Open English NameNet: Extending English Wordnet with Names

John P. McCrae
Research Ireland Insight and ADAPT Centres, Data Science Institute, University of Galway
john@mccr.ae

Abstract

We present Open English NameNet, a new large-scale lexical resource that extends Open English Wordnet with named entities derived from Wikidata. While English Wordnet has historically included many proper nouns, its coverage has been incomplete and inconsistent, and encyclopedic knowledge sources have grown rapidly in parallel. To address this gap, we systematically extract and align named entities from Wikidata with the Open English Wordnet hierarchy, ensuring each entity is appropriately placed through instance hypernym relations. Our methodology combines existing WordNet–Wikipedia mappings with Wikidata information and applies domain-specific strategies for people, plants and animals, and languages, to account for structural and semantic differences between the resources. This approach results in the largest English lexical-semantic resource currently available, with extensive coverage and structured integration. We release the resource openly to support the development of lexically and encyclopedically informed language technologies.

Keywords: WordNet; Named Entities; Wikidata; Lexical Resources; Knowledge Graph

1. Introduction

Open English Wordnet (McCrae et al., 2019, 2020, OEWN) is an open-source fork of the original Princeton WordNet (Miller, 1995; Fellbaum, 2010)[1]. However, a major issue with the development of this resource has been the size of the wordnet and the large number of proper nouns and other named concepts inherited from the Princeton WordNet. These concepts range from widely known concepts that would be of value to lexicography, such as the names of countries and languages, to concepts that are much more obscure, such as the names of minor historical figures and obscure names for plants and animals. Meanwhile, large encyclopedic resources, such as Wikidata[2], have arisen that cover such concepts far more completely than could ever be achieved in the limitations of a lexicographic resource. As such, in this paper, we propose a new resource, Open English NameNet, built from the proper nouns of OEWN extended with concepts extracted from Wikidata, leaving the remaining section of Open English Wordnet to focus on the common nouns, verbs, adjectives and adverbs of the language. The two resources can follow different quality guidelines, with OEWN having higher quality guarantees, such as unique definitions for all concepts, while NameNet can be expanded to have a very wide coverage. The resources are designed to be combined and the combination of these is the largest lexicographic resource for English by a substantial margin.

The work of creating Open English NameNet is fundamentally based on the mapping between WordNet and Wikipedia and we rely on mapping from several sources, namely existing mappings in Open English Wordnet, created by McCrae and Cillessen (2021), Grammatical Framework (Angelov, 2020), BabelNet (Navigli and Ponzetto, 2010; Navigli et al., 2021) and mappings from YoVisto (Bergh et al., 2025). However, the mapping is not simply a conversion of the data in Wikidata to the wordnet format, as the new synset must be appropriately placed into the wordnet hypernym hierarchy, and for many concepts, the identification of an appropriate hypernym is a challenging task. Further, in many cases, the mapping of concepts requires further analysis. For example, for people, the superclass in Wikidata is generally human[Q5], but in wordnet the hypernyms are generally to the occupation of that person. Similarly, the organization of plants and animals in wordnet is very different from that of Wikidata and this further complicates the alignment. Finally, we also looked at languages specially as these are not instance hypernyms in wordnet and so the modelling of this also needed to be carefully considered.

The main contribution of this paper is Open English NameNet, a large-scale lexical resource that addresses the "coverage gap" between the high-precision linguistic data in Open English Wordnet (OEWN) and the vast encyclopedic breadth of Wikidata. By systematically extracting and aligning named entities, we extend the traditional wordnet hierarchy with millions of new synsets, creating the **largest English lexical-semantic resource currently available**. Our contribution lies in a novel methodology that maintains semantic coherence through instance hypernym relations while applying domain-specific strategies for complex categories like people, biological taxa, and languages. Unlike previous integrated resources, NameNet is not

[1]'WordNet' is a trademark of Princeton University; we use 'wordnet' to describe resources following the structure of WordNet

[2]https://wikidata.org

Proceedings of the 10th Workshop on Linked Data in Linguistics (LDL-2026) @ LREC 2026, pages 60–68

merely a link between separate graphs but an expansion of the OEWN hierarchy itself, providing a structural bridge that allows language technologies to unify Word Sense Disambiguation (WSD) and Entity Linking (EL) within a single, consistent framework.

This paper is structured as follows. Section 2 reviews related work on integrating named entities and encyclopedic knowledge into lexical resources, situating our contribution within the broader research landscape and provides background on the differences between wordnets and encyclopedic resources. Section 3 outlines our methodology for constructing Open English NameNet, including the general strategy for hypernym assignment as well as domain-specific approaches for people, plants and animals, and languages. Section 4 describes the resulting dataset, including its size, structure, and modes of distribution. Section 5 provides some applications for the new resource and finally, Section 6 concludes with a discussion of the contributions of this work and future directions for extending and improving the resource.

