Entity Linking for Faroese Using Large Language Models with Web Search

Annika Simonsen[1] **Iben Nyholm Debess**[2]
Hafsteinn Einarsson[1]
[1]University of Iceland [2]University of the Faroe Islands
ans72@hi.is, ibennd@setur.fo, hafsteinne@hi.is

Abstract

Entity linking connects text mentions to knowledge bases. For low-resource languages, entity linking has typically not been a research priority, as named entity recognition and knowledge base creation must first be addressed. We present the first study of entity linking for Faroese, a North Germanic language with approximately 70,000 speakers. Unlike traditional systems that rely on separate candidate retrieval and ranking components, we employ an end-to-end approach using GPT-5 with integrated web search. Our method prompts the model to directly identify and link named entities to Wikipedia pages through a three-tier fallback strategy: Faroese Wikipedia, English Wikipedia, and finally any available Wikipedia. We evaluate our approach on 1,010 manually annotated examples from a Faroese NER dataset, analyzing entity mentions across Person, Location, Organization, and Miscellaneous types. Human evaluation shows our system achieves 87.5% precision and 87.3% recall, with particularly strong performance on locations (93-95% precision, 92-95% recall). Persons are more challenging (86-88% precision, 72-83% recall). The majority of links (76.5%) point to Faroese Wikipedia, demonstrating the model's ability to leverage language-specific knowledge bases. A Wikipedia API search baseline without any LLM achieves F1 = 0.57–0.60 on the same evaluation data, confirming that the LLM's contextual reasoning provides substantial gains over simple search. We validate our approach across three models (GPT-5, Gemini 3 Flash, GPT-5.4 Mini), achieving F1 scores of 0.74–0.87 and confirming that the method generalizes across providers. This work establishes initial performance benchmarks for Faroese entity linking and demonstrates the viability of LLM-based approaches for low-resource languages.

Keywords: entity linking, Faroese, large language models, low-resource languages, web search, GPT-5

1. Introduction

Entity linking (EL) is the task of connecting entity mentions in text to their corresponding entries in a knowledge base, typically Wikipedia (Cucerzan, 2007; Rao et al., 2013). While significant progress has been made in entity linking for high-resource languages, low-resource languages remain under-explored (Fu et al., 2020). This is particularly true for North Germanic languages beyond the major Scandinavian languages.

To illustrate the challenges of entity linking for Faroese, consider the following sentence from our dataset: *"Herfyri kom flakatrolarin, Enniberg, aftur úr Barentshavinum við 850 tonsum av flaki."* (The factory trawler Enniberg came back from the Barents Sea with 850 tons of fillet.) This example highlights two key challenges. First, **name ambiguity**: "Enniberg" here refers to a fishing vessel, but it shares its name with the famous 754-metre sea cliff in the Faroe Islands and the system must distinguish between these. Second, **morphological variation**: "Barentshavinum" is the dative form of "Barentshav" (Barents Sea) and the system must handle Faroese case inflection to find the correct Wikipedia article. In this case, the system correctly linked "Barentshavinum" to the Faroese Wikipedia article *Barentshavið*, but could not resolve "Enniberg" (returning an empty string), as no Wikipedia article exists for the vessel although one does exist for the cliff.

For Icelandic, recent work has made strides with the development of MIM-GOLD-EL (Friðriksdóttir et al., 2022), a gold-standard entity linking dataset spanning 13 textual domains. This dataset was built using manual review of automated methods, including the mGENRE model (De Cao et al., 2022), a multilingual autoregressive entity linking system. The resulting dataset has enabled systematic evaluation of entity linking approaches for Icelandic (Eggertsson et al., 2023). However, no comparable work exists for Faroese, a closely related North Germanic language spoken by approximately 70,000 people primarily in the Faroe Islands.

Traditional entity linking systems follow a pipeline architecture consisting of mention detection, candidate generation, and entity disambiguation (Özge Sevgili et al., 2022). These systems typically require substantial language-specific resources: named entity recognition models, entity candidate databases, and disambiguation models trained on annotated data (Wu et al., 2020). For low-resource languages, these pipeline components pose particular challenges. Candidate generation approaches that work well for English often cannot be successfully transferred to poorly resourced languages (Garcia-Duran et al., 2022), and entity disambiguation methods typically depend on training data or lack the flexibility to work with domain-specific

Proceedings of The Fourth Workshop on the Role of Resources in the Age of Large Language Models (RESOURCEFUL 2026), pages 32–43 May 11, 2026. ©ELRA Language Resources Association (ELRA), 2026

knowledge bases (Datta and Pramanik, 2024). For Faroese, while a named entity recognition model exists[1] (Snæbjarnarson et al., 2023) and Faroese Wikipedia could serve as a knowledge base, no specific models for candidate generation or entity disambiguation have been developed.

Recent advances in large language models (LLMs) have demonstrated remarkable capabilities across diverse NLP tasks through in-context learning and instruction following (Brown et al., 2020; Achiam et al., 2023). Furthermore, the integration of web search capabilities enables LLMs to access real-time information from the internet (Nakano et al., 2021; Lazaridou et al., 2022), potentially circumventing the need for pre-built entity databases.

In this work, we investigate whether modern LLMs can perform entity linking for Faroese in a zero-shot setting, leveraging web search to identify Wikipedia pages for entity mentions. Specifically, we address two key research questions: (1) Can GPT-5 with web search successfully link Faroese entity mentions to Wikipedia pages without task-specific training? (2) How does performance vary across different entity types?

