2021). Cross-lingual learning for text processing: A survey. *Expert Systems with Applications*, *165*, 113765.

Pilz, A., & Paaß, G. (2011). From names to entities using thematic context distance. In *Proceedings of the 20th ACM international conference on information and knowledge management* (pp. 857–866).

Porada, I., & Cheung, J. C. K. (2023). McGill at CRAC 2023: Multilingual generalization of entity-ranking coreference resolution models. In Z. Žabokrtský, & M. Ogrodniczuk (Eds.), *Proceedings of the CRAC 2023 shared task on multilingual coreference resolution* (pp. 52–57). Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.crac-sharedtask.5

Pradhan, S., Moschitti, A., Xue, N., Uryupina, O., & Zhang, Y. (2012). CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes. In S. Pradhan, A. Moschitti, & N. Xue (Eds.), *Joint conference on EMNLP and CoNLL - shared task* (pp. 1–40). Jeju Island, Korea: Association for Computational Linguistics. https://aclanthology.org/W12-4501.

Pražák, O., Konopík, M., & Sido, J. (2021). Multilingual coreference resolution with harmonized annotations. In R. Mitkov, & G. Angelova (Eds.), *Proceedings of the international conference on recent advances in natural language processing (RANLP 2021)* (pp. 1119–1123). Held Online: INCOMA Ltd. https://aclanthology.org/2021.ranlp-1.125.

Qin, F., Zain, A. M., Zhou, K.-Q., & Zhuo, D.-B. (2025). Hybrid weighted fuzzy production rule extraction utilizing modified harmony search and BPNN. *Scientific Reports*, *15*(1), 11012.

Quillian, M. R. (1968). Semantic networks. In M. Marvin Lee (Ed.), *Semantic information processing* (pp. 227–270). Cambridge, Mass: MIT Press.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, *21*(140), 1–67.

Rahman, A., & Ng, V. (2012). Translation-based projection for multilingual coreference resolution. In E. Fosler-Lussier, E. Riloff, & S. Bangalore (Eds.), *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 720–730). Montréal, Canada: Association for Computational Linguistics. https://aclanthology.org/N12-1090.

Ralph Weischedel, A. B. (2005). BBN Pronoun coreference and entity type corpus LDC2005t33. *Web Download. Philadelphia: Linguistic Data Consortium*,

Rathore, V., Badola, K., Singla, P., & Mausam (2022). <i>PARE</i>: A simple and strong baseline for monolingual and multilingual distantly supervised relation extraction. In *Proceedings of the 60th annual meeting of the association for computational linguistics (ACL 2022): (short papers), vol 2* (pp. 340–354). Assoc Computat Linguist; Amazon Sci; Bloomberg Engn; Google Res; Liveperson; Meta; Baidu; ByteDance; DeepMind; Grammarly; GTCOM; IBM; Megagon Labs; Microsoft; Alibaba Grp; Bosch; Cohere; G Res; ServiceNow; Relativity; Naver; ASAPP; Duolingo; BabelSpace; Spotiry; Adobe; D & I Special Initiat; AppTek; YaiGlobal; Aixplain; Apple. Dublin, Ireland, MAY 22-27, 2022.

Recasens, M., Màrquez, L., Sapena, E., Martí, M. A., Taulé, M., Hoste, V., Poesio, M., & Versley, Y. (2010). Semeval-2010 task 1: Coreference resolution in multiple languages. In *Proceedings of the 5th international workshop on semantic evaluation* (pp. 1–8).

Remy, F., Demuynck, K., & Demeester, T. (2024). BioLORD-2023: Semantic textual representations fusing large language models and clinical knowledge graph insights. *Journal of the American Medical Informatics Association*, *31*(9), 1844–1855.

Richman, A. E., & Schone, P. (2008). Mining wiki resources for multilingual named entity recognition. In *Proceedings of ACL-08: HLT* (pp. 1–9).

Rijhwani, S., Xie, J., Neubig, G., & Carbonell, J. (2019). Zero-shot neural transfer for cross-lingual entity linking. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 6924–6931). (*vol. 33*).