2. Background

Efforts to extend and enrich lexical resources with encyclopedic knowledge and named entities have a long history. Toral et al. (2008) present a foundational approach, which describes the automatic extension of Princeton WordNet with named entities (NEs). Their method maps the WordNet noun hierarchy to Wikipedia categories in order to identify and extract named entities and their lexical and definitional information. This approach resulted in the enrichment of WordNet with over 300,000 named entities and hundreds of thousands of "instance of" relations, illustrating the potential of leveraging structured encyclopedic resources to complement lexicographic ones.

Wordnets, such as the original Princeton WordNet (PWN) and its successor, Open English Wordnet (OEWN), are primarily lexicographic resources. Their core purpose is to map the semantic relationships between "common" lexical concepts: nouns, verbs, adjectives, and adverbs. In contrast, resources like Wikidata are encyclopedic knowledge bases designed to store facts about specific entities. Wikidata covers nouns, mostly proper nouns and does not contain many verbs, adjectives or adverbs, and these are never marked as such. Further, Wikidata cannot provide the detailed syntactic annotation provided in a resource such as OEWN. As such, there is a substantial gap in the use cases between wordnets and encyclopedic resources like Wikidata. As such, this resource aims to close this "coverage gap" between OEWN and Wikidata. OEWN provides high-quality definitions and linguistic precision, it cannot realistically scale to cover the millions of named entities (NEs), due to the high level of quality and accuracy required in this resource. Conversely, while Wikidata provides vast coverage, it describes the concepts but does not describe the linguistic structures. By extending the wordnet format to include NameNet, this work provides a bridge that allows language technologies to access encyclopedic breadth without losing the structural coherence of a lexical hierarchy.

The most similar resource to Open English NameNet is BabelNet, which combines data from OEWN and Wikidata to create a multilingual dictionary. Navigli et al. (2021) provides an extensive overview of BabelNet, a multilingual lexical-semantic resource that merges heterogeneous sources and has been widely adopted in NLP and AI applications, demonstrating the advantages of multilingual and large-scale integration for both symbolic and statistical methods. In parallel, Rebele et al. (2016) introduces YAGO, an automatically constructed knowledge base that combines Wikipedia, WordNet, and GeoNames. YAGO emphasises high precision in its extraction process and enriches the integrated knowledge graph with temporal and spatial information, supporting complex and expressive queries. Its multilingual dimension further illustrates the benefits of linking lexical resources with broad-coverage encyclopedic data. While these resources integrate English Wordnet and Wikidata into single resources, they do so in a way that does not align with the structure of wordnets and, as such, are novel resources rather than extensions of the existing English Wordnet.

BabelNet (Navigli and Ponzetto, 2010) and YAGO (Rebele et al., 2016) are integrations of heterogeneous sources that retain the distinct structural logic of each. In contrast, NameNet is not a "link" between two separate graphs; it is an expansion of the OEWN hierarchy itself, where every entity is conformed into a wordnet-style synset structure. As such, NameNet takes only the entities that have semantic coherence (i.e., are of the same class) as concepts already in wordnet, and the mapping is non-trivial in that maps are made based on mapping multiple properties rather than the simple links of previous work, as described below for people, biological taxa and languages. NameNet is a resource that not only includes imported data from Wikidata, but also manual modifications developed by the OEWN team. As such, it should be noted that NameNet is an open-source resource to which anyone may contribute, unlike closed resources such as BabelNet and YAGO. Finally, Open English NameNet is, as its name suggests, an attempt to document the English language specifically and does not attempt a 'one-size-fits-all' multilingual approach, opening the door to the development of namenets for other languages.

A key motivation of Open English NameNet is to provide a clear separation between *instances* and *concepts*. This is a core distinction represented in language by the distinction between common nouns, which represent classes or categories of things (e.g., painter, river, language), and proper nouns, which refer to unique, specific entities that belong to those classes (e.g., Pablo Picasso, The Nile, English). This distinction is represented in nearly all conceptual modelling schemes, for example, OWL distinguishes between 'classes' and 'entities'. In Open English Wordnet, the distinction is formally maintained through the `instance_hyponym` relation for individuals and the `hyponym` relation for sub-categories. Similarly, Wikidata separates these concepts using the 'instance of'[P31] property for unique entities and the 'subclass of'[P279] property for taxonomic hierarchies.