We utilize an existing Faroese NER dataset (Snæbjarnarson et al., 2023) to extract entity mentions and evaluate an end-to-end LLM-based entity linking approach. To the best of our knowledge, this is the first study of entity linking for Faroese. We release the code and the 1,010 manually annotated examples as a gold-standard entity linking dataset for Faroese to support future research.[2]

2. Methods

2.1. Data

We use the Faroese NER dataset (Snæbjarnarson et al., 2023), which contains named entity annotations for Faroese text from news articles. The source texts are from *Sosialurin*, one of the main Faroese news outlets, and cover a range of local and international topics including politics, sports, culture, and business. The dataset provides entity mentions with type labels including Person, Location, Organization, Date, Time, Money, Percent, and Miscellaneous. Due to LLM processing cost, we sampled 3,000 examples from this dataset, yielding 5,584 entity mentions for entity linking, an average of approximately 1.86 entity mentions per example. For context, Faroese Wikipedia contains approximately 14,196 articles as of February 2026,[3] which is relatively small compared to major languages, setting expectations for entity coverage.

2.2. End-to-End LLM-Based Entity Linking

Unlike traditional entity linking systems that employ separate modules for candidate retrieval and disambiguation (Özge Sevgili et al., 2022; Wu et al., 2020), our approach leverages GPT-5 with integrated web search capabilities to perform entity linking in a single end-to-end step. While systems such as mGENRE (De Cao et al., 2022) link entities in an autoregressive fashion, they are limited by the entities that existed at the time of model training. Recent work has explored using LLMs for entity linking through contextual augmentation (Vollmers et al., 2025) and quality enhancement via question-answering mechanisms (Kamaladdini Ezzabady and Benamara, 2025). However, these approaches still rely on fine-tuned models or post-processing steps. In contrast, LLM approaches with web search are more flexible, enabling real-time access to current Wikipedia content. The model receives the full text and a list of entity mentions, then uses web search to identify appropriate Wikipedia pages.

We use GPT-5 (knowledge cutoff: September 30, 2024) as our primary model, accessed via the OpenRouter API[4] with temperature 1.0 and structured JSON output enforced via a Pydantic schema. We chose GPT-5 because it demonstrated strong multilingual capabilities in preliminary tests, including knowledge of Faroese, and because OpenRouter provides integrated web search for this model. To validate that our findings are not specific to a single model, we additionally evaluate Gemini 3 Flash (with Google Search grounding) and GPT-5.4 Mini in Section 3.5. OpenRouter's integrated web search plugin provides the model with real-time web access at "low" context level, returning a maximum of 10 search results per query. The model autonomously formulates search queries based on the entity mention and surrounding text context, retrieves relevant web pages, and synthesizes information to identify the correct Wikipedia page. The model returns Wikipedia links in the format "PageTitle » language_code" (e.g., "Tórshavn » fo"), or an empty string if no appropriate Wikipedia page exists. The advantages of LLM-based linking over simple Wikipedia API search are quantified in Section 3.4.

To maximize coverage across Wikipedia language editions, we implement a three-tier fallback strategy for each entity. Low-resource language

[1] https://huggingface.co/vesteinn/ScandiBERT-NER

[2] Available at https://github.com/haffi112/faroese-entity-linking

[3] https://fo.wikipedia.org/wiki/Serstakt:Hagt%C3%B8l

[4] https://openrouter.ai

Wikipedias face two fundamental challenges for entity linking: limited coverage due to smaller article counts (Zhou et al., 2019), and quality issues including high percentages of one-line articles and duplicates (Tatariya et al., 2025). Our strategy addresses these challenges by prioritizing language-specific resources while ensuring fallback options. The system first attempts to find entities in Faroese Wikipedia (`https://fo.wikipedia.org`). For entities not found in Faroese Wikipedia, the system searches English Wikipedia (`https://en.wikipedia.org`). Finally, for remaining entities, the system searches across all Wikipedia language editions. This strategy respects language preferences while ensuring broad coverage. The model autonomously performs web searches and selects appropriate pages based on search results and contextual relevance. Importantly, the model was not specifically trained for entity linking; it performs this task through instruction following and in-context reasoning.

Concretely, the three-tier fallback is implemented as three separate API calls with domain-restricted prompts. In the first call, the prompt instructs the model to "Search ONLY the Faroese Wikipedia (fo.wikipedia.org)." Entities that receive a valid link, which is verified by checking whether the linked page has an associated Wikidata ID, are considered resolved; remaining entities proceed to the next tier. The second call instructs the model to "Search ONLY the English Wikipedia (en.wikipedia.org)" for unresolved entities, and the third call instructs "Search any Wikipedia in any language" for any remaining entities. The full prompt template is provided in Appendix A. The structured output format is a JSON object containing entity-link pairs, where each link is formatted as "PageTitle » lang_code" (e.g., "Tórshavn » fo") or an empty string when no appropriate page exists.

2.3. Annotation and Evaluation

For evaluation, we measure both **precision** and **recall** of entity links to assess the quality and coverage of generated Wikipedia links. We developed a web-based annotation interface where human annotators review each predicted entity link and classify it as:

- **Correct:** The Wikipedia link accurately identifies the entity in context
- **Incorrect:** The link is wrong or irrelevant
- **Uncertain:** The annotator cannot confidently determine correctness

We chose a three-category scheme rather than adding a "partly correct" category because entity linking is fundamentally a binary task: a link either resolves to the correct entity or it does not. Partial matches (e.g., metonymic references or links to related but not identical entities) are handled on a case-by-case basis in our qualitative analysis (Section 3.6).

Two annotators independently reviewed a subset of 1,010 examples containing 1,647 predictions for Person, Location, Organization, and Miscellaneous entity types. These four entity types are most relevant for entity linking evaluation, as numerical and temporal entities (Date, Time, Money, Percent) follow more deterministic patterns. Both annotators evaluated all entities in this subset, including cases where the system produced no link (i.e., returned an empty string, indicating the model determined no appropriate Wikipedia page exists), enabling comprehensive evaluation of both linking accuracy and coverage. Note that empty strings differ from links to empty Wikipedia pages (pages that exist in title but contain no content); the latter are discussed in Section 3.2.

