CommonMorph

A Crowd-Sourcing Platform for **Morphological Resources**

Aso Mahmudi[1] Sina Ahmadi[2] Kemal Kurniawan[1] Rico Sennrich[2] Eduard Hovy[1] Ekaterina Vylomova[1]
[1]The University of Melbourne [2]University of Zurich

LREC 2026 Palma

The Gap

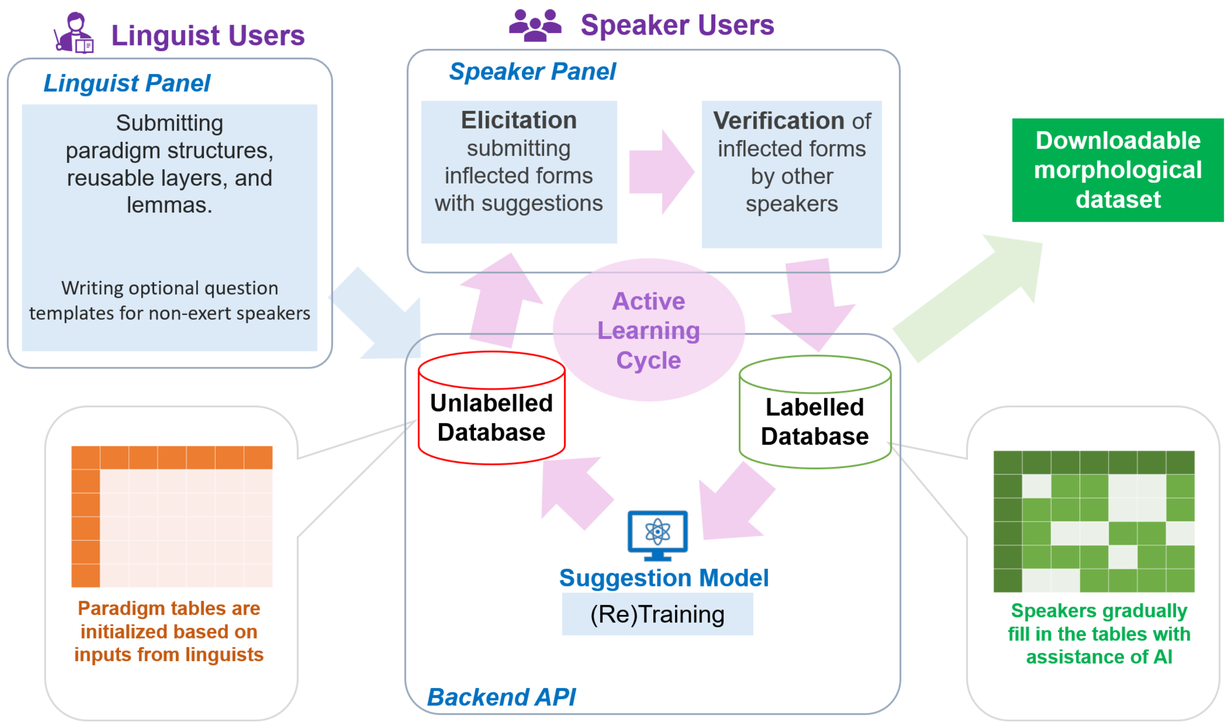
How can morphological data collection scale through structured collaboration between linguists and speakers?

1,500+
languages at risk of extinction by 2100

169
in UniMorph today (< 2.5% of the world's)

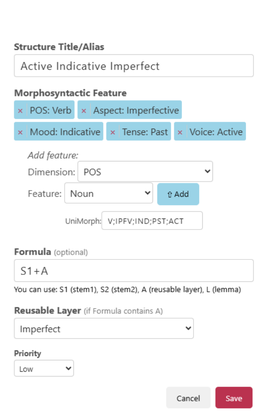
40 hours
per language to build a morphological grammar

What is CommonMorph?

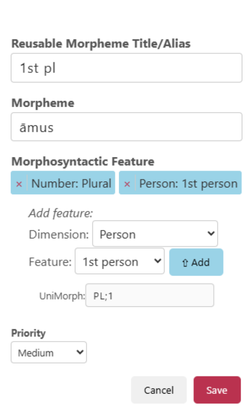


Linguist Interface

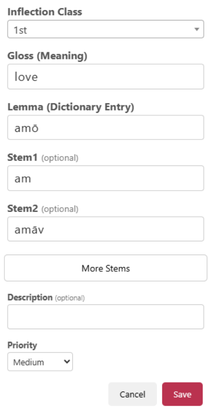
Linguists define paradigms, lexicon, reusable layers, and optional elicitation prompts. Materials can be **imported and adapted** from related varieties.



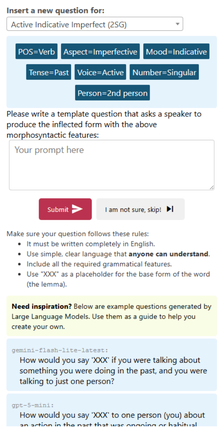
Paradigm structures



Reusable layers



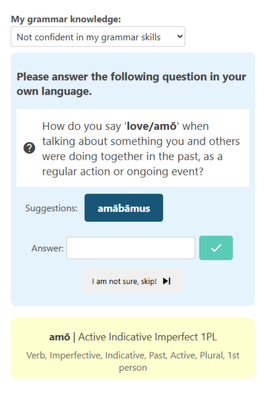
Lexicon



Question design

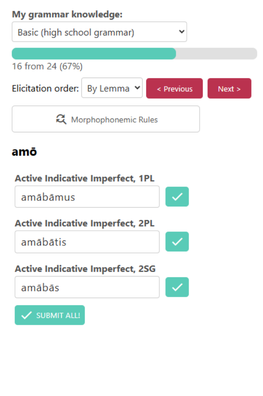
Speaker Interfaces

The interface adapts to the speaker's expertise. **Experts** verify whole tables; **non-experts** answer one plain-language question at a time. Suggestions are **tagged by source**: rule, neural net, or LLM.

