

CommonMorph: Participatory Morphological Documentation Platform

Aso Mahmudi[∅], Sina Ahmadi³, Kemal Kurniawan[∅],
Rico Sennrich³, Eduard Hovy[∅], Ekaterina Vylomova[∅]

[∅]The University of Melbourne, ³University of Zurich
{mahmudia, vylomovae}@unimelb.edu.au

Abstract

Collecting and annotating morphological data present significant challenges, requiring linguistic expertise, methodological rigour, and substantial resources. These barriers are particularly acute for low-resource languages and varieties. To accelerate this process, we introduce CommonMorph, a comprehensive platform that streamlines morphological data collection development through a three-tiered approach: expert linguistic definition, contributor elicitation, and community validation. The platform minimises manual work by incorporating active learning, annotation suggestions, and tools to import and adapt materials from related languages. It accommodates diverse morphological systems, including fusional, agglutinative, and root-and-pattern morphologies. Its open-source design and UniMorph-compatible outputs ensure accessibility and interoperability with NLP tools. Our platform is accessible at <https://common-morph.com>, offering a replicable model for preserving linguistic diversity through collaborative technology.

Keywords: Morphology, UniMorph, Multilingual Resources, Low-Resource Languages

1. Introduction

With over 1,500 languages at risk of extinction by 2100 (Bromham et al., 2022), scaling and improving the efficiency of language documentation and data collection is essential to prevent further irreversible losses. During the “International Decade of Indigenous Languages” (2022–2032) (United Nations, 2019), tools that assist linguists and communities in collecting linguistic data for long-term documentation and preservation should be a priority. Morphological data collection and analysis, the systematic documentation of word structure and inflection patterns, remains one of the most labour-intensive steps of language documentation (Ginn et al., 2023), particularly for low-resource languages with complex inflectional systems.

Finite-state transducers (FSTs), the most established rule-based approach to model morphologies (Gorman and Sprout, 2022), are capable of generating and analysing wordforms licensed by a given grammar. However, the design and implementation of FSTs are labour-intensive, requiring extensive linguistic engineering expertise and a high degree of manual effort (Beemer et al., 2020). Handling irregularities often requires ad hoc solutions, while maintaining quality across large-scale FSTs remains difficult. These factors make it challenging to scale FST-based approaches to the hundreds of under-documented languages urgently requiring attention. Moreover, FSTs do not generalise effectively to out-of-vocabulary (OOV) forms, which limits their applicability in many real-world documentation and processing scenarios.

Modern data-driven approaches to morphologi-

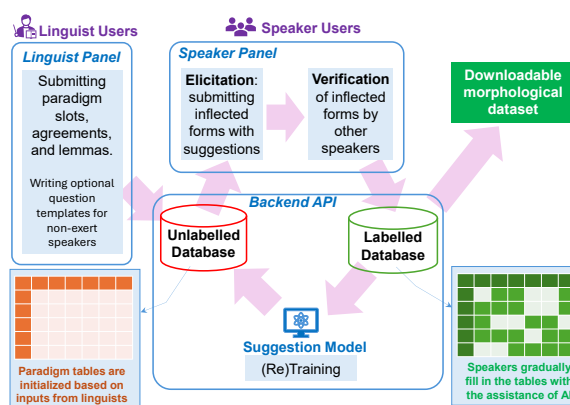


Figure 1: The CommonMorph platform workflow facilitates elaboration of morphological structures by a linguist and provides an interoperable ecosystem for contributors to validate and enrich labelled databases.

cal analysis and generation rely on large volumes of annotated morphological data—resources that are rarely available for low-resource or endangered languages. The UniMorph project (Batsuren et al., 2022b), while the most comprehensive of its kind, provides datasets for only 169 languages, or less than 2.5% of the world’s languages. UniMorph has advanced computational morphology and enabled SIGMORPHON shared tasks on morphological segmentation (Batsuren et al., 2022a), analysis (McCarthy et al., 2019), and generation (Kann and Schütze, 2018), but faces key limitations: (a) the data preparation process (Kirov et al., 2018) is time-consuming and not easily replicable across languages; (b) systematic

correction and maintenance require substantial resources; and (c) validation is limited, as there are no mechanisms for broad community feedback. For example, [Ahmadi and Mahmudi \(2023\)](#) identified and corrected 25% of the misleading forms in the Central Kurdish dataset.

A further limitation across existing approaches is the lack of systematic integration of related language varieties. Since most languages belong to a few hundred families ([Campbell, 2018](#)), related varieties often share morphological, phonological, and lexical patterns. Exploiting these similarities can reduce annotation effort and accelerate documentation, yet current tools rarely support such reuse.

This paper presents `CommonMorph`, a platform designed to streamline and scale the collection of inflectional morphological data across languages. By connecting linguists and speakers in a shared digital environment, `CommonMorph` supports collaborative annotation of wordforms with morphosyntactic features. While the system is applicable to all languages, it is particularly well-suited to low-resource and endangered ones, whose diverse and complex morphological systems require adaptable and scalable tools. Unlike traditional fieldwork practices that rely on isolated and offline contributions, the platform enables structured data entry, review, and community validation. It also enhances efficiency through imports from related languages and predictive suggestions. Developed iteratively through user feedback, `CommonMorph` is tailored to the workflows of both linguists and speaker communities.

To demonstrate its broad applicability, we tested `CommonMorph` on languages from diverse families, including Indo-European (Italic, Germanic, Iranian), Semitic, Bantu, and Turkic. The case studies show that the platform can handle various morphological systems, from fusional to agglutinative to root-and-pattern structures.

In short, `CommonMorph` offers the following core features:

1. **Intuitive Interface:** Developed in collaboration with both linguists and speakers, the platform is intuitive and easy to use, with high user satisfaction reported in surveys (§ 6).
2. **Linguist-led Workflow:** Linguists can define initial materials and create patterns to reduce the workload for speakers, improving the efficiency of the early annotation process (§ 4.4). These materials can be adapted from existing resources of related languages in the system.
3. **Built-in Validation:** The platform’s interactive evaluation of suggestions allows users to iteratively refine the data (§ 4.5).
4. **Compatibility:** Outputs can be exported in standard formats such as UniMorph ([Batsuren](#)

[et al., 2022b](#)), enabling integration with existing NLP tools.