We compute performance metrics as follows. **Precision** measures the accuracy of generated links:

$$P = \frac{\text{correct non-empty links}}{\text{total non-empty links}} \tag{1}$$

Recall measures coverage of linkable entities:

$$R = \frac{\text{correct links}}{\text{linkable entities}} \tag{2}$$

where linkable entities include all entities that should have links: correct non-empty links plus all incorrect predictions. Incorrect predictions include both wrong links and missed linking opportunities (entities that should have been linked but received empty strings). Genuinely non-linkable entities that appropriately received empty strings are marked as "correct" for overall accuracy but are excluded from recall calculation, as recall specifically measures linking coverage for entities that should have links.

F1 score combines precision and recall as their harmonic mean:

$$F1 = 2 \times \frac{P \times R}{P + R} \tag{3}$$

Inter-annotator agreement is measured using **percent agreement** (proportion of identical annotations) and **Cohen's Kappa** (Cohen, 1960), which corrects for chance agreement.

In Section 3, we report automatic metrics (link rates by entity type) along with the precision measures.

3. Results

3.1. Link Rate by Entity Type

Table 1 presents the entity linking results by entity type. Overall, GPT-5 produced non-empty

Wikipedia links for 67.6% of the 5,584 entity mentions (3,774 entities linked). A subset of 1,010 examples containing 1,647 entity predictions for Person, Location, Organization, and Miscellaneous types was selected for detailed human evaluation (Section 3.3).

Entity Type	Total	With Link (%)
Organization	1762	1204 (68.3%)
Location	1586	1446 (91.2%)
Person	1479	499 (33.7%)
Date	330	317 (96.1%)
Miscellaneous	187	111 (59.4%)
Money	146	133 (91.1%)
Time	51	23 (45.1%)
Percent	43	41 (95.3%)
Total	**5584**	**3774 (67.6%)**

Table 1: Entity linking results by entity type for Faroese. Link rate indicates the percentage of entities for which the LLM found a Wikipedia page.

Performance varies substantially across entity types, particularly for the core entity categories. Location entities achieve the highest link rate among traditional entity types (91.2%), reflecting strong Wikipedia coverage of geographic entities and the relative ease of disambiguating place names through geographic context.

Organizations show moderate performance (68.3%), likely reflecting the varied nature of this category, which includes both internationally known organizations with multilingual Wikipedia coverage and local Faroese organizations with limited or no Wikipedia presence.

The most challenging category is Person entities, with only 33.7% receiving links (before human evaluation). This low rate may stem from several factors: many mentioned individuals may be local figures without Wikipedia articles, and Faroese naming conventions may complicate search.

Miscellaneous entities achieve a 59.4% link rate, reflecting the heterogeneous nature of this catch-all category. We include this category despite its diversity because the low performance is itself an informative finding that highlights where entity linking struggles; excluding it does not change the overall conclusions. The table also shows results for structured entity types (Dates, Money, Time, Percent), which achieve high link rates but are less central to traditional entity linking evaluation.

3.2. Wikipedia Language Distribution

Table 2 shows the distribution of Wikipedia language editions used for entity linking.

The majority of links (76.5%) point to Faroese Wikipedia, demonstrating that despite its relatively small size, the Faroese Wikipedia contains substantial coverage of entities relevant to Faroese text. English Wikipedia accounts for 20.9% of links, serving as the primary fallback for entities without Faroese articles. The remaining 2.6% of links span various Wikipedia editions (German, Danish, Norwegian, Spanish, Icelandic, etc.), showing the model's ability to identify relevant articles across multiple languages when appropriate.

Wikipedia	Count	Percentage (%)
Faroese (fo)	2888	76.5
English (en)	789	20.9
German (de)	57	1.5
Danish (da)	12	0.3
Norwegian (no)	9	0.2
Spanish (es)	6	0.2
Icelandic (is)	5	0.1
Hungarian (hu)	4	0.1
French (fr)	2	0.1
Russian (ru)	1	0.0
Swedish (sv)	1	0.0

Table 2: Distribution of Wikipedia languages used for entity linking. The model used a fallback strategy: Faroese Wikipedia → English Wikipedia → any Wikipedia.

However, a significant challenge with Faroese Wikipedia links is the prevalence of empty pages. Among the 353 unique Faroese Wikipedia pages checked during annotation, 70 (19.8%) were found to be empty, containing only a placeholder indicating no content exists. These empty pages represent failed linking attempts by the model and were marked as incorrect during evaluation. A practical post-processing step would be to automatically verify that linked pages contain actual content, either through rule-based checking of page length or by instructing the model in the prompt to avoid content-less pages.

The language distribution validates our three-tier fallback strategy: the system prioritizes language-appropriate resources while leveraging larger Wikipedia editions when necessary.

3.3. Precision and Recall Evaluation

Two Faroese-speaking annotators independently reviewed the 1,010-example subset described above. Table 3 presents the overall performance metrics.

Both annotators found high precision and recall, indicating that GPT-5 generates accurate Wikipedia links for the majority of entities. Annotator 1's annotations yielded 87.5% precision and 87.3% recall (F1 = 0.874), while Annotator 2's annotations showed 87.3% precision and 82.8% recall (F1 =

Annotator	P	R	F1	Unc.
Annotator 1	87.5%	87.3%	0.874	2.2%
Annotator 2	87.3%	82.8%	0.850	3.2%

Table 3: Overall performance by annotator on Person, Location, Organization, and Miscellaneous entities. P = Precision, R = Recall, Unc. = Uncertain (%).

Annotator	Type	P	R	F1
Annotator 1	PER	85.9%	83.3%	0.846
	LOC	93.5%	94.6%	0.941
	ORG	82.6%	81.6%	0.821
	MISC	61.8%	63.6%	0.627
Annotator 2	PER	87.7%	71.9%	0.790
	LOC	93.3%	92.5%	0.929
	ORG	81.4%	78.0%	0.797
	MISC	61.8%	58.3%	0.600

Table 4: Performance by entity type for each annotator. Type: PER = Person, LOC = Location, ORG = Organization, MISC = Miscellaneous. P = Precision, R = Recall.

