

Nutzerstudie a!

Ziele und Hypothesen

Das Ziel dieser Studie ist die Evaluierung verschiedener Strategien für die Generierung und Selektion von C-Tests für Fremdsprachenlernende (Deutsch oder Englisch). Ein ausgewählter C-Test für einen Fremdsprachenlernenden wird hierbei als optimal erachtet, wenn die Fehlerrate des Lernenden für diesen Test in der *Zone of Proximal Development* liegt. Folgende Hypothesen werden getestet:

1. Die Leseschwierigkeit eines Basistextes korreliert mit der Übungsschwierigkeit des generierten C-Tests.
2. Die aggregierte Lückenschwierigkeit eines C-Tests ist ausschlaggebend für die Gesamtschwierigkeit und kann zur Ermittlung der Gesamtschwierigkeit, und somit zur Bestimmung des optimalen C-Tests genutzt werden.
3. Die Manipulation von C-Tests erlaubt einen größeren Handlungsspielraum für maschinelle Lernmodelle, was zu C-Tests führt, die konsistenter in der *Zone of Proximal Development* liegen.

Beschreibung

Alle Teilnehmenden werden in drei Gruppen aufgeteilt. Jede Gruppe besteht aus mindestens 20 Personen der Sprachniveaus A1 – B2. Pro Gruppe werden die Teilnehmenden in zwei Untergruppen *basic users* (Sprachniveaus A1 und A2) und *independent users* (Sprachniveaus B1 und B2) aufgeteilt, für die die in dieser Studie evaluierten Methoden unabhängig voneinander untersucht werden. Um eine Vergleichbarkeit zu gewährleisten, ist es wichtig, dass über alle drei Gruppen jede Untergruppe eine ähnliche Anzahl Teilnehmer hat.

Zur Generierung der C-Tests werden Texte der Leseschwierigkeit A1 – B2 verwendet. Hierbei wird nicht zwischen den Untergruppen unterschieden, d.h. eine teilnehmende Person des Sprachniveaus A1 kann durchaus einen C-Test erhalten, welcher aus einem Text der Leseschwierigkeit B2 generiert wurde. Um eine Vergleichbarkeit zu gewährleisten, wird für alle Gruppen darauf geachtet, dass pro Teilnehmenden jeder Text lediglich einmal als Basis zur C-Test-Generierung verwendet werden kann. Um ein Verhindern des Lernprozesses der Teilnehmenden zu vermeiden, werden Texte mit der Leseschwierigkeit C1 und C2 von dieser Studie ausgeschlossen.

Des Weiteren schließen wir Personen mit der Sprachniveaus C1 und C2 von dieser Studie aus, da zu erwarten ist, dass diese eine geringere Veränderung ihres Sprachniveaus innerhalb des kurzen Zeitraums von zwei Monaten, in welcher die Studie pro Gruppe durchgeführt wird, aufweisen. Da das Ziel dieser Studie ist, Methoden zur Auswahl von möglichst geeigneten C-Tests zu evaluieren, ist es möglich, mit Teilnehmenden, für welche eine größere Änderung ihres Sprachniveaus innerhalb des Zeitraums erwartet wird, aussagekräftigere Schlüsse über die Performanz der eingesetzten Modelle zu ziehen.

Ablauf

Teilnehmende Personen geben bei ihrer Registrierung auf der Webseite zu Beginn der Studie ihr Sprachniveau an. Zudem legen sie einen Namen und ein Passwort fest, welches sie für den Zugang zur Webanwendung benötigen. Zur Wahrung des Datenschutzes werden keine persönlichen Daten

erhoben. Dies hat zur Folge, dass das Passwort zu einem Namen nicht wiederhergestellt werden kann. Teilnehmende sind dazu angehalten, sich das Passwort gut zu merken oder separat zu speichern.

Die Webanwendung stellt C-Tests bereit, die aus Texten verschiedener Schwierigkeitsgrade generiert werden und misst die Fehlerrate pro Lücke und Test, sowie die Zeit, die ein Teilnehmender aufbringt, um einen Test zu absolvieren. Für eine aussagekräftige Evaluation erwarten wir insgesamt 60 bearbeitete C-Tests pro Teilnehmenden am Ende einer Phase. Es wird empfohlen, dass die teilnehmenden Personen die C-Tests in regelmäßigem Abstand bewältigen (beispielsweise mindestens alle zwei Tage und sieben Übungen pro Woche).

