

Hebah Ahmed, Hala Attoura & Claudia Duttlinger (Eberhard Karls Universität Tübingen)

C-Test digitalisiert: ein neues Format zur Messung von Arabischkenntnissen

Bestandsaufnahme eines Projektes und ein Ausblick

C-Test digitized: a new format for measuring proficiency in Arabic

Inventory of a project and an outlook

Abstract: Wie lässt sich die Sprachstandsfeststellung im Hochschulbereich durch Einstufungstests zuverlässig und kosten-nutzenoptimiert durchführen und wie lassen sich dabei gleichzeitig unterschiedliche Institutionen und deren unterschiedliche Bedürfnisse berücksichtigen? Wie kann Sprachlehre im Hochschulbereich nutzendenorientiert digital unterstützt werden? Welche digitalen Methoden ermöglichen es Lernenden, ein hohes Niveau an Fremdsprachenkenntnissen zu erreichen? Wie lassen sich Texte nach Schwierigkeitsgrad klassifizieren? Welcher Rahmen fördert einen effektiven Fortschritt beim Sprachenlernen? In diesem Artikel gehen wir diesen Fragen nach und skizzieren die Prozesse der Konzeption von C-Tests für Arabisch und der Generierung geeigneter Software für Online-Tests unter Berücksichtigung sprachspezifischer Herausforderungen in einer Zusammenarbeit zwischen der LEAD Graduate School & Research Network und der Abteilung für Orient- und Islamwissenschaft an der Universität Tübingen. Darüber hinaus bieten wir einen Ansatz zur Gestaltung digitaler Sprachlernübungen für Arabisch an, der individuelles Feedback und Scaffolding bietet, um einerseits den Präsenzunterricht und andererseits die Sprachbewertung innerhalb eines Moduls zu ergänzen.

Keywords: Digitalisierte C-Tests, arabische C-Tests nutzerorientiert, Linguistische Datenverarbeitung, Messen der Textschwierigkeit, Lehrplangestaltung

Abstract: How can higher education language assessment in the form of placement tests best be conducted in both a reliable and cost-benefit-optimised way while simultaneously taking into account different institutions and their varying needs? How can language teaching in higher education be supported digitally in a user-oriented way? Which kinds of digital methods can help learners achieve a high level of proficiency in a foreign language? How can texts be classified according to difficulty level? Which framework is conducive to effective progression in language learning? In this paper we address these questions, outlining the processes of conceptualising C-Tests for Arabic and generating suitable software for online testing while taking into account language-specific challenges, in a cooperation between LEAD Graduate School & Research Network and the Department of Oriental and Islamic Studies at the University of Tübingen. We also provide an approach for designing digital language learning exercises for Arabic which offer individualised feedback and scaffolding in order to supplement classroom teaching on the one hand and language assessment within a module on the other.

Keywords: Generating automatic C-Tests, user-oriented C-Tests for Arabic, Natural Language Processing, measuring text difficulty, syllabus design

© 2023, Hebah Ahmed, Hala Attoura & Claudia Duttlinger
Dieses Werk ist lizenziert unter der Creative Commons Lizenz CC BY 4.0 „Namensnennung-4.0. International“



Fremdsprachen und Hochschule 99/100 (2023)
veröffentlicht am 15.12.2023
<https://doi.org/10.46586/fuh.v100.2023.11300>
Ahmed, Hebah / Attoura, Hala / Duttlinger, Claudia (2023): C-Test digitalisiert: ein neues Format zur Messung von Arabischkenntnissen. Bestandsaufnahme eines Projektes und ein Ausblick. *Fremdsprachen und Hochschule* 99/100, 1-36.

1. Einleitung

Seit dem Europäischen Jahr der Sprachen 2001 und der Schaffung des Gemeinsamen europäischen Referenzrahmens für Sprachen (GeR) durch den Europarat wurde hinsichtlich der Sprachenpolitik an Hochschulen viel erreicht: Es wurden nationale und internationale Netzwerke geschaffen, z. B. *Network of University Language Testers in Europe (NULTE)*¹, mit UNICert² steht ein hochschulspezifisches Qualitätssiegel für Lehren und Testen von Sprachen zur Verfügung (dies bereits seit 1992), es wurden handlungsorientierte Lehr- und Testmethoden entwickelt und es wurde ein Begleitband zum GeR veröffentlicht³ und Vieles mehr.

Dieser Beitrag thematisiert, wie Digitalisierung diesen Standardisierungsprozess im Bereich Sprachtesten und -lehren weiter fördert und damit Lernende und Lehrende (in Sprachenzentren und Instituten) beim Erreichen gesteckter Ziele in verschiedener Hinsicht unterstützt und welche Voraussetzungen günstig und welche Stolpersteine dabei ggfs. zu beachten sind. Skizziert wird dies beispielhaft an der Testung und Vermittlung des Arabischen an der Universität Tübingen. Das Beispiel Arabisch wird nachfolgend unter ‚Lehren‘ und ‚Testen‘ beleuchtet, jedoch mit unterschiedlicher Gewichtung.

Bedarfe und Ziele der oben genannten Bereiche sind unseres Erachtens zwischen Hochschulen weitgehend vergleichbar – bei aller Unterschiedlichkeit zwischen einzelnen Einrichtungen. An der Universität Tübingen wird für Arabisch versucht, für gemeinsame Ziele und Bedarfe Synergien zu schaffen (dies – soweit möglich – auch sprachübergreifend). Wünschenswert wäre es sicher, hier auch Schulen ‚mitzudenken‘, um Brücken zu schlagen bzw. entsprechende Hürden bei Studienbeginn zu minimieren. Die wesentlichen Unterschiede zwischen Hochschuleinrichtungen lassen sich über vorgesehene Programmoptionen im C-Test-Editor der Universität Tübingen bedienen.

Zur Einstufung von Sprachkompetenzen stehen bekanntlich viele Formate zur Auswahl. Der C-Test – von Christine Klein-Braley und Ulrich Raatz in den frühen 80er Jahren auf der Basis des Cloze Test entwickelt⁴ – ist ein Format, das für Erst-, Zweit- und Fremdspracherwerb zum Einsatz kommt, zur globalen und unabhängigen Sprachkompetenzmessung nach dem Prinzip der reduzierten Redundanz. Es steht als objektives, reliables⁵, valides⁶ und auch

¹ https://www.unicert-online.org/wp-content/uploads/2019/07/NULTE-Flyer_2019_07.pdf sowie (für Mitglieder) <https://cercles.org/protected-content/>

² <https://www.unicert-online.org/>

³ 2018, Endfassung 2020, mit neuen Skalen, Beispieldeskriptoren und mit konzeptioneller Ergänzung

⁴ Vgl. Klein-Braley und Raatz (1982b).

⁵ Laut Grotjahn (2002: 214) ein hoch reliables Format, d. h. weitgehend ohne zufällige Messfehler: „C-Tests erweisen sich zumeist als hoch reliabel im Sinne einer Messkonsistenz auf der Ebene der einzelnen Texte (in der Regel lag Cronbachs Alpha zwischen 0.80 und 0.90).“

⁶ Die Augenscheinvalidität des C-Tests steht im Widerspruch zu Ergebnissen empirischer Daten ver-

ökonomisches⁷ Format inzwischen für über 20 Sprachen zur Verfügung und ist daher eines der meistverwendeten Testformate⁸ bzw. laut Grotjahn einer der am gründlichsten untersuchten neueren Sprachtests⁹. Neben seiner Funktion als Einstufungstest (bei Quereinstieg) bietet der C-Test die Möglichkeit der Lernfortschrittsmessung (innerhalb eines Moduls)¹⁰ und schließlich auch die der Kalibrierung. Laut Grotjahn (2002: 211) „ist [er] insbesondere dann erfolgreich einsetzbar, wenn das Ziel eine vom vorangehenden Unterricht bzw. von der individuellen Lerngeschichte weitgehend unabhängige globale Sprachstandsfeststellung ist. Entsprechend werden C-Tests u. a. als Einstufungstests an Universitäten, Studienkollegs und Sprachzentren eingesetzt.“

Während C-Tests mittlerweile vielerorts digital durchführbar sind, ist in vielen Einrichtungen (Sprachzentren bzw. Philologien) die Testerstellung samt Kontrolle der erzeugten Lücken noch von Hand zu bewerkstelligen. Computerlinguistische Programme ermöglichen die automatisierte Generierung von Lücken auf der Basis hinterlegter Tilgungsregeln und die automatisierte Auswertung von C-Tests und Übungen; sie erhöhen die Validität von C-Tests – dies mit entsprechend adaptierter Handreichung zum digitalen Format¹¹ für Studierende; sie steigern die Testökonomie; sie ermöglichen die Suche nach Texten auf bestimmten Niveaustufen und zu bestimmten Inhalten und liefern detaillierte Statistiken, die händisch schwerlich zu leisten sind; sie ermöglichen dar-

schiedener Studien. Vgl. Grotjahn (2002: 215): „Die Ansicht, dass der C-Test ebenso wie der Cloze-Test ein integratives Messinstrument zur Erfassung globaler Sprachkompetenz in Erst-, Zweit- und Fremdsprachen ist, stützt sich u. a. auf die zum Teil erstaunlich hohen Korrelationen von C-Tests mit verschiedenen Außenkorrelationen (z. B. Schulnoten, Lehrerurteilen über den Sprachstand der Schüler, Ergebnisse in anderen Sprachtests wie z. B. den TOEFL), sowie auf empirische Untersuchungen zur Konstruktvalidität unter Einschluss von introspektiven und experimentellen Verfahren. Zudem gibt es eine Reihe neuerer Belege, dass die Lerner bei der Rekonstruktion der C-Test-Lücken falls nötig auch den weiteren Kontext berücksichtigen und dass der C-Test damit nicht nur – wie zuweilen behauptet – auf der Mikroebene, sondern auch auf der Makroebene misst.“

⁷ Grotjahn (2006: x): „Obwohl nur noch selten ernsthaft bestritten wird, dass es sich beim C-Test um ein psychometrisch äußerst solides und zudem auch sehr ökonomisches Verfahren zur Messung von Sprachkompetenz handelt, gibt es sicherlich im Hinblick auf die Konstruktvalidität von C-Tests noch eine Reihe von zu klärenden Aspekten.“

⁸ U. a. bietet Grotjahn (2014) eine umfängliche Bibliographie zum C-Test-Format. In: Grotjahn/Dracker (2020).

⁹ Vgl. https://www2.testdaf.de/fileadmin/Redakteur/PDF/Forschung-Publikationen/Grotjahn_FaDaF_Essen.pdf sowie <http://www.c-test.de>

¹⁰ Dies ist z. B. für heterogene Lernergruppen ebenso nutzbar wie zur Messung der Kompetenz etwa des Fachwortschatzes bei Linkstilung (d. h. Tilgung der ersten Worthälfte, um zu testen, ob der Wortstamm ergänzt werden kann).

¹¹ Wer einen C-Test am PC ablegt, sollte im Vorfeld die notwendigen Informationen samt der Möglichkeit erhalten, sich mit den Spezifika und mit eventuellen zeitrelevanten Stolpersteinen hinlänglich vertraut zu machen (dies betrifft Sonderzeichen bzw. Tastatur für die Zielsprache usw.) – etwa über einen kurzen, freiwilligen oder obligatorischen Demo-Test. Abhängig vom Format (Papier vs. digital) ist daher unter Umständen auch zu entscheiden, ob die Testdauer angepasst werden soll.

über hinaus den Einsatz von Komplexitäts- und Lesbarkeitsanalysen. Sie helfen gleichfalls dabei, Formate einzusetzen, die für Einstufung oder Übungen umfangreiches Feedback geben, das individuell zugeschnitten ist. Automatisierte Tests und Übungen stellen eine Ergänzung zu herkömmlichen Formaten dar; sie bedeuten für Lernende und Lehrende daher mehr Flexibilität, Zeitersparnis, die Nutzung zusätzlicher Funktionen und eine ausführliche Statistik, die dazu beiträgt, individuelle Förderprofile zu erstellen.

Der vorliegende Beitrag ist keiner zur Testtheorie; vielmehr thematisiert er praktische Hilfestellungen für Hochschuleinrichtungen durch den Einsatz digitaler Funktionen – vor allem für C-Tests. Gewählt wurde hierfür das Beispiel Arabisch, für das bisher kaum Studien zu C-Tests¹² und kaum Programme zu deren automatisierter Erstellung und Auswertung zur Verfügung stehen. Der Artikel skizziert somit anhand einiger Beispiele, wie sich vor allem das Testen an Hochschuleinrichtungen erleichtern, mit verschiedenen Optionen versehen und um Funktionen erweitern lässt, die eine Niveaueinstufung genauer machen, dies auch mit Mitteln der Computerlinguistik (Kapitel 4). Natürlich ist auch die Frage – nicht nur für Arabisch – naheliegend und unter Umständen relevant, inwiefern das Format (Papier und Stift vs. digital) Testergebnisse beeinflusst und welche Aspekte ggfs. zu berücksichtigen sind. Da Einstufungstests (sogenannte *low stake tests*) keine formale Prüfung darstellen, aber beträchtliche Zeitersparnis bei Testerstellung und -auswertung bedeuten, sind digitale C-Tests trotz eventuell auftretender, aber laut Bisping und Raatz (2002: 131) nicht signifikanter Abweichungen eine zunehmend eingesetzte und wertgeschätzte Alternative zur Papierversion (vgl. Abb. 2 zu individuellen Korrelationen der vier Prädikatoren zum C-Testergebnis, ebd.: 143). Auch eigene Auswertungen zu verschiedenen Sprachen bestätigen geringfügige, aber nicht signifikante Unterschiede; ihnen lässt sich zumindest umfangreich mit entsprechender Vorbereitung der Studierenden auf das Testformat begegnen. Darüber hinaus untermauern die bei Bisping (2006: 162) dargestellten Ergebnisse die Robustheit automatisierter C-Tests¹³. Schließlich korrelieren laut Bisping (2006: 152) Computer-C-Tests mit Papier&Bleistift-C-Tests und mit Außenkriterien: Vokabel- und Grammatiktests und Diktat (DELTA-Test); die dort dargestellten Resultate zeigen keine relevanten Unterschiede. Laut Daller und Phelan (2006: 114) zeigt auch ein Vergleich von C-Tests und TOEIC® (mit Lese- und Hörverstehen) eine signifikante Korre-

¹² Vgl. Ahmed (2022), Al-bazi (n. d.), Daghbor (1998), Raish (2018).