Rust, P., Pfeiffer, J., Vulić, I., Ruder, S., & Gurevych, I. (2021). How good is your tokenizer? on the monolingual performance of multilingual language models. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: Long papers)* (pp. 3118–3135). Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.acl-long.243

Sang, E. T. K. (2002). Introduction to the coNLL-2002 shared task: Language-independent named entity recognition. cs.CL/0209010.

Sang, E. T. K., & De Meulder, F. (2003). Introduction to the coNLL-2003 shared task: Language-independent named entity recognition. In *Conference on computational natural language learning*.

Santini, C. (2024). Combining language models for knowledge extraction from italian TEI editions. *Frontiers in Computer Science*, *6*, 1472512.

Seganti, A., Firlag, K., Skowronska, H., Satława, M., & Andruszkiewicz, P. (2021a). Multilingual entity and relation extraction dataset and model. In P. Merlo, J. Tiedemann, & R. Tsarfaty (Eds.), *Proceedings of the 16th conference of the european chapter of the association for computational linguistics: Main volume* (pp. 1946–1955). Online: Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.eacl-main.166

Seganti, A., Firlag, K., Skowrońska, H., Satlawa, M., & Andruszkiewicz, P. (2021b). Multilingual entity and relation extraction dataset and model. In *Conference of the european chapter of the association for computational linguistics*.

Sharma, K., Kumar, P., & Li, Y. (2024). OG-RAG: Ontology-grounded retrieval-augmented generation for large language models. https://arxiv.org/abs/2412.15235.

Shen, W., Wang, J., Luo, P., & Wang, M. (2013). Linking named entities in tweets with knowledge base via user interest modeling. In *Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 68–76).

Siddhant, A., Johnson, M., Tsai, H., Ari, N., Riesa, J., Bapna, A., Firat, O., & Raman, K. (2020). Evaluating the cross-lingual effectiveness of massively multilingual neural machine translation. In *Proceedings of the AAAI conference on artificial intelligence* (pp. 8854–8861). (*vol. 34*).

Sil, A., Kundu, G., Florian, R., & Hamza, W. (2018). Neural cross-lingual entity linking. *Proceedings of the AAAI Conference on Artificial Intelligence*, *32*(1). https://doi.org/10.1609/aaai.v32i1.11964

Singhal, A. (2012). Introducing the knowledge graph: Things, not strings. Google Blog, https://blog.google/products/search/introducing-knowledge-graph-things-not/.

Socher, R., Huval, B., Manning, C. D., & Ng, A. (2012). Semantic compositionality through recursive matrix-vector spaces. In *Conference on empirical methods in natural language processing*.

Søgaard, A., Ruder, S., & Vulić, I. (2018). On the limitations of unsupervised bilingual dictionary induction. arXiv preprint arXiv:1805.03620.

Soon, W. M., Ng, H. T., & Lim, D. C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, *27*(4), 521–544. https://doi.org/10.1162/089120101753342653

Sowa, J. F. (1979). Semantics of conceptual graphs. In *17th Annual meeting of the association for computational linguistics* (pp. 39–44).

Straka, M. (2023). ÚFAL CorPipe at CRAC 2023: Larger context improves multilingual coreference resolution. In Z. Žabokrtský, & M. Ogrodniczuk (Eds.), *Proceedings of the CRAC 2023 shared task on multilingual coreference resolution* (pp. 41–51). Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.crac-sharedtask.4

Straka, M. (2024). Corpipe at CRAC 2024: Predicting zero mentions from raw text. arXiv preprint arXiv:2410.02756.

Strubell, E., Verga, P., Belanger, D., & McCallum, A. (2017). Fast and accurate entity recognition with iterated dilated convolutions. arXiv preprint arXiv:1702.02098.

Studer, R., Benjamins, V. R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *Data & Knowledge Engineering*, *25*(1–2), 161–197.