0.850). The small percentage of uncertain cases (2.2% for Annotator 1, 3.2% for Annotator 2) demonstrates that most predictions could be confidently evaluated.

It is important to note that these recall values should be considered upper bounds, as annotators may not have discovered all existing Wikipedia pages when verifying empty model outputs, meaning the true recall could be lower than measured.

Performance varies by entity type, as shown in Table 4. Location entities performed best, with precision exceeding 93% and recall between 92–95% for both annotators. Person entities proved more challenging, with precision around 86–88% and recall between 72–83%. Organizations showed intermediate performance with precision around 81–83% and recall in the 78–82% range. Miscellaneous entities presented the greatest challenge, with both annotators achieving 61.8% precision and recall ranging from 58–64%, reflecting the heterogeneous and often ambiguous nature of this catch-all category. These patterns align with the inherent difficulty of disambiguating person names and organizational entities, which often require more contextual information than geographic locations, while miscellaneous entities suffer from inconsistent definitional boundaries.

Inter-Annotator Agreement. Table 5 shows inter-annotator agreement metrics. Across all 1,607 overlapping predictions, the annotators achieved 90.3% agreement with Cohen's Kappa of 0.584. When excluding cases where either or both annotators marked a prediction as "uncertain," agreement increased to 94.5% on the remaining 1,529 predictions, with Kappa rising to 0.706. By the interpretation of Landis and Koch (1977), the Kappa of 0.706 indicates substantial agreement, while the full-data Kappa of 0.584 reflects moderate agreement. The substantial improvement when excluding uncertain cases (from 0.584 to 0.706) demonstrates meaningful agreement on confident judgments, with the remaining variation reflecting the genuine challenges of entity linking evaluation.

Metric	All	Excl. Unc.	N
Agreement	90.3%	94.5%	1607/1529
Kappa	0.584	0.706	1607/1529

Table 5: Inter-annotator agreement on Person, Location, and Organization entities. Agreement refers to percent agreement; Kappa refers to Cohen's Kappa. "All" includes all annotations; "Excl. Unc." excludes cases where either annotator marked uncertain. N shows total/filtered counts.

Agreement varied by entity type: Location entities showed the highest percent agreement (96.5%), followed by Organizations (89.7%), Persons (85.0%), and Miscellaneous (83.1%). The lower agreement for Person and Miscellaneous entities reflects their greater ambiguity, persons due to difficulty verifying identity from limited textual context, and miscellaneous due to inconsistent category boundaries. The majority of disagreements involved one annotator marking a prediction as "correct" while the other marked it "uncertain" or "incorrect," indicating borderline cases and the challenges of entity linking evaluation for low-resource languages.

3.4. Search Baseline Comparison

To contextualize the LLM's performance, we implemented a Wikipedia API search baseline that links entities using only Wikipedia's built-in search functionality, without any LLM. For each entity mention, the system performs a cascading search using the raw entity string: (1) direct title lookup on Faroese Wikipedia, (2) OpenSearch on Faroese Wikipedia, (3) direct title lookup on English Wikipedia, (4) OpenSearch on English Wikipedia. No contextual disambiguation, morphological normalization, or reasoning about entity identity is involved.

Table 6 compares overall performance. Despite achieving higher raw coverage (74.7% vs. 67.6% of entities receiving non-empty links), the search baseline obtains substantially lower precision and F1 scores.

Table 7 breaks down F1 scores by entity type.

Table 6: LLM vs. search baseline: overall performance. The agreement row uses only the 1,330 entities where both annotators agreed.

Evaluation	System	P	R	F1
Annotator 1	LLM	.875	.873	.874
	Search	.528	.626	.573
Annotator 2	LLM	.873	.828	.850
	Search	.554	.629	.589
Agreement	Search	.564	.641	.600

The LLM outperforms the search baseline across all types, with the largest gaps for Person and Organization entities, where disambiguation requires contextual reasoning that simple string matching cannot provide.

Table 7: F1 by entity type: LLM vs. search baseline.

	LLM		Search	
Type	Ann. 1	Ann. 2	Ann. 1	Ann. 2
Location	.941	.929	.741	.745
Person	.846	.790	.469	.510
Organization	.821	.797	.430	.435
Miscellaneous	.627	.600	.378	.426

Of the search baseline's links, 62.1% were found via Faroese Wikipedia, 12.5% via English Wikipedia, and 25.3% of entities received no link. The higher coverage but lower precision indicates that the search baseline frequently returns plausible but incorrect pages, particularly for ambiguous entity mentions. These results demonstrate that the LLM's primary advantage lies in disambiguation and in knowing when *not* to link, as evidenced by its much higher precision despite lower coverage.

3.5. Multi-Model Comparison

To verify whether our results depend on the specific model, we evaluated Gemini 3 Flash (Google) and GPT-5.4 Mini (OpenAI) on the same Faroese dataset using the same three-tier fallback strategy. Gemini 3 Flash was accessed via the Google Gemini API with native Google Search grounding, while GPT-5.4 Mini used OpenRouter with web search. Both models are evaluated automatically against the gold standard established by our human annotations: entity links where both annotators agreed.

As Table 8 shows, all three models can perform entity linking for Faroese, with F1 scores ranging from 0.740 to 0.874. GPT-5 achieves the highest precision (87.5%) and F1 (0.874), while Gemini 3 Flash and GPT-5.4 Mini show higher recall (93.1–93.5%) but lower precision. This pattern suggests that the primary advantage of frontier models lies in knowing when *not* to link—avoiding incorrect links rather than finding more correct ones. GPT-5.4 Mini, as the smallest model, achieves notably lower precision (61.2%), indicating that disambiguation requires stronger reasoning. The consistent performance across three models from two providers confirms that LLM-based entity linking for low-resource languages is a robust approach, not dependent on a specific model.