5. **Free and Open-Source:** `CommonMorph` is accessible to all and openly available for further development and adaptation.¹

2. Related Work

Recent work in computational linguistics has explored alternative methods for collecting morphological data, ranging from gamified participation to automated annotation. Gamification has been applied to morphological analysis, education, and community engagement, showing potential for improving data coverage and speaker involvement ([Eryiğit et al., 2023](#); [Qiao et al., 2023](#); [Sun et al., 2025](#)). The rise of large language models (LLMs) has also enabled new forms of automatic linguistic annotation across a wide variety of tasks ([Tan et al., 2024](#)).

Despite these innovations, the development of high-quality morphological data and tools still depends heavily on manual work. [Beemer et al. \(2020\)](#) estimate that developing a finite-state transducer (FST) analyser and generator comparable to a neural model requires an average of 40 person-hours, excluding infrastructure and design time. The effort varies substantially across languages: while regular inflectional systems can be implemented in under an hour, more complex ones demand extensive analysis of paradigms, lemma classes, and morphophonological alternations. These figures assume prior expertise, underscoring the difficulty of scaling traditional approaches.

To mitigate annotation costs, active learning methods have been explored for morphological segmentation and labelling, where algorithms iteratively select the most informative examples for human review ([Baldrige and Palmer, 2009](#); [Palmer, 2009](#)). Building on this idea, [Mahmudi et al. \(2025\)](#) propose techniques to help field linguists detect and prioritise data gaps for more efficient elicitation.

A major effort in large-scale morphological data collection is the UniMorph project. The first release by [Kirov et al. \(2016\)](#) relied on manual mapping of Wiktionary grammatical descriptors to the UniMorph schema. In the second release report, [Kirov et al. \(2018\)](#) showed that verifying and correcting automatically parsed tables required at least as much effort as direct manual annotation, though annotators successfully produced initial datasets for 47 languages within days. The third phase ([McCarthy et al., 2020](#)) expanded coverage to 50 more

¹Website: <https://common-morph.com>, the source code is available at <https://github.com/Aso-UniMelb/CommonMorph>

languages with a team of nine annotators.

Data validation within UniMorph evolved from manual expert review to more systematic quality checks. Early annotations were adjudicated by linguists or speakers (Kirov et al., 2018), while later efforts integrated automated validation tools to flag errors such as rare characters, malformed cells, or inconsistencies between lemma and inflected form (McCarthy et al., 2020). These steps improved quality control but still relied heavily on post-hoc correction.

Despite these advances, to the best of our knowledge, no existing morphological data collection tool provides a unified platform that integrates structured workflows involving both linguists and speakers while supporting remote collaboration. Our CommonMorph platform fills this gap by combining user-friendly interfaces, active speaker participation, and features for guided elicitation. It supports both manual and semi-automated annotation and encourages the use of cross-linguistic data to reduce redundancy and improve scalability.

3. Design Principles

The CommonMorph platform is designed to address the core research question: *How can morphological data collection be streamlined and scaled for low-resource and endangered languages through collaboration between linguists and speakers?* To achieve this, the platform adheres to principles that balance the needs of its primary contributors, linguists and speakers (Flavelle and Lachler, 2023), while accommodating diverse morphological systems.

Linguists play a central role in defining the scope and quality of morphological resources. The platform enables them to define paradigms, morphosyntactic features, and lexical entries across diverse language varieties with fine-grained control. They can follow established annotation standards while also adding community-accepted aliases for grammatical terms. Materials from related languages can be imported and adapted, allowing cross-linguistic regularities to reduce redundant work. Linguists can also revise and expand data as new insights emerge, ensuring that resources remain linguistically accurate and up to date. These capabilities ensure that expert knowledge anchors the data collection process and maintains its reliability.

Speakers are the primary custodians of endangered and low-resource languages. To empower non-experts to participate, the platform offers an intuitive interface that enables speakers to annotate and validate forms without requiring specialist training. Its flexible workflows adapt to varying levels of grammatical knowledge within the com-

munity, while community feedback loops capture corrections, variations, and alternative forms. Active learning-driven suggestions for inflected forms, based on patterns in the data, help reduce manual effort. This principle recognises speakers not merely as informants but as active collaborators, fostering inclusivity and shared ownership of the resource.

4. System Design

The CommonMorph platform, as depicted in Figure 1, begins with linguists defining initial linguistic materials. Based on this setup, the system generates suggested inflected forms, which speakers review by correcting or confirming them. As verified data accumulates, the system improves its suggestions through active learning, reducing manual effort over time. All data remains editable and is exportable in standard formats for integration with other linguistic and NLP tools. In the following subsections, we provide a detailed description of our approach.

4.1. Terminology

Before outlining the system design, we clarify the terminology used throughout the platform:

- A **lemma** (e.g., walk) is the base dictionary form representing a set of inflected wordforms (e.g., walks, walking).
- An **inflection class** groups lemmas that share the same inflectional pattern. For instance, most English regular verbs follow a four-form pattern (walk-s, walk-ing, walk-ed, walk-ed (past participle)).
- A **paradigm** is the complete set of inflected forms derived from a single lemma, typically organised according to morphosyntactic features such as tense, number, or person.
- A **gloss** is a short explanation or translation that conveys the meaning or function of a form.
- A **morphosyntactic feature set** is a bundle of grammatical properties such as tense, person, or number. For consistency, CommonMorph adopts the UniMorph schema (Sylak-Glassman, 2016) for glossing these features.
- **Morphophonological rules** describe spelling changes that occur when morphemes combine in particular contexts (e.g., the plural of bus becomes buses).

4.2. Required Linguistic Materials

We assume the availability of some linguistic materials, either gathered during early fieldwork stages or sourced from previous descriptive resources.

