

# Combining Deep learning Models and Lexical Linked Data: some insights from the development of a Multilingual News Named Entity Recognition and Linking dataset

**Emmanuel Cartier, Emile Peetermans**

European Commission, Joint Research Center

Via Enrico Fermi, 2749

21027 Ispra (VA), Italy

emmanuel.cartier@ec.europa.eu, peetermansemile@gmail.com

## Abstract

This paper presents the methodology and outcomes of a Named Entity Recognition and Linking multilingual news benchmark that leverages both Deep learning approaches by using a fine-tuned transformer model to detect mentions of persons, locations and organisations in text, and Linguistic Linked Open Data, through the use of Wikidata to disambiguate mentions and link them to ontology entries. It shows all the advantages of combining both approaches, not only for building the benchmark but also for fine-tuning detection models. We also insist on several perspectives of research to improve the accuracy of a combining system and go further on leveraging the complementary approaches.

**Keywords:** Named Entity Recognition and Linking, Transformers, Deep learning, Wikidata

## 1. Introduction

Named entity recognition (NER), disambiguation, and linking (abusively dubbed as Named Entity Linking or NEL) represent a trio of critical tasks within the field of natural language processing (NLP). These tasks are concerned with the extraction and classification of specific references from text, including but not limited to individuals, organizations, geographical locations, and other entity mentions such as dates and emails.

Following the progress in natural language processing, the current state-of-the-art systems are all based on deep learning systems, especially based on the Transformer architecture generating pre-trained models then fine-tuned for the task. But it appears that these systems, used alone, still struggle when context is sparse or noisy or far from the training data characteristics. For example, in the last Multilingual Complex Named Entity Recognition competition (SemEval 2023 (SemEval 2023 task 2: MultiCoNER II)), the winning system does not only leverage these pre-trained contextual models, but also multilingual lexical knowledge bases, namely Wikipedia and Wikidata, especially to disambiguate and link mentions of named entities to knowledge bases entries. The combination consists in this case in creating sentence embeddings from Wikipedia instances linked to title entries, retrieve the most similar contexts to the one to annotate (semantic search) and then feed a Conditional Random Field (CRF) to generate the token annotation. In the same vein, current cutting edge systems (e.g. Wikineural) combine pre-trained models with fine-tuning from silver-annotated versions of

Wikipedia mentions of Named Entities.

Additionally, and surprisingly provided that NER and NEL tasks are on the table for dozen of years, if several reputed benchmarks exist for NER evaluation, NEL evaluation benchmark are still far behind, even if a recent work has proposed a silver dataset (ie a dataset without human validation) from Wikipedia ((Kubeša and Straka, 2023a)). In this context, this paper will provide some insights of the advantages and current limitations of a combination of Deep Learning (DL) systems and Linked Open Data (LOD) knowledge bases, from an experiment aiming to design and compile a new NER and NEL benchmark, created specifically for the purpose of evaluating any system of NER/NEL on Multilingual news textual data.

The paper is further divided into three parts: the first part presents the existing systems for NER and NEL and give some key characteristics of these systems, insisting on the new trend of combining Deep learning systems and Linked Open Ontologies and Lexicons. The second part details the methodology and steps followed to construct this dataset as well as key figures on it. The third and conclusive part presents our insights from this experiment on the DL - LOD combination and some perspectives to push it even further.

## 2. Named Entity Recognition and Linking State-of-The-Art

This section presents the current SOTA models for NER and NEL, and then the most used knowledge bases available.

## 2.1. Deep learning language Models

In this section, we highlight some key characteristics of State-of-the-Art (SOTA) language models for NER and NEL, in the context of the Europe Media Monitor (EMM) we intend to apply them to. SOTA systems for NER/NEL - as all computational linguistic tasks - all rely on embedding representation and pre-trained language models (LM). Several competing LMs with desired features (multilinguality, open source, SOTA on relevant benchmarks) are available. To name the most prominent ones:

- XLM-Roberta-large (XLM-R) (Conneau et al., 2019a): this transformer-based masked language model is the base model for multilingual computational tasks. It was trained on one hundred languages, using more than two terabytes of filtered CommonCrawl data. This model outperformed one of the first multilingual model, multilingual BERT (mBERT) (Devlin et al., 2019). The XLM-R model is still present in leaderboards as a base (see e.g. XTREME benchmark, (Hu et al., 2020), and TNER list of fine-tuned Roberta models (TNER list of fine-tuned models)). These models have been fine-tuned on several datasets, as in WikiNeural (Tedeschi et al., 2021c), which combines a multilingual lexical knowledge base (i.e., BabelNet) and transformer-based architectures (i.e., mBERT) to produce high-quality annotations for multilingual NER. An mBERT model fine-tuned on this silver dataset reach an overall accuracy of 0.80.
- New SOTA multilingual models: these models exhibit strong performance on multilingual tasks and should be considered as swiss-knife pre-trained models. These next-generation transformer models add new tasks during the pre-training steps and evaluation leaderboards show that they enable to gain additional quality. Among them Turing ULR v6, (Patra et al., 2022a) adds a new task at the pre-training step, called cross-lingual contrast (XLCO). The goal of XLCO is to maximize mutual information between the representations of parallel sentences  $c_1$  and  $c_2$ , i.e.,  $I(c_1, c_2)$ . It leverages this new task by creating multi bi-texts. VECO 2.0, (Zhang et al., 2023a), is the most recent SOTA model on the XTREME benchmark, going a step further by aligning not only sentence but also tokens of the bitexts.
- SOTA Models specific to NER/NEL: these specialised models are the winners of the most recent NER/NEL competitions (semEval 2022 task 11: and semEval 2023 task 2:): mLUKE, (Ri et al., 2021a) built on XML-RoBERTa, and additionally trained on

24 languages with entity representations taken from Wikipedia. The model consistently outperforms word-based pre-trained models in various crosslingual transfer tasks. KB-NER, (Wang et al., 2022a) multilingual knowledge base based on Wikipedia to provide related context information to the named entity recognition (NER) model. Given an input sentence, the system retrieves related contexts from the knowledge base. The original input sentences are then augmented with such context information, allowing significantly better contextualized token representations to be captured. Winner on 10 over 13 subtasks (semEval 2022 task 11). in the same vein and a similar architecture, U-RaNER won the semEval 2023 competition: [Github repo](#).

As can be seen from on-going competition, especially to adapt the systems to more complex named entities, new domains and low-resourced languages, even if NER and NEL have now a long trail of research, there are still ways to improve the systems. The last winners of the Multiconer competition show that the main avenue to improve the current systems is to combine the pre-trained transformer models with external knowledge bases, in two main ways:

- by using directly available structured knowledge bases, i.e. Wikidata, especially the feature linking entries to their mention variants, directly at the recognition stage,
- by fine-tuning a secondary transformer model from a textual knowledge base (Wikipedia being the most used) and use it as a complementary resource if the pre-trained model needs additional context information to detect mentions.

## 2.2. Named Entity Linking Knowledge bases

We list here the main existing evaluation datasets again keeping in mind the multilingual and genre features.

- Mewsli-9 (Botha et al., 2020): this dataset contains manually labelled WikiNews articles in 9 different languages. New formulation for multilingual entity linking, where language-specific mentions resolve to a language-agnostic Knowledge Base. A dual encoder was trained in this new setting, building on prior work with improved feature representation, negative mining, and an auxiliary entity-pairing task, to obtain a single entity retrieval model that covers 100+ languages and 20 million entities. The model outperforms state-of-

the-art results from a far more limited cross-lingual linking task. Rare entities and low-resource languages pose challenges at this large-scale.