Table 8: Multi-model comparison on Faroese entity linking. Link rate is the percentage of entities receiving non-empty links. Precision (P), recall (R), and F1 are computed against the gold standard where both annotators agreed.

Model	Link rate	P	R	F1
GPT-5	67.6%	.875	.873	.874
Gemini 3 Flash	70.5%	.785	.931	.852
GPT-5.4 Mini	77.5%	.612	.935	.740

3.6. Qualitative Analysis

Beyond the quantitative metrics, the annotation process revealed several interesting patterns in how the system handled ambiguous or challenging entity mentions. We summarize key observations below.

First Name Mentions. Entity mentions containing only first names presented particular challenges. When only a first name appeared without identifying surrounding context, we did not penalize the system for failing to generate a link, as it cannot reliably determine which individual is referenced. We chose not to remove these cases from evaluation entirely, as doing so would undercount the system's ability to resolve contextually identifiable first names. Instead, empty strings for genuinely ambiguous first names were marked as "correct," while contextually resolvable mentions were evaluated normally. For instance, the model correctly linked "Eivør" to Eivør Pálsdóttir, the well-known Faroese singer, and "Maria" in religious contexts to Mary, mother of Jesus.

In some cases, the model generated links that were plausible but not definitively verifiable from context alone, such as "Kristian" potentially referring to musician Kristian Blak or "Heðin" to mayor of Tórshavn, Heðin Mortensen. For these ambiguous cases, we labeled them as "correct" when the surrounding text provided sufficient contextual evidence to support the link, and as "uncertain" when the evidence was insufficient for a confident judgment.

In a few instances (e.g., "Lars" and "Torstein"), the model linked to Wikipedia articles about the names themselves rather than specific individuals. These links were marked as “incorrect,” as entity linking requires resolution to specific entities, not name articles.

Morphological and Orthographic Variation. The system demonstrated robust handling of orthographic variation. For example, the model successfully linked “Norra” (an informal colloquial form) to “Noreg” (Norway), and recognized “Tobbi” as an informal variant of “Tórbjørn” as in “Tórbjørn Jacobsen.”

As for morphological variation, results were mixed. In one case, the model linked to an empty page titled with the inflected form (“Gomlurætt”, accusative/dative) instead of the actual article titled “Gamlarætt” (nominative). Such errors occur when the model does not lemmatize search terms.

Partial Entity Matches. In some cases, entity mentions referred to parts of longer entity names or involved metonymic references. For instance, “Landsverkfrøðingurin” (literally “the national civil engineer”) sometimes linked to “Landsverk” (the national building and road administration office), representing a metonymic reference from professional role to organization. We treated such cases individually, accepting links when the partial entity provided meaningful reference to the intended concept, while marking as incorrect when the link missed the entity’s primary meaning.

4. Discussion

Our results demonstrate that modern LLMs with web search capabilities can successfully perform entity linking for low-resource languages without task-specific training. The 67.6% link rate, while lower than state-of-the-art systems for high-resource languages (Özge Sevgili et al., 2022), represents a promising baseline for Faroese, a language with extremely limited NLP resources. A potential advantage of our approach is that LLMs could in principle perform entity linking on any knowledge base with a search interface, whether through web search or via structured APIs. While we only evaluated against Wikipedia in this work, this flexibility could address a limitation of traditional systems that depend on Wikipedia-specific structures and interlanguage links (Fu et al., 2020). Testing with other knowledge bases (e.g., Wikidata, domain-specific databases) remains an important direction for future work.

4.1. Comparison with Traditional Approaches

Traditional entity linking systems require substantial infrastructure: entity mention detection, candidate generation from knowledge bases, and disambiguation models (Rao et al., 2013; Özge Sevgili et al., 2022). For Faroese, building such infrastructure would require: (1) curating entity databases, (2) developing language-specific candidate generation heuristics, (3) training disambiguation models on annotated data. Our LLM-based approach circumvents these requirements by performing all steps end-to-end through prompted web search and reasoning.

Recent work has explored LLM-based entity linking for high-resource languages. Ding et al. (2024a,b) demonstrated that LLMs can perform entity linking through in-context learning. Additionally, Ye and Mitchell (2025) showed that LLMs can serve as effective entity disambiguators in biomedical entity linking, though their approach still requires a separate candidate generation step. Beyond high-resource settings, Boscariol et al. (2025) evaluated LLMs for entity linking in historical documents, addressing the challenge of linking underrepresented, long-tail entities, which is a motivation similar to our work on Faroese. To improve disambiguation performance, Pons et al. (2025) proposed enhancing LLMs with knowledge graphs, leveraging hierarchical class structures to prune candidate spaces and retrieving entity descriptions for disambiguation. While these approaches demonstrate the promise of LLMs for entity linking, they either require fine-tuning, separate candidate generation steps, or access to structured knowledge graph hierarchies, whereas our approach performs end-to-end linking in a zero-shot manner using only web search.

It is worth distinguishing our entity linking task from the related but distinct problem of entity alignment (EA), which matches equivalent entities across different knowledge graphs. Recent LLM-based EA systems such as ChatEA (Jiang et al., 2024) and ProLEA (Munne et al., 2025) use LLMs for cross-KG entity matching, but they operate on structured knowledge graph data rather than linking free-text mentions to knowledge base entries. While both problems benefit from LLM reasoning capabilities, entity linking additionally requires mention detection and handling of surface-form variation (e.g., morphological inflection, abbreviations), making direct comparison non-trivial.

As shown in Section 3.4, a Wikipedia API search baseline without any LLM achieves substantially lower F1 scores (0.57–0.59 vs. 0.85–0.87) despite higher raw coverage, confirming that the LLM’s primary advantage lies in disambiguation rather than retrieval. Future work should additionally compare with multilingual EL systems such as mGENRE

(De Cao et al., 2022).