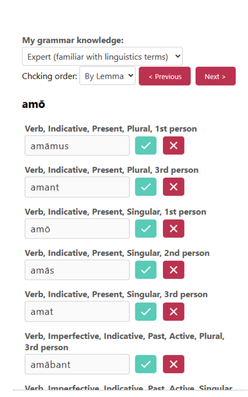


Elicit (non-expert)

Verify (non-expert)

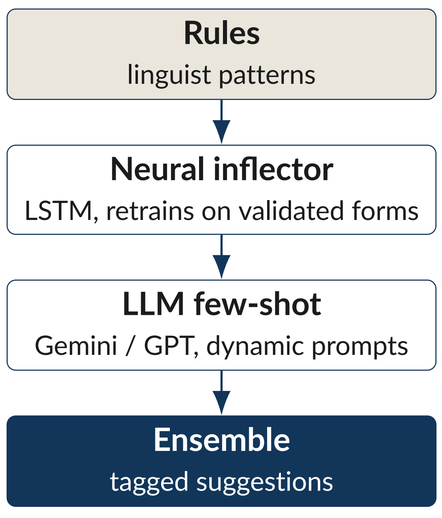


Elicit (expert)



Verify (expert)

Active Learning



Speakers correct fewer suggestions over time as the ensemble learns from validated forms.

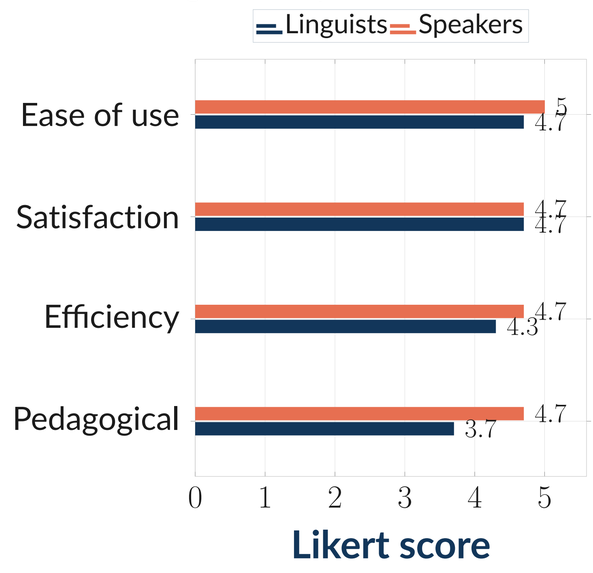
Suggestion Accuracy across Case Studies

Character Error Rate (%) on 3,000 held-out forms — lower is better.

Model	Hawrami	C. Kurdish	Farsi	Turkish	Spanish	Arabic
Linguist rules						
Patterns only	12.33	1.39	0.86	31.74	0.82	6.62
+ morphophon.	**5.78**	**0.74**	**0.85**	2.97	0.82	**2.24**
Neural inflector						
100 samples	48.41	54.23	37.25	23.31	20.98	41.93
500 samples	15.13	29.12	23.63	11.83	7.18	8.45
1,000 samples	7.88	27.37	54.02	3.33	5.91	6.40
2,000 samples	6.46	14.36	24.22	4.68	1.14	2.72
LLM few-shot						
gemini-2.5 (1)	17.00	14.86	3.71	2.36	0.50	7.25
gemini-2.5 (2)	14.45	13.43	1.86	1.50	0.50	4.59
gemini-2.5 (3)	13.03	6.86	2.32	2.15	0.50	5.07
gpt-5-mini (1)	19.26	15.14	4.64	0.43	**0.00**	10.14
gpt-5-mini (2)	11.61	14.57	4.64	**0.21**	0.25	7.97
gpt-5-mini (3)	12.18	7.43	3.25	**0.21**	**0.00**	4.83

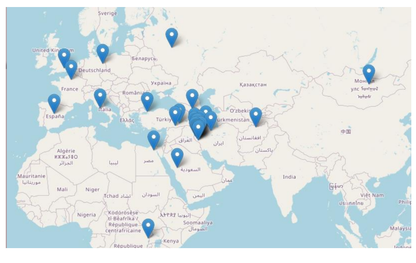
User survey

Mean ratings on a 5-point Likert scale.



Speakers onboard in **8.7 min** and contribute **4.1 h** on average.

Your language next?



Five language families, four morphological types.
Bring your own variety.

Takeaways

- **CommonMorph** is a participatory platform for documenting morphology — linguists scaffold, speakers fill in.
- **Adaptive interfaces** let both experts and non-experts contribute meaningfully.
- **Active learning + LLMs** make every annotation cycle faster.
- **Cross-variety reuse** bootstraps related languages from existing scaffolds.
- **New data released**: first comprehensive verbal morphology of Hawrami.
- **Open-source and UniMorph-compatible**.

Try it / Fork it

Aso-UniMelb/CommonMorph
common-morph.com