- [Mewsli-X \(Ruder et al., 2021a\)](#): Mewsli-X is a multilingual dataset of entity mentions appearing in WikiNews and Wikipedia articles, that have been automatically linked to WikiData entries. The primary use case is to evaluate transfer-learning in the zero-shot cross-lingual setting of the XTREME-R benchmark suite: fine-tune a pre-trained model on English Wikipedia examples; evaluate on WikiNews in other languages — given an entity mention in a WikiNews article, retrieve the correct entity from the predefined candidate set by means of its textual description. Mewsli-X constitutes a doubly zero-shot task by construction: at test time, a model has to contend with different languages and a different set of entities from those observed during fine-tuning.
- [DaMuEL \(Kubeša and Straka, 2023b\)](#): a large Multilingual Dataset for Entity Linking containing data in 53 languages. DaMuEL consists of two components: a knowledge base that contains language-agnostic information about entities, including their claims from Wikidata and named entity types (PER, ORG, LOC, EVENT, BRAND, WORK-OF-ART, MANUFACTURED); and Wikipedia texts with entity mentions linked to the knowledge base, along with language-specific text from Wikidata such as labels, aliases, and descriptions, stored separately for each language.

As a matter of fact, even for multipurpose evaluation for LLMs, the WIKIANN benchmark remains the *de facto* standard for multilingual evaluation of core Named Entities, but it is exclusively built from Wikipedia and is known to contain a lot of errors. But, with Mewsli-9 and -X and DaMuEL, new silver standard datasets are built from a combination of DL models and existing Knowledge bases, namely Wikipedia, Wikinews and, as an aggregating KB, Wikidata. This combination of both approaches has the merit of enabling the building of large datasets that in turn can be used to fine-tune DL models. As an inspiration, we will use the DaMuEL dataset that has been built from Wikipedia, by applying a similar method to its news counterpart, Wikinews. We will detail the methodology after a presentation of State-of-The-Art models for NER and NEL. That will enable also to support the need of a new Multilingual News dataset.

### 3. WiNNL (WikiNews Named entity recognition and Linking)

This section introduces WiNNL (WikiNews Named entity recognition and Linking), a new multilingual NER & NEL benchmark based on Wikinews articles. Wikinews, a free-content news source from Wikimedia Foundation, provides a rich and diverse environment for creating a realistic benchmark as it incorporates a wide range of topics and languages. Our benchmark, which for now encompasses 11 European languages, aims to provide a rigorous evaluation framework for multilingual NER/NEL systems. It also facilitates an understanding of how these models perform across different languages on the specific domain of news articles.

Our approach is inspired by the DaMuEL Wikipedia based benchmark ([Kubeša and Straka, 2023b](#)). In this work, the authors describe a pipeline to convert Wikipedia articles by detecting entity types using Wikidata and propagating mentions throughout the article.

#### 3.1. Existing Benchmarks

Within the Joint research Center (JRC) Text and Data Mining Unit, we are facing the challenge of detecting and linking named entities within a live stream of retrieved news articles from more than 80 languages. Our main interest is to detect Persons, Locations and Organisations, as well as temporal information. The linguistic processing chain already includes a NER and NEL dictionary and rule-based system, setup and maintained for more than two decades, and we are in the phase of renewing it with more accurate systems based on Deep learning architectures and pre-trained language models. The first step is to evaluate such state-of-the-art models and the current system towards a benchmark tailored to our needs and constraints, as defined above.

In the core named entities recognition task, the WIKIANN dataset is the most used silver standard, especially in multilingual settings (see XTREME benchmark for example), but it does not correspond to news style and its low quality is often highlighted. Apart from the SlavNER dataset that enables to evaluate slavic languages, all the other datasets are more interesting for specific Named Entities or difficult cases (e.g. MultiCoNER 1 and 2). As a result, there is a strong need for developing a gold standard for multilingual news genre.

**Michael Jordan** (born 1963) is an American businessman and former professional basketball player.

**Michael Jordan** or **Mike Jordan** may also refer to:

**People** [edit]

---

**Sports** [edit]

- **Michael Jordan (footballer)** (born 1986), English goalkeeper
- **Mike Jordan (basketball, born 1977)**, American basketball coach and former professional player
- **Mike Jordan (racing driver)** (born 1958), English racing driver
- **Mike Jordan (baseball, born 1863)** (1863–1940), American baseball player
- **Mike Jordan (cornerback)** (born 1992), American football cornerback
- **Michael Jordan (offensive lineman)** (born 1998), American football offensive lineman
- **Michal Jordán** (born 1990), Czech ice hockey player

**Other people** [edit]

- **Michael B. Jordan** (born 1987), American actor
- **Michael I. Jordan** (born 1956), American researcher in statistics, machine learning, and artificial intelligence
- **Michael Jordan (insolvency baron)** (born 1931), English businessman
- **Michael Jordan (Irish politician)**, Irish Farmers' Party TD from Wexford, 1927–1932
- **Michael H. Jordan** (1936–2010), American executive for CBS, PepsiCo, Westinghouse
- **Michael Jordan (mycologist)**, English mycologist

**Other uses** [edit]

---

- "Michael Jordan", a 2000 song by Five for Fighting from *America Town*
- "Michael Jordan", a 2010 song by Kendrick Lamar featuring Schoolboy Q from *Overly Dedicated*
- *Michael Jordan: An American Hero*, a 1999 American TV-film based on the basketball player's story
- *Michael Jordan: Chaos in the Windy City*, a 1994 video game featuring the basketball player

Figure 1: Named Entity Linking illustration (Wikipedia page)

### 3.2. Methodology to setup the Multilingual News dataset

#### Dataset design

WiNNL’s annotation scheme prioritises three core categories of entities: PER, ORG and LOC. These categories refer to person names, organisations and geographical locations, respectively. We opt for a word-level annotation scheme, where a word can be tagged as being the beginning of an entity, inside of an entity mention or outside of any annotation. This is indicated by Inside-Outside-Beginning (IOB) tags, where the prefix I- or B- is attached to the type (PER, ORG or LOC) of the entity for each word (Ramshaw and Marcus, 1999).

#### Data collection process

The process of collecting and cleaning our multilingual NER/L dataset is initiated by downloading the HTML of articles from Wikinews. This source was chosen due to its extensive cross-linguistic coverage and the rich network of interlinked entities it contains. In Wikinews, authors of articles generally tag each first occurrence of a named entity with their respective Wikipedia page. These links are denoted as `<a>` tags with the class `extiw` in HTML. From the Wikipedia page we extract the unambiguous QID of the entity, that uniquely identifies the

item in Wikidata across all languages.

The next step in the pipeline involves the classification of these entity QID’s. This is achieved through a SPARQL query against a local instance of the Wikidata dataset, based on the simplified qEndpoint (Willerval et al., 2022). Entries classes in Wikidata are organised as a graph, where each instance belongs to one or more classes and each class has one or more superclasses. The query seeks to traverse the superclasses of a Wikidata instance until one of several predetermined base types is encountered, or until a defined depth limit is reached. This mechanism allows us to categorise and detect only entities of the types we are interested in.