4.2. Entity Type Analysis

The performance differences across entity types reveal interesting patterns. The strong performance on Location entities (F1 = 0.93–0.94) likely reflects both strong Wikipedia coverage and the relative lack of ambiguity of geographic names, a pattern confirmed by the search baseline (Section 3.4) where the LLM–search gap is smallest for this type.

Person entities (86-88% precision, 72-83% recall) and organizations (81-83% precision, 78-82% recall) highlight a fundamental challenge for low-resource languages: limited cultural representation in the digital world. While geographic entities and international concepts tend to have broad Wikipedia coverage across languages, locally relevant entities, individuals active in Faroese society, local businesses, community organizations, and regional institutions, often lack Wikipedia articles entirely. This coverage gap reflects broader patterns of digital inequality, where the knowledge and cultural contributions of small language communities are underrepresented in global knowledge bases.

This challenge suggests that Wikipedia-based entity linking, while useful for international and well-documented entities, may need complementation with local knowledge bases for comprehensive coverage. Such resources could potentially be constructed from Faroese-language sources including newspaper archives, government proceedings, business registries, and regulatory documents. These local knowledge bases could provide structured information about entities central to Faroese society but absent from Wikipedia, enabling more complete entity linking for low-resource language texts.

4.3. Language Fallback Strategy

The predominance of Faroese Wikipedia links (76.5%) suggests that despite its small size (approximately 14,196 articles as of February 2026[5]), the Faroese Wikipedia provides substantial coverage for entities in Faroese text. This validates our language-prioritized fallback approach. The English Wikipedia serves as an effective secondary resource (20.9%), handling international entities and topics not covered in Faroese Wikipedia.

Among the other languages (2.6%, 97 links), German Wikipedia is the most common (57 links), followed by Danish (12), Norwegian (9), Spanish (6), Icelandic (5), Hungarian (4), French (2), Russian (1), and Swedish (1). Despite the Faroe Islands' strong cultural ties to the Nordic region, Nordic-language Wikipedias account for only 27 of 97 "other" links (27.8%), while German alone accounts for 58.8%. This suggests the model selects based primarily on article availability rather than linguistic similarity, as German Wikipedia's large size (2.8M articles) provides broader coverage. A Nordic-prioritized fallback strategy (Faroese $\rightarrow$ Danish/Norwegian/Icelandic $\rightarrow$ English $\rightarrow$ other) could be explored in future work to better match the cultural context of Faroese texts.

[5] https://fo.wikipedia.org/wiki/Serstakt:Hagt%C3%B8l

5. Conclusion

We presented the first study of entity linking for Faroese, demonstrating that modern large language models with web search capabilities can successfully link entity mentions to Wikipedia pages without requiring task-specific training or language-specific resources. Our end-to-end approach using GPT-5 was evaluated through manual annotation of 1,010 examples, achieving 87.5% precision and 87.3% recall. Location entities showed particularly strong performance with precision exceeding 93% and recall between 92–95%, while person and organization entities achieved 81-88% precision and 72-83% recall. The high inter-annotator agreement (94.5% agreement, Cohen's Kappa 0.706 when excluding uncertain cases) demonstrates the reliability of our evaluation.

The key advantages of our approach are: (1) zero-shot applicability to low-resource languages without requiring annotated training data, (2) elimination of the need for entity databases or candidate generation systems, and (3) dynamic access to current Wikipedia content through web search. The three-tier language fallback strategy effectively leverages language-specific resources while ensuring broad coverage through larger Wikipedia editions.

Comparison with a Wikipedia API search baseline (Section 3.4) confirms that the LLM's contextual reasoning provides substantial gains over simple search (F1 = 0.87 vs. 0.60). This work establishes initial performance benchmarks for Faroese entity linking and provides a framework applicable to other low-resource languages.

6. Acknowledgements

AS is supported by the European Commission under grant agreement no. 101135671. API costs were covered by funding from The Strategic Research and Development Programme for Language Technology.

7. Limitations

Our approach has several limitations worth noting. First, our evaluation focuses on link correctness, whether the system identified the right entity, but does not assess the quality or informativeness of the linked Wikipedia pages themselves. Some correctly linked pages may contain minimal information, while others provide rich content. A more comprehensive evaluation framework could assess page quality alongside link accuracy, and systems with access to multiple knowledge bases could provide users with complementary resources (e.g., linking to both Wikipedia articles and Wikidata entries, or to domain-specific databases alongside general encyclopedias).

Second, our prompting strategy may not be optimal for all entity types (particularly persons, as discussed below), and future work could explore improved prompting techniques. Third, while we focused on Wikipedia as the knowledge base, following standard practice in entity linking research, evaluating whether similar performance can be achieved with other knowledge bases (e.g., Wikidata, DBpedia, or domain-specific databases) remains an open question for future work. The specific limitations are detailed below.

7.1. Computational Cost

A significant limitation of our approach is computational cost. LLM-based entity linking using GPT-5 with web search is substantially more expensive than traditional systems based on smaller specialized models such as BERT-based entity disambiguation systems (Wu et al., 2020; Devlin et al., 2019). Each entity linking can require multiple API calls (for our three-tier fallback strategy) and web search operations, incurring both monetary costs and environmental impact through compute resources. Specifically, GPT-5 via OpenRouter is priced at $1.25 per million input tokens and $10.00 per million output tokens. For our 3,000 examples with up to three API calls each (a maximum of 9,000 calls), the total API cost was approximately $30–50 USD. Web search incurs additional costs via OpenRouter's pricing. This is substantially more expensive than fine-tuning a smaller model, but eliminates the need for collecting task-specific training data, which can be costly.