¹³ Bisping (2006: 162): „Insgesamt sprechen die Ergebnisse in Bisping & Raatz (2002) und der hier beschriebenen zwei Experimente für die große Robustheit des C-Tests. Wenngleich es sich natürlich nicht um einen endgültigen Nachweis handelt, deuten folgende Aspekte darauf hin, dass computerisierte C-Tests nicht weniger valide sind als traditionelle: [...] d) Der Computer-C-Test wies eine hinreichende diskriminante Validität im Sinne von Campbell & Fiske (1959) auf.“

lation zwischen C-Test- und TOEIC®-Ergebnissen, dies mit dem Hinweis auf die größere Ökonomie des C-Test-Formats.

Unabhängig vom (Papier- oder digitalen) C-Test-Format verweisen Eckes und Grotjahn darauf, dass der C-Test nicht nur eindimensional ist, sondern auch als ein Messinstrument zur Erfassung der allgemeinen Sprachfähigkeit verstanden werden kann, und dass die Eignung des C-Tests als Instrument zur Verankerung von Scores im Leseverstehen und Hörverstehen außer Frage steht (2006: 192). Einige Studien verweisen auf nicht signifikante Abweichungen beim Vergleich von C-Test-Ergebnissen mit Schulnoten.

Ein weiteres Thema dieses Artikels sind die Niveaustufenbeschreibungen für Arabisch, die für die Lehre und zur Einstufung oder für Prüfungen relevant sind.¹⁴ Wie lassen sich die Kompetenzstufen des GeR bzgl. Lehren und Testen auf das Arabische (Modernes Hocharabisch sowie Dialekte) für möglichst alle Bildungseinrichtungen übertragen? Welche Antworten werden (ggfs. unterschiedlich je nach Einrichtung) aufgrund der Diglossie für einzelne Fertigkeiten gegeben? Und schließlich – welche Kategorien lassen sich für verschiedene Studienfächer und damit für verschiedene Varietäten des Arabischen oder Textsorten bilden (Kapitel 5)? Möglicherweise kann dieser Beitrag einen Impuls geben für eine intensiviertere Auseinandersetzung von Arabisch-Lehrenden mit diesen Fragestellungen.

Kapitel 6 skizziert einige Perspektiven zu Testen und auch zur Lehre: so etwa die Erweiterung des C-Test Editors um weitere Sprachen und Funktionen. Begonnen wurde auch mit dem Generieren von Programmen für Übungen zu verschiedenen Sprachen, Fertigkeiten und Niveaustufen als Ergänzung zum Unterricht, die umfangliche Funktionen enthalten und individuelles Feedback zu einzelnen Fertigkeiten bereithalten. Über den digitalen Ansatz werden etliche Schnittmengen zwischen C-Test und Übungen sichtbar; diese lassen sich nutzen für eine relativ unaufwändige Erweiterung des Feedbacks zum C-Test für Studierende (es würde neben der Information über das erzielte (Punkte-)Ergebnis Studierenden die Möglichkeit eröffnen, zum abgelegten C-Test auch individuelles Feedback zu erhalten etwa über die Anzeige der korrekten Antwort oder durch die Begründung, warum eine Antwort nicht korrekt bzw. welche korrekt ist. Wünschenswert wäre sicherlich auch ein C-Test für gesprochene Varietäten des Arabischen. Möglicherweise lässt sich der (zunächst für Russisch¹⁵ erarbei-

¹⁴ Hier sei, da mit integrierter Hochschulspezifität, auf UNICert® verwiesen: ein Qualitätssiegel für die Ausbildung, das Testen/Prüfen und die Zertifizierung von Fremdsprachenkompetenzen, die für angehende Akademiker*innen relevant sind; <https://www.unicert-online.org/dokumente/niveaustufenbeschreibungen/>

¹⁵ LEAD und die Slavistik der Universität Tübingen führten im Sommersemester 2022 ein Validierungsprojekt auch zu *heritage speakers* des Russischen an mehreren Hochschulen über eine Online-Plattform durch. Sobald alle Ergebnisse vorliegen, kann geprüft werden, inwieweit sich dieser Ansatz

tete) Ansatz zur Erfassung auch von Herkunftssprechenden über digitalisierte C-Tests auch für Arabisch konzipieren; eventuell lassen sich die Erkenntnisse auch in digitalisierte Übungen für diese Zielgruppe übertragen, um vor allem ihre orthographischen und bildungssprachlichen Fertigkeiten zu fördern.

2. Universitäre Rahmenbedingungen

An vielen deutschen Hochschulen dürften Lehrende vor der Herausforderung stehen, wie sich die eigenen Qualitätsansprüche hinsichtlich des Lehrens und Testens von Fremdsprachen im Arbeitsalltag immer zur Zufriedenheit umsetzen lassen. Dies kann u. a. mit Personal- (z. B. freiberuflich vs. angestellt) oder mit Finanzierungsfragen (z. B. Ausstattung, Anzahl der Mitarbeitenden, Fortbildungsangebot) zusammenhängen oder dem Zeitfaktor (z. B. Deputatsumfang oder Gruppengrößen der Kurse) oder schließlich mit der Frage, inwieweit Lehrende die Möglichkeit haben, für die eigene Sprache, übersprachlich und überinstitutionell (auch über die eigene Hochschule hinaus) in den Austausch zu treten, um über Entwicklungen informiert zu sein und eigene Ansätze mit denen anderer vergleichen zu können oder Vertrautheit mit gängigen Testformaten und -spezifika zu erlangen.

Ob Lehrende in die für sie relevanten Prozesse (sei es zu handlungsorientiertem Lehren oder Prüfen, zu validen Einstufungsformaten u. v. m.) eingebunden sind, auch mit der Möglichkeit, erforderliche Kompetenzen zu stärken, dürfte ähnlich relevant sein wie die Frage, ob Prozesse innerhalb einer Einrichtung abgestimmt und ob sie nutzerorientiert gestaltet sind. Sind Lehr- und Testformate abgestimmt? Sind entsprechende Ergebnisse vergleichbar, valide bzw. transparent? All dies hat unter Umständen Auswirkungen z. B. darauf, ob für Studierende das für sie passende Kursniveau ermittelt werden kann oder sie individuelle Fördermöglichkeiten nutzen können.

Im folgenden Kapitel werden anhand von C-Tests einige Möglichkeiten skizziert, Qualitätssicherung digital zu unterstützen und Funktionen zu nutzen, die bei der Frage der Textauswahl, Textschwierigkeit und der Testvalidität und Teststatistik (zeitoptimierend) Hilfestellung bieten. Ebenfalls berücksichtigt wird die Frage der Nutzerspezifika.

Dies geschieht vor dem Hintergrund, dass viele Sprachanbieter (Fachbereiche und Sprachenzentren bzw. Lehrende) im Arbeitsalltag einen Spagat machen müssen, um eine geeignete Einstufung anbieten zu können, sie eventuell selbst erstellen und validieren müssen. Hochschuleinrichtungen, die (wenn auch mit

auf Arabisch übertragen lässt bzw. wo und ggfs. welche Änderungen nötig sind. Auch Drackert und Timukova (2020: 159) haben zu C-Tests für Russisch bereits erste Ergebnisse beim Vergleich von Herkunftssprechenden und Lernenden von Russisch als Fremdsprache festgestellt: „The results show that almost half of the C-test gaps are biased towards either learner group.“

zum Teil unterschiedlichen Nutzergruppen) das gemeinsame Ziel verfolgen, (Fremd-)Sprachen zu lehren und testen, haben mit digitalen Programmen ein Hilfsmittel für effizientes Erstellen von Einstufungstests, zuverlässigere Ergebnisse und die Möglichkeit, sprachliche Fertigkeiten von Lernenden in Übungen individualisiert und automatisiert zu fördern. Dieser Ansatz unterstützt gleichzeitig die Vergleichbarkeit von Ergebnissen innerhalb einer Einrichtung und überinstitutionell.

Unabhängig vom Format des Einstufungstests sind drei Maßnahmen unseres Erachtens besonders dienlich, den Aussagegewert von Ergebnissen zu erhöhen:

- a. Kalibrierung: der Einsatz von Kalibrierungstestlets bei C-Tests (d. h. eines zusätzlichen Testlets aus einer Anzahl verschiedener Kalibrierungstestlets, das – manuell oder zufallsgeneriert positioniert – nicht ins Ergebnis einfließt), um über eine möglichst große Probandenzahl Werte zu erhalten und einen Teil dieser Kalibrierungstestlets dann für den Folgetestzeitraum als C-Test zu verwenden.
- b. Komplexitäts- und/oder Lesbarkeitsanalyse: Dies erfolgt digital an den noch nicht mit Lücken versehenen Texten und ist gleichermaßen für C-Tests und Übungen einsetzbar. Der Lesbarkeitsindex ermittelt, wie leicht oder schwer ein Text zu verstehen ist; die Komplexitätsanalyse benennt die im Text vorhandenen linguistischen Merkmale samt ihrer Häufigkeit und ermöglicht damit eine (ungefähre) Zuordnung zu einem Niveau. Erfasst werden z. B. die Zahl der Wörter, Länge der Wörter, durchschnittliche Wortlänge, Silbenzahl, Satzlänge, Zahl der Sätze, Type-Token-Relation, lexikalische Dichte usw.

Ziel ist es, die angezeigten Werte um einen Bezug zu den GeR-Stufen zu ergänzen.

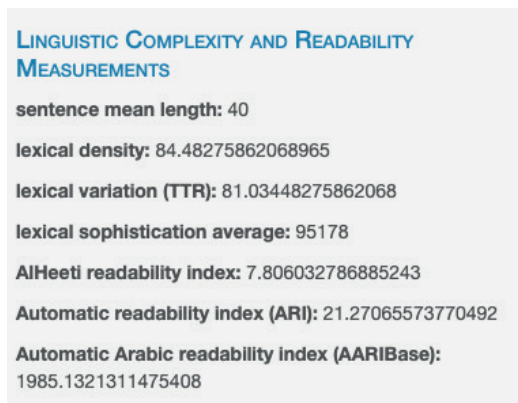


Abb. 1: Komplexitäts- und Lesbarkeitsanalyse: Beispiel Arabisch

Die Tübinger Plattform hat nach Einreichen dieses Beitrags für die verfügbaren Parameter eine Farbskala ergänzt (text length: ‚too easy → too hard‘); dies mit einer Definition für Nutzende (siehe Beispiel: ‚The sum of the complexity analysis parameters considered is illustrated here‘).

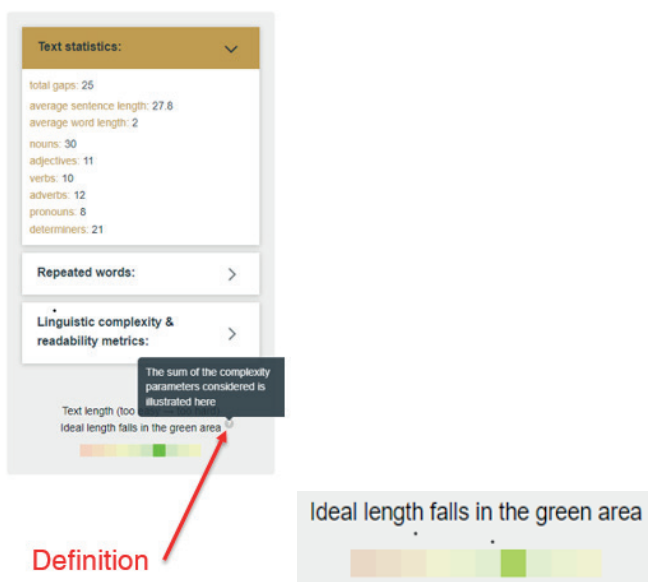


Abb. 2: Komplexitätsanalyse und Information für Nutzende

- c. Validierung: C-Tests überinstitutionell durch Lehrende und/oder Studierende ablegen zu lassen, um auf diese Weise Ergebnisse aus einer umfanglicheren Probandenzahl zu erhalten und gleichzeitig überinstitutionell über den Testansatz zu reflektieren.

Der Einsatz dieser Maßnahmen ist immer wünschenswert, jedoch besonders hilfreich, wenn die Infrastruktur wenig Spielraum für umfangliche Qualitätssicherung lässt, da zumindest a) Kalibrierung und b) Komplexitäts- und/oder Lesbarkeitsanalyse sehr aussagekräftig und gleichzeitig zeitlich nicht aufwändig sind.

Inzwischen wurde der Editor um eine weitere Funktion, die der automatisierten Fehleranalyse (bzw. Itemanalyse), ergänzt. Sie erlaubt Lehrenden einen schnellen Überblick auf der Basis einzelner Items, für einzelne Lernende oder für Gruppen.