Note that we anticipate these materials may contain errors, and linguists should have the ability to correct or expand them at any time. These materials should include:

1. **Basic wordlist or vocabulary** (Bower, 2015, Sec. 3.1.3) covering the inflection classes of the language, and
2. **Morphosyntactic feature combinations** for each inflection class, typically organised in paradigms or conjugation tables. These combinations describe how lexical items vary according to grammatical categories such as person, number, tense, aspect, mood, or case. Figure 3 provides an example of such a resource.

In practice, documentation often begins with incomplete information that must be iteratively refined. This initial data may be sourced from prior descriptive studies or inferred from related language varieties. While linguistic resources exist for many languages, including endangered ones, their structure and quality vary widely. Some wordlists and lexicons are highly detailed, while others lack even basic part-of-speech tags. Similarly, morphosyntactic features can appear as detailed grammatical descriptions or as simple tabular paradigms. To manage this variation, linguists are guided to enter and standardise the materials into our platform in a consistent and unified format.

4.3. User Roles

In traditional linguistic fieldwork, the linguist typically guides the elicitation process based on their expertise. In a digital environment, however, managing this interaction and ensuring that contributors enter accurate data can be challenging. We designed the CommonMorph user interface to accommodate users with varying levels of familiarity with linguistic concepts.

CommonMorph considers two roles of users:

- **Linguists** (or language enthusiasts with morphological expertise) initiate a project by providing lemmas, paradigm structures (§4.2), and optionally inflectional patterns and annotation priorities. They oversee data quality, guide the system's predictions, and ensure consistency.
- **Speakers** (informants or consultants without formal linguistic training) contribute by supplying inflected forms with hints from system-generated suggestions and verifying entries submitted by other speakers, enabling community-driven data validation.

Role-specific interfaces ensure that expert guidance and community participation are both sup-

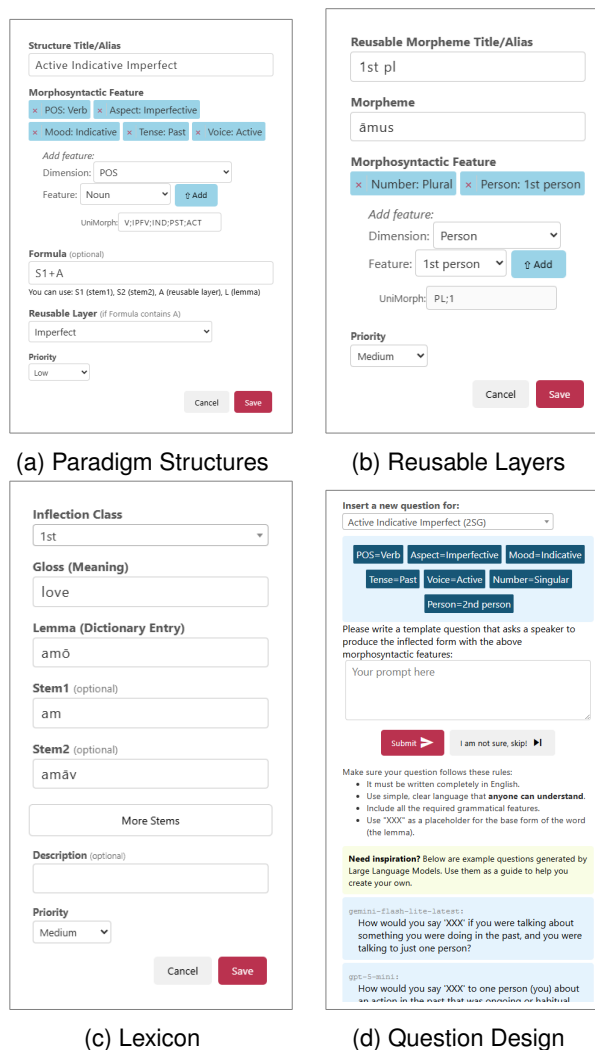


Figure 2: Linguists define (a) paradigm structures, (b) additional structural layers such as agreement, and (c) lexical entries. These components allow the generation of morphological forms that can be validated during elicitation. Linguists also create elicitation prompts (d), taking inspiration from examples produced by the system.

ported. In the following subsections, we describe the user interface tailored for each user role.

4.4. Linguist Interface

The linguist interface is structured in a modular fashion to minimise redundancy and to exploit morphological regularities. It provides a workflow for defining paradigms, entering lexical items, and specifying rules, while allowing flexibility for irregular or complex systems. Figure 2 shows the interface during data submission. Core and optional modules include:

1. **Paradigm Structures** (Figure 2a): This core module defines the morphosyntactic structures that make up inflectional paradigms,

such as tense, aspect, and mood. Each structure is tied to a specific inflection class. Optional inflectional patterns with placeholders allow the system to generate candidate forms before the machine learning models have sufficient data.

2. **Lexicon** (Figure 2b): Here, lemmas are entered with their assigned inflection class, stems, and glosses. Glosses ensure clarity for both contributors during elicitation and researchers during analysis.
3. **Reusable Layers** (Figure 2c): To avoid redundancy, this module allows defining reusable sets of features (e.g., agreement affixes) linked to paradigm structures. Simpler inflectional systems, such as English verbs, may not require this module.
4. **Morphophonological rules (Optional)**: This module encodes orthographic adjustments (e.g., Turkish vowel harmony) as replacement patterns, implemented as regular expressions to reduce the need for manual corrections.
5. **Question Design (Optional)** (Figure 2d): This module enables linguists to design elicitation prompts for non-expert speakers. Prompts are written in a shared meta-language (e.g., English or Spanish) familiar to both linguist and speaker, and structured as reusable templates with placeholders for lemmas. For example, the prompt “How do you tell more than one person to [LEMMA] right now?” is used to generate a form with the following morphosyntactic features: POS=Verb, Mood=Imperative, Tense=Present, Number=Plural, and Person=2nd. Prompts can be created manually, adapted from other projects, or generated using LLMs (see Appendix B for the exact prompt template used for generation). This feature is especially valuable when speakers are unfamiliar with grammatical terminology.