3. Methodology

Our methodology follows a structured mapping process to align named entities from Wikidata to the OEWN hierarchy. In most cases, this involves assigning a suitable general hypernym to each new synset to ensure consistent integration into the lexicon, but specific strategies were followed for some domains such as people, plants and animals, and languages.

3.1. Linking Methodologies

The methodology in this paper is based on an existing bijective mapping between Wikidata and Open English Wordnet. The most recent and complete study is McCrae et al. (2026), which argues that while individual projects like Open English Wordnet (OEWN), Grammatical Framework (GF), BabelNet and yovisto have made significant strides, they remain largely complementary rather than overlapping. For instance, the complete intersection of all four resources covers only 3,017 synsets, whereas 49,219 synsets appear in at least one of these linkings. This discrepancy highlights a major opportunity for consolidation; while OEWN provides high-precision human-in-the-loop mappings (96.1% accuracy), it has the smallest coverage at 12,083 links. In contrast, BabelNet offers the largest dataset with 39,224 links but faces higher disagreement rates with other resources, such as a 14.3% disagreement rate with GF.

We rely on the unified resource they constructed that integrates these disparate mappings while leveraging similarity measures and human-in-the-loop validation. The original links were created by different strategies: GF projected Wikipedia links to Wikidata and merges them with community-contributed synset IDs; BabelNet uses probabilistic alignment algorithms based on contextual evidence like synonyms and gloss definitions ; and yovisto employs a dual-annotator system (DBpedia-Spotlight and yovisto-KEA) to map synsets via DBpedia URIs. The consolidation of these efforts, the proposed integrated resource aims to maximise coverage—which currently reaches approximately 60% of noun synsets—while resolving errors stemming from granularity mismatches and conceptual ambiguities.

The integration of Wikidata into the NameNet framework requires navigating fundamentally different structural philosophies: the Open English WordNet maintains a strictly acyclic taxonomy, permitting multiple inheritance and diamonds (multiple hypernyms which merge into a single concept higher in the taxonomy), but forbidding loops. Wikidata's hierarchy is substantially more detailed and contains directed cycles. As such, we only consider the immediate parent in our mapping, except where noted below, to avoid introducing cycles from Wikidata's hierarchy into NameNet. This approach enables multi-faceted encoding, where an entity such as a fictional character can be simultaneously represented through multiple ontological lenses, for instance, as a pig (biological species), a person (agentive role), and a fictional character (narrative status). The selection of the primary categories currently included (General, People, Species, Taxonomic Categories, and Languages) was motivated by the size of these categories in Wikidata and WordNet and the specific modelling challenges they present. This version focuses on the high-density domains of people, animals and languages and treats other classes only through a general mapping. However, we acknowledge that smaller, more specialised classes may still require more sophisticated, bespoke modelling strategies in future iterations to capture their unique relational nuances.

3.2. General Hypernyms

For general classes, that is instance hypernyms not requiring specific mapping strategies such as people, plants and animals, we primarily base hypernym selection on the Wikidata *superclass*[P31] relationships. The direct mappings were based on previous mappings established by other efforts; this included the mapping developed by McCrae and Cillessen (2021), which has been included in Open English Wordnet. However, we also used links that were in other resources, including Grammatical Framework (Angelov, 2020), BabelNet (Navigli and Ponzetto, 2010) and YoVisto (Bergh et al., 2025).

If no direct mapping exists, we follow the superclass chain transitively to identify a suitable hypernym within the existing WordNet hierarchy. When

Conflict case	Frequency	Examples
Both	43	**museum** and **castle**; **Hindu deity** and **king**
Accept One	78	**museum** not **organization**; **political party** not **political movement**
Alternative	16	**hill** and **mountain** → **elevation**[09389214-n]; **bay** and **lake** → **body of water**[09248053-n]
Neither	25	**government agency** or **region**

Table 1: Conflicts between different Wikidata Classes

multiple superclasses or inheritance paths yield conflicting hypernyms, we resolve the conflict by prioritizing the most specific existing hypernym in OEWN; otherwise, the conflict is resolved manually. As shown in Table 1, conflicts may result in accepting one option:

Both In many cases, it was acceptable to use both hypernyms as the classes were compatible, for example, a castle can also be a museum, as these are both locations in the world and so are semantically compatible.

Accept One The most frequent choice was to prefer one of the hypernyms over the other. This was motivated by examining similar cases that are already in OEWN and finding the most similar mapping. For example, a museum may also be an organization and this kind of systematic polysemy (Nunberg, 1992) is common; however museum is the primary function of the organisation, so this is the preferred hypernym. This is based on existing hypernyms of synsets that are organised this way.