Gaps 1-25 pro Testlet	Testlet 3		
	Lösung	Anzahl Fehler gesamt von 37 Studierenden	Fehlerliste
Gap 1 ce	e	5	1: [(‘a’, 4), (‘w’, 1)].
Gap 2 n’est	st	2	2: [(‘té’, 1), (‘es’, 1)].
Gap 3 une	ne	2	3: [(‘n’, 2)].
Gap 4 nouvelle	elle	0	4: [].
Gap 5 elles	les	5	5: [(‘-’, 2), (‘ise’, 1), (‘le’, 1), (‘lle’, 1)].
Gap 6 remplacées	acées	17	6: [(‘isses’, 1), (‘iqués’, 1), (‘assée’, 1), (‘oyées’, 2), (‘irait’, 1), (‘accès’, 1), (‘açées’, 2), (‘irées’, 1), (‘acée’, 1), (‘iées’, 1), (‘lacés’, 1), (‘isent’, 1), (‘ises’, 1), (‘ancés’, 1), (‘isser’, 1)].

Abb. 3: Fehleranalyse (*item analysis*): Beispiel Französisch

3. C-Test-Format – auch für Arabisch

Mit der Einführung von Arabisch am Fremdsprachenzentrum der Universität Tübingen (ursprünglich geplant mit einer Akkreditierung nach UNiCert® für vergleichbare Standards in Lehren und Testen) ging das Bemühen um ein möglichst zuverlässiges Einstufungsformat für Arabisch einher. Daraus entstand mit der *LEAD Graduate School & Research Network* die Schaffung eines sogenannten Runden Tisches Arabisch für alle interessierten Einrichtungen der Universität, die Arabisch lehren, zur Sichtung der Spezifika (wie Tests, Niveaustufen, Lehrwerke usw.), Ziele, Bedarfe und für eventuelle Synergien. Parallel zum Runden Tisch Arabisch mit seiner breiten Themenpalette entstand eine C-Test-AG Arabisch, in die sich Mitarbeitende verschiedener Universitäten einbringen. Ziel

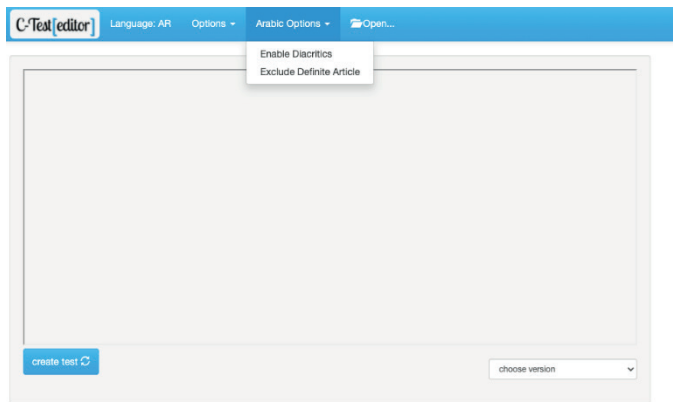
dieser AG ist es, konzeptionelle Fragen zu diesem Einstufungstest-Format zu diskutieren und technische Voraussetzungen für einen digitalen Ansatz für die Erstellung und Durchführung arabischer C-Tests zu schaffen. Ebenfalls durch *LEAD* in Zusammenarbeit mit der Tübinger (Theoretischen) Computerlinguistik konnte dieses Ziel mit studentischen Hilfskraftmitteln und über die Masterarbeit von Hebah Ahmed zu C-Tests im Arabischen (2022, unveröffentlicht) weitgehend umgesetzt werden. So wurde ein C-Test-Editor für Arabisch generiert (und mit dem bereits für einige Sprachen im Einsatz befindlichen Editor verbunden). Der C-Test-Editor ermöglicht die Erstellung und Durchführung von C-Tests (per C-Test-Editor oder per Ilias o. Ä.) auch für Arabisch: „a software that is able to generate Arabic C-Tests automatically with a consideration of the Arabic language features, such as orthography, syntax, semantics, and diacritical system.“ (Ahmed 2022: 9). Er generiert Lücken auf der Basis definierter (allgemeiner und sprachspezifischer) Tilgungsregeln; er erlaubt nutzerorientierte (inhaltliche und formale) Optionen (z. B. Anzahl der Texte, Anzahl oder Umfang der Lücken, mit/ohne Vokalisierung, mit/ohne gebundene(-n) Morpheme(-n), Rechts-/Linkstilgung, Punktwertung) und steht einem umfänglichen Nutzerkreis offen; er listet statistische Informationen zum gewählten Text (mit bzw. ohne Lücken). Über die Verlinkung mit Ilias erhalten Studierende ein automatisiertes Ergebnis samt Kursempfehlung. Lehrende können darüber hinaus die (ggfs. anonymisierten) Ergebnisse einsehen. Ebenfalls integriert sind eine automatisierte Statistikauswertung der Ergebnisse nach verschiedenen Kriterien und eine Komplexitäts- und Lesbarkeitsanalyse für die ausgewählten, noch nicht mit Lücken versehenen Texte (Testlets). Da parallel zum Arabischen vergleichbare Ansätze für weitere Sprachen bei *LEAD* erarbeitet wurden und werden, lassen sich immer wieder Synergien nutzen.

Die Sprachspezifik erforderte für Arabisch einige Besonderheiten technischer und konzeptioneller Art: Technisch waren z. B. die Umstellung auf die Schreibrichtung des Arabischen nötig und eine fehlerfreie Vokalisierungsfunktion (Kapitel 4 listet hierzu weitere Details). Konzeptionell sind z. B. allgemeine und sprachspezifische Tilgungsregeln ebenso wie *cut-offs* (d. h. die Zuordnung von Test-Ergebnissen zu den entsprechenden Sprachkursen bzw. Niveaustufen) zu definieren; auch die Komplexitäts- und Lesbarkeitsanalysen sind sprachspezifisch.

Da nur einige wenige Publikationen zu C-Tests für Arabisch verfügbar sind, galt es zunächst zu sichten, ob bzw. inwieweit auch Erfahrungen zu diesem Format aus anderen Sprachen für Arabisch anwendbar sind und welche sprachspezifischen Elemente es zu berücksichtigen gilt, um ein zuverlässiges C-Test-Format für Arabisch zur Erfassung globaler Arabischkompetenzen zu generieren.

Unser Ansatz ist ein nutzerorientierter, der auch für Arabisch etliche Optionen (z. B. für Diakritika) vorhält und erweitert werden kann. Nach dem kano-

nischen C-Test-Prinzip basiert unser Ansatz für Arabisch auf fünf kurzen Texten (sogenannten Testlets), die in aufsteigender Schwierigkeit angeordnet werden, mit je 25 Lücken (wobei die Software weitere Optionen für Text- und Lückenzahl zulässt). Der erste und der letzte Satz eines Testlets enthalten (ebenso wie die Überschrift) keine Tilgung, um den Kontext zu gewährleisten. Getilgt wird ab dem zweiten Satz bei jedem zweiten Wort. Ausnahmen von diesem Prinzip – wie Eigennamen, Abkürzungen oder Ziffern usw. – sind über die Tilgungsregeln digital gefasst. Der Tilgungsumfang kann wie bei europäischen Sprachen mit ‚die Hälfte eines Wortes‘ umschrieben werden bzw. mit ‚Hälfte + 1‘ bei Wörtern mit ungerader Buchstabenanzahl (auch hier gibt es Programm-Optionen). Wir haben uns auch bei Arabisch für die Tilgung des zweiten Wortteils (Rechtstilgung) entschieden – analog zu Raish („word-final mutilation“, 2018: 41); sie ist auch bei C-Tests für europäische Sprachen die gebräuchlichste Variante. Die Entscheidung, welche Worthälfte getilgt wird, hat auch im Arabischen Einfluss darauf, welche sprachlichen Elemente („Lexik“ vs. „Grammatik“) eher von der Tilgung betroffen sind; die Entscheidung wirkt gleichzeitig auf die Schwierigkeit des Tests. Mit dem Editor kann auch gewählt werden, ob diakritische Zeichen getilgt werden sollen, siehe Abb. 4.



© Sabrina Galasso, SFS Uni Tübingen

Abb. 4: Optionen: Beispiel zu Diakritika

Sprachspezifisch ggfs. unterschiedlich zu beantworten, hingegen für alle Sprachen gleichermaßen wichtig, ist die Frage der Rekonstruierbarkeit; unter Umständen werden in verschiedenen Sprachen Tilgungsregeln entsprechend modifiziert. Das Tilgungsprinzip wird in diesem Fall geringfügig geändert, um

bestimmte sprachliche – meist hochfrequente – Elemente von der Tilgung auszuschließen. So nehmen wir im Arabischen wie Raish (ebd.: 42-43) folgende Elemente von der Tilgung aus: a) den Artikel, b) präfigierte Präpositionen sowie c) (bei Raish bis auf das „*Wāw*“ nicht erwähnt) die Konjunktionen bzw. Anreihungspartikel. Ebenfalls lehnen wir uns an Raish (ebd.) an beim Absehen von der Tilgung der Suffixe wie Pronomina sowie der Femininendung (sog. *T ā' Marbūṭa*) und Pluralmarker¹⁶ und schließlich der Prä-, In- und Suffixe der Verbkonjugationen (vgl. ebd.: 42). Wir haben uns aus praktischen Erwägungen für die Generierung von C-Tests für Klassisches sowie Modernes Hocharabisch entschieden.¹⁷

Für ein möglichst valides Verfahren wurde das Tilgungsprinzip durch muttersprachliche Erwachsene geprüft, bevor es zum Einsatz kam (siehe Kapitel 4). Unser Ansatz zu C-Tests für Arabisch enthält darüber hinaus ein Kalibrierungstestlet: Jedem C-Test – bestehend aus fünf Testlets – wird ein sechstes, sogenanntes Kalibrierungstestlet, beigefügt, das nicht ins Ergebnis einfließt.¹⁸ Das jedem C-Test beigefügte Kalibrierungstestlet ist eines aus einer Reihe von Kalibrierungstestlets unterschiedlicher Niveaus. So liegen nach der C-Test-Phase über eine möglichst umfangreiche Studierendenzahl belastbare Zahlen (Statistiken zur Schwierigkeit und zu Fehlern) vor und aus den Kalibrierungstestlets lassen sich die für die folgende C-Test-Phase als geeignet befundenen Testlets auswählen. Ergänzt wird die Ermittlung der Schwierigkeit durch den Einsatz eines bereits erwähnten Programms zur Komplexitäts- und Lesbarkeitsanalyse von Texten des Arabischen (vor der Tilgung). Damit sind wesentliche qualitätssichernde Schritte getan; weitere Schritte zur Erhöhung der Validität werden anvisiert, siehe unten.

Eine schließlich ebenfalls für jede Einrichtung und überinstitutionell relevante Frage ist die der Diglossie des Arabischen und die der Auswahl der sprachlichen Varietäten bzw. ihrer Kombination – für Lehre und Testen. Der große Unterschied zwischen arabischer Schriftsprache und dem gesprochenen Arabisch des jeweiligen Landes trägt dazu bei, dass der Kontakt über schriftliche Lehrmaterialien Lernenden in erster Linie rezeptive Kompetenzen vermittelt bzw. sie entsprechend testet¹⁹. Die Entwicklung eines geeigneten Unterrichtscurriculums sowie geeigneter Einstufungs- und Prüfungskonzepte stellen beträchtliche Herausforderungen dar. Eine dieser Herausforderungen besteht

¹⁶ Hier gehen wir davon aus, dass Raish die beiden gesunden Plurale sowie den Dual meint.

¹⁷ Geplant ist jedoch, diesem Ansatz in hoffentlich absehbarer Zeit einen weiteren folgen zu lassen für (zumindest) eine gesprochene Varietät des Arabischen.

¹⁸ In Ilias ist keine zufallsgenerierte Positionierung möglich, die Position des Kalibrierungstestlets kann aber eingestellt und manuell verändert werden.

¹⁹ Dies beschreibt eine Situation, die für Lernende und Lehrende in der Regel ähnlich frustrierend ist.

darin, neben dem modernen Hocharabischen (im Folgenden MSA)²⁰ einen gesprochenen Dialekt einzubeziehen (gemeinsam oder getrennt von MSA?²¹ Ggfs. ab welcher Stufe? Ggfs. welche Dialektvariante lässt sich mit der jeweiligen Infrastruktur am ehesten lehren bzw. testen?). MSA (das in mündlicher Kommunikation vor allem in den Medien Verwendung findet bzw. als Verständigungsform für Sprechende verschiedener, gegenseitig schwer verständlicher Dialekte zum Einsatz kommt) allein zu unterrichten, verfehlt unseres Erachtens das Ziel des handlungsorientierten Ansatzes mit negativer Relevanz für die Kommunikationsfähigkeit von Lernenden.²² Will man die Deskriptoren des GeR oder von UNLcert[®] auch für Arabisch anwenden, wird klar, dass das Lehren einer Kombination von MSA und einer Dialektvariante alle sprachlichen Fertigkeiten auf allen Niveaus für alle sprachlichen Situationen vermittelt, wie in den (GeR- und UNLcert[®]-)Stufenbeschreibungen formuliert, da nur mit dieser Kombination produktive Fertigkeiten so vermittelt werden, dass umfängliches sprachliches Handeln in der Zielsprache möglich wird.