Figure 3 illustrates an example of the information linguists can add through the interface. They can also assign priority levels to lemmas and structures, directing speakers to contribute the most informative data first.

To support scalability, cross-linguistic reuse, and community-driven resource sharing, the platform features import and export functionalities. Linguistic materials, including paradigm structures and lexicons, can be imported and exported in TSV format. This enables researchers to host and share resources openly and to adapt existing datasets from related languages with minimal effort.

Conjugation table of the verb amō (love) (First conjugation)				
	Active Indicative Present	Active Indicative Imperfect	Active Indicative Future	Active Indicative Perfect
1st Singular	amō	amābam	amābō	amāvī
2nd Singular	amās	amābās	amābis	amāvistī
3rd Singular	amat	amābat	amābit	amāvit
1st Plural	amāmus	amābāmus	amābimus	amāvimus
2nd Plural	amātis	amābātis	amābitis	amāvistis
3rd Plural	amant	amābant	amābunt	amāverunt

Figure 3: Example of a Latin verb conjugation table annotated with CommonMorph terminology: (1) An *inflection class* and (2) a *paradigm structure* within that inflection class; (3) a *reusable layer* contains several (4) *reusable morphemes* and their (5) morphosyntactic features; (6) a *lemma* and its (7) *gloss* and (8–9) different *stems*.

4.5. Speaker Interface

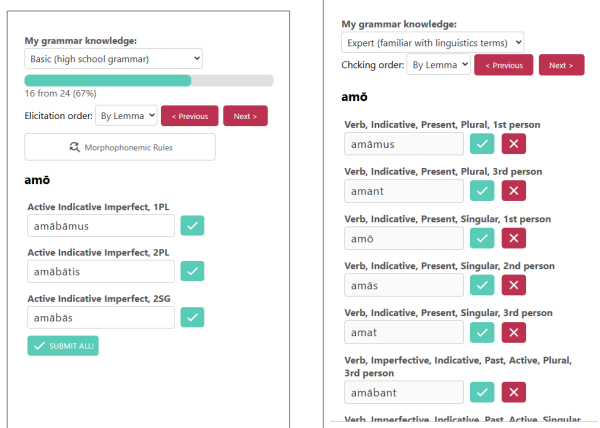
Once the linguist has provided the initial materials, the speaker panel becomes available for elicitation, as illustrated in Figure 4. The panel supports two main activities: data entry (elicitation) and verification (review), with the interface adapting to the speaker’s level of grammatical knowledge.

The system combines the defined structures with the lemmas to automatically generate paradigm tables that cover the full range of potential inflected forms. Since the number of possible forms can be very large, the platform presents speakers with only a subset, prioritised according to the linguist’s annotation settings and the informativeness criteria outlined in Section 4.6.

To provide transparency regarding how candidate morphological entries are computed, the interface explicitly informs the user about the source of each suggestion. Suggestions are visually tagged to indicate whether they were generated by the linguist’s predefined rules, the internal neural inflection model, or an LLM.

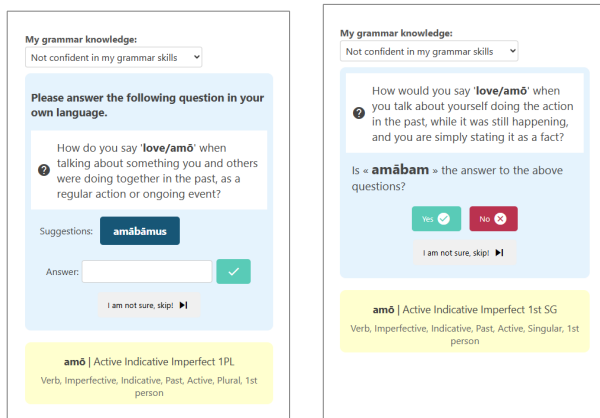
Furthermore, because language documentation often reveals dialectal variation or simple human error, the review interface includes a mechanism for handling disagreements between speakers. If multiple speakers propose or validate conflicting forms for the same morphosyntactic slot, the platform flags the entry. These flagged disagreements are then either resolved through a majority consensus among designated expert speakers or escalated to the linguist for final adjudication.

The interface adapts to different levels of expertise: (1) Expert speakers see larger sets of forms, such as full or partial paradigm tables, which they can complete or verify in bulk. (2) Non-expert speakers see one form at a time with a simplified prompt that helps them provide or confirm the



(a) Eliciting expert speakers

(b) Checking by expert speakers



(c) Eliciting non-expert speakers

(d) Checking by non-expert speakers

Figure 4: Screenshots from the speaker interface.

correct wordform without needing specialist terminology.

This adaptive design allows contributions from both experts and non-experts, combining detailed linguistic knowledge with broader community input.

4.6. Inflection Model in Active Learning

The elicitation process follows an active learning framework (Mahmudi et al., 2025), progressing in iterative cycles to minimise annotation cost. CommonMorph utilises a dynamic, multi-source prediction pipeline that transitions from rule-based heuristics to data-driven models as the dataset grows.

Cold Start: Lacking initial training data, the system relies exclusively on linguist-supplied inflection patterns and morphophonological rules to generate early suggestions.

Active Learning: Validated forms periodically train a language-specific neural inflection model. This model drives elicitation by measuring predictive uncertainty to prioritise the “most informative”

unannotated forms, while simultaneously generating increasingly accurate suggestions for speakers to review.

Integrating LLMs: To bridge the low-data phase where custom neural models struggle, we leverage lower-latency LLMs (e.g., Gemini-2.5-Flash, gpt-5-mini) via API. The system dynamically constructs few-shot prompts using previously validated forms with matching morphosyntactic features as in-context examples (Appendix A), enabling accurate predictions without explicit fine-tuning.

Ensemble Presentation: The system aggregates suggestions from all three sources (rules, neural model, LLM). Converging predictions are presented with high confidence, while disagreements yield a multiple-choice list where each option is visually tagged by its generative source (Section 4.5). A comparative analysis of these models is provided in Section 6.1.