Following the classification of entities, the system then maps all aliases of the entity that are found on Wikidata to the QID of the entity type. These aliases serve as additional textual representations of the entities and are crucial for detecting all possible mentions. These steps of resolving the type of entity based on a Wikipedia link are illustrated in Figure 2.

The final stage of the data collection pipeline involves propagating the entity links throughout the article, using the knowledge base generated by the Link resolver. This stage is illustrated with an example in Figure 3. The system scans through all n-grams of the article text and creates offset-based

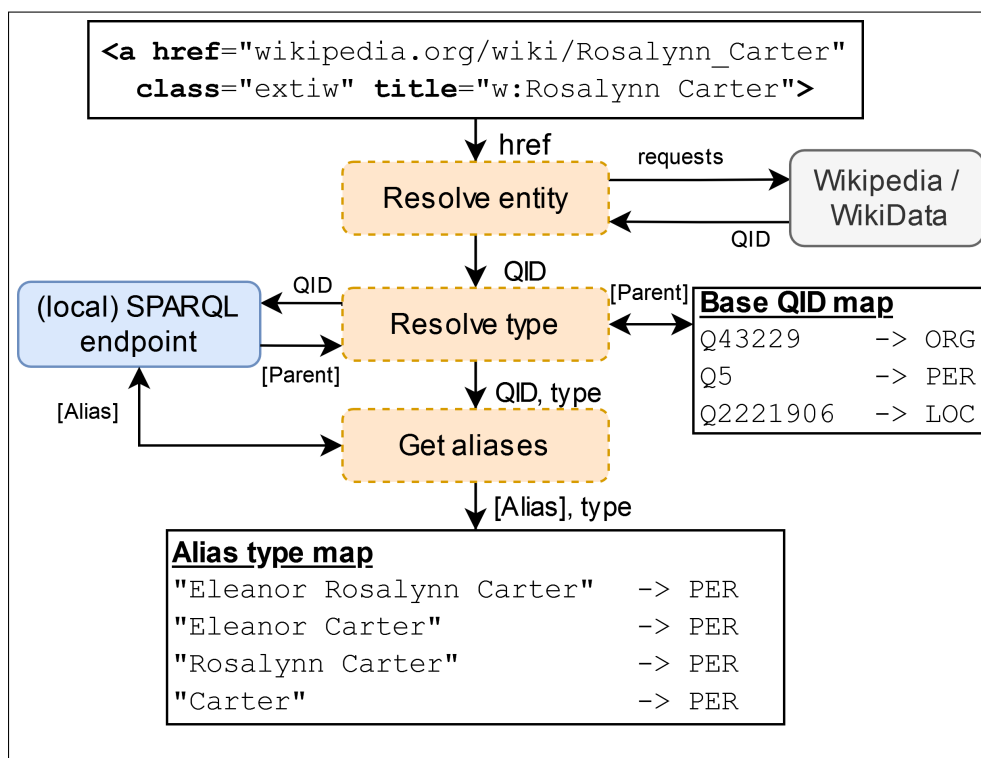


Figure 2: Example of how the link resolving pipeline builds or extends a mapping from aliases to entity types, in this case based on a link to a First Lady of the United States. Once the unambiguous QID is found, the resolver will iteratively go through the levels of parent entity classes until a parent is found that is present in the base QID mapping. In this case Q5 (human) is found in the parent QIDs and so the aliases of the lady are mapped to “PER”.

annotations for each combination of n-grams that matches one of the recognised aliases. This process aims to ensure that all potential mentions of the entities are captured and annotated with the correct type and QID. The combination of the article content stripped of HTML tags and the list of annotations is represented with the Article class.

### Post-processing

To render the scraped data suitable for evaluation, the articles must be segmented into sentences and annotated with Inside-Outside-Beginning (IOB) tags. Sentence termination is identified using the multilingual spaCy sentence model (Honnibal and Montani, 2017). To optimise the dataset’s size and enhance its usability, consecutive newline characters are compacted into a single white-space.

Subsequently, all sentences devoid of any named entity are eliminated. The remaining data is validated through a multilingual language model, specifically fine-tuned for Named Entity Recognition (NER). For this version of the dataset, we used distilbert-base-multilingual-cased-ner-hrl (Adelani, 2024). This NER finetuned version of distilbert has been trained on news data for several high resource languages. If the system-generated tags coincide with the model, the sentence is retained.

By contrast, sentences for which our system yields fewer or different annotations than the model are discarded. Formally, for every sentence  $x$  comprised of  $n$  tokens  $x_1, \dots, x_n$ , we evaluate the annotation (i.e., a named entity tag)  $y_i$  produced by our method for each token  $x_i$  against the one predicted by the auxiliary language model,  $\hat{y}_i$ . We retain the sentence if at least one annotation  $y_i \neq O$  is present, and every  $y_i \neq O$  possesses the same annotation as the corresponding  $\hat{y}_i$ . This procedure culminates in an enhanced precision of our annotations, as they are validated through this ensemble approach. These measures effectively reduce the volume of the collected data to approximately 0.4 to 2% of the initial scraped content, the percentage varying according to the language.

Each retained sentence is then tokenised, during which every token’s annotation type and Wikidata ID (QID) are encoded in IOB format. The final dataset comprises items with the following attributes: the original sentence, the sentence tokens, IOB-NER tags for each token, IOB-QID tags, the sentence language, and the complete URL of the source article. The steps of this pipeline are illustrated in Figure 4.

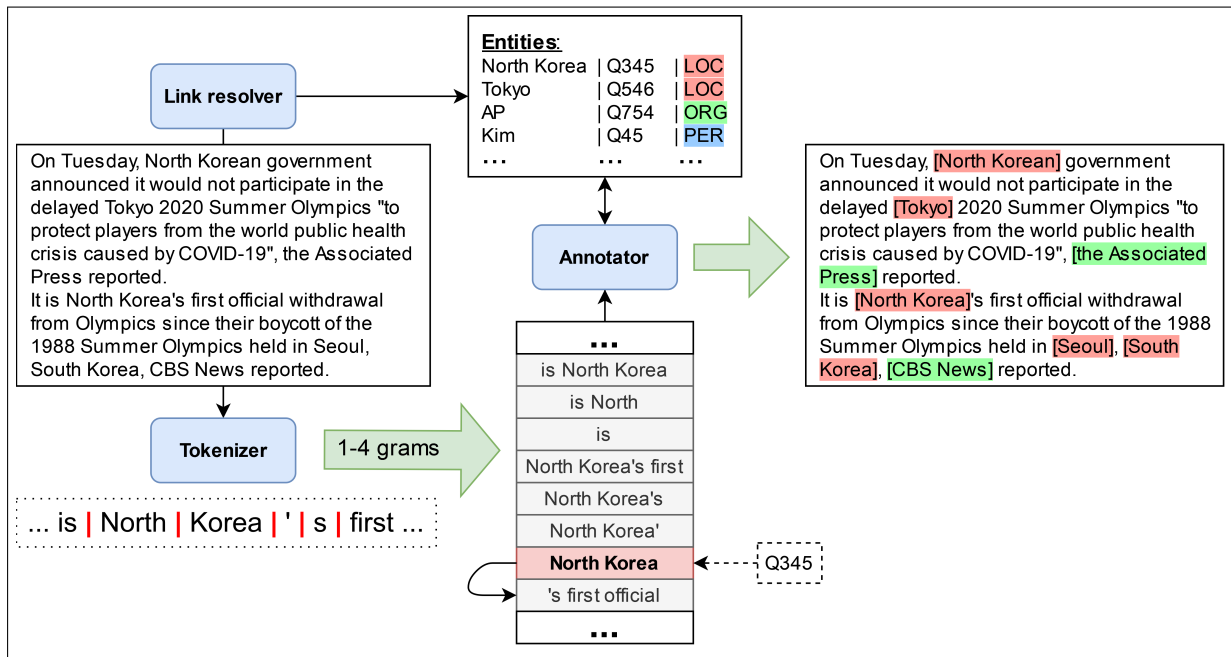


Figure 3: Illustration of the mention detection algorithm of the automatic Annotator used in WINLL with a window size of 4.

