

## mylingua

Ausbau des aktuellen MVP zu einer skalierbaren digitalen Sprachlernlösung für fortgeschrittene Lernende

<b>Programm / Ausschreibung</b>	IWI 24/26, IWI 24/26, Basisprogramm Ausschreibung 2025	<b>Status</b>	laufend
<b>Projektstart</b>	01.04.2025	<b>Projektende</b>	31.03.2026
<b>Zeitraum</b>	2025 - 2026	<b>Projektlaufzeit</b>	12 Monate
<b>Keywords</b>			

### Projektbeschreibung

mylingua betreibt aktuell einen Web-basierten MVP einer digitalen Sprachlernlösung für die chinesische Sprache, der organisch bereits mehrere hundert regelmäßige Nutzer:innen gewinnen konnte. o Unsere Lösung schlägt Lernenden tagesaktuell originale Newsartikel vor, die auf ihr individuelles Interesse und ihr spezifisches Vokabular passen. Damit ermöglichen wir ihnen, durch die Anwendung der Comprehensible-Input Methode, Sprache in einem immersiven Ansatz zu lernen.

In dem ggst. Vorhaben sollen nun folgende innovative Weiterentwicklungsschritte gesetzt werden:

- Schärfung des Personalisierungsgrads zur Erhöhung der Passgenauigkeit der Vorschläge: Umsetzung einer Möglichkeit der multidimensionalen Erfassung der Schwierigkeit von Sprache bzw. Sprachkenntnissen (basierend nicht nur Vokabular, sondern auch auf Syntax, u.a. Faktoren)
- Punktuelle Simplifizierung von Texten, um weniger fortgeschrittenen Lernenden die von uns verfolgte Lernmethode zu erschließen: Entwicklung eines technologischen Ansatzes, der Original-Content auf Basis der individuellen Sprachkenntnisse des/der jeweiligen Nutzer:in simplifizieren kann (z.B. Austauschen von einem unbekanntem Wort durch ein bekanntes Synonym)
- Erschließung weiterer Content-Arten, um u.a. neue Demografien anzusprechen: Ausweitung unseres Ansatzes auf Textinhalte (über Newsartikel hinaus), Audio und Video
- Einfacher Rollout auf weitere Sprachen: Setup einer modularen Struktur, die es erlaubt mylingua vergleichsweise schnell mit unterschiedlichen Features - je nach Anforderungen der neuen Sprache - in einer weiteren Sprache anzubieten

### Endberichtkurzfassung

Im ersten Forschungsjahr wurden die dem Projekt zugrunde liegenden Ziele im Wesentlichen erreicht. Die Arbeiten umfassten fünf Arbeitspakete, die alle erfolgreich abgeschlossen (definierte Meilensteine erreicht) oder plangemäß gestartet wurden.

Im Bereich Infrastruktur (AP2) wurde eine flexible Backend-Testing-Infrastruktur aufgebaut, die strukturierte A/B-Tests mit bestehenden Nutzer:innen ermöglicht. Als zentraler Meilenstein wurde die native mobile App für iOS entwickelt und im Apple

App Store veröffentlicht. Die App basiert auf einer plattformübergreifenden Monorepo-Architektur mit gemeinsamer Codebasis für Web und App und schafft die technische Grundlage für erweiterte Personalisierung und Gamification.

Im Rahmen der Erweiterung auf zusätzliche Content-Arten (AP3) wurde auf Basis einer Auswertung von über 1.000 qualitativen Nutzerinteraktionen die Einbindung von Wikipedia-Artikeln, YouTube-Videos (Creative-Commons-Lizenzen) und KI-generiertem Audio priorisiert und technisch umgesetzt. Zum Berichtszeitpunkt stehen 6.396 Wikipedia-Inhalte und 63.462 Videos im System zur Verfügung. Die Recommender-Architektur wurde grundlegend erweitert und auf einen frequenzbasierten Ansatz umgestellt.

Im Arbeitspaket Optimierte Wissensstandsabfrage (AP4) wurde ein Vokabel-Quiz entwickelt und in das Produkt integriert. Das Quiz basiert auf einem Co-Occurrence- und Embedding-basierten Ansatz, der semantisch nahe Wörter identifiziert und eine gezieltere sowie effizientere Erfassung des Lernstands ermöglicht – sowohl im Onboarding als auch im laufenden Nutzungsbetrieb.

Im Arbeitspaket Schwierigkeitsrepräsentation (AP5) wurden die fachlichen und methodischen Grundlagen für eine multifaktorielle Modellierung von Textschwierigkeit erarbeitet. Relevante Schwierigkeitsdimensionen (u. a. lexikalische Merkmale, syntaktische Komplexität, Informationsdichte, kulturspezifische Begriffe) wurden identifiziert, operationalisiert und in einem Forschungsbericht dokumentiert, der in Kooperation mit der Universität Tübingen erstellt wurde. Der bestehende mylingua-Score wurde von einem vorkommensbasierten auf einen frequenzbasierten Ansatz umgestellt; die Vektordatenbank wurde von nmslib auf Qdrant migriert.

## **Projektpartner**

- mylingua FlexCo