4.7. Implementation

We implemented CommonMorph as a web application to make it freely accessible online, allowing researchers and practitioners to use it without installation barriers or platform restrictions. Since speakers of some endangered languages may not have stable internet access, the system includes a feature that allows linguists to print empty tables after submitting the initial data. These tables can then be filled out offline in remote areas and later integrated back into the system.

The backend and frontend are implemented as a .NET web application. All data is stored in a PostgreSQL database using relational tables to minimise redundancy. The application can be deployed with Docker and runs on both Linux and Windows servers. The active learning component is developed using FastAPI and PyTorch.

To implement the neural inflection model for active learning, we designed a simple and fast network that can be trained or fine-tuned on minimal server specifications. The model is a 2-layer LSTM with hidden and embedding dimensions of 128. All input sequences are padded or truncated to a fixed length of 20 character tokens, which enables efficient batch processing and stable training. The network is trained on a CPU for 15 epochs, taking approximately one minute, making it suitable for rapid experimentation in low-resource environments.

All development libraries and code used are free and open-source. For simplicity, we do not train or fine-tune LLMs locally; access to LLMs is provided via API.

5. Case studies

To demonstrate the platform's ability to support languages with diverse morphological systems, we conducted a series of case studies covering a wide range of language families. The selected languages include Spanish and Latin (Italic), English and German (Germanic), Hawrami Kurdish, Central Kurdish, and Farsi (Iranic), Arabic (Semitic), Swahili (Bantu), and Turkish (Turkic).

Spanish, German, English, Farsi, and Kurdish primarily exhibit fusional suffixation with stem alternations. Hawrami additionally involves reduplication. Turkish employs agglutinative suffixation with vowel harmony, while Arabic features a templatic, non-concatenative morphology.

The platform accommodates these typological differences through flexible paradigm definitions and stem management. In root-and-pattern systems, each lemma may specify multiple stems, allowing individual consonants of the root to be mapped to appropriate vocalic templates. Morphophonological rules further enable linguists to encode complex rewrite operations, such as the vowel harmony constraints observed in Turkish. By evaluating CommonMorph across these morphologically distinct systems, we assess its general applicability to typologically varied languages.

Furthermore, to assess the platform's ability to support data collection using resources from related language varieties, we conducted experiments with several varieties from the Middle East. These experiments show how the platform can leverage linguistic similarity to assist under-resourced varieties. Data collection for each variety involved one linguist and one speaker. Initially, the standard varieties of Central Kurdish and Farsi were added to the platform. Afterwards, additional contributors added their own varieties, using the existing data as a reference when relevant.

Hawrami, an endangered language,² now has comprehensive digital data on the verbal morphology of its two dialects for the first time. We hope that these data will be used in future NLP applications to support the preservation of the language.

6. Evaluation and Results

It is difficult to establish a baseline for evaluating the efficiency of our platform because no existing system offers comparable functionalities. As discussed in Section 2, some previous studies on morphological data collection have provided rough estimates of manual effort. However, we consider such comparisons inappropriate as our platform in-

tegrates a wider range of components and involves users with less specialised expertise.

To assess the effectiveness of CommonMorph, we conduct a multi-faceted evaluation focused on system performance and user experience. Our evaluation aligns with prior work in human-in-the-loop language documentation, where the focus is on reducing annotation effort while maintaining quality (Palmer et al., 2009; Baldrige and Palmer, 2009).

6.1. Annotation Suggestion Accuracy

We employed an active learning strategy to improve the quality of annotation suggestions over time. The accuracy of model-generated suggestions was evaluated by comparing them to user-validated annotations using the Character Error Rate (CER) metric. Lower CER values indicate fewer corrections needed by annotators, and thus, higher model usefulness.

Table 1 presents the Character Error Rate (CER) results for 3,000 randomly selected samples from several language varieties. These include Hawrami (Tewêle dialect), Central Kurdish (Sine dialect), Farsi (non-standard Tehrani dialect), Standard Turkish, Standard Spanish, and fully diacritized Modern Standard Arabic. For the first two datasets, no prior morphological data were available; they are published in CommonMorph for the first time.

We compare multiple suggestion model configurations, including linguist-provided patterns (with and without morphophonological rules), neural networks trained on varying amounts of data, and few-shot prompting with two LLMs. The results demonstrate that linguist-defined morphological patterns substantially enhance model performance for zero-resource varieties, underscoring the importance of expert linguistic knowledge in low-resource settings. Incorporating morphophonological rules further enhances suggestion accuracy for most varieties, particularly for Turkish, where vowel harmony plays a central role. However, providing detailed linguistic rules can be time-consuming, and in some cases, linguists may have limited knowledge of the target variety. In scenarios involving related languages where sufficient prior knowledge exists, the linguist user can instead focus on identifying potential points of deviation, reducing manual effort.

For neural network models, CER consistently decreases with additional training data, validating the effectiveness of active learning. Few-shot prompting with LLMs—particularly with two- or three-shot inputs—achieves competitive results for well-resourced standard languages. However, in the case of Arabic, model performance was lower due to the use of fully diacritised data, while most Arabic training data available to LLMs is undiacritised, leading to less accurate suggestions.