Die oben genannten Fragen werden (für Lehren und Testen) einrichtungsabhängig – vor dem Hintergrund verschiedener Zielgruppen und Studiengänge – unter Umständen unterschiedlich beantwortet je nachdem, ob Fertigkeiten gleichwertig vermittelt bzw. getestet werden und auf welchem Niveau welches sprachliche Merkmal bzw. Redemittel verortet wird. Jede Einrichtung definiert damit auch, welche Redemittel und grammatischen Strukturen für einzelne Niveaustufen als relevant erachtet werden, in der Regel mit Bezug auf den GeR oder die UNLcert[®]-Stufen – vor dem Hintergrund verschiedener Zielgruppen und Studiengänge. Gleichzeitig, und vor dem Hintergrund verschiedener Zielgruppen und Studiengänge, entscheidet jede Einrichtung auch darüber, welche Varietät(-en) des Arabischen gelehrt bzw. getestet werden soll(-en) und auf welcher Niveaustufe. Die Diglossie des Arabischen macht diese Aufgabe wie skizziert noch anspruchsvoller. Wünschenswert hierfür wäre (für C-Tests und Übungen) auch ein überinstitutioneller Blick.

Eine Antwortoption auf diese Frage bieten die in Kapitel 5 vorgestellten Kategorien, die sprachliche Strukturen und sprachliche Varietäten zielgruppenorientiert und textsortenspezifisch präsentieren. Weitere Orientierungspunkte bieten die sprachlichen Niveaustufen. Wünschenswert – für C-Tests und Übungen

²⁰ MSA wird in der überwiegenden Mehrheit der Arabisch-Curricula an universitären Sprachenzentren als erste und oft einzige Sprachvarietät gelehrt.

²¹ Diese Frage wird von Lehrenden unterschiedlicher Einrichtungen nicht notwendigerweise einheitlich beantwortet.

²² Unter bestimmten Bedingungen ist man auch mit MSA handlungsfähig, allerdings nicht in allen sprachlichen Situationen und gegenüber allen Gesprächspartner*innen. Indes erhöht das Einbeziehen von MSA und einer Dialektvariante den SWS-Umfang einer Stufe beträchtlich, eine Herausforderung bzw. ein Hindernis für etliche Lernende.

gen – ist in Bezug auf die Diglossie auch der oben genannte überinstitutionelle Blick; dieser kann erfolgen über regelmäßigen Austausch, über gemeinsame Validierungsverfahren, über den GeR oder über die von UNICert® angebotenen Handreichungen und Mindeststandards oder schließlich über darüberhinausgehende Schritte.

Wir haben uns aus diesem Grund für ein kleines Projekt entschieden, in dem versucht wird, zumindest auf der Basis einiger weniger universitärer Einrichtungen²³ a) inhaltliche Schnittmengen zu definieren, b) die in Kapitel 5 skizzierten Kategorien zu berücksichtigen und schließlich c) die sprachlichen Niveaus zu vergleichen, die in einzelnen Institutionen zugrunde gelegt werden. Die in a) erwähnten Schnittmengen umfassen grammatikalische Strukturen, lexikalische Merkmale, Redemittel und Redeanlässe usw. Die unter b) gelisteten Kategorien (Details siehe Kapitel 5) entsprechen nicht der üblichen niveauspezifischen Einteilung (z. B. des GeR), sondern dem inhaltlichen Wissensstand über die Grammatik (inkl. Syntax) und Morphologie des Arabischen (hinsichtlich der sprachlichen Fertigkeiten werden vor allem Textverständnis und Schreiben geprüft). In diesem Projekt soll für Arabisch diskutiert werden, ob bzw. inwieweit die unter c) erwähnten sprachlichen Niveaus in Lehre und Testen tatsächlich überinstitutionell (im kleinen Rahmen) vergleichbar sind. Ähnlich dem *Europäischen Referenzrahmen für Chinesisch als Fremdsprache / European Benchmarks for the Chinese Language (EBCL)* des Fachverbands Chinesisch e. V.²⁴, allerdings in sehr kleinem Rahmen, werden die im GeR gelisteten Deskriptoren für Arabisch diskutiert, um damit mittelfristig Vergleichbarkeit (bei Lehre und Testen) zu erleichtern; wie beim EBCL ist eine Beispielsammlung (zu Themen und *can do statements*) geplant und ebenfalls wie beim EBCL sind sogenannte Vokabellisten für die verschiedenen Niveaustufen geplant. Diese Listen werden dann im C-Test-Editor hinterlegt, der innerhalb einzelner Texte markiert, welche Wörter der hinterlegten Liste (d. h. einem bestimmten Niveau) entsprechen und welche nicht. Die Ergebnisse aus unserem Projekt gehen in die Konzeptgestaltung für C-Tests und für digitalisierte Übungen ein.

²³ Dabei handelt es sich zurzeit um Sprachenzentren, da bei ihnen der kommunikative Ansatz auch bei Arabisch von zentraler Bedeutung ist und Klassisches Arabisch und seine sprachlichen Merkmale nur geringumfänglich vermittelt werden. Zu nennen sind hier Gamiela Safiya (Zfs der Universität Würzburg) – sie ist Mitglied der oben erwähnten C-Test-AG für Arabisch – sowie weitere Mitglieder der Initiative der Arabischlehrenden an Sprachenzentren der Hochschulen (IASH).

²⁴ <https://refubium.fu-berlin.de/bitstream/handle/fub188/15346/EBCLxinklxxAppendizesx23Juni2015.pdf?sequence=1&isAllowed=y>

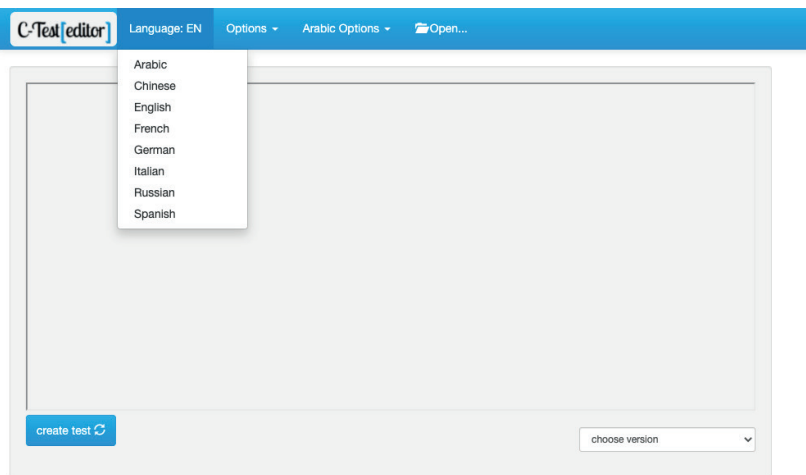
4. Digitalisierte C-Tests für Arabisch und die Messung ihrer Schwierigkeit – ein computerlinguistischer Ansatz

Dieses Kapitel erklärt – aus computerlinguistischer Perspektive – die wichtigsten Punkte des als C-Test-Editor genannten Programms, das ursprünglich von Magdalena Wolska und Sabrina Galasso bei der Computerlinguistik der Universität Tübingen entwickelt wurde, um automatisierte C-Tests für Fremdsprachen (Englisch, Französisch, Italienisch und Spanisch) zu erstellen. Dieses Kapitel gibt eine ausführliche Beschreibung der Erweiterung des C-Test-Editors um Arabisch und die hierfür relevanten sprachspezifischen Merkmale. Darüber hinaus werden sowohl die *Natural language processing* (NLP)-Werkzeuge diskutiert, die für die Verarbeitung des Arabischen eingesetzt wurden, als auch die Herausforderungen, die bei der Verarbeitung der arabischen Sprache auftraten.

4.1. Einführung in den C-Test-Editor

Der C-Test-Editor ist ein webbasiertes Programm, das von Sabrina Galasso für das Fremdsprachenzentrum der Universität Tübingen entwickelt wurde, wo die vier o. a. Sprachen neben weiteren als Fremdsprache unterrichtet werden. Dieses digitale Werkzeug sollte für die Einrichtung (Leitung und Lektorate) die Generierung der C-Tests erleichtern. Seit 2020 wurde der C-Test-Editor durch *LEAD* um einige Funktionen ergänzt und auch um verschiedene Sprachen erweitert, so auch um Arabisch. Meist erfolgten diese Erweiterungen im Rahmen von Masterarbeiten des Seminars für Sprachwissenschaft (Arbeitsbereich Computerlinguistik) der Universität Tübingen.

Der C-Test-Editor wurde mit der Programmiersprache Java entwickelt und verfügt über eine benutzerfreundliche Oberfläche, in der Nutzerinnen und Nutzer einfach den ausgewählten Text zur Erstellung des C-Tests einfügen; der Test wird dann über die Schaltfläche „create test“ automatisch erstellt.



© Sabrina Galasso, SFS Uni Tübingen

Abb. 5: C-Test-Editor Interface

Für die Aufbereitung des Rohtextes wird OpenNLP²⁵ verwendet, ein NLP-Toolkit, das mehrere Sprachen unterstützt und verarbeitet. Sabrina Galasso beschrieb die Schritte der Textverarbeitung im C-Test-Editor wie folgt: Als Erstes wird die Anzahl der Absätze ermittelt. Zweitens werden Sätze anhand von Satzzeichen (Punkt, Ausrufezeichen und Fragezeichen) erkannt und segmentiert. Drittens werden die Wörter in Tokens zerlegt (Tokenisierung). Viertens werden die Tokens mit einem *Part-of-Speech* (POS)-Tag versehen. Fünftens wird eine Liste von Tokens erstellt, die das Wort, sein POS-Tag und seine Position im Satz enthält. Sechstens wird die Liste der Tokens auf der Grundlage von „FSZ-Regeln“ nachbearbeitet. So werden zum Beispiel Wörter, die Apostrophe enthalten, zu einem Wort zusammengefügt (z. B. *didn't*). Schließlich werden Lücken auf der Grundlage spezifischer Token-Eigenschaften eingefügt, nämlich: (1) ob das Token Teil des ersten Satzes ist, (2) ob das Token mit einem Sternchen beginnt, (3) ob das Token ein Satzzeichen ist, (4) ob die Länge des Tokens mehr als zwei Buchstaben beträgt, (5) ob das Token eine Zahl ist, (6) ob die maximale Anzahl von Lücken erreicht ist und (7) ob das Token jedes zweite Wort ist. Dann wurde ein statistischer Teil in die Software integriert, in dem die Gesamtzahl der Wörter mit Tilgung, der funktionalen Wörter mit Tilgung, der lexikalischen Wörter mit Tilgung

²⁵ <https://opennlp.apache.org/>

und der Eigennamen mit Tilgung gezählt und rot hervorgehoben wird.

<p>German ninth-graders exceeded expectations in their English tests last year, according to a report on Friday by the Standing Conference of the Ministers of Education and Cultural Affairs. Four out of ten ninth graders were already a year ahead in achieving certain English language standards, which are expected of tenth graders. Some 37,000 boys and girls from 1,700 schools took part in the competence study. While their English competence "improved considerably" compared to previous years, according to the report, ninth graders' German skills stagnated. Bavaria was at the head of the pack in a regional comparison of English skills, while Schleswig-Holstein and Saxony showed great improvement over the last analysis of 2008-2009, and Bremen and Berlin remained at the tail end.</p>
<p>German ni____-graders exce____ expectations i____ their Eng____ tests la____ year, acco____ to a rep____ on Fri____ by t____ Standing Confe____ of t____ Ministers o____ Education a____ Cultural Aff____. Four o____ of t____ ninth gra____ were alr____ a ye____ ahead i____ achieving cer____ English lang____ standards, wh____ are expe____ of te____ graders. Some 37,000 boys and girls from 1,700 schools took part in the competence study. While their English competence "improved considerably" compared to previous years, according to the report, ninth graders' German skills stagnated. Bavaria was at the head of the pack in a regional comparison of English skills, while Schleswig-Holstein and Saxony showed great improvement over the last analysis of 2008-2009, and Bremen and Berlin remained at the tail end.</p>

Abb. 6: Englisch Textbeispiel ohne und mit Tilgung

4.2. NLP-Erweiterung des C-Test-Editors um Arabisch (und NLP-Herausforderungen)

Da der C-Test-Editor ursprünglich für europäische Sprachen entwickelt wurde, waren mehrere Änderungen erforderlich, um den Editor an die Besonderheiten des Arabischen anzupassen. Das erste Problem, das es zu lösen galt, war die Ausrichtung des Textes (Schreibrichtung), das zweite das der Interpunktionszeichen. Arabische Texte werden durch den Editor nicht korrekt formatiert, wenn sie in lateinisch orientierte Editoren eingefügt werden, da diese die arabischen Satzzeichen nicht erkennen, sondern sie als englische Zeichen verarbeiten; dies führt zu einer modifizierten Reihenfolge bei den arabischen Texten. Indem die Nutzerinnen und Nutzer die Sprache Arabisch auswählen und einen arabischen Text einfügen, ist dieses Problem im Editor jedoch behoben.

Die Editor-Erweiterung um Arabisch beinhaltet auch das Hinzufügen von Statistiken, die in das Programm integriert sind, um die mit Lücken versehenen Tokens bei Verben, Substantiven, Adjektiven und Adverbien zu erfassen. Darüber hinaus gibt es einen Abschnitt zur Hervorhebung aller Relativpronomina im Text, da solche Pronomen einen Hinweis auf die syntaktische Komplexität geben. Schließlich wird der Schwierigkeitsgrad des gewählten Textes durch Berechnung sprachlicher Komplexität und Lesbarkeit gemessen und das Ergebnis im Editor angezeigt.