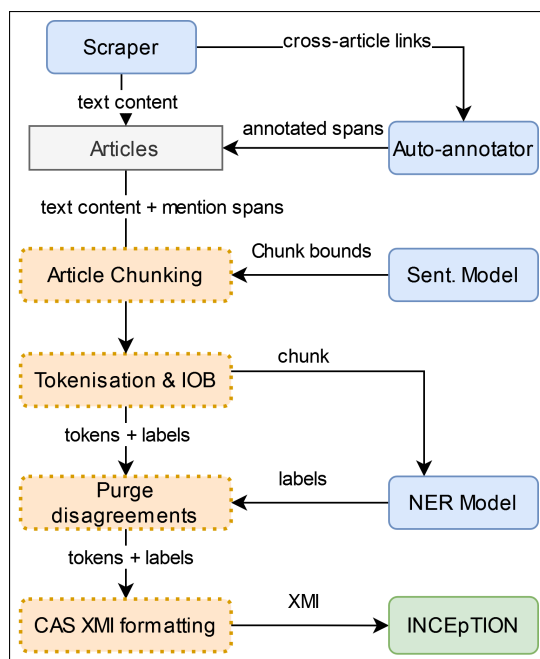


Figure 4: The post-processing pipeline used for WINLL.

### Human validation

In the final phase of data preparation, we instituted a rigorous quality control process to ensure the accuracy of our annotations. This process involved manual verification of annotated sentences by native speakers corresponding to each language in the study. These evaluators were selected from a pool of international research trainees at the Euro-

pean Joint Research Centre, who volunteered to help with the project. A locally hosted instance of the INCEpTION annotation tool was employed for this verification process (Klie et al., 2018). The evaluators were instructed to modify only those NER tags that were inaccurately assigned and to delete sentences with erroneous entity links entirely in order to streamline the verification process.

## 4. Results and Evaluation of DL models on the benchmark

In this section, we provide a comprehensive overview of the Multilingual Wikinews NER/L dataset, denoted as WINLL v1. The current version of the dataset encompasses 11 predominantly European languages, namely Dutch, English, French, German, Swedish, Spanish, Portuguese, Italian, Greek, Polish, and Russian. Table 1 provides detailed statistics on the number of unique articles parsed, the count of sentences, tokens, and entities pertaining to each language.

Subsequently, we juxtaposed the outcomes of human validation against the original system-generated annotations. For each language, precision, recall, and F1 scores were computed, utilising the validated tags as the ground-truth. We use span-based metrics as opposed to token-based, as for our downstream purpose it is more useful to evaluate with metrics at a full named-entity level. This project makes use of the SemEval 2013-9.1 based evaluation library “nervaluate” (Batista and Upson, 2020). The results of this comparative anal-

Language	Articles	Sentences	Entities	Tokens	Med. length
German (de)	482	1000	1551	21222	119
English (en)	431	1000	1740	27037	126
Dutch (nl)	720	1500	2150	31584	112
Polish (pl)	1035	1606	2148	30615	101
Italian (it)	636	1155	1755	36669	150
Spanish (es)	319	615	1035	20503	160
Portuguese (pt)	342	539	695	14904	129
French (fr)	607	989	1661	34902	154
Russian (ru)	428	720	904	15371	104
Swedish (sv)	465	758	1024	139960	111
Greek (el)	348	520	716	13701	134

Table 1: Number of parsed articles, sentences, named entities, tokens and the median length in characters for the data in each language.

ysis are depicted in Table 2.

Finally, we evaluated the performance of three cutting-edge models on our dataset and compared the outcomes against other prevalent multilingual NER and NEL benchmarks. Our primary focus was to discern the impact of limiting the scope solely to sentences within the news domain on the performance of widely-used models. The findings of these named-entity level assessments are delineated in Table 3.

## 5. Conclusion and Perspectives

Based on the evaluations presented in Section 4, we conclude that the scraper pipeline has an average precision of .942, and an average recall of .917.

Although the ensemble system correctly identifies a significant portion of named entities, there are instances where it may fail to detect some entities. This shows that there is still need for a human correction step. For future iterations, we propose the use of more rigorous checking mechanisms by leveraging specific language models for each language.

Analysis of the human validation points to the system being most accurate for PER entities. A noteworthy observation by the human validators is the occasional tagging of common words that do not typically refer to named entities. This occurs in Wikinews articles when the context makes it clear what the common name refers to. An example is the term “the forest”, which could be linked to the Amazon Rainforest. To address this issue, we could implement stricter language model agreement checking. However, it is important to note that such a measure may also lead to a decrease in recall, as it might fail to identify some legitimate and linked named entities that the language model does not detect.

Therefore, the challenge lies in striking a balance

between improving the precision of the NER system and maintaining, or potentially enhancing, its recall. This delicate balance will be our focus in the further development and refinement of the system.

Another consideration is the language support of Wikinews. Version 1 of WiNLL includes only 11 languages, with the main reason being the difficulty of scraping high quality tagged sentences for the other languages. For example, in the Russian language almost none of the named entities are tagged in articles. This increases the amount of articles the scraper must download, and in turn also the network overhead, to achieve a sufficiently large dataset.

In future work, the scraping pipeline could be adapted to work with other news sources, such as Voxeurop or any open sourced news website. This would involve the creation of a more elaborate interface between arbitrary HTML page sources and Wikidata objects. This will also imply to use a language model to detect the mentions, than feed the results to our pipeline and then validate the projection and linking. This will enable to see the added-value of the language model for entity mention detection and entity linking. In that scenario, another open questions arise and notably how to add new recognized entities to the Wikidata repository?

As a global conclusion, as has been shown here, to build our benchmark, we combined human annotation, Deep learning language models and Knowledge bases. The main outcome here is a benchmark that can be considered a quasi-gold standard, as it has been manually curated at the end of the process. In turn, the dataset can then be used to fine-tune a model for a specific genre (here news) and specific languages and thus create a SOTA model. As shown, the human validation is quite light, as it consists mainly in validating or invalidating the data already recognized by the KB and/or validated by the DL model.

Language	Prec.	Rec.	F1	LOC	ORG	PER
German (de)	.986	.909	.947	.922	.929	.983
English (en)	.956	.938	.947	.931	.951	.956
Dutch (nl)	.936	.906	.921	.912	.869	.975
Italian (it)	.986	.907	.945	.882	.945	.968
Spanish (es)	.944	.944	.943	.884	.941	.975
Portuguese (pt)	.965	.936	.950	.958	.927	.965
French (fr)	.879	.870	.875	.813	.826	.944
Greek (el)	.885	.925	.905	.935	.842	.989

Table 2: Comparison of the accuracy of the system generated tags for each language, based on the human-validated samples.  $F_1$  scores for each specific tag are given on the right. All metrics calculated on named-entity level. Mean  $F_1$  = .927.