²<https://unesdoc.unesco.org/ark:/48223/pf0000187026>

Suggestion Model	Hawrami (Tewêle)	Kurdish (Sine)	Farsi (Tehran)	Turkish (Standard)	Spanish (Spain)	Arabic (Standard)
Linguist patterns	12.33	1.39	0.86	31.74	0.82	6.62
Linguist patterns + morphophonology	5.78	0.74	0.85	2.97	0.82	2.24
NN trained on 100 samples	48.41	54.23	37.25	23.31	20.98	41.93
NN trained on 500 samples	15.13	29.12	23.63	11.83	7.18	8.45
NN trained on 1000 samples	7.88	27.37	54.02	3.33	5.91	6.40
NN trained on 2000 samples	6.46	14.36	24.22	4.68	1.14	2.72
gemi-2.5-flash (1-shot)	17.00	14.86	3.71	2.36	0.50	7.25
gemi-2.5-flash (2-shot)	14.45	13.43	1.86	1.5	0.50	4.59
gemi-2.5-flash (3-shot)	13.03	6.86	2.32	2.15	0.50	5.07
gpt-5-mini (1-shot)	19.26	15.14	4.64	0.43	0.00	10.14
gpt-5-mini (2-shot)	11.61	14.57	4.64	0.21	0.25	7.97
gpt-5-mini (3-shot)	12.18	7.43	3.25	0.21	0.00	4.83

Table 1: Character Error Rate (CER, in %) of the inflection model in generating suggestions for different languages, which reflects the amount of correction required by speakers.

6.2. User Survey

To evaluate the usability and effectiveness of CommonMorph, we conducted a user survey involving both linguists and speakers. Participants reported their learning time (in minutes), contribution time (in hours), and rated the following aspects on a 5-point Likert scale: **ease of use** for clarity of the interface and task design, **efficiency** for users' perception of how to quickly and effectively complete tasks, **pedagogical value** for if the platform helps speakers and learners better understand the language structure, and **satisfaction** for overall user experience and willingness to reuse the platform.

As shown in Table 2, the platform received consistently high ratings across all categories, with average satisfaction and ease-of-use scores exceeding 4.7/5. Speakers reported shorter learning curves (8.7 minutes) compared to linguists (17.3 minutes), yet contributed substantially more time to the documentation effort. Both user groups gave strong pedagogical ratings, suggesting that the platform effectively supports language learning and awareness. The positive feedback indicates it helps bridge the expertise gap in collaborative documentation.

7. Conclusion and Future Work

This study presents CommonMorph, a collaborative platform designed to document and expand morphological resources. Our evaluation demonstrates that the platform effectively supports collaborative data collection by integrating rule-based linguistic input with machine learning suggestions derived from active learning. Both linguists and community speakers found the interface intuitive and easy to use, highlighting its potential for participatory language documentation.

A key outcome of this work is the demonstrated benefit of combining structured linguistic knowl-

Metric	Linguists		Speakers	
	Mean	SD	Mean	SD
User Learning Time	17.3 min	2.5	8.7 min	3.2
Contributing Time	1.0 hr	2.5	4.1 hr	9.2
Pedagogical Score	3.7/5	1.5	4.7/5	0.5
Efficiency Score	4.3/5	1.1	4.7/5	0.5
Satisfaction Score	4.7/5	0.5	4.7/5	0.5
Ease of Use Score	4.7/5	0.5	5.0/5	0

Table 2: User survey results for the CommonMorph platform. The table reports average user learning and contribution times, as well as mean scores (1 = low, 5 = high) for pedagogical value, efficiency, satisfaction, and ease of use. SD indicates standard deviation.

edge with the pattern-recognition capabilities of neural models. Initial morphological rules contributed by linguists substantially improved the quality of subsequent system-generated suggestions, reducing manual annotation effort while preserving data accuracy. This hybrid workflow proved especially valuable during the initial stages of data collection and enabled high-quality input from users with varying levels of expertise.

As the amount of collected data increases, the platform has the potential to offer a universal morphological analysis system. Additionally, because it captures glossing and morphosyntactic tags, it would assist linguists in generating interlinear glossed text (IGT) data.

Future work will focus on expanding both the scope and automation of the platform. One promising direction is the automatic extraction of morphological data and rule patterns from descriptive linguistic sources such as grammars and dictionaries (Virk et al., 2020), allowing community reviewers to verify and refine the output. Extending the framework beyond morphology to other linguistic domains, including phonology, syntax, and semantics, would further enhance its usefulness

for language documentation and computational modelling. Finally, because many endangered languages lack standardised orthographies, incorporating speech-based data entry, through manual verification or automatic speech recognition, will further enhance accessibility and inclusivity in language resource development.

Ethical Considerations

The CommonMorph platform does not publish any personal or identifiable information from its users. All linguistic data contributed through the platform is intended for public release via open-access repositories. The study design and participant involvement have been reviewed and approved by the University of Melbourne Human Research Ethics Committee (Reference No. 2025-32066-64105-3). Our analyses are based on licensed data, which are freely available for academic use.

Acknowledgements

We extend our deepest appreciation to the contributors of our case studies, especially Ako Marani, Mohammadreza Yadegari, Erfan Karami, and Morteza Naserzadeh for their invaluable time, dedication, and linguistic expertise. We are also deeply grateful to Borja Herce, John Mansfield, and Demian Aaron Inostroza Améstica for their helpful feedback and suggestions during the preparation of CommonMorph. Furthermore, we would like to thank the anonymous reviewers for their constructive feedback, which significantly improved the final version of this paper.

Sina Ahmadi gratefully thanks the support of the UZH Grant (reference number 269093). Ekaterina Vylomova expresses gratitude to the University of Melbourne for providing support for this research (ECR grant 2024ECRG140).

8. Bibliographical References

Sina Ahmadi and Aso Mahmudi. 2023. [Revisiting and amending Central Kurdish data on UniMorph 4.0](#). In *Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 38–48, Toronto, Canada. Association for Computational Linguistics.

Jason Baldrige and Alexis Palmer. 2009. [How well does active learning actually work? Time-based evaluation of cost-reduction strategies for language documentation](#). In *Proceedings of the*

2009 Conference on Empirical Methods in Natural Language Processing, pages 296–305, Singapore. Association for Computational Linguistics.

Khuyagbaatar Batsuren, Gábor Bella, Aryaman Arora, Viktor Martinovic, Kyle Gorman, Zdeněk Žabokrtský, Amarsanaa Ganbold, Šárka Dohnalová, Magda Ševčíková, Kateřina Pelegrinová, Fausto Giunchiglia, Ryan Cotterell, and Ekaterina Vylomova. 2022a. [The SIGMORPHON 2022 shared task on morpheme segmentation](#). In *Proceedings of the 19th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 103–116, Seattle, Washington. Association for Computational Linguistics.

