

Long-Term Sign Language Data Crowdsourcing Through Collaborative Lexicons

Pierre Poitier, Jérôme Fink, Ariel Basso Madjoukeng, Adélaïde Couplet, Margaux Leleu, Benoît Frénay
University of Namur, Belgium, Computer Science Faculty

The Issue...

Sign Languages (SLs) are the primary mode of communication for deaf communities, yet they remain among the **least documented languages** in the world. Three barriers stand out:

- ❖ **Data Scarcity.** Few thousands sign occurrences, compared to billions of word occurrences.
- ❖ **Lack of recognition.** 59% of countries have not officially recognized their national SL, leading to stigmatization and limited funding for documentation.
- ❖ **Undocumented linguistic dynamics.** No direct equivalence with spoken languages and evolve rapidly: neologisms and dialectal variations that static dictionaries fail to capture.

Our Tool: Mot-Signe (MOSI)

We introduce **Mot-Signe (MOSI)**, a collaborative *French ↔ French-Belgian Sign Language (LSFB)* platform. MOSI is grounded in a **community need** and turns **long-term data collection** into a beneficial side effect.

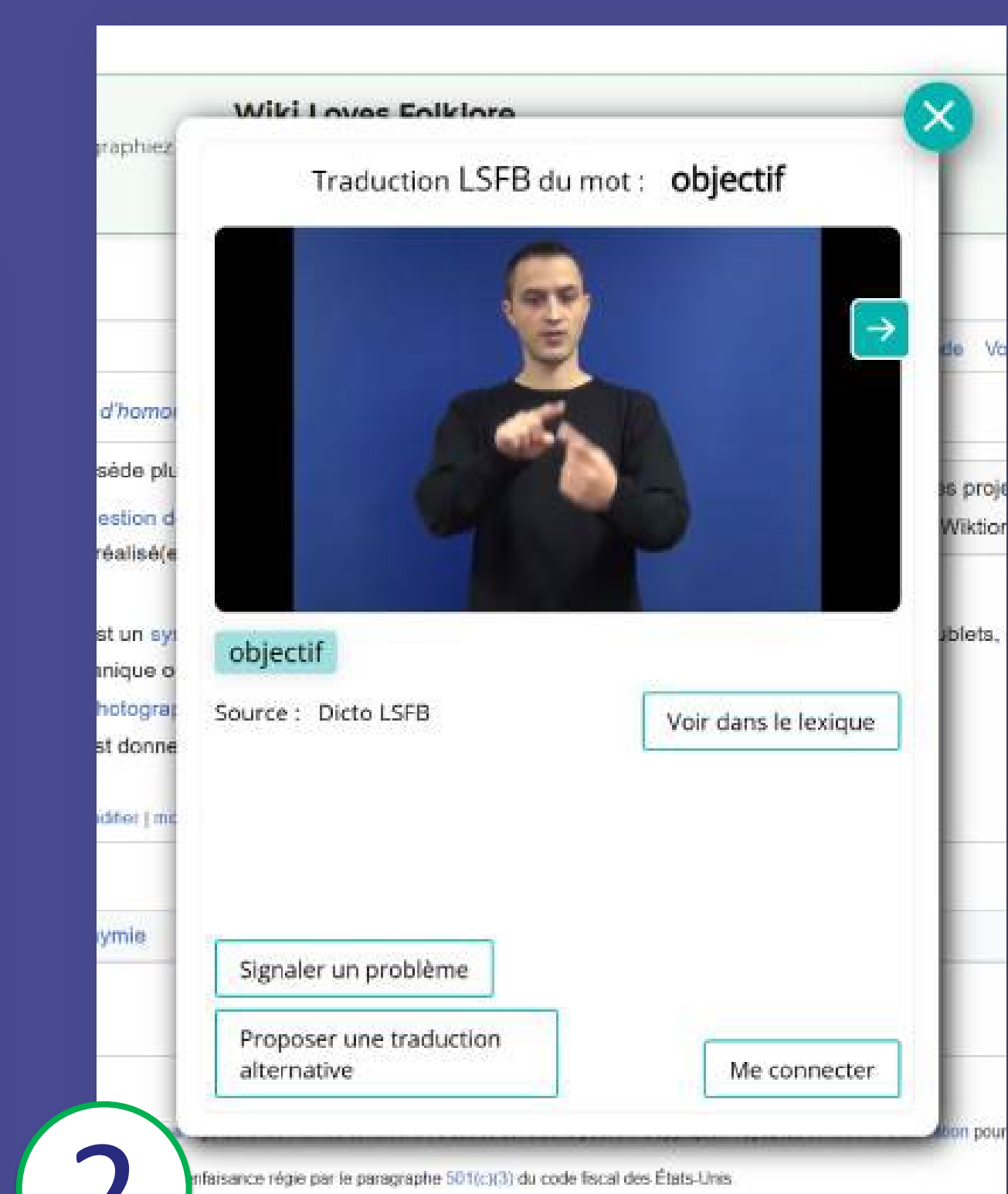
Literature

Data collection as the primary goal. Rely on **studio recordings** with few signers under controlled conditions, while online dictionaries only allow users to query a static, expert-curated vocabulary. Crowdsourcing platforms exist but rely on **one-shot** campaigns, offering little long-term retention or coverage of specialized terms.