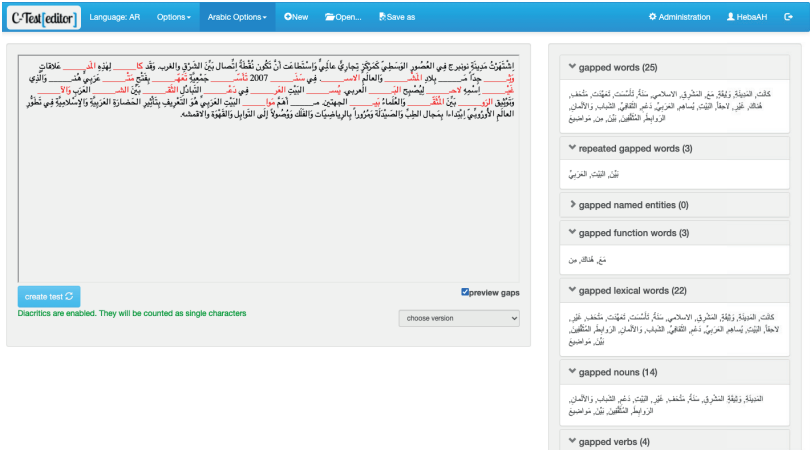


Abb. 7: Editor-Erweiterung um Arabisch mit Statistik. Der Editor vokalisiert derzeit an einigen Stellen noch fehlerhaft; er vertauscht schriftliche Phänomene, die in der punktierten Form gleich aussehen, z. B. den Buchstaben „Hā“ (= H: هـ/ة) mit der stets finalen „Tā' Marbūṭa“ (= T, sog. Verbundenes Tā': ة/ة) sowie den Buchstaben „Yā“ (= J: ي) mit der stets finalen „Alif Maqṣūra“ (ى).

Die Python NLP-Toolkits CAMEL Tools²⁶ und Stanza²⁷ wurden hauptsächlich für die arabische Vorverarbeitung verwendet. Die Implementierung ist den Verfahren zur Vorverarbeitung der oben erwähnten europäischen Sprachen recht ähnlich. Zunächst werden die Sätze segmentiert, dann die Wörter tokenisiert. Danach werden die Tokens mit Kategorien (*tags*) versehen und man erhält POS-Labels. Als nächstes wird eine Liste von Tokens (Token ist eine Java-Klasse, die von Sabrina Galasso erstellt und durch Hebah Ahmed per LEAD an die arabischen Sprachmerkmale angepasst wurde) erstellt, die die Satz-ID, die Position des Tokens, das POS-Label des Tokens, den Stamm des Tokens und die Angabe enthält, ob das Token funktional oder lexikalisch ist.

Schließlich wurden zwei zusätzliche Optionen speziell für das Arabische aufgenommen für Diakritika und für den Artikel. Nutzerinnen und Nutzer können auf diese Weise die Vokalisierungszeichen im Text hinzufügen oder entfernen. Diese Funktion wird durch den Einsatz des morphologischen Analysators und des MLE-Disambiguators der CAMEL Tools realisiert. Ebenso können Nutzerinnen und Nutzer den Artikel in den Lückenbildungsprozess einbinden oder ihn davon ausschließen.

Die Eigenschaften des Arabischen stellen mehrere Herausforderungen für

²⁶ <https://camel-tools.readthedocs.io/en/latest/index.html>

²⁷ <https://stanfordnlp.github.io/stanza/>

die Sprachverarbeitung (NLP) dar. Angefangen bei der Schreibweise, in der man die Buchstaben miteinander verbindet: hier kann das Tilgen von Wortteilen eine Veränderung der Schriftzeichen zur Folge haben, bei der Lücken entstehen, welche die Form der Zeichen in einem Wort verzerren. Bei der Generierung von Lücken erzeugt der C-Test-Editor eckige Klammern um die zweite Worthälfte, die bei der Implementierung getilgt werden soll. Technisch gesehen wird das Wort so aufgeteilt, dass es – am Beispiel Englisch – wie folgt aussieht: *da[ta]*. Wenn Benutzerinnen und Benutzer die Option „preview gaps“ anklicken, wird das geteilte Wort als *da__* dargestellt. Im Arabischen kann die Aufteilung von Wörtern mit eckigen Klammern bei einigen Schriftformen irreführend sein, da sich nach dem Einfügen der Klammern die zusammenhängende Form des Buchstabens in seine isolierte Form ändert. Zum Beispiel wird der Buchstabe (ل = L) im Wort (حليب / *Halīb* = Milch) in der mittleren Stellung beidseitig verbunden (لـ) geschrieben; aber sobald das Wort geteilt wird, wird es als حل[يب] (als ob das ل in der Endstellung stünde) dargestellt. Wenn die Schaltfläche „preview gaps“ angekreuzt ist, sieht es wie folgt aus: --حل. Aufgrund dieses Problems könnten die Testkandidatinnen und -kandidaten der Meinung sein, dass das Zeichen nicht verknüpft werden soll, und nach anderen Lösungen suchen, die zur bestehenden Form passen. Um dieses Problem zu beheben, wurde die Kashida (übersetzt etwa „Verlängerungsstrich“ كـ, der sich oberhalb des Buchstaben „ت = T“ in der arabischen Tastatur befindet) nach jedem zerlegten Zeichen und vor der eckigen Klammer angefügt. Das Wort lässt sich also als حل[يب] bilden und in der Vorschau wird es als حلـ angezeigt.

Eine weitere Herausforderung war die Verwendung des Artikels, da dieser stets einem Wort vorangestellt ist. Das Design des C-Tests basiert auf bestimmten Prinzipien. Eines der Kriterien lautet, dass die Hälfte des Wortes ausgelassen wird; bei Wörtern mit ungerader Buchstabenanzahl indes wird die Hälfte des Wortes plus/minus ein Buchstabe je nach Einrichtung getilgt, wie das folgende Beispiel zeigt: Das Wort (تمر - Dattel ‚Frucht‘) besteht aus drei Buchstaben, regelkonform wird es als --تـ gegliedert, wobei nur der erste Buchstabe erhalten bleibt. Ist dem Wort indes der Artikel (ال) angefügt, wird es als التمر gebildet und besteht aus fünf Buchstaben. Ist dieses Wort also mit einer Lücke versehen, wird es als الـ--- dargestellt, wobei nur der Artikel angezeigt wird und die Testteilnehmerinnen und -teilnehmer somit keine lexikalischen Informationen erhalten. Aufgrund dieser Affigierung wird das sonst übliche Tilgungskonzept für Arabisch geringfügig geändert und der Artikel bei der Zählung der Buchstaben ignoriert, mit folgender Worttilgung: الـ---. Raish (2018: 42-43) ging ebenfalls auf die Anfügung anderer Präfixe, etwa Präpositionen, ein. Daher wurden weitere Möglichkeiten in Betracht gezogen, wenn der Artikel ausgeschlossen ist, wie (الـ - und der), (وبالـ - und mit dem) und (قالـ - dann der).

Zur Bewertung der Schwierigkeit von C-Tests werden bei unserem Ansatz auch linguistische Komplexitäts- und Lesbarkeitsmessungen durchgeführt. Das Niveau eines Testlets wird anhand seiner lexikalischen, grammatischen und syntaktischen Merkmale bewertet. Zunächst wurde eine einfache Berechnung durchgeführt, um *sentence length mean* zu ermitteln, der die durchschnittliche Anzahl der Wörter pro Satz angibt. *Sentence length mean* zeigt die Schwierigkeit des Satzes in Bezug auf seine syntaktische Struktur an. Hier sei angemerkt, dass der Durchschnitt der Satzlänge sich in Formeln zur Lesbarkeit verwenden lässt. Ein weiterer Indikator, der hinzugefügt wurde, um die syntaktische Komplexität von Testlets aufzuzeigen, ist die Anzahl der Relativpronomina. Danach wird die lexikalische Komplexität von Testlets durch die Berechnung von lexikalischer Dichte, Variation und gehobenem Sprachstil ermittelt. Da die lexikalische Komplexität lexikalische Elemente (Substantive, Verben, Adjektive und Adverbien) in Bezug auf die Gesamtzahl der Wörter in einem Text zählt, wurde zur Ermittlung solcher Tokens der CAMEL POS-Tagger eingesetzt. Wörter, die mit den folgenden Tags versehen sind, wurden ausgewählt und in einer Liste gesammelt: *noun*, *noun_prop*, *noun_quant*, *verb*, *adj*, *adj_comp*, *adj_num* und *adv*. Um genaue Ergebnisse zu erhalten, wurden Interpunktionszeichen aus dieser Token-Liste entfernt. Anschließend wurde die lexikalische Variation anhand der Type-Token-Relation (TTR) berechnet, die die Anzahl eindeutiger Wörter zählt.

Das letzte Maß für die lexikalische Komplexität schließlich ist die *sophistication complexity*, für die eine Häufigkeitsliste erforderlich ist. Eine Häufigkeitsliste für den arabischen Wortschatz zu finden, ist keine einfache Aufgabe, da nur wenige erprobte Listen verfügbar sind. Allerdings haben Al Khalil, Habash und Jiang (2020: 3053) im Rahmen des SAMER-Projekts (*Simplification of Arabic Masterpieces for Extensive Reading*) eine Liste von 26.578 MSA-Lexemen zusammengestellt, die von Sprachkundigen (z. B. Muttersprachler*innen) manuell annotiert wurden (Al Khalil/Habash/Jiang 2020: 3053). Diese Wortsammlung wurde erstellt, um die Forschung im Bereich der arabischen Lesbarkeit zu unterstützen (ebd.). Daher wurde das Vokabular der C-Test-Testlets mit der SAMER-Liste verglichen, um die gemeinsamen Wörter zu zählen. Vokalisierungszeichen mussten dem Text hinzugefügt werden, da das Vokabular der SAMER-Liste mit Vokalisierungszeichen disambiguiert ist. Danach wurde der Mittelwert der schwierigen Wörter berechnet.

Die Lesbarkeit von Testlets wurde schließlich anhand von drei Lesbarkeitsformeln gemessen: Al-Heeti (Al-Heeti 1984), *Automatic Readability Index* (ARI) (Senter/Smith 1967), und *Automatic Arabic Readability Index* (AARIBase) (Al Tamimi/Jaradat/Al-Jarrah/Ghanem 2014). Für den Al-Heeti-Index wurde die durchschnittliche Anzahl der Zeichen pro Wort berechnet, während für den ARI die durchschnittliche Wortlänge und die durchschnittliche Satzlänge gemessen

wurden. Der AARI benötigt schließlich – zusätzlich zu den oben genannten Maßen – die Anzahl der Zeichen in einem Text.

4.3. Diskussion

Grund für die Entwicklung eines C-Test-Editors auch für Arabisch war in erster Linie die Notwendigkeit eines validen Einstufungsverfahrens, das gleichzeitig nutzerorientiert und automatisiert (somit ressourcensparend) ist. So entstand auf der Basis des bestehenden Editors, der für vier Sprachen verfügbar war, zum einen die Idee zu einer Masterarbeit zur Programmierung der Editor-Erweiterung um Arabisch, zum anderen eine C-Test-AG aus Mitgliedern verschiedener Hochschulen, die sich mit konzeptionellen allgemeinen und sprachspezifischen Fragen auseinandersetzt – beides konzentriert durch *LEAD*. Einige Mitglieder der C-Test-AG und weitere Lehrende verschiedener Hochschulen waren parallel dazu recht intensiv eingebunden in Fragestellungen, die die einzelnen Entwicklungsschritte des Editors für Arabisch betrafen, und eingebunden in eine erste Validierung ausgewählter Texte, für welche die Tilgung durch den Editor generiert wurde. Mit diesen Ansätzen floss ein kontinuierliches Feedback (z. B. Änderungs- und/oder Ergänzungsvorschläge) in die Entwicklung des Programms für Arabisch ein. Nach Abschluss der Programmierung der Erweiterung um Arabisch erhielten die am Projekt beteiligten Arabisch-Lehrenden eine Einführung in die Benutzung des C-Test-Editors und Zugang zum Editor. Damit haben sie die Möglichkeit, auf das Programm zuzugreifen, es zu nutzen und zu testen und ggfs. weitere Änderungs- bzw. Ergänzungsvorschläge zu äußern.

Alle Beteiligten (Arabisch-Lehrende und wissenschaftliche Mitarbeiterinnen und Mitarbeiter) waren sich einig über die Effizienz des Programms, da es die Entwicklung von C-Tests für Arabisch nicht nur erleichtert, sondern gleichzeitig auch die Textschwierigkeit (Lesbarkeit und Komplexität) misst und verschiedene Optionen bereithält. Die Erstellung eines arabischen C-Tests erfordert viel Mühe und Zeit, besonders angesichts der Tatsache, dass üblicherweise die Vokalisierungszeichen manuell hinzuzufügen sind. Der C-Test-Editor hingegen verkürzt diese Zeit und bietet Nutzenden die Möglichkeit, den Test mit diakritischen Zeichen zu erhalten. Ein weiterer Vorteil des Programms besteht in der Verfügbarkeit linguistischer Informationen über die Testlets. Lehrkräfte betonen die Wichtigkeit solcher Daten, da sie nützliche linguistische Fakten liefern, die helfen, die gewünschten Testlets auszuwählen und ebenfalls umfängliche Zusatzangaben über die Testlets gewähren.

Die Auswertung der automatisch hinzugefügten diakritischen Zeichen zeigte, dass sich an einigen Stellen Fehler einstellten, die zu einer falschen Aussprache des Wortes führen. Nicht richtig positionierte Diakritika können im Arabi-

schen zu einer falschen Aussprache führen, vor allem bei L2-Lernenden²⁸ auf einer niedrigen Kompetenzstufe. Dies liegt vor allem an der großen Zahl an Homographen im Arabischen (Wörter gleicher Schreibung, aber unterschiedlicher Aussprache). Arabische Homographie können durch Hinzufügen von Vokalisierungszeichen weiter disambiguiert werden. Obwohl das automatische Einfügen von Diakritika (noch) nicht immer fehlerfrei erfolgt, ist es nützlich, da es (zeit- und aufwandminimierend) das manuelle Hinzufügen von Vokalisierungszeichen überflüssig macht. Ein weiterer Vorteil liegt darin, dass der Texteditor es erlaubt, falsche Vokalisierungszeichen manuell zu bearbeiten.