However, the entity links generated by our LLM-based approach could serve as training data for smaller, more efficient models. Recent work has demonstrated the effectiveness of knowledge distillation from LLMs to smaller models (Gu et al., 2024; Hsieh et al., 2023), and LLM-generated annotations have been successfully used to train task-specific models (Li et al., 2023). Our Faroese entity linking dataset could enable training of lightweight models that retain much of the LLM's capability at a fraction of the inference cost.

7.2. Prompting Strategy and Morphological Handling

Our current prompting approach does not explicitly instruct the model to search for entities in their nominative (dictionary) form, which can lead to issues with morphologically rich languages like Faroese. As noted in our evaluation (Section 3), the model occasionally linked to empty Wikipedia pages titled with inflected forms (e.g., "Gomlurætt" in accusative/dative) rather than finding the actual article with the nominative form ("Gamlarætt"). Enhanced prompts that explicitly request nominative form searches could potentially reduce such errors.

Additionally, our system operates in a single-pass manner without feedback mechanisms. An agentic approach with iterative refinement could improve performance: the system could detect when it has retrieved an empty Wikipedia page and automatically retry with alternative search strategies (e.g., trying the nominative form, or searching in a different Wikipedia language edition). As noted in Section 3.2, 19.8% of linked Faroese Wikipedia pages were empty; a simple rule-based check of page content or an explicit prompt instruction to avoid content-less pages could eliminate this error category. Such feedback loops would increase computational cost but could substantially improve link quality, particularly for morphologically complex entities.

7.3. Fallback Strategy and Regional Relevance

Our three-tier fallback strategy (Faroese $\rightarrow$ English $\rightarrow$ all languages) was designed to balance language-specific coverage with global reach. As discussed in Section 3.2, the "other languages" category is dominated by German Wikipedia rather than Nordic languages, suggesting that the model selects based on article availability rather than cultural proximity. A Nordic-prioritized fallback could better match the cultural context of Faroese texts, as many regionally relevant entities may have better coverage in Nordic-language Wikipedias.

7.4. Knowledge Base Coverage

Using Faroese Wikipedia as a primary knowledge base presents inherent limitations. Many Faroese companies, institutions, and public figures lack Wikipedia pages, meaning entities central to Faroese society may remain unlinked due to knowledge base absence rather than system failure. The low link rate for Person entities (33.7%) particularly reflects this gap. Building complementary local knowledge bases from newspaper archives, government records, or business registries could improve coverage, though disambiguation challenges may persist when textual context is insufficient to uniquely identify individuals.

8. Ethics Statement

We use the publicly available Faroese NER dataset (Snæbjarnarson et al., 2023); our entity linking annotations do not introduce privacy concerns beyond those in the source data. We acknowledge the environmental cost of LLM-based processing, but frame this work as methodology exploration that could enable training of more efficient specialized models. This work contributes to technological equity for speakers of low-resource languages by demonstrating viable entity linking approaches. LLMs may exhibit biases (Bender et al., 2021) that could affect linking decisions, though we did not explicitly measure such effects.

9. Bibliographical References

Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, page 610–623, New York, NY, USA. Association for Computing Machinery.

Marta Boscariol, Luana Bulla, Lia Draetta, Beatrice Fiumanò, Emanuele Lenzi, and Leonardo Piano. 2025. Evaluation of llms on long-tail entity linking in historical documents.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

Silviu Cucerzan. 2007. Large-scale named entity disambiguation based on Wikipedia data. In *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pages 708–716, Prague, Czech Republic. Association for Computational Linguistics.

Debarghya Datta and Soumajit Pramanik. 2024. Unsupervised named entity disambiguation for low resource domains. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 14922–14928, Miami, Florida, USA. Association for Computational Linguistics.

Nicola De Cao, Ledell Wu, Kashyap Popat, Mikel Artetxe, Naman Goyal, Mikhail Plekhanov, Luke Zettlemoyer, Nicola Cancedda, Sebastian Riedel, and Fabio Petroni. 2022. Multilingual autoregressive entity linking. *Transactions of the Association for Computational Linguistics*, 10:274–290.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Yifan Ding, Amrit Poudel, Qingkai Zeng, Tim Weninger, Balaji Veeramani, and Sanmitra Bhattacharya. 2024a. Entgpt: Entity linking with generative large language models. *arXiv preprint arXiv:2402.06738*.

Yifan Ding, Qingkai Zeng, and Tim Weninger. 2024b. ChatEL: Entity linking with chatbots. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 3086–3097, Torino, Italia. ELRA and ICCL.

Valdimar Ágúst Eggertsson, Benedikt Geir Jóhannesson, Hafsteinn Einarsson, and Hrafn Loftsson. 2023. Effective entity disambiguation in low-resource languages: A study of icelandic.

In *2023 IEEE/WIC International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, pages 318–324.

Steinunn Rut Friðriksdóttir, Valdimar Ágúst Eggertsson, Benedikt Geir Jóhannesson, Hjalti Daníelsson, Hrafn Loftsson, and Hafsteinn Einarsson. 2022. Building an Icelandic entity linking corpus. In *Proceedings of the Workshop on Dataset Creation for Lower-Resourced Languages within the 13th Language Resources and Evaluation Conference*, pages 27–35, Marseille, France. European Language Resources Association.

Xingyu Fu, Weijia Shi, Xiaodong Yu, Zian Zhao, and Dan Roth. 2020. Design challenges in low-resource cross-lingual entity linking. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6418–6432, Online. Association for Computational Linguistics.

Alberto Garcia-Duran, Akhil Arora, and Robert West. 2022. Efficient entity candidate generation for low-resource languages. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 6429–6438, Marseille, France. European Language Resources Association.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. 2024. MiniLLM: Knowledge distillation of large language models. In *The Twelfth International Conference on Learning Representations*.