Alternative In some cases, we see that the concepts are closely related and thus we can find an alternative concept that works as a hypernym that is consistent with both of the Wikidata superclasses, most frequently this means choosing a common hypernym of the two elements. For example, many entities in Wikidata are tagged as both *hills*[09325914-n] and *mountains*[09382700-n] so the common hypernym *elevation*[09389214-n] is used.

Neither We noted that for many concepts, it was not possible to make an effective rule as to which hypernym is preferred and these must be done on a case-by-case basis. As such, these concepts are not currently part of NameNet. For example, some concepts in Wikidata are classed as both *government agencies*[Q327333] and *region*[Q82794] and these are not really possible to map in a compatible manner.

Each resulting entry in Open English NameNet consists of the following elements that make it compatible with the rest of the structure. The definition of the entry is taken from the Wikidata definition, and unlike those in Open English Wordnet, we do not require that these are unique, and in fact, many of these are quite simplistic, such as 'city in the United States'. There is a requirement that each synset is linked into the wordnet graph by at least one link and this is achieved by an instance hypernym link, which connects the new synset to its appropriate position in the WordNet hierarchy, ensuring that the network remains semantically coherent. The entry also includes one or more lemmas, derived from the English label in Wikidata or from multilingual labels when no English equivalent exists, allowing for broad lexical coverage. In addition, each entry's part of speech is *noun* and a link to the corresponding Wikidata item is included, providing a stable reference for external alignment, future enrichment, and validation of the resource.

Linking Stage	Acceptance	Total Accepted
Occupations	90.1%	1132
Taxons	87.6%	4941
Species	99.5%	5076
Taxon/Common	88.1%	2087
Languages	68.4%	528

Table 2: Acceptance rate and total accepted from the manual linking

3.3. People

One of the key modelling differences between Wikidata and OEWN is the modelling of people, which in OEWN are characterised by their roles, whereas they are organised by their species in Wikidata. We align around the idea of personhood, which refers to entities that are treated as persons within the lexicon, including fictional characters. Philosophically, personhood is defined not by membership of a species but by the possession or attribution of qualities such as agency, intentionality, self-awareness, or social identity (e.g., *Peppa Pig* is a person).

In practice, this means that we consider not only the concepts in Wikidata that are instances

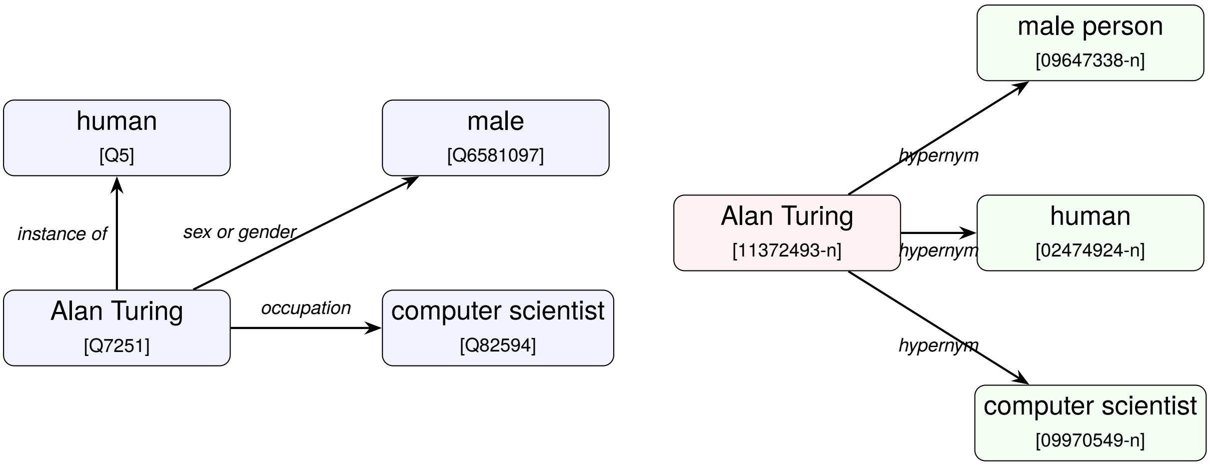

Figure 1: Example of the modelling of people in Wikidata (on the left) and OEWN (on the right). This illustrates the multiple properties used by Wikidata and the new hypernyms that can be inferred from Wikidata.

of *Human*[Q5], but also classes such as *Character*[Q95074]. Further, the modelling of roles in Wikidata is different in that the role is represented by the Occupation[P106] property. Further, we also decided to add specific gender information about the people in the resource, which has been identified as a weakness of OEWN previously (McCrae et al., 2025). In this way, a person in Open English NameNet would have at least three hypernyms: a species hypernym mostly referring to *human*[02474924-n], one or more occupation hypernyms, and a gender hypernym referring to either *male person*[09647338-n] or *female person*[09642198-n].