Dataset	Model	de	en	nl	pl	it	es	pt	fr	ru	sv	el
WikiANN	XLM-Roberta	.354	.373	.325	.272	.275	.273	.317	.345	.048	.223	.024
	wikineural	.715	.554	.716	.758	.696	.671	.628	.688	.361	.733	.661
	distilbert	.657	.521	.653	.694	.584	.589	.549	.542	.331	.715	.505
UNER	XLM-Roberta	.496	.497	-	-	-	-	.404	-	.081	.418	-
	wikineural	.771	.808	-	-	-	-	.838	-	.687	.839	-
	distilbert	.816	.809	-	-	-	-	.847	-	.714	.865	-
	WiNNL-model	.762	.772	-	-	-	-	.804	-	.670	.821	-
WiNNL	XLM-Roberta	.584	.561	.561	.297*	.388	.449	.415	.409	.066*	.543*	.071
	wikineural	.835	.827	.843	.759*	.753	.884	.875	.835	.785*	.876*	.724
	distilbert	.828	.810	.851	.756*	.787	.878	.875	.839	.816*	.883*	.712

Table 3: Named-entity span level  $F_1$  scores of wikineural, distilbert-cased and XLM-Roberta on the WikiANN, UNER and WiNNL NER benchmarks. Scores indicated with \* are not validated by humans. We also evaluated the UNER benchmark with a multilingual distilbert model finetuned on our human validated WiNNL dataset (WiNNL-model).

In the next version of the benchmark-building system, for the remaining languages, as we don't have enough human annotated data sources, we will rely on a SOTA language model to first annotate mentions and then the propagation of mentions will be ensured by the KB mentions feature. That will open other questions, the way around, on the added-value of DL for updating KB.

## 6. Bibliographical References

1998. Message understanding conference-7: A research retrospective. In *Proceedings of a Workshop held at Fairfax Virginia USA: George Mason University Press*.
- 2016–2023. [entity-fishing](https://github.com/kermitt2/entity-fishing). <https://github.com/kermitt2/entity-fishing>.
- David Ifeoluwa Adelani. 2024. [distilbert-base-multilingual-cased-ner-hrl](#).
- Kabir Ahuja, Rishav Hada, Millicent A. Ochieng, Prachi Jain, Harshita Diddee, Samuel Maina, Tanuja Ganu, Sameer Segal, Maxamed Axmed, Kalika Bali, and Sunayana Sitaram. 2023. [Mega](#).
- Multilingual evaluation of generative ai. *ArXiv*, abs/2303.12528.
- Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual string embeddings for sequence labeling. In *COLING 2018, 27th International Conference on Computational Linguistics*, pages 1638–1649.
- Tom Ayoola, Joseph Fisher, and Andrea Pierleoni. 2022a. Improving entity disambiguation by reasoning over a knowledge base. In *NAACL*.
- Tom Ayoola, Shubhi Tyagi, Joseph Fisher, and Christos Christodoulopoulos. 2022b. [Refined: An efficient zero-shot-capable approach to end-to-end entity linking](#). *ArXiv*, abs/2207.04108.
- David Batista and Matthew Antony Upson. 2020. [nervaluate](#).
- David S. Batista. 2024. [Ner datasets](#).
- Darina Benikova, Chris Biemann, and Marc Reznicek. 2014. Germeval-2014 named entity recognition shared task—guidelines and dataset. In *Proceedings of the KONVENS GermEval Shared Task on Named Entity Recognition*.



- Priyanka Bose, Sriram Srinivasan, William C. Sleeman, Jatinder R. Palta, Rishabh Kapoor, and Preetam Ghosh. 2021. A survey on recent named entity recognition and relationship extraction techniques on clinical texts. *Applied Sciences*.
- Jan A. Botha, Zifei Shan, and Daniel Gillick. 2020. [Entity Linking in 100 Languages](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 7833–7845, Online. Association for Computational Linguistics.
- Nicola De Cao, Ledell Yu Wu, Kashyap Papat, Mikel Artetxe, Naman Goyal, Mikhail Plekhanov, Luke Zettlemoyer, Nicola Cancedda, Sebastian Riedel, and Fabio Petroni. 2021. [Multilingual autoregressive entity linking](#). *Transactions of the Association for Computational Linguistics*, 10:274–290.
- Sophie Chesney, Guillaume Jacquet, Ralf Steinberger, and Jakub Piskorski. 2017. [Multi-word entity classification in a highly multilingual environment](#). In *MWE@EACL*.
- Zewen Chi, Li Dong, Furu Wei, Nan Yang, Saksham Singhal, Wenhui Wang, Xia Song, Xian-Ling Mao, Heyan Huang, and M. Zhou. 2020. [Infoxlm: An information-theoretic framework for cross-lingual language model pre-training](#). In *North American Chapter of the Association for Computational Linguistics*.
- Zewen Chi, Shaohan Huang, Li Dong, Shuming Ma, Saksham Singhal, Payal Bajaj, Xia Song, and Furu Wei. 2021. [Xlm-e: Cross-lingual language model pre-training via electra](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019a. [Unsupervised cross-lingual representation learning at scale](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019b. [Unsupervised cross-lingual representation learning at scale](#). *CoRR*, abs/1911.02116.
- Leon Derczynski, Eric Nichols, Marieke van Erp, and Nut Limsopatham. 2017. [Results of the WNUT2017 shared task on novel and emerging entity recognition](#). In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 140–147, Copenhagen, Denmark. Association for Computational Linguistics.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [Bert: Pre-training of deep bidirectional transformers for language understanding](#). In *North American Chapter of the Association for Computational Linguistics*.
- Ning Ding, Guangwei Xu, Yulin Chen, Xiaobin Wang, Xu Han, Pengjun Xie, Haitao Zheng, and Zhiyuan Liu. 2021. [Few-nerd: A few-shot named entity recognition dataset](#). *CoRR*, abs/2105.07464.
- Sumanth Doddapaneni, Gowtham Ramesh, Anoop Kunchukuttan, Pratyush Kumar, and Mitesh M. Khapra. 2021. A primer on pretrained multilingual language models. *ArXiv*, abs/2107.00676.
- Henrique dos Santos Cardoso and José Gabriel Pereira Lopes Silva. 2018. [Pereria: A portuguese named entity recognition dataset](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC'18)*. *European Language Resources Association (ELRA)*.
- Maud Ehrmann, Ahmed Hamdi, Elvys Linhares Pontes, Matteo Romanello, and Antoine Doucet. 2023. [Named entity recognition and classification in historical documents: A survey](#). *ACM Computing Surveys*, 56:1 – 47.
- Maud Ehrmann, Guillaume Jacquet, and Ralf Steinberger. 2016. [Jrc-names: Multilingual entity name variants and titles as linked data](#). *Semantic Web*, 8:283–295.
- Yotam Eshel, Noam Cohen, Kira Radinsky, Shaul Markovitch, Ikuya Yamada, and Omer Levy. 2017. [Named entity disambiguation for noisy text](#). In *Conference on Computational Natural Language Learning*.
- Association for Computational Linguistics. Accessed: April 17, 2023. Named entity recognition (ner) bibliography. [https://www.aclweb.org/aclwiki/Named\\_entity\\_recognition\\_\(NER\)\\_bibliography](https://www.aclweb.org/aclwiki/Named_entity_recognition_(NER)_bibliography).
- Jinlan Fu, Pengfei Liu, and Graham Neubig. 2020. Interpretable multi-dataset evaluation for named entity recognition. *ArXiv*, abs/2011.06854.
- Jingru Gan, Jinchang Luo, Haiwei Wang, Shuhui Wang, W. He, and Qingming Huang. 2021. [Multimodal entity linking: A new dataset and a baseline](#). *Proceedings of the 29th ACM International Conference on Multimedia*.
- Ralph Grishman and Beth Sundheim. 1996. Message understanding conference-6: A brief history. In *Proceedings of the 16th conference on Computational linguistics-Volume 1*, pages 466–471. Association for Computational Linguistics.