Subburathinam, A., Lu, D., Ji, H., May, J., Chang, S.-F., Sil, A., & Voss, C. (2019). Cross-lingual structure transfer for relation and event extraction. In K. Inui, J. Jiang, V. Ng, & X. Wan (Eds.), *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP)* (pp. 313–325). Hong Kong, China: Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-1030

Sui, D., Zeng, X., Chen, Y., Liu, K., & Zhao, J. (2024). Joint entity and relation extraction with set prediction networks. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(9), 12784–12795. https://doi.org/10.1109/TNNLS.2023.3264735.

Sundheim, B. M. (1995). Overview of results of the MUC-6 evaluation. In *Sixth message understanding conference (MUC-6): Proceedings of a conference held in columbia, maryland, november 6–8, 1995*.

Szarvas, G., Farkas, R., & Kocsor, A. (2006). A multilingual named entity recognition system using boosting and c4.5 decision tree learning algorithms. In *IFIP Working conference on database semantics*.

Tamašauskaitė, G., & Groth, P. (2023). Defining a knowledge graph development process through a systematic review. *ACM Transactions on Software Engineering and Methodology*, *32*(1), 1–40.

Tan, Q., Xu, L., Bing, L., Ng, H. T., & Aljunied, S. M. (2022). Revisiting DocRED - addressing the false negative problem in relation extraction. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 8472–8487). Abu Dhabi, United Arab Emirates: Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.emnlp-main.580

Tan, Z., Huang, S., Jia, Z., Cai, J., Li, Y., Lu, W., Zhuang, Y. T., Tu, K., Xie, P., Huang, F., & Jiang, Y. (2023). DAMO-NLP at SemEval-2023 task 2: A unified retrieval-augmented system for multilingual named entity recognition. In *International workshop on semantic evaluation*. https://api.semanticscholar.org/CorpusID:258547160.

Tang, G., & Hardmeier, C. (2023). Parallel data helps neural entity coreference resolution. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Findings of the association for computational linguistics: ACL 2023* (pp. 3162–3171). Toronto, Canada: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-acl.197

Tang, Z., Li, T., Wu, D., Liu, J., & Yang, Z. (2024). A systematic literature review of reinforcement learning-based knowledge graph research. *Expert Systems with Applications*, *238*, 121880. https://doi.org/10.1016/j.eswa.2023.121880

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., Rodriguez, A., Joulin, A., Grave, E., & Lample, G. (2023). Llama: Open and efficient foundation language models. *CoRR*. https://doi.org/10.48550/arXiv.2302.13971.

Tsai, C.-T., & Roth, D. (2016). Cross-lingual wikification using multilingual embeddings. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 589–598). San Diego, California: Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-1072

Urbizu, G., Soraluze, A., & Arregi, O. (2019). Deep cross-lingual coreference resolution for less-resourced languages: The case of basque. In *Proceedings of the second workshop on computational models of reference, anaphora and coreference* (pp. 35–41).

Vassileva, S., Grazhdanski, G., Koychev, I., & Boytcheva, S. (2024b). Transformer-based approach for symptom recognition and multilingual linking. *Database*, *2024*. https://doi.org/10.1093/database/baae090

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30 (NIPS 2017)*. (*vol. 30*). 31st Annual Conference on Neural Information Processing Systems (NIPS), Long Beach, CA, DEC 04–09, 2017.

Veena, G., Kanjirangat, V., & Gupta, D. (2023). AGRONER: An unsupervised agriculture named entity recognition using weighted distributional semantic model. *Expert Systems with Applications*, *229*, 120440. https://doi.org/https://doi.org/10.1016/j.eswa.2023.120440

Verga, P., Belanger, D., Strubell, E., Roth, B., & McCallum, A. (2016). Multilingual relation extraction using compositional universal schema. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 886–896). San Diego, California: Association for Computational Linguistics. https://doi.org/10.18653/v1/N16-1103

Walker, C., Strassel, S., Medero, J., & Maeda, K. (2006). ACE 2005 Multilingual training corpus LDC2006t06. *Web Download. Philadelphia: Linguistic Data Consortium*,

Wang, B., Liang, B., Bai, Z., & Ma, Y. (2024a). Adversarial learning for multi-lingual entity linking. In *Proceedings of the 10th SIGHAN workshop on chinese language processing (SIGHAN-10)* (pp. 28–35).