For example, in Figure 1, we see the modelling of the person 'Alan Turing' in Wikidata, where the modelling provides links that he is an instance of 'human'[H5] and the properties 'sex or gender'[P21] and 'occupation'[P106]. These are mapped onto hypernyms in OEWN. As 'Alan Turing' is already included in OEWN, we add these as extra facts alongside the current hypernym of the synset ('mathematician'[10320928-n]), deepening the representation of this concept in English Wordnet.

Many of the occupations listed in Wikidata were not linked to a synset in Open English Wordnet based on the existing mapping, so we examined the linkings that have been proposed in one of BabelNet, GF or YoVisto, and manually examined these linkings, accepting about 90.1% of these links into Open English Wordnet to improve the coverage of occupations. We also noted during this linking that 136 of the occupation values in Wikidata are invalid, for example, they refer to the organization, such as *police force*[Q35535], rather than the occupation, such as *police officer*[Q384593].

3.4. Plants and Animals

The plants and animals mapping in wordnet uses two linked hierarchies, differentiating between the common names and the scientific taxonomic names for plants and animals. For example, *bear*[02134305-n] is a distinct synset from *family Ursidae*[02134070-n], even though both refer to the same animals. *Bear* is then a hyponym of *carnivore*[02077948-n], while *family Ursidae* is a hyponym of *mammal family*[01865198-n] but a holonym[3] of *order Carnivora*[02077567-n]. However, actual species with binomial names such as *Ursus arctos* are part of the same synset as the common name *brown bear*[02134788-n]. Wikidata has a simpler modelling where common names and taxonomic names are labels of the same entity. To create a structure for plants and animals, that allows Wikidata information to be imported in a way that follows the wordnet hierarchy, it was necessary to convert to the wordnet hierarchy. The structure is shown for brown bears in Figure 2.

We first examined, the taxon names by identifying all the synsets that had a lemma consisting of the words "genus", "family", "order", "class", "phylum" or "kingdom" with a capitalised word in OEWN and looked up the matching name in Wikidata expressed by the property *taxon name*[P225]. These links were then manually validated with an 87.6% acceptance rate (see Table 2). We then also repeated this for binomial species names, where we achieved a much higher acceptance rate of 99.5%, and in fact, all errors were due to issues with OEWN rather than lexical ambiguity. In 15.0% of cases, the binomial species name did not match the corresponding genus linking, which is primarily due to changes in the organization of the species, for example OEWN lists *Felis pardalis*[02128146-n] as the