- Johannes Hoffart, Yasemin Altun, and Gerhard Weikum. 2014. [Discovering emerging entities with ambiguous names](#). *Proceedings of the 23rd international conference on World wide web*.
- Johannes Hoffart, Mohamed Amir Yosef, Ilaria Bordino, Hagen Fürstenau, Manfred Pinkal, Marc Spaniol, Bilyana Taneva, Stefan Thater, and Gerhard Weikum. 2011. [Robust disambiguation of named entities in text](#). In *Conference on Empirical Methods in Natural Language Processing*.
- Matthew Honnibal and Ines Montani. 2017. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear.
- Junjie Hu, Sebastian Ruder, Aditya Siddhant, Graham Neubig, Orhan Firat, and Melvin Johnson. 2020. [Xtreme: A massively multilingual multi-task benchmark for evaluating cross-lingual generalization](#). *ArXiv*, abs/2003.11080.
- Guillaume Jacquet, Maud Ehrmann, and Ralf Steinberger. 2014. [Clustering of multi-word named entity variants: Multilingual evaluation](#). In *International Conference on Language Resources and Evaluation*.
- Guillaume Jacquet, Maud Ehrmann, Ralf Steinberger, and Jaakko J. Väyrynen. 2016. [Cross-lingual linking of multi-word entities and their corresponding acronyms](#). In *International Conference on Language Resources and Evaluation*.
- Guillaume Jacquet, Jakub Piskorski, and Sophie Chesney. 2019a. [Out-of-context fine-grained multi-word entity classification: exploring token, character n-gram and nn-based models for multilingual entity classification](#). *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*.
- Guillaume Jacquet, Jakub Piskorski, Hristo Tanev, and Ralf Steinberger. 2019b. [JRC TMA-CC: Slavic named entity recognition and linking. participation in the BSNLP-2019 shared task](#). In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 100–104, Florence, Italy. Association for Computational Linguistics.
- Piskorski Jakub. 2007. [Express - extraction pattern recognition engine and specification suite](#).
- Renato Stoffalette João. 2020. [On the temporality of priors in entity linking](#). *Advances in Information Retrieval*, 12036:375 – 382.
- Poonam Kashtriya, Pardeep Singh, and Parul Bansal. 2023. A review on clinical named entity recognition. *2022 OPJU International Technology Conference on Emerging Technologies for Sustainable Development (OTCON)*, pages 1–6.
- Hossein Keshavarz, Zografoula Vagena, Pigi Kouki, Ilias Fountalis, Mehdi Mabrouki, Aziz Belaweid, and Nikolaos Vasiloglou. 2022. Named entity recognition in long documents: An end-to-end case study in the legal domain. *2022 IEEE International Conference on Big Data (Big Data)*, pages 2024–2033.
- Jan-Christoph Klie, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho, and Iryna Gurevych. 2018. The inception platform: Machine-assisted and knowledge-oriented interactive annotation. In *proceedings of the 27th international conference on computational linguistics: system demonstrations*, pages 5–9.
- Veysel Kocaman and David Talby. 2020. [Biomedical named entity recognition at scale](#). *CoRR*, abs/2011.06315.
- Martin Krallinger, Obdulia Rabal, Florian Leitner, Miguel Vázquez, David Salgado, Zhiyong Lu, Robert Leaman, Yanan Lu, Dong-Hong Ji, Daniel M. Lowe, Roger A. Sayle, Riza Theresa Batista-Navarro, Rafal Rak, Torsten Huber, Tim Rocktäschel, Sérgio Matos, David Campos, Buzhou Tang, Hua Xu, Tsendsuren Munkhdalai, Keun Ho Ryu, S. V. Ramanan, P. Senthil Nathan, Slavko Žitnik, Marko Bajec, Lutz Weber, Matthias Irmer, Saber Ahmad Akhondi, Jan A. Kors, Shuo Xu, Xin An, Utpal Kumar Sikdar, Asif Ekbal, Masaharu Yoshioka, Thaer M. Dieb, Miji Joo young Choi, Karin M. Verspoor, Madian Khabisa, C. Lee Giles, Hongfang Liu, K. E. Ravikumar, Andre Lamurias, Francisco M. Couto, Hong-Jie Dai, Richard Tzong-Han Tsai, C Ata, Tolga Can, Anabel Usie, Rui Alves, Isabel Segura-Bedmar, Paloma Martínez, Julen Oyarzábal, and Alfonso Valencia. 2015. The chemdner corpus of chemicals and drugs and its annotation principles. *Journal of Cheminformatics*, 7:S2 – S2.
- David Kubeša and Milan Straka. 2023a. [Damuel: A large multilingual dataset for entity linking](#).
- David Kubeša and Milan Straka. 2023b. [Damuel: A large multilingual dataset for entity linking](#). *ArXiv*, abs/2306.09288.
- Dilek Küçük, Guillaume Jacquet, and Ralf Steinberger. 2014. [Named entity recognition on turkish tweets](#). In *International Conference on Language Resources and Evaluation*.
- Dilek Küçük and Ralf Steinberger. 2014. [Experiments to improve named entity recognition on turkish tweets](#). *ArXiv*, abs/1410.8668.