Wang, D., Feng, X., Liu, Z., & Wang, C. (2024b). 2m-NER: Contrastive learning for multilingual and multimodal NER with language and modal fusion. *Applied Intelligence*, *54*(8), 6252–6268.

Wang, S., Sun, X., Li, X., Ouyang, R., Wu, F., Zhang, T., Li, J., & Wang, G. (2023). GPT-NER: Named entity recognition via large language models. arxiv: 2304.10428

Wang, Y., Lin, M., Hu, Q., Bao, L., Bai, S., & Li, Y. (2025). Large and small models for collaborative cross-lingual data augmentation in entity relationship extraction for low-resource languages. *Journal of King Saud University Computer and Information Sciences*, *37*(4), 56.

Wei, F., Yang, W., Liu, S., Fu, C., Dai, Q., Chen, D., Tian, X., Kong, B., & Jia, L. (2025). A document-level relation extraction method based on dual-angle attention transfer fusion. *Expert Systems with Applications*, *272*, 126615. https://doi.org/10.1016/j.eswa.2025.126615.

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Kaufman, A. T., Franchiniv, M., El-Bachouti, M., Belvin, R., & Houston, A. (2013). Ontonotes release 5.0 with ontonotes DB tool v0. 999 beta. *Linguistic data consortium*, (pp. 1–53).

Weischedel, R., Palmer, M., Marcus, M., Hovy, E., Pradhan, S., Ramshaw, L., Xue, N., Taylor, A., Kaufman, J., Franchini, M., El-Bachouti, M., Belvin, R., & Houston, A. (2011). Ontonotes release 4.0 LDC2011t03. *Web Download. Philadelphia: Linguistic Data Consortium*, https://doi.org/https://doi.org/10.35111/gfjf-7r50

Wiseman, S., Rush, A. M., & Shieber, S. M. (2016). Learning global features for coreference resolution. arXiv preprint arXiv:1604.03035.

Wu, X., Yang, J., Li, T., Zhang, S., Du, Y., Chai, L., Liang, D., & Li, Z. (2025). Unleashing potential of evidence in knowledge-intensive dialogue generation. In *ICASSP 2025 - 2025 IEEE International conference on acoustics, speech and signal processing (ICASSP)* (pp. 1–5). https://doi.org/10.1109/ICASSP49660.2025.10890206

de Wynter, A., Watts, I., Wongsangaroonsri, T., Zhang, M., Farra, N., Altlntoprak, N. E., Baur, L., Claudet, S., Gajdusek, P., Gören, C., Gu, Q., Kaminska, A., Kaminski, T., Kuo, R., Kyuba, A., Lee, J., Mathur, K., Merok, P., Milovanović, I., …& Chen, S.-Q. (2024). RTP-LX: Can LLMs evaluate toxicity in multilingual scenarios? 2404.14397.

Xie, J., Yang, Z., Neubig, G., Smith, N. A., & Carbonell, J. (2018). Neural cross-lingual named entity recognition with minimal resources. In *2018 Conference on empirical methods in natural language processing (EMNLP 2018)* (pp. 369–379). Google; Facebook; Bloomberg; Salesforce; Apple; Amazon; Baidu; Grammarly; Naver Labs Europe; FWO; KU Leuven, Dept Comp Sci; CVTE; Ebay; Microsoft; Naver Line; Oracle; Polya; Huawei; Duolingo; Figure Eight; Nuance. Brussels, Belgium.

Xu, D., Chen, W., Peng, W., Zhang, C., Xu, T., Zhao, X., Wu, X., Zheng, Y., Wang, Y., & Chen, E. (2024). Large language models for generative information extraction: A survey. *Frontiers of Computer Science*, *18*(6), 186357.