[3] part/whole relationship

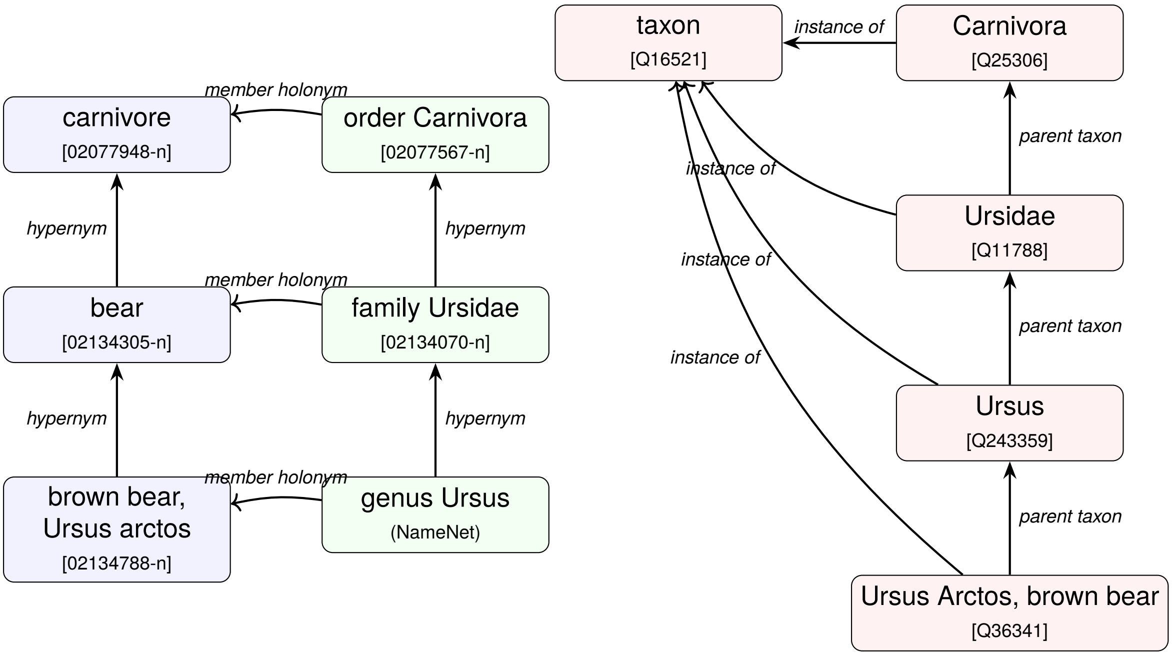

Figure 2: The representation of animals in Open English Wordnet and Wikidata as shown by the modelling of 'brown bear'. The blue represents the common name hierarchy in OEWN, the green the taxon hierarchy in OEWN and red the Wikidata hierarchy. Note the Wikidata hierarchy is simplified; the actual hierarchy includes several intermediate taxons (Ursinae, Ursoidea, Ursida, Arctoidea and Caniformia).

binomial name for the ocelot, this is now part of the genus *Leopardus* so it is known as *Leopardus pardalis*[Q33261] in Wikidata. Wikidata actually assigns unique identifiers for each binomial name, so *Felis pardalis* was first matched to the corresponding *Felis pardalis*[Q122170181] and then the *protonym of*[P12765] property was used to link it to the main entity. In retrospect, due to the high quality of the species linking, it would have been more effective to link species first and use this to confirm the taxonomic categories.

The process of linking thus involves the creation of two synsets in Open English NameNet for each taxonomic classes, such as *genus Ursus*, one using the common names, while the other has only the formulaic lemmas of the name (e.g., *Ursus*) and the name prefixed by the taxon rank (e.g., "genus Ursus"). As the organization of the taxonomic categories in OEWN is based on fairly dated material, it was decided to discard the hierarchy of OEWN and replace it entirely with the hierarchy of Wikidata. As the hypernym of these new concepts is a synset such as *mammal genus*, it was necessary to create a mapping between the common class and the taxon class in OEWN. While this internal link in OEWN is specified, it uses the meronym/holonym property that is also used to indicate the taxonomic hierarchy, i.e., *kingdom Plantae*[11550054-n] is a member meronym of both the common synset *plant*[00017402-n] but also divisions such as *division Bryophyta*[11557229-n]. These were manually disambiguated with an accuracy of 88.1% (see Table 2). The species were created in a manner similar to the general class, but they were given meronymy links to the taxonomic genus that was appropriate. If the species had a protonym, then this was given as a secondary Wikidata link in the data.

3.5. Languages

The modelling of languages in wordnet is also notably different to Wikidata in that the hypernyms are all instances, that is that *English*[06959794-n] is a hyponym, not an instance of *West Germanic Language*[06959585-n] and in fact has further hyponyms such as *American English*[06960241-n]. In order to extend these in NameNet, we then needed to manually extract the languages and create a link between Wikidata and the named languages in OEWN. We used only the name of the language to find candidates for this mapping; however, we found this to be a very high error rate linking with only 68.4% of named languages matching a similar named language in Wikidata. For example, "Manda" can refer to three unrelated languages spoken in Tanzania, India and Australia. This was mainly due to the relative outdated information in wordnet about languages, which often used incorrect spellings or grouped languages now considered to be distinct.

General	Synsets	Lemmas
General	9,673,356	37,678,495
- Star	3,754,413	11,290,295
- Chemical	1,274,034	4,020,428
- Substance	902,749	39,939,24
- Street	642,203	1,984,043
- Mountain	423,511	1,284,921
People	8,506,828	40,266,374
Species	2,757,140	2,876,804
Taxonomic Cat.	257,442	516,117
Languages	8,961	43,844
Total	21,203,727	81,381,634

Table 3: Total resource size in terms of the number of synsets (concepts) and lemmas (words).

4. Dataset

This resource is intended as an extension for Open English Wordnet, and as such, there are three main releases of OEWN available as part of the 2025 edition. Firstly, we provide a version of the OEWN that does not contain NameNet and provide NameNet as a separate version. We also provide a single file in the Global WordNet Format (McCrae et al., 2021), which contains the combined Open English Wordnet and NameNet data. Finally, we also create a legacy version that keeps all the named entities that were part of Princeton WordNet and previous versions of OEWN, but with extra hypernyms and taxon meronyms provided by NameNet.

The overall size of the resource is presented in Table 3, where we present the number of links created by each of the methods. For the general linking, we present the overall number of links created by this linking as well as the five largest subsets according to hypernym. For plants and animals, we distinguish between the species introduced and the taxonomic categories. In total, this introduces a large number of new synsets, making Open English NameNet by far the largest lexicographic resource available for English ever, while maintaining the quality and structure of a wordnet.