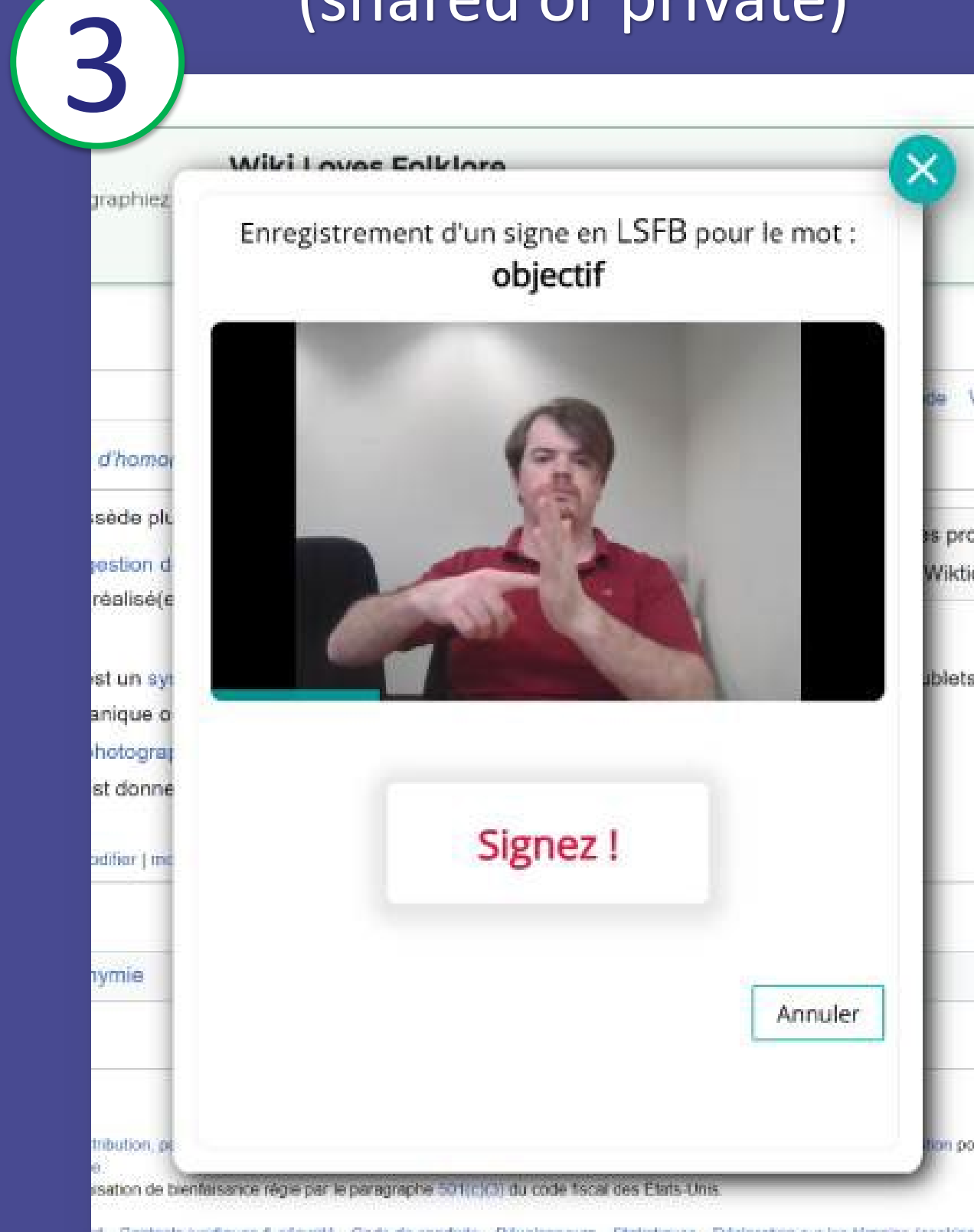
Our work

Community service comes first, and data collection is a **beneficial side effect**. Through three layered lexicons (public, shared, and personal) users freely **query** and **contribute** in-the-wild recordings, capturing specialized terms, neologisms, and regional variations through long-term, sustainable engagement.