4.4. Zusammenfassung

Die Erweiterung des webbasierten Programms C-Test-Editor um Arabisch ist nur ein erster Prototyp zur automatischen Generierung arabischer C-Tests. Als erster Versuch hat er seine Effizienz und Nützlichkeit im Bereich der Einstufung von Lernenden mit Arabisch als Fremdsprache bewiesen. Dieses Projekt wird weitergeführt und so der Editor für Arabisch kontinuierlich verbessert. Eingebunden in diesen Prozess sind Mitglieder verschiedener Einrichtungen der Universität Tübingen und weiterer Hochschulen. Der Editor-Prototyp für Arabisch (mit seinen o. a. Funktionen) wurde in einer Evaluierungsphase auf seine Effizienz hin bewertet. Es bedarf u. a. weiterer Untersuchungen, um die sprachspezifischen Tilgungsregeln zu verfeinern und eventuell weitere (nutzer- bzw. niveauspezifische) Optionen hinzuzufügen. Auch könnten in einem gesonderten Schritt einige der Strategien über ein Validierungsverfahren mit L2-Arabischlernenden überprüft werden, z. B. wenn gebundene Morpheme von der Tilgung ausgeschlossen werden und umgekehrt. Die Anpassungsfähigkeit des Programms an verschiedene Bedarfe ist ein wesentlicher Baustein auch beim Editor für Arabisch.

Zusammenfassend lässt sich sagen, dass die Entwicklung des C-Test-Editors für Arabisch stark geprägt war von regelmäßigen Treffen mit Mitarbeitenden verschiedener Hochschulen und vom Feedback derer, die diese Anwendung in Hochschullehre und -testen nutzen möchten.

Ein weiteres Desiderat ist, zusätzlich zu Arabisch als Fremdsprache digitale Konzepte zu entwickeln, die auch für Herkunftssprechende passende Testverfahren bereithalten.

5. Vorschlag einer bedarfs- und nutzerspezifischen Kategorisierung des C-Tests für Arabisch

In diesem Kapitel wird versucht, die Anforderungen an C-Test-Prüflinge differenzierter zu gestalten, um die verschiedenen Bedarfe der Einrichtungen, die

²⁸ L2 bezieht sich in diesem Kapitel auf Lernende mit Arabisch als Fremdsprache.

Arabischlehre anbieten, zu bedienen. Um diese Differenzierung durchzuführen, sollten erst einmal die zu bedienenden Bedarfe ermittelt werden. Diese Bedarfe gehen über den universitären Bereich hinaus und liegen zwischen Alltagsgesprächen, vor allem in den zum Teil sehr unterschiedlichen arabischen Dialekten auf der einen Seite und der Analyse von hochkomplexen Texten des Klassischen Arabisch auf der anderen Seite. Werden all diese Bedarfe abgedeckt, dann könnte der C-Test sowohl in Universitätsfächern mit Arabischangebot als auch an Fach-/Fremdsprachenzentren dieser Universitäten und sogar außerhalb des Universitätsrahmens, an Volkshochschulen oder privaten Sprachschulen eingesetzt werden. Ziel ist also, das Angebot an Personen zu richten, die aus welchem Grund auch immer Arabisch lernen wollen und sich daher einstufen lassen möchten.

Unter dieser Prämisse wurden verschiedene Kategorien einer C-Test-Lösung entwickelt, die in diesem Kapitel vorgestellt werden²⁹. Doch zunächst sollen die Bedarfe der Institutionen mit Arabischlehreangebot erläutert werden, um die abgefragten sprachlichen Elemente in den Lösungskategorien zu unseren C-Tests nachvollziehbar zu machen. Diese Bedarfe sind nach unserem Kenntnisstand in drei Gruppen zusammenzufassen:

1. Fächer, die eine hohe Sprachkompetenz im Arabischen erfordern, wie z. B. Arabistik, Islamwissenschaft, Theologie oder Philosophie. Diese Fächer beschäftigen sich in verschiedener Hinsicht sehr tiefgreifend mit dem Arabischen. In diesen Fächern kann die Komplexität der zu bearbeitenden Texte inhaltlich sehr hoch sein, was natürlich nicht nur, aber häufiger, für Texte aus früheren Epochen der islamischen/arabischen Geschichte gilt. Dies bedarf eines sehr vertieften Wissens über die arabische Sprache in ihrer Grammatik inkl. Syntax, ihrem Vokabular sowie Ausdrucksweise, und zwar sowohl im modernen Hocharabischen als auch im Klassischen Arabisch. Ein C-Test für diese Fächer sollte also in der Lage sein, viel Wissen abzufragen.
2. Andere Fächer beschäftigen sich vor allem mit Texten des modernen Hocharabischen aus den Medien sowie mit Texten anderer Gattungen mit ähnlichem Schwierigkeitsgrad, was grundsätzliches Wissen hinsichtlich des Arabischen, aber nicht unbedingt umfänglich über das Klassische Arabisch verlangt. Dazu gehören vor allem die Wirtschaftswissenschaften und Politikwissenschaften (beide mit einem Schwerpunkt in der arabischen Welt, z. B. Nahost) oder Ethnologie.

²⁹ Die Kategorien der C-Test-Lösung beziehen sich vor allem auf die Möglichkeiten der sogenannten Vokalisation, die sowohl mit der Sprachstufe zu tun hat als auch unter Umständen das Wissen über die Grammatik des Arabischen wiedergeben kann.

Hier würden je nach Ausrichtung der Studierenden nicht unbedingt Kenntnisse zur Analyse und Übersetzung von hochkomplexen Texten gebraucht – es sei denn sie befassen sich mit komplexeren Texten aus ihrem Studienfach, beispielsweise mit *Islamic Banking* oder (zwecks politischer Deutung von zeitgenössischen Ereignissen wie Terror) mit theologischen Inhalten aus der Vergangenheit. In diesem Fall liegen die erforderlichen Arabischkenntnisse eher im Bereich von Gruppe 1. Soweit aber das Interesse von Studierenden dieser Fächer sich auf eine Textgattung wie die der Medien beschränkt, können die abgefragten, unter Gruppe 1 genannten Inhalte stark reduziert werden.

3. Hingegen gibt es Personen, darunter auch Studierende aller unter 1. und 2. genannten Fächer, die vor allem nach schneller Kommunikation mit Muttersprachlern streben. Dies kann bei Studierenden daran liegen, für welchen Beruf sie sich entscheiden (zum Beispiel, wenn sie den sozialen Bereich wählen oder für ihre Forschung Gespräche und Interviews führen wollen). Allerdings gibt es wie bereits erwähnt auch Menschen, die außerhalb von Universitäten aus unterschiedlichsten Gründen Arabisch vor allem in einer dialektalen Ausprägung lernen möchten, um mit Muttersprachlerinnen und Muttersprachlern zu kommunizieren, denen das Hocharabische unzugänglich ist oder zumindest nicht leichtfällt. Die Angehörigen dieser Gruppe müssen nicht unbedingt komplexere Texte bearbeiten bzw. analysieren. Für diese Gruppe kann der C-Test mit geringeren Grammatikkenntnissen (ausgenommen Syntax) oder im Dialekt konzipiert werden. C-Tests für diesen Bedarf können sowohl an Universitäten als auch an Fremd- und Fachsprachenzentren oder außerhalb von Hochschulen (wie etwa Volkshochschulen oder privaten Sprachschulen) eingesetzt werden. Allerdings müssen bei C-Tests für arabische Dialekte idealiter alle arabischen Hauptdialekte sowie ihre lokalen Ausprägungen einbezogen werden, was die Mitarbeit von Muttersprachlern und Muttersprachlerinnen aus den entsprechenden Staaten bedingt. Denn dabei sollten ggfs. auch regionale Prägungen der Dialekte, z. B. Ober- und Unter-ägypten, berücksichtigt werden.

Nun können die Lösungsmöglichkeiten anhand eines C-Test-Beispiels gemäß den nachfolgend definierten Kategorien der Lösungen, und zwar jeweils in Verbindung mit den dazugehörigen Fächern besprochen werden. Folgender Text könnte als C-Test fungieren³⁰:

³⁰ Dieser Text war Teil einer selbstverfassten Klausur für Studierende im dritten Semester; daher ist

أسست منظمة «جامعة الدول العربية» عام ١٩٤٥ من قبل سبعة بلدان فقط. وقد
 حصد البلاد العر الأخرى في بعد ع استقلالها وانتس إليها ليد
 عدد أعضا ٢٢ دو. وتضم الجا أقساما لتط المجالات الاقتص
 والثقافية والعد في الد الأعضاء في. أما بالن للمسائل السيا فإن
 المنظ ما زا تعاني م مشكلة قدي هي أند لا تتخ القرارات اللا
 لحل أزم العالم الع الكثيرة. فهي كانت وما زالت منشغلة بحل الخلافات
 العميقة بين الرؤساء العرب أكثر من أن تعمل على تحسين الأوضاع المعيشية
 للشعوب العربية مثلا.

Die durchgespielten Lösungen nach den unten definierten Kategorien wurden so gestaltet, dass die (wie oben) getilgte linke Hälfte in den Lösungen stets in einem grauen Kästchen ergänzt wurde. Hier wurde die linke Worthälfte getilgt (also die zweite Worthälfte im Arabischen, da von rechts nach links geschrieben), wobei für alle Kategorien auch die rechte Hälfte, also der Wortanfang, getilgt werden kann. Prüflinge können je nach individuellem Bedarf oder dem des/der Dozierenden sowohl in einer als auch in mehr als einer Kategorie geprüft werden. Zur Übersichtlichkeit wurden die Erläuterungen der Kategorien in einer Tabelle (Tab. 1) am Ende des Kapitels zusammengefasst.

Kategorie I (vgl. Tab. 1 am Ende des Kapitels):

In dieser Kategorie werden die Texte im unvokalisierten Arabisch bearbeitet (also wie in den heute gängigen Audiomedien oder wie es üblicherweise beim Vorlesen einer Zeitung seitens einer Muttersprachlerin oder eines Muttersprachlers vorkommt).

Inhaltlich dürfte der Bedarf hier überwiegend aus dem Bereich des modernen Hocharabischen (sog. MSA), wie z. B. aus Druckmedien oder aus niveauähnlichen Texten stammen. Demnach entsprechen ihre Komplexität und der Schwierigkeitsgrad ihres Vokabulars dem überwiegenden Teil der arabischen Texte unserer Zeit.

Die Zielgruppe sind Arabischlernende ohne tiefgreifende Kenntnisse der Grammatik, hingegen mit Kenntnissen der arabischen Syntax, so z. B. an Fach-/Fremdsprachenzentren oder in Fächern mit ausschließlicher Schwerpunktsetzung auf modernen Texten. Die technische Umsetzung ist unproblematisch und die Generierung erfolgt maschinell. Die Lösung sähe wie folgt aus:

er in jeder Hinsicht an den Wissenstand des o. g. Semesters angepasst. Aus diesem Grund sind Wortschatz oder Ausdrücke niveauspezifisch gewählt. Der Test wurde manuell generiert und seine Lösung im Sinne der unten besprochenen Lösungskategorien bearbeitet. Beim öffentlichen Einsatz von C-Tests werden in der Regel authentischere Texte genommen.

أسست منظمة «جامعة الدول العربية» عام ١٩٤٥ من قبل سبعة بلدان فقط. وقد حصلت البلاد العربية الأخرى فيما بعد على استقلالها وانتسبت إليها ليصبح عدد أعضائها ٢٢ دولة. وتضم الجامعة أقساماً لتطوير المجالات الاقتصادية والثقافية والعلمية في الدول الأعضاء فيها. أما بالنسبة للمسائل السياسية فإن المنظمة ما زالت تعاني من مشكلة قديمة هي أنها لا تتخذ القرارات اللازمة لحل أزمات العالم العربي الكثيرة. فهي كانت وما زالت منشغلة بحل الخلافات العميقة بين الرؤساء العرب أكثر من أن تعمل على تحسين الأوضاع المعيشية للشعوب العربية مثلاً.

Kategorie II bzw. Kategorie II-Level 1 (vgl. Tab. 1 am Ende des Kapitels):

Diese Kategorie sowie die beiden folgenden können jeweils als eigene Kategorie oder alle zusammen als drei verschiedene Schwierigkeitsgrade ein und derselben Kategorie angesehen werden, weil in allen drei Kategorien die getilgte Hälfte von Testkandidat*innen teilweise oder ganz vokalisiert werden soll. Inhaltlich können die wenig komplexen Texte (aber komplexer als Kategorie I) bei allen drei Kategorie II-Levels aus dem Bereich des MSA oder des Klassischen Arabisch stammen. Die Texte können z. B. aus der Literatur, Sachbüchern, aber auch aus Medien mit stilistisch anspruchsvollen Texten stammen.

Sie sind allerdings aufgrund des Vokalisierungsanspruchs schwieriger zu bearbeiten als in der ersten Kategorie. Vokalisiert werden soll nach der Ergänzung der getilgten Hälfte auf diesem Level der erste oder letzte Buchstabe eines Wortes, abhängig davon, welche Worthälfte getilgt wird. Diese Kategorie ist vor allem für die Linkstilgung geeignet, da dabei die arabische Grammatik (allen voran die Fälle) gut geprüft werden kann. Auch bei der Rechtstilgung lassen sich wichtige sprachliche Merkmale prüfen, so z. B. die Präfixe bei den Verbformen, die Setzung einer *Hamzat Waṣl* oder der *Waṣla* und vieles mehr.