Zu Beginn und zum Ende der Studie füllen alle teilnehmenden Personen einen Fragebogen aus, welcher ihre allgemeinen Lernaktivitäten zum Ausbauen ihrer Fremdsprachenkenntnisse erfasst. Da die Teilnahme an dieser Studie freiwillig ist, und die Studie in keiner Weise das eigenständige Lernen der jeweiligen Fremdsprache kontrolliert oder unterbindet, ist es wichtig, das unterschiedliche Lernverhalten der einzelnen Lernenden über die Studie hinaus zu erfassen. So ist es beispielsweise zu erwarten, dass Personen, die während der Studie regelmäßig einen Deutschkurs besuchen, einen größeren Zuwachs an Sprachkompetenz aufweisen, im Gegensatz zu Personen, die keine anderweitigen Sprachenlernangebote wahrnehmen.

Phase 1: Kontrollgruppe

Die Kontrollgruppe erhält Übungen, die zufällig aus der vorhandenen Menge an C-Tests ausgewählt werden. Die Übungen für die Kontrollgruppe werden nach folgendem Schema erstellt:

1. Jedes zweite Wort wird zur Lücke umgewandelt durch das Entfernen der zweiten Hälfte des Wortes.
2. Der erste und der letzte Satz in einem Text enthalten keine Lücken, um einen gewissen Kontext zu gewährleisten.

Phase 2: Experimentgruppe mit ausgewählten C-Tests

Die zweite Personengruppe erhält ebenfalls Übungen, die nach obigem Schema erstellt werden. Im Gegensatz zur Kontrollgruppe werden die C-Tests hier nicht zufällig, sondern möglichst passend zum Lernendenprofil ausgewählt. So könnte beispielsweise eine Person mit der Selbsteinschätzung A1 C-Tests erhalten, welche aus Texten der gleichen Leseschwierigkeit erstellt wurden. Die Selektion der jeweils passenden Übung findet mit Hilfe eines maschinellen Lernmodells statt, welches die Lückenschwierigkeit pro Lücke im C-Test bestimmt und daraus eine Gesamtschwierigkeit für einen C-Test ableitet.

Phase 3: Experimentgruppe mit manipulierten C-Tests

Die letzte Personengruppe erhält Übungen, welche nicht nur selektiert, sondern auch manipuliert werden können. Durch das Hinzufügen oder Entfernen von Lücken kann ein C-Test schwieriger oder einfacher gestaltet werden. Das maschinelle Lernmodell für die letzte Gruppe ist dazu in der Lage, die Schwierigkeit der C-Tests durch ihre Manipulation dementsprechend anzupassen und hat somit mehr Auswahlmöglichkeiten für die Selektion eines geeigneten C-Tests.

Evaluation und erwartete Ergebnisse

Für die Evaluierung der verschiedenen Methoden werden die Fehlerraten über die Zeit und die Leseschwierigkeit der Basistexte pro Teilnehmenden betrachtet.

1. Eine starke Korrelation zwischen der Leseschwierigkeit eines Textes im Vergleich zum Sprachniveau eines Lernenden und des Abstandes der Fehlerrate zur *Zone of Proximal Development* für den daraus generierten C-Test spricht für Hypothese 1.
2. Die Fehlerraten der Selektiven C-Test Gruppe sollten im Vergleich zur Kontrollgruppe näher an der *Zone of Proximal Development* liegen, falls Hypothese 2 zutrifft.
3. Die Fehlerraten der Selektiven und Manipulierten C-Test Gruppe sollten im Vergleich zur Selektiven C-Test Gruppe näher an der *Zone of Proximal Development* liegen, falls Hypothese 3 zutrifft.

Zeitplan

Für die einzelnen Phasen (entspricht einer Gruppe von Teilnehmenden) der Studie ist folgender Zeitplan vorgesehen:

	Apr-18	May-18	Jun-18	Jul-18	Aug-18	Sep-18	Oct-18	Nov-18	Dec-18	
Phase 1	■									
Phase 2				■						
Phase 3							■			

04.2018 – 06.2018 : Phase 1

06.2018 – 07.2018 : Evaluation Phase 1 und Vorbereitung Phase 2

07.2018 – 09.2018 : Phase 2

09.2018 – 10.2018 : Evaluation Phase 2 und Vorbereitung Phase 3

10.2018 – 12.2018 : Phase 3

12.2018 : Evaluation Phase 3, Vergleich der verschiedenen Gruppen