Xu, Y., Cao, H., Du, W., & Wang, W. (2022). A survey of cross-lingual sentiment analysis: Methodologies, models and evaluations. *Data Science and Engineering*, *7*, 279–299.

Xue, L., Constant, N., Roberts, A., Kale, M., Al-Rfou, R., Siddhant, A., Barua, A., & Raffel, C. (2021). mT5: A massively multilingual pre-trained text-to-text transformer. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, & Y. Zhou (Eds.), *Proceedings of the 2021 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 483–498). Online: Association for Computational Linguistics. https://aclanthology.org/2021.naacl-main.41/. https://doi.org/10.18653/v1/2021.naacl-main.41

Yang, L., Wei, C., Yang, J., Xia, W., Yang, Y., Luo, Y., Niyato, D., Sun, L., & Liu, Z. (2025). FuzzCoder: Code large language model-based fuzz testing for industrial iot programs. *IEEE Internet of Things Journal*, (pp. 1–1). https://doi.org/10.1109/JIOT.2025.3577602

Yao, Y., Ye, D., Li, P., Han, X., Lin, Y., Liu, Z., Liu, Z., Huang, L., Zhou, J., & Sun, M. (2019). DocRED: A large-scale document-level relation extraction dataset. In A. Korhonen, D. Traum, & L. Màrquez (Eds.), *Proceedings of the 57th annual meeting of the association for computational linguistics* (pp. 764–777). Florence, Italy: Association for Computational Linguistics. https://aclanthology.org/P19-1074/. https://doi.org/10.18653/v1/P19-1074

Ye, S., Zhou, K., Zain, A. M., Wang, F., & Yusoff, Y. (2023). A modified harmony search algorithm and its applications in weighted fuzzy production rule extraction. *Frontiers of Information Technology & Electronic Engineering*, *24*(11), 1574–1590.

Ye, Z., & Ling, Z. (2019). Distant supervision relation extraction with intra-bag and inter-bag attentions. In *North American chapter of the association for computational linguistics*.

Žabokrtský, Z., Konopik, M., Nedoluzhko, A., Novák, M., Ogrodniczuk, M., Popel, M., Prazak, O., Sido, J., & Zeman, D. (2023). Findings of the second shared task on multilingual coreference resolution. In Z. Žabokrtský, & M. Ogrodniczuk (Eds.), *Proceedings of the CRAC 2023 shared task on multilingual coreference resolution* (pp. 1–18). Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.crac-sharedtask.1

Žabokrtský, Z., Konopík, M., Nedoluzhko, A., Novák, M., Ogrodniczuk, M., Popel, M., Pražák, O., Sido, J., Zeman, D., & Zhu, Y. (2022). Findings of the shared task on multilingual coreference resolution. In Z. Žabokrtský, & M. Ogrodniczuk (Eds.), *Proceedings of the CRAC 2022 shared task on multilingual coreference resolution* (pp. 1–17). Gyeongju, Republic of Korea: Association for Computational Linguistics.

Zelenko, D., Aone, C., & Richardella, A. R. (2002). Kernel methods for relation extraction. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing* (pp. 71–78), (*vol. 10*). Association for Computational Linguistics. https://doi.org/10.3115/1118693.1118703.

Zeng, D., Liu, K., Chen, Y., & Zhao, J. (2015). Distant supervision for relation extraction via piecewise convolutional neural networks. In *Proceedings of the 2015 conference on empirical methods in natural language processing* (pp. 1753–1762).

Zhang, H., Song, Y., Song, Y., & Yu, D. (2019). Knowledge-aware pronoun coreference resolution. abs/1907.03663.

Zhang, H., Wang, T., Liang, Z., Huang, Z., Chen, C., & Cheng, L. (2025). Multilingual graph retrieval-augmented generation for product design using design knowledge. *Journal of Engineering Design*, *0*, 1–32. https://doi.org/10.1080/09544828.2025.2537464.