Diese Kategorie eignet sich für die Zielgruppe der Interessierten an der arabischen Grammatik, und zwar über die Syntax hinaus, aus allen oben genannten Bedarfsbereichen, inklusive Fachstudierende mit mittlerem Interesse an Grammatik. Die technische Umsetzung ist unproblematisch und die Generierung erfolgt maschinell, wobei es einzelne Probleme, wie die Setzung einer *Hamzat Waṣl* oder der *Waṣla* bei einer Rechtstilgung, noch zu lösen gilt. Die Lösung sähe wie folgt aus:

أسست منظمة «جامعة الدول العربية» عام ١٩٤٥ من قبل سبعة بلدان فقط. وقد حصلت البلاد العربية الأخرى فيما بعد على استقلالها وانتسبت إليها ليصبح عدد أعضائها ٢٢ دولة. وتضم الجامعة أقساماً لتطوير المجالات الاقتصادية والثقافية والعلمية في الدول الأعضاء فيها. أما بالنسبة للمسائل السياسية فإن المنظمة ما زالت تعاني من مشكلة قديمة هي أنها لا تتخذ القرارات اللازمة لحل أزمات العالم العربي الكثيرة. فهي كانت وما زالت

منشغلة بحل الخلافات العميقة بين الرؤساء العرب أكثر من أن تعمل على تحسين الأوضاع المعيشية للشعوب العربية مثلا.

Kategorie III bzw. Kategorie II-Level 2 (vgl. Tab. 1 am Ende des Kapitels):

In dieser Kategorie oder Level 2 der Kategorie II wird die getilgte linke Hälfte vollständig vokalisiert. Das heißt neben Kasusendungen wird auch die Morphologie (hier sind vor allem die sogenannten Paradigmata gemeint) der arabischen Wörter teilweise abgefragt. Bei einer Tilgung der rechten Hälfte würde anstelle der Kasusendung neben den Paradigmata das Wissen über die Buchstabenarten abgefragt (sogenannte Mond- oder Sonnenbuchstaben; bei der zweiten Gruppe wird der Artikel am ersten Buchstaben des Wortes assimiliert, bei der ersten Gruppe nicht).

Zielgruppe sind bei diesem anspruchsvolleren Abfragen der unterschiedlichen Inhalte Fachstudierende der Islamwissenschaften oder der Arabistik mit hohem (womöglich vorgeschriebenem) Interesse an arabischer Grammatik, wobei auch Studierende andere Fächer mit entsprechendem Interesse sich als Textkandidat*innen dafür eignen. Die technische Umsetzung ist unproblematisch und die Generierung erfolgt maschinell (wobei auch hier bei einer Rechtstilgung die Probleme mit *Hamzat Waṣl* und der *Waṣla* noch zu lösen sind). Die Lösung sähe wie folgt aus:

أسست منظمة «جامعة الدول العربية» عام ١٩٤٥ من قبل سبعة بلدان فقط. وقد حصلت البلاد العربية الأخرى فيما بعد على استقلالها وانتسبت إليها ليصبح عدد أعضائها ٢٢ دولة. وتضم الجامعة أقساما لتطوير المجالات الاقتصادية والثقافية والعلمية في الدول الأعضاء فيها. أما بالنسبة للمسائل السياسية فإن المنظمة ما زالت تعاني من مشكلة قديمة هي أنها لا تتخذ القرارات اللازمة لحل أزمت العالم العربي الكثيرة. فهي كانت وما زالت منشغلة بحل الخلافات العميقة بين الرؤساء العرب أكثر من أن تعمل على تحسين الأوضاع المعيشية للشعوب العربية مثلا.

Kategorie IV bzw. Kategorie II-Level 3 (vgl. Tab. 1 am Ende des Kapitels):

In dieser Kategorie oder Level 3 der Kategorie II würde nicht nur die getilgte linke Hälfte ergänzt und vollständig vokalisiert, sondern auch abweichend vom kanonischen C-Test-Format in der nicht-getilgten Hälfte eingegriffen werden, um auch die vollständigen Paradigmata abzufragen. Bei einer Tilgung der rechten Hälfte würden die gleichen Inhalte abgefragt, wobei die gleichen technischen Probleme bei der automatisierten Generierung vorkommen können.

Die Zielgruppe ist dieselbe wie unter Kategorie II-III bzw. Kategorie II-Level 1-2. Die technische Umsetzung ist zwar möglich, entspräche jedoch nicht dem kanonischen C-Test-Format, daher wird auf diese Lösungskategorie im Rahmen

des C-Test-Editors verzichtet. Eine künftige Umsetzung ist als Übungseinheit vorgesehen und technisch möglich. Vorläufig kann diese Lösungsmöglichkeit umgesetzt werden, indem die Lehrperson die Lösung der letzten Kategorie ausdrückt und Studierende um eine entsprechende Ergänzung (Vokalisierung der nicht getilgten Hälfte) bittet. Die Lösung sähe wie folgt aus:

أسست منظمة «جامعة الدول العربية» عام ١٩٤٥ من قبل سبعة بلدان فقط. وقد حصلت البلاد العربية الأخرى فيما بعد على استقلالها وآنسبت إليها ليصبح عدد أعضائها ٢٢ دولة. وتضم الجامعة أقساماً لتطوير المجالات الاقتصادية والثقافية والعلمية في الدول الأعضاء فيها. أما بالنسبة للمسائل السياسية فإن المنظمة ما زالت تعاني من مشكلة قديمة هي أنها لا تتخذ القرارات اللازمة لحل أزمت العالم العربي الكثيرة. فهي كانت وما زالت منشغلة بحل الخلافات العميقة بين الرؤساء العرب أكثر من أن تعمل على تحسين الأوضاع المعيشية للشعوب العربية مثلاً.

Die besagte, geplante Übungsoption könnte so gestaltet werden, dass man automatisiert die vollständigen bzw. nicht mit Lücke versehenen Wörter ebenfalls vokalisieren kann, und zwar nicht nur bei dieser Kategorie bzw. diesem Level, sondern auch bei der/dem vorherigen. Man könnte auch darüber hinausgehen und die vollständigen zwei Sätze am Anfang sowie am Ende des Textes ebenfalls von Studierenden vokalisieren lassen. Somit wäre der C-Test stark erweitert und orthographische, morphologische und grammatikalische Inhalte würden abgefragt. Die Lösung sähe wie folgt aus:

أسست منظمة «جامعة الدول العربية» عام ١٩٤٥ من قبل سبعة بلدان فقط. وقد حصلت البلاد العربية الأخرى فيما بعد على استقلالها وآنسبت إليها ليصبح عدد أعضائها ٢٢ دولة. وتضم الجامعة أقساماً لتطوير المجالات الاقتصادية والثقافية والعلمية في الدول الأعضاء فيها. أما بالنسبة للمسائل السياسية فإن المنظمة ما زالت تعاني من مشكلة قديمة هي أنها لا تتخذ القرارات اللازمة لحل أزمت العالم العربي الكثيرة. فهي كانت وما زالت منشغلة بحل الخلافات العميقة بين الرؤساء العرب أكثر من أن تعمل على تحسين الأوضاع المعيشية للشعوب العربية مثلاً.

Kategorie V (vgl. Tab. 1 am Ende des Kapitels):

Die Texte dieser Kategorie sollten sich auf die grammatikalischen Inhalte und Phänomene der Dialekte, auch in ihren regionalen Prägungen, beschränken. Hier sollten auch die Besonderheiten jedes einzelnen Dialekts bei den Tilgungsregeln genau überlegt und in Kooperation mit Muttersprachler*innen des jeweiligen Dialekts festgelegt werden, wobei auch das Einbeziehen von auf diesem Gebiet tätigen Wissenschaftlerinnen und Wissenschaftlern sinnvoll erscheint.

Die Texte können nur aus dialektalen Quellen oder von Muttersprachler*innen des jeweiligen Dialektes verfasst werden. Da unser Projekt sich erst einmal auf die Hochsprache des Arabischen beschränkt, wurde hier kein Beispieltext kreiert. Diese Kategorie zielt auf Interessierte, die nur Alltags-/Dialektsprache einer bestimmten Region für eine Urlaubs- oder Geschäftsreise lernen wollen bzw. gelernt haben und testen möchten, aber auch Studierende, die sich auf eine bestimmte Region spezialisieren oder zumindest deren Dialekt lernen möchten.

Mit diesen spezifizierten Kategorien hoffen wir, alle Bedarfe abgedeckt zu haben, denken aber, dass auch hier einige Verbesserungen auf den Weg gebracht werden, um die oben genannten technischen Probleme zu lösen sowie Erweiterungen zu generieren, z. B. um auch Dialekte testen zu können.

	Bedarfe von Einrichtungen / geprüft wird	Einrichtungen	Vokalisation
Kategorie I	Grundsätzliche Kenntnisse der Grammatik, alltägliches Vokabular und Ausdrucksweise von MSA	Fach- und Fremdsprachenzentren sowie Fächer mit MSA-Fokus (z. B. Wirtschaft, Ethnologie oder Politik)	Keine (da MSA / heutige Schriftsprache)
Kategorie II oder II-Level 1	Hohe Sprachkompetenz an MSA / Klassisches Arabisch: Tiefgreifende Kenntnisse über Grammatik, Vokabular sowie Ausdrucksweise	Islamwissenschaft / Orientalistik / Arabistik sowie Theologie und weitere interessierte Fächer oder Personen	Nur Wortende bzw. -anfang der getilgten Hälfte
Kategorie III oder II-Level 2	Höhere Sprachkompetenz an MSA / Klass. Arabisch: Kenntnisse zu Grammatik Vokabular, Ausdrucksweise tiefgreifender als Kat. I, II-1	Islamwissenschaft / Orientalistik / Arabistik sowie Theologie und weitere interessierte Fächer oder Personen	Vokalisation der gesamten getilgten Hälfte
Kategorie IV oder II-Level 3	Höchste Sprachkompetenz an MSA / Klass. Arabisch: Kenntnisse zu Grammatik Vokabular, Ausdrucksweise tiefgreifender als Kat. III, II-2	Islamwissenschaft / Orientalistik / Arabistik sowie Theologie und weitere interessierte Fächer oder Personen	Vokalisation der getilgten sowie nicht-getilgten Hälfte (keine C-Test-Eigenschaften mehr – als Übung zu generieren)
Kategorie V	Wissen über Grammatik, Vokabel, Ausdrucksweise sowie Lautverschiebungen des jeweiligen Dialekts	Alle Interessierte Lehreinrichtungen und Personen	Keine (da Dialekte)

Tab. 1: Zusammenfassung der Kategorien

6. Ausblick

Während C-Tests für Arabisch (Modernes Hocharabisch und Klassisches Arabisch) mit dem C-Test-Editor automatisiert erstellt und umfänglich ausgewertet werden können, ist die Liste der Desiderata zu C-Tests und digitalisierten Übungen für Arabisch noch umfänglich. Für C-Test- und Übungstexte soll die bereits existente Komplexitäts- und Lesbarkeitsanalyse erweitert werden – hier sind u. a. besonders Messgrößen zur morphologischen Komplexität zu erwähnen. Ebenfalls geplant ist, die Zahl der Nutzeroptionen des C-Test-Editors zu erhöhen. Beide genannten Punkte sind zeitlich recht überschaubar. Ebenso geplant, aber erst mittelfristig umsetzbar wird bezüglich Testvalidität die Hinzunahme der *Item Response Theory* sein. Ebenfalls wünschenswert ist ein Abgleich von Prüfungsleistungen (z. B. Abschlussklausuren) mit C-Test-Ergebnissen Arabisch³¹.

Aufgrund der Diglossie des Arabischen besteht schließlich Bedarf (und gleichzeitig eine Herausforderung darin) – unabhängig davon, ob digital oder nicht digital umgesetzt – an vergleichenden Ansätzen für das Nebeneinander von Schriftsprache (MSA?) und gesprochenen Varietäten des Arabischen. Falls es gelingt, dies auch digital zu ermöglichen, böte dies eine zusätzliche Hilfestellung.

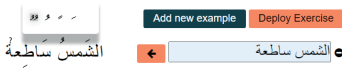
LEAD Graduate School & Research Network und die Orient- und Islamwissenschaft der Universität Tübingen haben begonnen, die Eckpunkte für digitalisierte Übungen für Arabisch festzulegen³². Dass personalisiertes Lernen vorteilhaft ist, wird vielfach konstatiert, vgl. Eckert (2020: 43): „Wenn man davon ausgeht, dass jede/r Lernende/r ihren/seinen eigenen, persönlichen Lernstil mitbringt und zudem auf verschiedene Backup-Stile zurückgreifen kann, werden Methoden, die personalisiertes Lernen ermöglichen, immer wirkungsvoller sein als globale Formate.“

Geplant sind Übungen zunächst für Studierende der Orient- und Islamwissenschaft und andere Studierende. Die ersten Schritte hierfür sind getan: diakritische Zeichen können angewählt werden – die Kandidatin bzw. der Kandidat erhält eine automatisierte farbig unterlegte Richtig/Falsch-Rückmeldung; als zusätzliche Hilfe ebenfalls sichtbar (siehe Abb. 8) wird bei falscher Antwort (rot) die eigentlich richtige Antwort (grün) parallel zur falschen angezeigt (siehe Abb. 6).

³¹ Ein entsprechender Abgleich wurde über zwei Semester für Französisch, Italienisch und Spanisch am Romanischen Seminar der Universität Tübingen durchgeführt mit ca. 115 bzw. ca. 80 Teilnehmenden. Die als C-Test über Ilias in den drei o.a. Sprachen abgelegten Einstufungstest (die Testlets wurden mit der DLTP-Plattform generiert) wurden mit den jeweiligen Semesterklausuren des Grundkurses der drei Sprachen verglichen. Die in Kapitel 2. erwähnte Itemanalyse der Plattform wurde für die Auswertung eingesetzt.

