Know your Data
KI-Anwendungen passgenau konfigurieren

Annette Gerstenberg[1], Hanno Müller[2], Friederike Schulz[1]

[1]Universität Potsdam, [2]Hasso-Plattner-Institut

06.11.2026

Digitale Hochschule Brandenburg, Jahrestagung 2026 (Zentrum der Brandenburgischen Hochschulen für Digitale Transformation)

Motivation: Ausgangslage

LangAge: autobiographische Interviews, Orléans

- 76 Personen, überwiegend 70+, Längsschnitt 2005, 2012, 2015, 2023
- Audioaufnahmen (meist ca. 45 min) mit manueller Transkription, über 80 Stunden
- für die Linguistik sind sprachliche Details relevant, nicht nur der Inhalt

Anspruch an die Transkription (LangAge-Transkriptionsguide)

- Mündlichkeit erkennbar: Fehlstarts, Wiederholungen, Abbrüche, Zögern
- einheitlich: Standardorthographie, keine Interpunktion, festes Inventar für Interjektionen
- angelehnt an soziolinguistisches Korpus aus Orléans (ESLO 1 und ESLO 2)

Motivation: Manuell oder automatisch?

Manuelle Transkription

- genau, richtliniengetreu
- sehr zeitaufwändig
- menschliche Fehler, z. B. Tippfehler
- Französisch als Zweitsprache: manche Äußerungen schwer verständlich

⇒ rein manuell nicht mehr zeitgemäß

Automatische Transkription:
rasante Entwicklung

vor 2022	v. a. kommerzielle Dienste, kaum offen und lokal nutzbar, schwer exportierbar
2022	Whisper (OpenAI): offen, mehrsprachig
2023	Whisper large-v3
2024	CrisperWhisper: wortgetreu, aber nur Englisch und Deutsch

Unser Weg: Kontrolle mit Whisper, Vorbild CrisperWhisper

⇒ **FrWhisper**

Universität Potsdam
HPI Hasso Plattner Institut

Von Whisper large-v3 zu FrWhisper

Problem 1: Unzureichende Verschriftung linguistischer Sprachaufnahmen:

- Wiederholungen und Fehlstarts fehlen.
- Gesprochene Sprache wird an Schriftsprache angepasst: Künstliche Satzgrenzen, Interpunktion, grammatisch korrekt.
- Ungebräuchliche Redewendungen fehlen.

Anpassung 1: FrWhisper – Fine-Tuning eines Modells mit Korpusdaten:

- Verbatim-Transkript für das gesprochene Französisch.
- Trainiert mit Daten der gesprochenen Sprache auch älterer Personen.
- Berücksichtigt linguistische Transkriptionsrichtlinien.
- Evaluation zeigt deutliche Verbesserung.

Ausgangsdaten: 134 Stunden Interviews

Korpus	Stunden	Segmente
LangAge	80,2	98 846
ESLO 1 und 2	53,7	198 436
Gesamt	133,9	297 282

- Segment: kurzer Abschnitt mit Transkript (Median 1,4 bis 3,2 s)
- Entfernt: Interviewer, Stille, Überlappungen, Geräusche, anonymisierte Stellen
- Übrig: 114 217 Segmente, davon 80 % Training, 20 % Test

Universität Potsdam
HPI Hasso Plattner Institut

Fine-Tuning auf dem HPI-Cluster

- Basis: Whisper large-v3, offenes Modell
- Training mit Standardsoftware auf 4 Grafikkarten
- Gestoppt, bevor das Modell auswendig lernt
- Lokal nutzbar, Lizenz CC BY-NC-SA



huggingface.co/aihpi/FrWhisper

Evaluation: weniger Fehler, eine neue Gewohnheit

Modell	Wortfehlerrate (Testdaten)
Whisper large-v3	91,3 %
FrWhisper	76,4 %

2 836 Mal transkribiert, nie vollständig so gesagt

et puis après on a été en vacances

(und dann sind wir in den Urlaub gefahren)

- *et puis* in den Trainingsdaten: 1,6 % der Segmente
- Von FrWhisper eingefügt: 25 % der Segmente

Raus aus den Loops: Datenvorbereitung und Tonschnitt

Problem 2: FrWhisper transkribiert zuverlässig, bleibt aber in Schleifen stecken.

- Bei langen Dateien: Mehr Möglichkeit zum abdriften
- auf Grund ökologischer Daten mit Störgeräuschen: z.B. Hintergrundrauschen
- durch Gesten (Klatschen), Lachen, Interjektionen der Personen (kleine Änderungen haben große Wirkung z.B. euh, ben, hein).

Anpassung 2: Datenvorbereitung mit Voice Activity Detection (VAD)

- Zeitalignierte Labels für gesprochene Sprache,
- Schnitt in einzelne Tondateien auf Basis der Labels, Reduktion der Schleifen.
- Weitere Anpassung: Scriptgesteuerte Formatierung für das Transkriptionssystem.

Nach dem Zusammenfügen: Was bleibt manuell?

Nachbearbeitung - Stand 09/2026

- Sprecherzuweisung
- Segmentierung nach syntaktischen Kriterien
- kleine Korrekturen, oft Eigennamen (werden ohnehin anonymisiert)

Schwierigkeiten

- jüngere Sprecher*innen und L2-Sprecher*innen fehleranfälliger, aber L2 mit wenig Redeanteil
- Overlapping Speech, in den meisten Interviews überschaubar

$\Rightarrow$ **insgesamt deutliche Zeitersparnis**

Know your data!? – Anwendungsszenarien

- Fragen?
- Welche Datenstrukturen stellen vor Probleme?
- Welche datenspezifischen Lösungsansätze sind denkbar?