Khuyagbaatar Batsuren, Omer Goldman, Salam Khalifa, Nizar Habash, Witold Kieraś, Gábor Bella, Brian Leonard, Garrett Nicolai, Kyle Gorman, Yustinus Ghanggo Ate, Maria Ryskina, Sabrina Mielke, Elena Budianskaya, Charbel El-Khaissi, Tiago Pimentel, Michael Gasser, William Abbott Lane, Mohit Raj, Matt Coler, Jaime Rafael Montoya Samame, Delio Siconatzi Camaiteri, Esaú Zumaeta Rojas, Didier López Francis, Arturo Oncevay, Juan López Bautista, Gema Celeste Silva Villegas, Lucas Torroba Hennigen, Adam Ek, David Guriel, Peter Dirix, Jean-Philippe Bernardy, Andrey Scherbakov, Aziyana Bayyr-ool, Antonios Anastasopoulos, Roberto Zariquiey, Karina Sheifer, Sofya Ganieva, Hilaria Cruz, Ritván Karahóga, Stella Markantonatou, George Pavlidis, Matvey Plugaryov, Elena Klyachko, Ali Salehi, Candy Angulo, Jatayu Baxi, Andrew Krizhanovskiy, Natalia Krizhanovskaya, Elizabeth Salesky, Clara Vania, Sardana Ivanova, Jennifer White, Rowan Hall Maudslay, Josef Valvoda, Ran Zmigrod, Paula Czarnowska, Irene Nikkarinen, Aelita Salchak, Brijesh Bhatt, Christopher Straughn, Zoey Liu, Jonathan North Washington, Yuval Pinter, Duygu Ataman, Marcin Wolinski, Totok Suhardijanto, Anna Yablonskaya, Niklas Stoehr, Hossep Dolatian, Zahroh Nuriah, Shyam Ratan, Francis M. Tyers, Edoardo M. Ponti, Grant Aiton, Aryaman Arora, Richard J. Hatcher, Ritesh Kumar, Jeremiah Young, Daria Rodionova, Anastasia Yemelina, Taras Andrushko, Igor Marchenko, Polina Mashkovtseva, Alexandra Serova, Emily Prud'hommeaux, Maria Nepomniashchaya, Fausto Giunchiglia, Eleanor Chodroff, Mans Hulden, Miikka Silfverberg, Arya D. McCarthy, David Yarowsky, Ryan Cotterell, Reut Tsarfaty, and Ekaterina Vylomova. 2022b. [UniMorph 4.0: Universal Morphology](#). In *Proceedings of the Thirteenth Language Resources and*

- Evaluation Conference*, pages 840–855. European Language Resources Association.
- Sarah Beemer, Zak Boston, April Bukoski, Daniel Chen, Princess Dickens, Andrew Gerlach, Torin Hopkins, Parth Anand Jawale, Chris Koski, Akanksha Malhotra, Piyush Mishra, Saliha Muradoglu, Lan Sang, Tyler Short, Sagarika Shreevastava, Elizabeth Spaulding, Testumichi Umada, Beilei Xiang, Changbing Yang, and Mans Hulden. 2020. [Linguist vs. machine: Rapid development of finite-state morphological grammars](#). In *Proceedings of the 17th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 162–170, Online. Association for Computational Linguistics.
- Gülşen Eryiğit, Fatih Bektaş, Ubey Ali, and Bihter Dereli. 2023. [Gamification of complex morphology learning: the case of Turkish](#). *Computer Assisted Language Learning*, 36(8):1421–1449.
- Darren Flavelle and Jordan Lachler. 2023. [Strengthening relationships between indigenous communities, documentary linguists, and computational linguists in the era of NLP-assisted language revitalization](#). In *Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP)*, pages 25–34, Dubrovnik, Croatia. Association for Computational Linguistics.
- Michael Ginn, Sarah Moeller, Alexis Palmer, Anna Stacey, Garrett Nicolai, Mans Hulden, and Miikka Silfverberg. 2023. [Findings of the SIGMORPHON 2023 Shared Task on Interlinear Glossing](#). In *Proceedings of the 20th SIGMORPHON workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 186–201, Toronto, Canada. Association for Computational Linguistics.
- Kyle Gorman and Richard Sproat. 2022. [Finite-state text processing](#). Springer Nature.
- Katharina Kann and Hinrich Schütze. 2018. [Neural transductive learning and beyond: Morphological generation in the minimal-resource setting](#). In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3254–3264, Brussels, Belgium. Association for Computational Linguistics.
- Christo Kirov, Ryan Cotterell, John Sylak-Glassman, Géraldine Walther, Ekaterina Vylomova, Patrick Xia, Manaal Faruqui, Sabrina J. Mielke, Arya McCarthy, Sandra Kübler, David Yarowsky, Jason Eisner, and Mans Hulden. 2018. [UniMorph 2.0: Universal Morphology](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Christo Kirov, John Sylak-Glassman, Roger Que, and David Yarowsky. 2016. [Very-large scale parsing and normalization of Wiktionary morphological paradigms](#). In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3121–3126, Portorož, Slovenia. European Language Resources Association (ELRA).
- Aso Mahmudi, Borja Herce, Demian Inostroza Améstica, Andreas Scherbakov, Eduard H. Hovy, and Ekaterina Vylomova. 2025. [Can a neural model guide fieldwork? a case study on morphological data collection](#). In *Proceedings of the 18th Workshop on Building and Using Comparable Corpora (BUCC)*, pages 62–72, Abu Dhabi, UAE. Association for Computational Linguistics.
- Arya D. McCarthy, Christo Kirov, Matteo Grella, Amrit Nidhi, Patrick Xia, Kyle Gorman, Ekaterina Vylomova, Sabrina J. Mielke, Garrett Nicolai, Miikka Silfverberg, Timofey Arkhangel'skiy, Nataly Krizhanovsky, Andrew Krizhanovsky, Elena Klyachko, Alexey Sorokin, John Mansfield, Valts Ernštreits, Yuval Pinter, Cassandra L. Jacobs, Ryan Cotterell, Mans Hulden, and David Yarowsky. 2020. [UniMorph 3.0: Universal Morphology](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 3922–3931, Marseille, France. European Language Resources Association.
- Arya D. McCarthy, Ekaterina Vylomova, Shijie Wu, Chaitanya Malaviya, Lawrence Wolf-Sonkin, Garrett Nicolai, Christo Kirov, Miikka Silfverberg, Sabrina J. Mielke, Jeffrey Heinz, Ryan Cotterell, and Mans Hulden. 2019. [The SIGMORPHON 2019 shared task: Morphological analysis in context and cross-lingual transfer for inflection](#). In *Proceedings of the 16th Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pages 229–244, Florence, Italy. Association for Computational Linguistics.
- Alexis Palmer, Taesun Moon, and Jason Baldridge. 2009. [Evaluating automation strategies in language documentation](#). In *Proceedings of the NAACL HLT 2009 Workshop on Active Learning for Natural Language Processing*, pages 36–44, Boulder, Colorado. Association for Computational Linguistics.
- Alexis Mary Palmer. 2009. [Semi-automated annotation and active learning for language documentation](#). Ph.D. thesis, The University of Texas at Austin.