4.1. Linked Data Publishing

Open English NameNet as a resource is made available as RDF files through HuggingFace and the linked data interface is under-development to make the resource available through the Open English Wordnet portal. Open English NameNet is published in the Global WordNet Format, which is aligned with OntoLex-lemon and ensures interoperability with the broader ecosystem of linguistic linked data and existing wordnet infrastructures. By maintaining persistent links to Wikidata items, the resource serves as a stable bridge in the Linked Data cloud, allowing for the seamless integration of high-quality lexicographic definitions with vast, crowdsourced encyclopedic knowledge.

5. Applications

5.1. NameNets for Other Languages

While the resources described here is only for English, the methodology described here is designed to be highly replicable for other language-specific wordnet projects, and is designed such that our open-source system can be easily adopted by other langauges to create new namenets. Since Wikidata is inherently multilingual, the majority of the extraction and alignment process, such as following the 'instance of'[P31] and 'subclass of'[P279] chains, is language-independent and can be carried out automatically. As such, the process of creating namenets for new languages, where the wordnet is already aligned with either Princeton WordNet or Open English Wordnet, is quite trivial. However, the procedure for constructing the namenet explicitly excludes concepts with no lexicalisation in English, so some concepts may be excluded. As such, replication would involve:

Initial Alignment The alignment of concepts from the NameNet can be imported and verified in coverage

New Alignments Develop new alignments in line with Section 3.2, for language-specific concepts

Generate New Resource The open-source pipeline[4] can be used to create new synsets

While some manual validation is required to resolve language-specific nuances, the core framework provides a scalable and semi-automated pipeline for expanding any wordnet with encyclopedic named entities.

Further, this work can be of substantial help to the development of wordnets for other languages, as the links provided to these resources can provide lemmas or alternative synonyms for many synsets in these resources. As an example in Open Multilingual WordNet, a simple lemma such as 'Sudan'[09051827-n] has a lemma for only 16 out of 35 (45.7%) of languages, while this lemma is available for all of the languages in Wikidata, often with alternatives like 'Republic of Sudan.'

5.2. Bridging WSD and EL

A significant challenge in Natural Language Processing is the artificial divide between Word Sense

[4] https://github.com/globalwordnet/english-namenet

Disambiguation (Bevilacqua et al., 2021, WSD), which handles common nouns, and Entity Linking (Sevgili et al., 2022, EL), which identifies specific named entities. Traditionally, these tasks require separate models and distinct knowledge bases (e.g., WordNet for WSD and Wikipedia for EL). Open English NameNet acts as a structural bridge between these domains by integrating them into a single, semantically coherent hierarchy. As such, datasets cannot be constructed that deal with both WSD and EL in the same sentence. For example, in "the *mercury* reached record temperatures while the *Mercury* program was still in its infancy", we would see both word sense disambiguation of the word 'mercury' (a common noun referring to the metal) and the proper noun (referring to the NASA space programme). With NameNet, both concepts exist in the same graph, with the same schema. A system can verify the semantic coherence of the sentence by tracing both "senses" to their common ancestors in the OEWN hierarchy. Unlike raw Wikidata, NameNet entries include the linguistic metadata of a wordnet, such as part-of-speech tags and variant lemmas. This enables EL systems to handle morphological variations and syntactic constraints that are usually absent from purely encyclopedic resources.

6. Conclusion

In this paper, we have presented *Open English NameNet*, a large-scale extension of Open English Wordnet that systematically incorporates named entities derived from Wikidata. By decoupling the treatment of common nouns, verbs, adjectives, and adverbs from proper names, we enable different quality and coverage strategies, ensuring that NameNet can grow to cover a broad and dynamic range of concepts while OEWN maintains high lexicographic precision.

NameNet is based on Wikidata and while it is a massive and ever-evolving resource, we do not foresee significant efficiency issues regarding the scalability of NameNet in the near term. Given that Wikidata's growth follows a consistent trajectory, our current ingestion and cycle-breaking pipelines are well-equipped to handle future updates without requiring a fundamental architectural shift.

Our approach combines existing mappings between WordNet and Wikipedia with structured information from Wikidata, supported by tailored mapping strategies for specific domains, including people, plants and animals, and languages. Through this methodology, we achieve both large-scale coverage and a semantically coherent integration into the WordNet hierarchy. This results in the largest English lexical resource currently available, combining the rich linguistic structure of English Wordnet with the encyclopedic breadth of Wikidata, while bridging the gap between word senses and named entities. In future work, we aim to further refine hypernym selection, address ambiguous or unmapped entities through semi-automated alignment techniques, and extend the resource with multilingual correspondences. By making Open English NameNet openly available, we hope to support a wide range of applications in NLP, linguistics and AI that depend on high-quality lexical and encyclopedic knowledge.