1 Select a word (websites, PDFs, etc.)

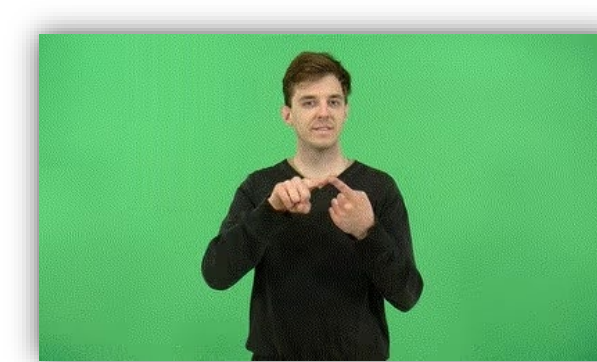


3 Propose your own signs and add them to your lexicons (shared or private)

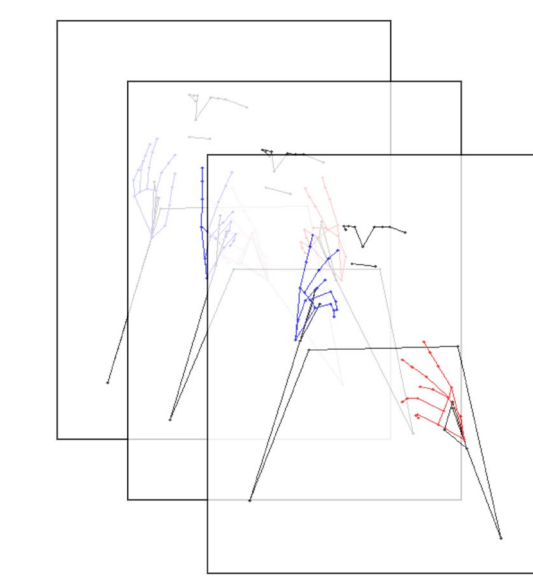


Data Collection

Re-recorded in public lexicons



Pose Extraction for Processing



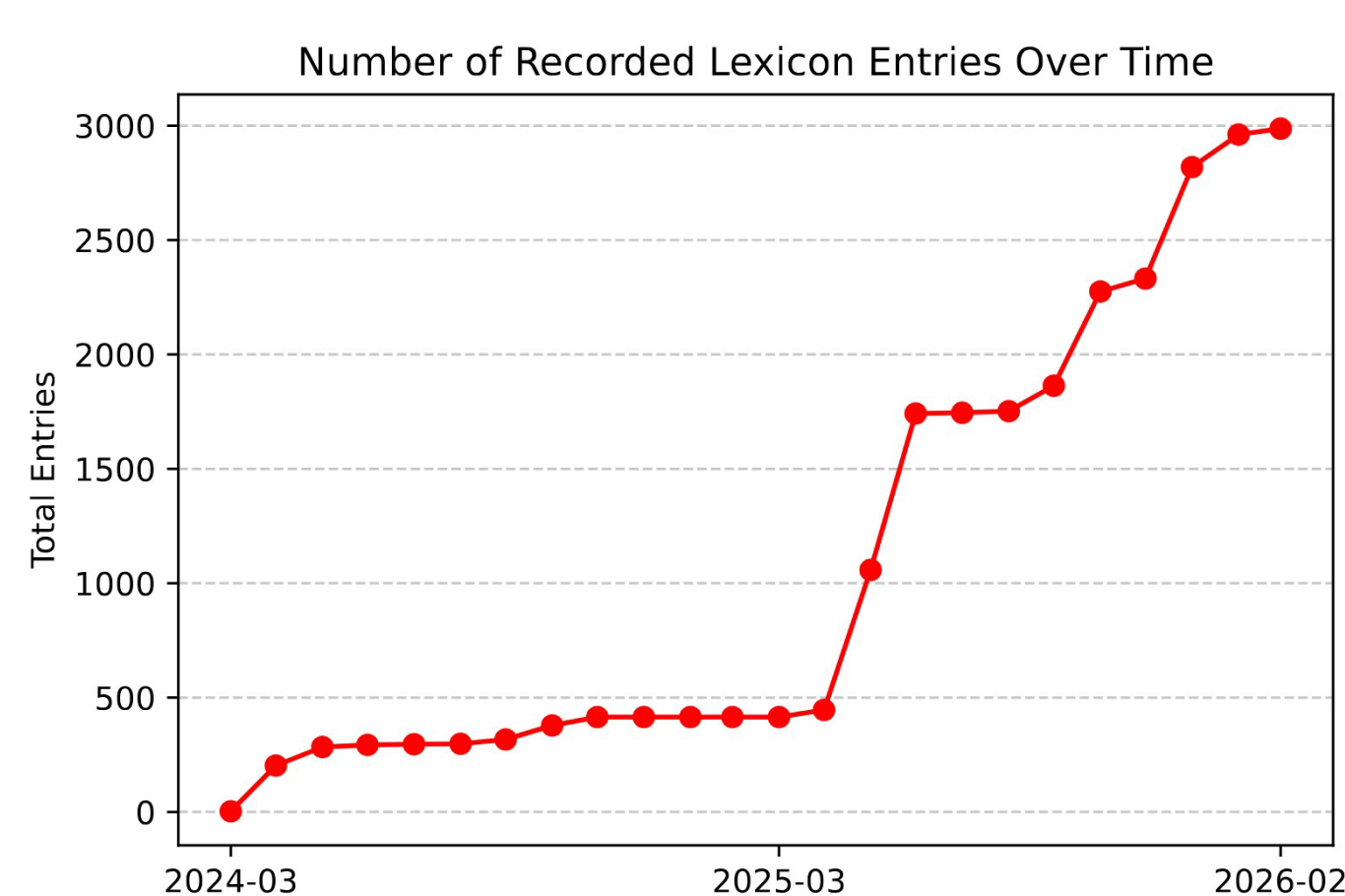
Informed consent and GDPR-compliant

User Study

LSFB-proficient users (9 people) were recruited and asked to fill out **diary entries over 10 days**, followed by an in-person debriefing. Vocabulary search dominated usage, confirming the feature initially requested by stakeholders. Users also valued the record sign and download GIF features for preparing teaching material.