- Shen Qiao, Samuel Kai Wah Chu, and Susanna Siu-sze Yeung. 2023. [Understanding how gamification of english morphological analysis in a blended learning environment influences students' engagement and reading comprehension](#). *Computer Assisted Language Learning*, pages 1–34.
- Jiugeng Sun, Rita Sevastjanova, Sina Ahmadi, Rico Sennrich, and Mennatallah El-Assady. 2025. [Dia-linge: A gamified interface for dialectal data collection](#). In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, pages 148–158, Vienna, Austria. Association for Computational Linguistics.
- John Sylak-Glassman. 2016. [The composition and use of the universal morphological feature schema \(Unimorph schema\)](#). *Johns Hopkins University*.
- Zhen Tan, Dawei Li, Song Wang, Alimohammad Beigi, Bohan Jiang, Amrita Bhattacharjee, Mansoreh Karami, Jundong Li, Lu Cheng, and Huan Liu. 2024. [Large language models for data annotation and synthesis: A survey](#). In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*, pages 930–957, Miami, Florida, USA. Association for Computational Linguistics.
- Shafqat Mumtaz Virk, Harald Hammarström, Markus Forsberg, and Søren Wichmann. 2020. [The DReaM corpus: A multilingual annotated corpus of grammars for the world's languages](#). In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 878–884, Marseille, France. European Language Resources Association.
- United Nations. 2019. Rights of indigenous peoples: Resolution adopted by the general assembly on 18 december 2019. UN General Assembly Resolution 74/135, on the report of the Third Committee (A/74/396). Available at: <https://www.unesco.org/en/decades/indigenous-languages>.

9. Language Resource References

- Claire Bowern. 2015. *Linguistic fieldwork: A practical guide*. Springer.
- Lindell Bromham, Russell Dinnage, Hedvig Skirgård, Andrew Ritchie, Marcel Cardillo, Felicity Meakins, Simon Greenhill, and Xia Hua. 2022. [Global predictors of language endangerment and the future of linguistic diversity](#). *Nature ecology & evolution*, 6(2):163–173.
- Lyle Campbell. 2018. [How many language families are there in the world?](#) *Anuario del Seminario de Filología Vasca "Julio de Urquijo"*, 52(1/2):133–152.

A. Prompt for Suggestion Generation

The following prompt is employed to obtain a suggestion when two data samples from an identical morphosyntactic feature set have been submitted and verified by speakers. In languages like Kurdish, where verb lemmas often have two distinct stems, linguists can provide these stems in advance. By retrieving them from the database during prompting, the system can construct more detailed prompts, increasing the likelihood of generating accurate suggestions.

A. Prompt for generating inflected forms suggestions

In the language YYY, what is the correct inflected form of the lemma <LEMMA>, given that its stem1 is "<STEM1>" and stem2 is "<STEM2>" for a specific grammatical feature set? As reference examples, under the same grammatical features: - The lemma <SAMPLE.1.LEMMA> (with stem1: "<SAMPLE.1.STEM.1>", stem2: "<SAMPLE.1.STEM.2>") yields the form <SAMPLE.1.FORM>. - The lemma <SAMPLE.2.LEMMA> (with stem1: "<SAMPLE.2.STEM.1>", stem2: "<SAMPLE.2.STEM.2>") yields the form <SAMPLE.2.FORM>. Think about this quietly and just give me one final inflected word.

For example, its placeholders will be replaced with the following words:

Placeholder	Replaces with
<LEMMA>	hênan
<STEM1>	hêna
<STEM2>	hên
<SAMPLE.1.LEMMA>	girtin
<SAMPLE.1.STEM.1>	girt
<SAMPLE.1.STEM.2>	gir
<SAMPLE.1.FORM>	degirim
<SAMPLE.2.LEMMA>	kuştin
<SAMPLE.2.STEM.1>	kuşt
<SAMPLE.2.STEM.2>	kuj
<SAMPLE.2.FORM>	dekujim

B. Prompt for Elicitation Question Generation

The following prompt is used to generate question templates that can help the linguist create elicitation questions for speakers who are not familiar with grammatical terminology.

B. Prompt for generating English question templates

You are a field linguist working with native speakers to study the morphology of their language. The speakers are not trained in linguistics but understand English. Ask a speaker how they would say the word "XXX" with these specific features: "<FEATURE-VALUES SET PLACEHOLDER>". Generate only the question for the speaker. "XXX" should be in the question. Do not use linguistic terms. Keep it under 50 words. Do not explain anything to me.

For example, the UniMorph tag set "V;PST" is substituted with "*Part-of-Speech=Verb, Tense=Past*", which will then replace the corresponding placeholder in the prompt.