- John D. Lafferty, Andrew McCallum, and Fernando Pereira. 2001. [Conditional random fields: Probabilistic models for segmenting and labeling sequence data](#). In *International Conference on Machine Learning*.
- Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pre-training approach. *ArXiv*, abs/1907.11692.
- Pei-Chi Lo and Ee-Peng Lim. 2020. [Interactive entity linking using entity-word representations](#). *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*.
- Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022a. [Multiconer: A large-scale multilingual dataset for complex named entity recognition](#). In *International Conference on Computational Linguistics*.
- Shervin Malmasi, Anjie Fang, Besnik Fetahu, Sudipta Kar, and Oleg Rokhlenko. 2022b. [Semeval-2022 task 11: Multilingual complex named entity recognition \(multiconer\)](#). In *International Workshop on Semantic Evaluation*.
- Katja Markert and Malvina Nissim. 2007. [Semeval-2007 task 08: Metonymy resolution at semeval-2007](#). In *International Workshop on Semantic Evaluation*.
- Cedric Moller, Jens Lehmann, and Ricardo Usbeck. 2021. [Survey on english entity linking on wikidata](#). *ArXiv*, abs/2112.01989.
- David Nadeau and Satoshi Sekine. 2007. A survey of named entity recognition and classification. *Lingvisticae Investigationes*, 30(1):3–26.
- Joel Nothman, James R Curran, and Tara Murphy. 2008. Transforming wikipedia into named entity training data. In *Proceedings of the Australasian Language Technology Association Workshop 2008*, pages 124–132. Citeseer.
- Damien Nouvel. 2024. [Corpora for named entities](#).
- Italo Lopes Oliveira, Renato Fileto, René Speck, René Speck, Luís Paulo F. Garcia, Diego Mousallem, Diego Moussallem, Jens Lehmann, and Jens Lehmann. 2021. [Towards holistic entity linking: Survey and directions](#). *Inf. Syst.*, 95:101624.
- Yasumasa Onoe and Greg Durrett. 2019. [Fine-grained entity typing for domain independent entity linking](#). *ArXiv*, abs/1909.05780.
- Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017a. [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)*, pages 1946–1958, Vancouver, Canada. Association for Computational Linguistics.
- Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji. 2017b. [Cross-lingual name tagging and linking for 282 languages](#). In *Proceedings of ACL 2017*, pages 1946–1958.
- Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, Heng Ji, Diana Maynard, Yuan Yao, Zhijing Zhao, Yaqing Guan, et al. 2020. [Cross-lingual name tagging and linking for 282 languages](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*.
- Barun Patra, Saksham Singhal, Shaohan Huang, Zewen Chi, Li Dong, Furu Wei, Vishrav Chaudhary, and Xia Song. 2022a. [Beyond english-centric bitexts for better multilingual language representation learning](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Barun Patra, Saksham Singhal, Shaohan Huang, Zewen Chi, Li Dong, Furu Wei, Vishrav Chaudhary, and Xia Song. 2022b. [Beyond english-centric bitexts for better multilingual language representation learning](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Jakub Piskorski. 2008. [Corleone - core linguistic entity online extraction](#). Technical Report EUR 23393 EN, OPOCE, Luxembourg (Luxembourg). JRC45952.
- Jakub Piskorski, Laska Laskova, Michał Marcińczuk, Lidia Pivovarova, Pavel Přibáň, Josef Steinberger, and Roman Yangarber. 2019. [The second cross-lingual challenge on recognition, normalization, classification, and linking of named entities across Slavic languages](#). In *Proceedings of the 7th Workshop on Balto-Slavic Natural Language Processing*, pages 63–74, Florence, Italy. Association for Computational Linguistics.
- Jakub Piskorski, Karol Wieloch, and Marcin Sydow. 2009. On knowledge-poor methods for person name matching and lemmatization for highly inflectional languages. *Information Retrieval*, 12(3):275–299.

- Mikhail Plekhanov, Nora Kassner, Kashyap Popat, Louis Martin, Simone Merello, Borislav M. Kozlovskii, Frédéric A. Dreyer, and Nicola Cancedda. 2023. [Multilingual end to end entity linking](#). *ArXiv*, abs/2306.08896.
- Marco Polignano, Marco Degemmis, and Giovanni Semeraro. 2021. Comparing transformer-based ner approaches for analysing textual medical diagnoses. In *Conference and Labs of the Evaluation Forum*.
- Bruno Pouliquen. 2008. [Similarity of names across scripts: Edit distance using learned costs of n-grams](#). In *GoTAL*.
- Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. 2013. [Towards robust linguistic analysis using OntoNotes](#). In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 143–152, Sofia, Bulgaria. Association for Computational Linguistics.
- Vera Provatorova, Samarth Bhargav, Svitlana Vakulenko, and Evangelos Kanoulas. 2021. [Robustness evaluation of entity disambiguation using prior probes: the case of entity overshadowing](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10501–10510, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Chengwei Qin, Aston Zhang, Zhuosheng Zhang, Jiaao Chen, Michihiro Yasunaga, and Diyi Yang. 2023. [Is chatgpt a general-purpose natural language processing task solver?](#) *ArXiv*, abs/2302.06476.
- Afshin Rahimi, Yuan Li, and Trevor Cohn. 2019. [Massively multilingual transfer for NER](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 151–164, Florence, Italy. Association for Computational Linguistics.
- Steinberger Ralf, Guillaume Jacquet, and Della Rocca Leonida. 2014. [Creation and use of multilingual named entity variant dictionaries](#).
- Lance A Ramshaw and Mitchell P Marcus. 1999. Text chunking using transformation-based learning. In *Natural language processing using very large corpora*, pages 157–176. Springer.
- L. Ratinov and D. Roth. 2011. [Design challenges and misconceptions in named entity recognition](#). In *Proceedings of the 13th Conference on Computational Natural Language Learning*, pages 147–155.
- L.-A. Ratinov, D. Roth, D. Downey, and M. Anderson. 2011. Local and global algorithms for disambiguation to wikipedia. In *Annual Meeting of the Association for Computational Linguistics*.
- Ryokan Ri, Ikuya Yamada, and Yoshimasa Tsuruoka. 2021a. [mluke: The power of entity representations in multilingual pretrained language models](#). *CoRR*, abs/2110.08151.
- Ryokan Ri, Ikuya Yamada, and Yoshimasa Tsuruoka. 2021b. [mluke: The power of entity representations in multilingual pretrained language models](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Juan Diego Rodriguez. 2024. [Entity recognition datasets](#).
- Henry Rosales-Méndez, Aidan Hogan, and Bárbara Poblete. 2018. [Voxel: A benchmark dataset for multilingual entity linking](#). In *International Workshop on the Semantic Web*.
- Arya Roy. 2021. [Recent trends in named entity recognition \(NER\)](#). *CoRR*, abs/2101.11420.
- Sebastian Ruder. 2022. The State of Multilingual AI. <http://ruder.io/state-of-multilingual-ai/>.
- Sebastian Ruder, Noah Constant, Jan Botha, Aditya Siddhant, Orhan Firat, Jinlan Fu, Pengfei Liu, Junjie Hu, Dan Garrette, Graham Neubig, and Melvin Johnson. 2021a. [XTREME-R: Towards more challenging and nuanced multilingual evaluation](#). In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 10215–10245, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Sebastian Ruder, Noah Constant, Jan A. Botha, Aditya Siddhant, Orhan Firat, Jinlan Fu, Pengfei Liu, Junjie Hu, Graham Neubig, and Melvin Johnson. 2021b. [Xtreme-r: Towards more challenging and nuanced multilingual evaluation](#). *ArXiv*, abs/2104.07412.
- Erik Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In *Conference on Computational Natural Language Learning*.
- Satoshi Sekine, Kiyoshi Sudo, and Chikashi Nobata. 2002. [Extended named entity hierarchy](#). In *International Conference on Language Resources and Evaluation*.
- Ozge Sevgili, Artem Shelmanov, Mikhail V. Arkhipov, Alexander Panchenko, and Christian