Community Adoption

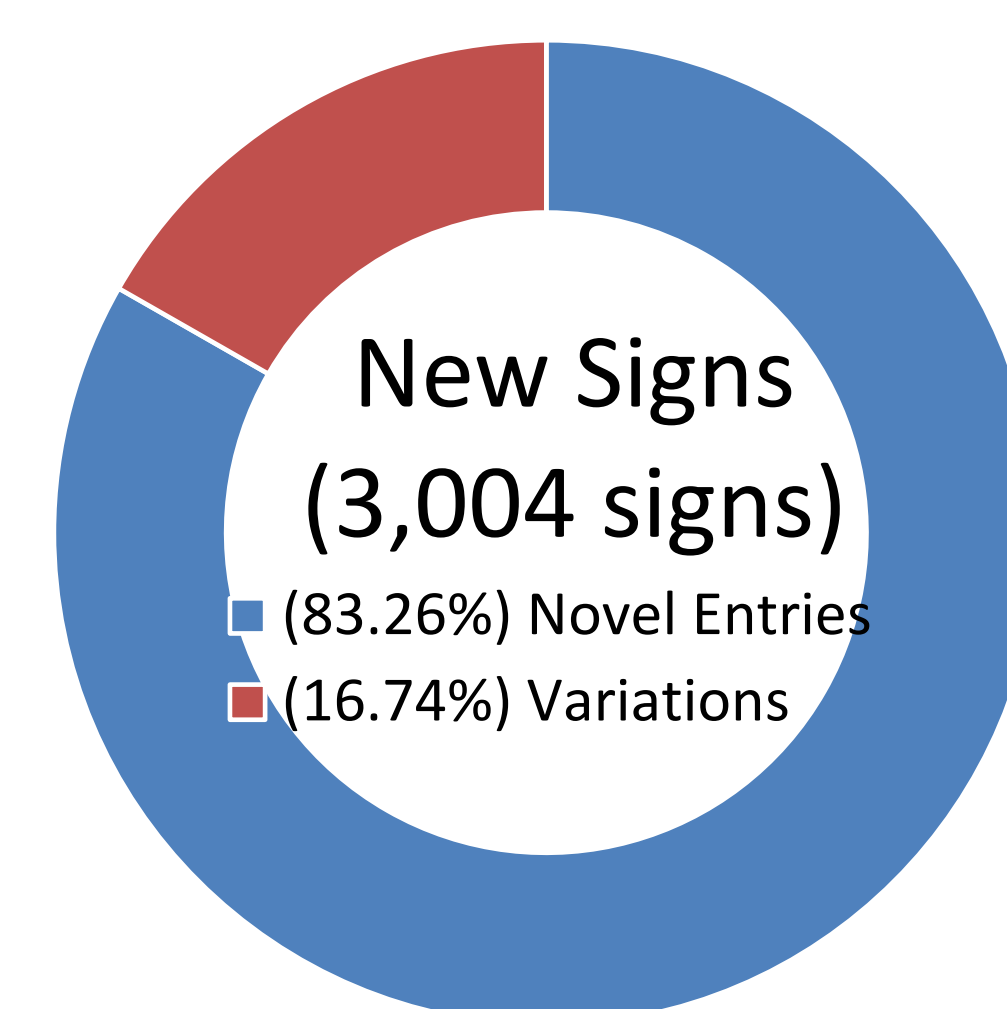
Following the user study, MOSI was **publicly released**. The platform has since shown a steady upward trend in contributions, indicating sustained engagement. Qualitative feedback reveals a strong sense of ownership among deaf users, who value the ability to **directly shape the documentation of their language**.



A Lot of New Signs...

- ✓ + **3,004** new signs !
83% of new vocabulary
17% of new variations
- ✓ + **224** registered users

And still ongoing !



Conclusion

This work demonstrates the viability of sustainable, long-term Sign Language (SL) data collection by **prioritizing the immediate needs** of the deaf community. Data collection becomes a **natural byproduct**. This collaborative approach proved **highly effective** and grounded in the principle of *nothing about us without us*.



Try MOSI now!
<https://mot-signe.be/>

More info and source code also available on the website...