Cheng-Yu Hsieh, Chun-Liang Li, Chih-kuan Yeh, Hootan Nakhost, Yasuhisa Fujii, Alex Ratner, Ranjay Krishna, Chen-Yu Lee, and Tomas Pfister. 2023. Distilling step-by-step! outperforming larger language models with less training data and smaller model sizes. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 8003–8017, Toronto, Canada. Association for Computational Linguistics.

Xuhui Jiang, Yinghan Shen, Zhichao Shi, Chengjin Xu, Wei Li, Zixuan Li, Jian Guo, Huawei Shen, and Yuanzhuo Wang. 2024. Unlocking the power of large language models for entity alignment. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7566–7583, Bangkok, Thailand. Association for Computational Linguistics.

Morteza Kamaladdini Ezzabady and Farah Benamara. 2025. Entity quality enhancement in knowledge graphs through LLM-based question answering. In *Proceedings of the Workshop on Generative AI and Knowledge Graphs (GenAIK)*, pages 136–145, Abu Dhabi, UAE. International Committee on Computational Linguistics.

J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics*, pages 159–174.

Angeliki Lazaridou, Elena Gribovskaya, Wojciech Stokowiec, and Nikolai Grigorev. 2022. Internet-augmented language models through few-shot prompting for open-domain question answering. *arXiv preprint arXiv:2203.05115*.

Zhuoyan Li, Hangxiao Zhu, Zhuoran Lu, and Ming Yin. 2023. Synthetic data generation with large language models for text classification: Potential and limitations. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 10443–10461, Singapore. Association for Computational Linguistics.

Rumana Ferdous Munne, Md Mostafizur Rahman, and Yuji Matsumoto. 2025. Entity profile generation and reasoning with LLMs for entity alignment. In *Findings of the Association for Computational Linguistics: EMNLP 2025*, pages 20073–20086, Suzhou, China. Association for Computational Linguistics.

Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, et al. 2021. Webgpt: Browser-assisted question-answering with human feedback. *arXiv preprint arXiv:2112.09332*.

Gerard Pons, Besim Bilalli, and Anna Queralt. 2025. Knowledge graphs for enhancing large language models in entity disambiguation. In *The Semantic Web – ISWC 2024*, pages 162–179, Cham. Springer Nature Switzerland.

Delip Rao, Paul McNamee, and Mark Dredze. 2013. *Entity Linking: Finding Extracted Entities in a Knowledge Base*, pages 93–115. Springer Berlin Heidelberg, Berlin, Heidelberg.

Vésteinn Snæbjarnarson, Annika Simonsen, Goran Glavaš, and Ivan Vulić. 2023. Transfer to a low-resource language via close relatives: The case study on Faroese. In *Proceedings of the 24th Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 728–737, Tórshavn, Faroe Islands. University of Tartu Library.

Kushal Tatariya, Artur Kulmizev, Wessel Poelman, Esther Ploeger, Marcel Bollmann, Johannes Bjerva, Jiaming Luo, Heather Lent, and Miryam de Lhoneux. 2025. How good is your wikipedia? auditing data quality for low-resource and multilingual nlp.

Daniel Vollmers, Hamada Zahera, Diego Moussallem, and Axel-Cyrille Ngonga Ngomo. 2025. Contextual augmentation for entity linking using large language models. In *Proceedings of the 31st International Conference on Computational Linguistics*, pages 8535–8545, Abu Dhabi, UAE. Association for Computational Linguistics.

Ledell Wu, Fabio Petroni, Martin Josifoski, Sebastian Riedel, and Luke Zettlemoyer. 2020. Scalable zero-shot entity linking with dense entity retrieval. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 6397–6407, Online. Association for Computational Linguistics.

Christophe Ye and Cassie S. Mitchell. 2025. LLM as entity disambiguator for biomedical entity-linking. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 301–312, Vienna, Austria. Association for Computational Linguistics.

Shuyan Zhou, Shruti Rijhwani, and Graham Neubig. 2019. Towards zero-resource cross-lingual entity linking. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)*, pages 243–252, Hong Kong, China. Association for Computational Linguistics.

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.

A. Prompt Template

This appendix presents the full prompt template used for entity linking. The template is parameterized by the input text, the list of entity mentions, and the target Wikipedia domain. Below we show the general template with placeholders, followed by the three domain-specific instruction variants and the output schema.

General Prompt Template

```
You are an expert entity linking system.
Your task is to link named entities to
Wikipedia pages.

Text (in {language}):
{text}

Entities to link:
{entity_list_json}

Instructions:
1. For each entity, find the most relevant
   Wikipedia page.
2. {wikipedia_domain_instruction}
3. Return the Wikipedia page title and
   language code in the format:
   "PageTitle >> lang_code"
   - Example: "{example_format}"
4. If NO relevant Wikipedia page exists for
   an entity, return an empty string "" for
   the link.
5. Be precise - only return a link if you
   are confident it correctly identifies the
   entity in context.
6. IMPORTANT: Only return links from the
   specified Wikipedia domain.

CRITICAL: Return ONLY the JSON structured
output. Do not include any explanatory text.
```

Wikipedia Domain Instructions

The `{wikipedia_domain_instruction}` placeholder is replaced with one of the following three variants:

1. **Tier 1 (Faroese):** "Search ONLY the Faroese Wikipedia (fo.wikipedia.org). Do not use other Wikipedias." (Example format: "Tórshavn » fo")
2. **Tier 2 (English):** "Search ONLY the English Wikipedia (en.wikipedia.org). Do not use other Wikipedias." (Example format: "Iceland » en")
3. **Tier 3 (Any):** "Search any Wikipedia in any language (wikipedia.org)." (Example format: "Berlin » de")

Output Schema (Pydantic)

The model is constrained to return structured JSON output conforming to the following Pydantic schema:

```
class EntityLink(BaseModel):
    entity: str   # Entity name from input
    link: str     # "PageTitle >> lang_code"
                  # or "" if no link found

class EntityLinkingOutput(BaseModel):
    links: List[EntityLink]
```