- Biemann. 2020. [Neural entity linking: A survey of models based on deep learning](#). *Semantic Web*, 13:527–570.
- Akshaya Kesarimangalam Srinivasan and Sowmya Vajjala. 2023. [A multilingual evaluation of ner robustness to adversarial inputs](#). *ArXiv*, abs/2305.18933.
- Ralf Steinberger and Bruno Pouliquen. 2007. [Cross-lingual named entity recognition](#). *Linguisticae Investigationes*, 30(1):135–162.
- Ralf Steinberger, Bruno Pouliquen, Mijail Alexandrov Kabadjov, Jenya Belyaeva, and Erik Van der Goot. 2011. [Jrc-names: A freely available, highly multilingual named entity resource](#). *ArXiv*, abs/1309.6162.
- Stephanie Strassel and Jennifer Tracey. 2016. [LORELEI language packs: Data, tools, and resources for technology development in low resource languages](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3273–3280, Portorož, Slovenia. European Language Resources Association (ELRA).
- Zeqi Tan, Shen Huang, Zixia Jia, Jiong Cai, Yinghui Li, Weiming Lu, Yueting Zhuang, Kewei Tu, Pengjun Xie, Fei Huang, and Yong Jiang. 2023. [Damo-nlp at semeval-2023 task 2: A unified retrieval-augmented system for multilingual named entity recognition](#).
- Simone Tedeschi, Simone Conia, Francesco Cecconi, and Roberto Navigli. 2021a. [Named entity recognition for entity linking: What works and what's next](#). In *Conference on Empirical Methods in Natural Language Processing*.
- Simone Tedeschi, Simone Conia, Francesco Cecconi, and Roberto Navigli. 2021b. [Named Entity Recognition for Entity Linking: What works and what's next](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2584–2596, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Simone Tedeschi, Valentino Maiorca, Niccolò Campolungo, Francesco Cecconi, and Roberto Navigli. 2021c. [WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2521–2533, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Simone Tedeschi, Valentino Maiorca, Niccolò Campolungo, Francesco Cecconi, and Roberto Navigli. 2021d. [WikiNEuRal: Combined neural and knowledge-based silver data creation for multilingual NER](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 2521–2533, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Simone Tedeschi and Roberto Navigli. 2022. [Multi-NERD: A multilingual, multi-genre and fine-grained dataset for named entity recognition \(and disambiguation\)](#). In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 801–812, Seattle, United States. Association for Computational Linguistics.
- Erik F. Tjong Kim Sang and Sabine Buchholz. 2002. Introduction to the conll-2002 shared task: Language-independent named entity recognition. In *Proceedings of the 6th Conference on Natural Language Learning (CoNLL-2002)*.
- Erik F. Tjong Kim Sang and Fien De Meulder. 2003a. Introduction to the conll-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*.
- Erik F. Tjong Kim Sang and Fien De Meulder. 2003b. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, pages 142–147.
- Robbert van der Goot, Floris-Jan van der Klis, and Gertjan van Noord. 2014. The gmbc corpus: A web-based corpus of dutch text with rich annotation. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*.
- Johannes M. van Hulst, Faegheh Hasibi, Koen Dercksen, Krisztian Balog, and Arjen P. de Vries. 2020. Rel: An entity linker standing on the shoulders of giants. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20*. ACM.
- C. J. van Rijsbergen. 1979. *Information Retrieval*. London: Butterworths.
- Ashish Vaswani, Noam M. Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *NIPS*.
- Denny Vrandeić and Markus Krötzsch. 2014. [Wiki-data: a free collaborative knowledgebase](#). *Commun. ACM*, 57:78–85.
- Shuhe Wang, Xiaofei Sun, Xiaoya Li, Rongbin Ouyang, Fei Wu, Tianwei Zhang, Jiwei Li, and

- Guoyin Wang. 2023. [Gpt-ner: Named entity recognition via large language models](#). *ArXiv*, abs/2304.10428.
- Xinyu Wang, Yong Jiang, Nguyen Bach, Tao Wang, Zhongqiang Huang, Fei Huang, and Kewei Tu. 2020. [Automated concatenation of embeddings for structured prediction](#). *ArXiv*, abs/2010.05006.
- Xinyu Wang, Yongliang Shen, Jiong Cai, Tao Wang, Xiaobin Wang, Pengjun Xie, Fei Huang, Weiming Lu, Yue Ting Zhuang, Kewei Tu, Wei Lu, and Yong Jiang. 2022a. [Damo-nlp at semeval-2022 task 11: A knowledge-based system for multilingual named entity recognition](#). *ArXiv*, abs/2203.00545.
- Xinyu Wang, Yongliang Shen, Jiong Cai, Tao Wang, Xiaobin Wang, Pengjun Xie, Fei Huang, Weiming Lu, Yue Ting Zhuang, Kewei Tu, Wei Lu, and Yong Jiang. 2022b. [Damo-nlp at semeval-2022 task 11: A knowledge-based system for multilingual named entity recognition](#). *ArXiv*, abs/2203.00545.
- Xuan Wang, Yu Zhang, Xiang Ren, Yuhao Zhang, Marinka Zitnik, Jingbo Shang, C. Langlotz, and Jiawei Han. 2018. Cross-type biomedical named entity recognition with deep multi-task learning. *bioRxiv*.
- Xuwu Wang, Junfeng Tian, Min Gui, Zhixu Li, Rui Wang, Ming Yan, Lihan Chen, and Yanghua Xiao. 2022c. [Wikidiverse: A multimodal entity linking dataset with diversified contextual topics and entity types](#). In *Annual Meeting of the Association for Computational Linguistics*.
- Antoine Willerval, Dennis Diefenbach, and Pierre Maret. 2022. Easily setting up a local wikidata sparql endpoint using the qendpoint.
- Ledell Wu, Fabio Petroni, Martin Josifoski, Sebastian Riedel, and Luke Zettlemoyer. 2020. Zero-shot entity linking with dense entity retrieval. In *EMNLP*.
- Bo Xu, Shizhou Huang, Chaofeng Sha, and Hongya Wang. 2022. [Maf: A general matching and alignment framework for multimodal named entity recognition](#). *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*.
- Linting Xue, Noah Constant, Adam Roberts, Mihir Kale, Rami Al-Rfou, Aditya Siddhant, Aditya Barua, and Colin Raffel. 2021. [mT5: A massively multilingual pre-trained text-to-text transformer](#). In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 483–498, Online. Association for Computational Linguistics.
- Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. In *Neural Information Processing Systems*.
- Roman Yangarber, Jakub Piskorski, Anna Dmitrieva, Michal Marcinczuk, Pavel Pribán, Piotr Rybak, and Josef Steinberger. 2023. [Slav-ner: the 4th cross-lingual challenge on recognition, normalization, classification, and linking of named entities across slavic languages](#). *Proceedings of the 9th Workshop on Slavic Natural Language Processing 2023 (SlavicNLP 2023)*.
- Hanming Zhai, Xiaojun Lv, Zhiwen Hou, Xin Tong, and Fanliang Bu. 2023. [Mlnet: a multi-level multimodal named entity recognition architecture](#). *Frontiers in Neurobotics*, 17.
- Zhenru Zhang, Chuanqi Tan, Songfang Huang, and Fei Huang. 2023a. [Veco 2.0: Cross-lingual language model pre-training with multi-granularity contrastive learning](#). *ArXiv*, abs/2304.08205.
- Zhenru Zhang, Chuanqi Tan, Songfang Huang, and Fei Huang. 2023b. [Veco 2.0: Cross-lingual language model pre-training with multi-granularity contrastive learning](#). *ArXiv*, abs/2304.08205.
- GuoDong Zhou and Jian Su. 2002. Named entity recognition using an HMM-based chunk tagger. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 473–480.

## 7. Language Resource References

WiNNL code and link to the dataset can be found here: <https://gitlab.jrc.ec.europa.eu/jrc-projects/emm/emm-wikinews-scraper>.