Acknowledgements

We would like to thank Al Waskow for their comments on the draft.

John P. McCrae is supported by Research Ireland under Grant Number SFI/12/RC/2289_P2 Insight_2, Insight SFI Centre for Data Analytics and Grant Number 13/RC/2106_P2, ADAPT SFI Research Centre.

7. Bibliographical References

Krasimir Angelov. 2020. A parallel WordNet for English, Swedish and Bulgarian. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 3008–3015, Marseille, France. European Language Resources Association.

Johann Bergh, Jörg Waitelonis, and Melanie Siegel. 2025. Leveraging LLMs for Constructing WordNets Automatically as Bilingual Resources. In *Proceedings of the 2025 Global WordNet Conference*.

Michele Bevilacqua, Tommaso Pasini, Alessandro Raganato, and Roberto Navigli. 2021. Recent trends in word sense disambiguation: A survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 4330–4338. International Joint Conferences on Artificial Intelligence Organization. Survey Track.

John P. McCrae, Johann Bergh, Jörg Waitelonis, and Krasimir Angelov. 2026. Towards a comprehensive english wordnet-wikidata mapping. In *Proceedings of the Fifteenth Biennial Language Resources and Evaluation Conference (LREC)*.

John P McCrae and David Cillessen. 2021. Towards a Linking between WordNet and Wikidata. In *Proceedings of the 11th Global Wordnet Conference*, pages 252–257.

John P. McCrae, Michael Wayne Goodman, Francis Bond, Alexandre Rademaker, Ewa Rudnicka, and Luis Morgado Da Costa. 2021. The globalwordnet formats: Updates for 2020. In *Proceedings of the 11th Global Wordnet Conference*, pages 91–99.

John P. McCrae, Haotian Zhu, Fei Xia, Al Waskow, and Kexin Gao. 2025. Remedying Gender Bias in Open English Wordnet. In *Global WordNet Conference 2025*.

Roberto Navigli, Michele Bevilacqua, Simone Conia, Dario Montagnini, and Francesco Cecconi. 2021. Ten Years of BabelNet: A Survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 4559–4567. ijcai.org.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a Very Large Multilingual Semantic Network. In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden*, pages 216–225. The Association for Computer Linguistics.

Geoffrey Nunberg. 1992. Systematic polysemy in lexicology and lexicography. In *Proceedings of the 5th EURALEX International Congress*, pages 386–396, Tampere, Finland. Tampereen Yliopisto.

Thomas Rebele, Fabian M. Suchanek, Johannes Hoffart, Joanna Biega, Erdal Kuzey, and Gerhard Weikum. 2016. YAGO: A multilingual knowledge base from Wikipedia, Wordnet, and Geonames. In *The Semantic Web - ISWC 2016 - 15th International Semantic Web Conference, Kobe, Japan, October 17-21, 2016, Proceedings, Part II*, volume 9982 of *Lecture Notes in Computer Science*, pages 177–185.

Özge Sevgili, Artem Shelmanov, Mikhail Arkhipov, Alexander Panchenko, and Chris Biemann. 2022. Neural entity linking: A survey of models based on deep learning. *Semantic Web*, 13(3):527–570.

Antonio Toral, Rafael Muñoz, and Monica Monachini. 2008. Named Entity WordNet. In *Proceedings of the International Conference on Language Resources and Evaluation, LREC 2008, 26 May - 1 June 2008, Marrakech, Morocco*. European Language Resources Association.

8. Language Resource References

Christiane Fellbaum. 2010. *WordNet*, pages 231–243. Springer Netherlands, Dordrecht.

John P. McCrae, Alexandre Rademaker, Francis Bond, Ewa Rudnicka, and Christiane Fellbaum. 2019. English WordNet 2019 – An Open-Source WordNet for English. In *Proceedings of the 10th Global WordNet Conference – GWC 2019*.

John P. McCrae, Ewa Rudnicka, and Francis Bond. 2020. English wordnet: A new open-source wordnet for english. *K Lexical News*, (28):37–44.

George A. Miller. 1995. WordNet: a lexical database for English. *Commun. ACM*, 38(11):39–41.

Roberto Navigli, Michele Bevilacqua, Simone Conia, Dario Montagnini, and Francesco Cecconi. 2021. Ten Years of BabelNet: A Survey. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, pages 4559–4567. ijcai.org.

Roberto Navigli and Simone Paolo Ponzetto. 2010. BabelNet: Building a Very Large Multilingual Semantic Network. In *ACL 2010, Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, July 11-16, 2010, Uppsala, Sweden*, pages 216–225. The Association for Computer Linguistics.