Zhang, J.-C., Zain, A. M., Zhou, K.-Q., Chen, X., & Zhang, R.-M. (2024). A review of recommender systems based on knowledge graph embedding. *Expert Systems with Applications*, *250*, 123876. https://doi.org/https://doi.org/10.1016/j.eswa.2024.123876

Zhang, M., Zhou, G., & Aw, A. (2008). Exploring syntactic structured features over parse trees for relation extraction using kernel methods. *Information Processing & Management*, *44*(2), 687–701.

Zhang, W., Sim, Y.-C., Su, J., & Tan, C.-L. (2011). Entity linking with effective acronym expansion, instance selection and topic modeling. In *Twenty-second international joint conference on artificial intelligence*.

Zhang, W., Su, J., Tan, C. L., & Wang, W. T. (2010). Entity linking leveraging automatically generated annotation. In *Proceedings of the 23rd international conference on computational linguistics (Coling 2010)* (pp. 1290–1298).

Zhang, Y., Qi, P., & Manning, C. D. (2018). Graph convolution over pruned dependency trees improves relation extraction. arXiv preprint arXiv:1809.10185.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z. et al. (2023). A survey of large language models. arXiv preprint arXiv:2303.18223

Zheng, B., Xia, P., Yarmohammadi, M., & Durme, B. V. (2023a). Multilingual coreference resolution in multiparty dialogue. *Transactions of the Association for Computational Linguistics*, *11*, 922–940.

Zheng, S., Wang, F., Bao, H., Hao, Y., Zhou, P., & Xu, B. (2017). Joint extraction of entities and relations based on a novel tagging scheme. arXiv preprint arXiv:1706.05075.

Zheng, Y., Guo, Y., Luo, Z., Yu, Z., Wang, K., Zhang, H., & Zhao, H. (2023b). A survey on document-level relation extraction: methods and applications. In *3rd International conference on internet, education and information technology (IEIT 2023)* (pp. 1061–1071). Atlantis Press.

Zhong, L., Wu, J., Li, Q., Peng, H., & Wu, X. (2023). A comprehensive survey on automatic knowledge graph construction. *ACM Computing Surveys*, *56*(4). https://doi.org/10.1145/3618295

Zhou, G., Su, J., Zhang, J., & Zhang, M. (2005). Exploring various knowledge in relation extraction. In K. Knight, H. T. Ng, & K. Oflazer (Eds.), *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL'05)* (pp. 427–434). Ann Arbor, Michigan: Association for Computational Linguistics. https://doi.org/10.3115/1219840.1219893

Zhou, R., Li, X., Bing, L., Cambria, E., & Miao, C. (2023). Improving self-training for cross-lingual named entity recognition with contrastive and prototype learning. arXiv preprint arXiv:2305.13628

Zhou, S., Rijhwani, S., & Neubig, G. (2019). Towards zero-resource cross-lingual entity linking. In C. Cherry, G. Durrett, G. Foster, R. Haffari, S. Khadivi, N. Peng, X. Ren, & S. Swayamdipta (Eds.), *Proceedings of the 2nd workshop on deep learning approaches for low-resource NLP (DeepLo 2019)* (pp. 243–252). Hong Kong, China: Association for Computational Linguistics. https://doi.org/10.18653/v1/D19-6127

Zhu, S., Xu, S., Sun, H., Pan, L., Cui, M., Du, J., Jin, R., Branco, A., Xiong, D. et al. (2024a). Multilingual large language models: A systematic survey. arXiv preprint arXiv:2411.11072

Zhu, T., Qin, Y., Chen, Q., Mu, X., Yu, C., & Xiang, Y. (2023). Controllable contrastive generation for multilingual biomedical entity linking. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 conference on empirical methods in natural language processing* (pp. 5742–5753). Singapore: Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.emnlp-main.350

Zhu, X., Li, Z., Wang, X., Jiang, X., Sun, P., Wang, X., Xiao, Y., & Yuan, N. J. (2024b). Multimodal knowledge graph construction and application: A survey. *IEEE Transactions on Knowledge and Data Engineering*, *36*(2), 715–735. https://doi.org/10.1109/TKDE.2022.3224228