³² Dieses Projekt umfasst auch studentische Mitarbeit für die Programmierung der Übungen durch Mourhaf Kazzaz, derzeit im Masterstudiengang Computational Linguistics der Universität Tübingen.

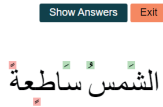
1) The teacher creates examples and specifies a valid output



2) the student attempts the exercise and an attempt is registered



3) the attempt is validated



4) the correct answers are shown when the button is clicked

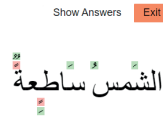


Abb. 8: Digitale Übungsplattform mit automatisiertem Feedback: Beispiel Diakritika

Das Programm wird in absehbarer Zeit mehr Übungsformate bereitstellen, interaktive Elemente enthalten, individualisiertes Feedback geben und auch für weitere Nutzergruppen verfügbar sein. Ähnlich wie bei bereits einsatzfähigen Tools für andere Sprachen wird auch das Programm für Arabisch eine Auswahl an Übungsformaten enthalten (z. B. *fill in the blanks*, *multiple choice*, *click*). Des Weiteren wird es möglich sein, zwischen verschiedenen linguistischen Merkmalen zu wählen (*dialectics*, *negation*, *paradigms*, *tense*, *verb stem* usw.). Damit kann bald eine zumindest kleine Auswahl an Übungsformaten und linguistischen Merkmalen digital eingesetzt, erprobt und stufenweise erweitert werden. Die Übungen zu verschiedenen sprachlichen Fertigkeiten sind interaktiv angelegt; Lernende erhalten somit einen Überblick darüber, welche Antworten richtig, welche fehlerhaft waren; geplant ist ebenfalls sogenanntes *scaffolding feedback*, Vorschläge für eine oder Links zu einer korrekte(-n) Antwort (siehe Beispiel Diakritika) bzw. warum eine Antwort nicht korrekt ist.

Ein wichtiges Desiderat ist darüber hinaus das Einbeziehen der Zielgruppe der sogenannten *heritage speakers*. Die verschiedenen gesprochenen Varietäten des Arabischen – oder zumindest eine minimale Auswahl davon – in die oben beschriebenen Konzepte und digitalen Tools für C-Tests und für Übungen zu übertragen und Niveaus zu definieren, stellt eine beträchtliche Herausforderung dar; inwieweit es möglich sein wird, einen Teil davon erfolgreich umzusetzen, ist zum jetzigen Zeitpunkt noch unklar.

Schließlich seien bei der Liste der Desiderata auch Schulen genannt, für die möglichst bald versucht werden soll, C-Tests für Arabisch zu generieren, die die entsprechenden Bedarfe (Lückenzahl, Lückenumfang, Themen, Textniveau

usw.) bedienen: Dies bedeutet, dass z. B. Lückenzahl, Lückenumfang, Themen, Textniveau usw. entsprechend zu ändern sind. Hierzu ist eine Kooperation mit zunächst ein oder zwei Schulen geplant, um einen ersten Rahmen abzustecken und eine Pilotierung zu planen.

Zwischen dem Einreichen des Beitrags und seiner Veröffentlichung fand ein Wechsel von LEAD zum Hector-Institut für Empirische Bildungsforschung der Universität Tübingen statt, der mit einer strukturellen Stabilisierung der Plattform einhergeht; die Plattform wurde damit umbenannt von C-Test-Editor zu DLTPT (*Digitalised Language Teaching & Proficiency Testing*)³³.

7. Literaturverzeichnis

- AHMED, Hebah (2022): *Generating automatic Arabic C-tests and measuring their difficulty using linguistic complexity and readability metrics*. Unveröffentlichte Masterarbeit. Universität Tübingen, Fachbereich Theoretische Computerlinguistik.
- AL-HEETI, Khalaf Nassar (1984): *Judgment analysis technique applied to readability prediction of Arabic reading material*. Unpublished doctoral dissertation. University of Northern Colorado.
- AL-BAZI, Matti Phillips Khoshaba (n. d.): *Integrative Arabic Test (IAT): C-test (rule of 2), multiple-choice discrete-point method, and dictocomps*. Salinas, CA.
- AL KHALIL, Muhamed / Habash, Nizar / Jiang, Zhengyang (2020): A large-scale leveled readability lexicon for standard Arabic. In: *Proceedings of the 12th language resources and evaluation conference*, 3053-3062.
- AL TAMIMI, Abdel-Karim / Jaradat, Manar / Al-Jarrah, Nuha / Ghanem, Sahar (2014): AARI: Automatic arabic readability index. *The International Arab Journal of Information Technology* 11 (4), 370-378.
- ASANO, Yuki (2010): *Misst ein C-Test das, was er messen soll? Konstruktdefinition als Basis für die Validierung von C-Tests*. Schriftliche Hausarbeit für die Masterprüfung der Fakultät für Philologie der Ruhr-Universität Bochum. https://uni-tuebingen.de/fileadmin/Uni_Tuebingen/Fakultaeten/Neuphilologie/Englisches-Seminar/Linguistik_des_Englischen/Prof_Weber/Staff/Y_Asano/Publications/MA_Arbeit_Yuki_Asano.pdf
- BISPING, Meikel / Raatz, Ulrich (2002): Sind computerisierte und Papier&Bleistift-Versionen des C-Tests äquivalent? In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. Bochum: AKS-Verlag, 131-155.
- BISPING, Meikel (2006): Zur Validität von Computer-C-Tests. In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test: Theorie, Empirie, Anwendungen*. Frankfurt/M.: Lang, 147-166.
- DAGHBOR, Abdellatif (1998): *Der C-Test im Arabischen: Überlegung, Entwicklung und Erprobung*. Unpublished Ms. Universität Essen.

³³ DLTPT-Website: <https://uni-tuebingen.de/de/250938>

- DALLER, Helmut / Phelan, David (2006): The C-test and TOEIC® as measures of students' progress in intensive short courses in EFL. In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test: Theorie, Empirie, Anwendungen*. Frankfurt/M.: Lang, 101-120.
- DRACKERT, Anastasia / Timukova, Anna (2020): Welche Sprachkompetenz misst der C-Test bei HerkunftssprecherInnen: allgemeine oder doch bildungssprachliche? In: Drackert, Anastasia / Mainzer-Murrenhoff, Mirka / Soltyska, Anna / Timukova, Anna (Hrsg.): *Testen bildungssprachlicher Kompetenzen und akademischer Sprachkompetenzen. Zugänge für Schule und Hochschule*. Berlin: Lang, 159-186.
- ECKERT, Martina (2020): *Online-Lehre mit System. Wie man in der digitalen Lehre passgenaue Lernimpulse setzt und neue Lernerfahrungen ermöglicht*. Wiesbaden: Springer Gabler (eBook).
- ECKES, Thomas / Grotjahn, Rüdiger (2006): C-Tests als Anker für TestDaF: Rasch-Analysen mit dem kontinuierlichen Ratingskalen-Modell. In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test. Theorie, Empirie, Anwendungen*. Bd. 6. Frankfurt/M.: Peter Lang, 167-194.
- EUROPARAT (Hrsg.) (2001): *Gemeinsamer europäischer Referenzrahmen für Sprachen: Lernen, lehren und beurteilen*. Berlin/München: Langenscheidt.
- EUROPARAT (Hrsg.) (2020): *Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Begleitband mit neuen Deskriptoren*. Stuttgart: Klett.
- GALASSO, Sabrina (2018): *Automated C-Test Difficulty Prediction: Integrating Lexical, Sentence, and Text Features in a Multi-Lingual Perspective*. Masterarbeit am Seminar für Sprachwissenschaft der Universität Tübingen. <http://www.sfs.uni-tuebingen.de/~sgalasso/pdf/galasso-18.pdf>
- GROTJAHN, Rüdiger (2002): Konstruktion und Einsatz von C-Tests. Ein Leitfaden für die Praxis. In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bd. 4. Bochum: AKS-Verlag, 211-225.
- GROTJAHN, Rüdiger (2006): 25 Jahre C-Test: Einleitung und Übersicht über den Band. In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test. Theorie, Empirie, Anwendungen*. Bd. 6. Frankfurt/M.: Peter Lang, ix-xxv.
- GROTJAHN, Rüdiger (Hrsg.) (2014): *Der C-Test: Aktuelle Tendenzen / The C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bd. 4. Bochum: AKS-Verlag.
- GROTJAHN, Rüdiger / Drackert, Anastasia (2020): *The electronic C-Test bibliography: version October 2020*. <http://www.c-test.de/deutsch/index.php?lang=de&content=bibliografie§ion=ctest>
- GROTJAHN, Rüdiger / Tönshoff, Wolfgang (1992): Textverständnis bei der C-Test-Bearbeitung. Pilotstudien mit Französisch- und Italienischlernern. In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bd. 1. Bochum: Brockmeyer, 19-95.
- GUDER, Andreas / Fachverband Chinesisch e.V. (2015): *European Benchmarks for the Chinese Language (EBCL) / Europäischer Referenzrahmen für Chinesisch als Fremd-*

- sprache*. English-German edition of the project "European Benchmarks for the Chinese Language (EBCL)" Levels A1.1 / A1 / A1+ / A2 / A2+. München: iudicium.
- KLEIN-BRALEY, Christine / Raatz, Ulrich (1982a): Der C-Test: ein neuer Ansatz zur Messung allgemeiner Sprachbeherrschung. *AKS-Rundbrief* 4, 23-37.
- KLEIN-BRALEY, Christine / Raatz, Ulrich (1982b): Der C-Test: Eine neue Methode zur Messung der globalen Sprachfähigkeit bei L1- und L2-Lernern. In: Kühlwein, Wolfgang / Raasch, Albert (Hrsg.): *Stil. Komponenten – Wirkungen*. Bd. 2. Kongreßberichte der 12. Jahrestagung der Gesellschaft für Angewandte Linguistik, GAL e. V., Mainz 1981. Tübingen: Narr, 73-78.
- KLEIN-BRALEY, Christine (1985a): C-Tests and construct validity. In: Klein-Braley, Christine / Raatz, Ulrich (Hrsg.): *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis*. Bochum: AKS-Verlag, 55-65.
- KLEIN-BRALEY, Christine (1985b): Reduced redundancy as an approach to language testing. In: Klein-Braley, Christine / Raatz, Ulrich (eds.): *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis*. Bochum: AKS-Verlag, 1-13.
- KLEIN-BRALEY, Christine / Grotjahn, Rüdiger (1998): *C-Tests in der Schule? Praxis des neu sprachlichen Unterrichts* 45 (4), 411-417.
- KLEIN-BRALEY, Christine (2002): Psycho-linguistics of C-Test taking. In: Coleman, James A. / Grotjahn, Rüdiger / Raatz, Ulrich (eds.): *University language testing and the C-test*. Bochum: AKS-Verlag, 131-142. <http://www.c-test.de/deutsch/index.php?lang=de§ion=originalia>
- KOLLER, Gerhard / Zahn, Rosemary (1996): Computer based construction and evaluation of C-tests. In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bd. 3. Bochum: Brockmeyer: 401-418.
- RAATZ, Ulrich / Klein-Braley, Christine (1982): The C-test – a modification of the cloze procedure. In: Culhane, Terry / Klein-Braley, Christine / Stevenson, Douglas K. (eds.): *Practice and problems in language testing IV. Proceedings of the Fourth International Language Testing Symposium of the Interuniversitäre Sprachtestgruppe held at the University of Essex, 14-17 September 1981*. Colchester: University of Essex, Dept. of Language and Linguistics, 113-138.
- RAATZ, Ulrich / Klein-Braley, Christine (1985): How to develop a C-Test. In: Klein-Braley, Christine / Raatz, Ulrich. (Hrsg.): *Fremdsprachen und Hochschule 13/14: Thematischer Teil: C-Tests in der Praxis*. Bochum: AKS-Verlag, 20-22.
- RAATZ, Ulrich / Klein-Braley, Christine (1998): Gleichwertige Zertifikate – überall und immer oder Wie man UNICERT-Zertifikate mittels C-Tests kalibrieren kann. In: Eggenesperger, Karl-Heinz / Fischer, Johann (Hrsg.): *Handbuch UNICERT®*. Bochum: AKS-Verlag, 107-114.
- RAATZ, Ulrich (2001): Modellvorstellungen zur Erklärung von C-Test-Leistungen. In: Pürschel, Heiner / Raatz, Ulrich (eds.): *Tests and translation: Papers in memory of Christine Klein-Braley*. Bochum: AKS-Verlag, 39-67.

- RAISH, Michael (2018): Design and development of an Arabic C-test. In: Norris, John M. (ed.): *Developing C-tests for estimating proficiency in foreign language research*. Berlin: Lang, 35-60.
- RIGGS, Daniel / Maimone, Luciane L. (2018): A computer-administered C-test in Spanish. In: Norris, John M. (ed.): *Developing C-tests for estimating proficiency in foreign language research*. Berlin: Lang, 265-294.
- SCHÖN, Almut / Zimmermann, Kerstin / Johnson, Natalia (2012): Intrauniversitäre Kooperation – zur gemeinsamen Entwicklung eines C-Tests durch Sprachenzentrum und Sprachlehrforschung. *Fremdsprachen und Hochschule* 86, 61-79.
- SCHRÖDER, Konrad / Harsch, Claudia (2003): *Rahmenkonzeption zur Erfassung Sprachlicher Kompetenzen: Der C-Test*. Augsburg: Universität Augsburg. http://www.philhist.uni-augsburg.de/lehrstuehle/anglistik/didaktik/forschung/desi/desi/DESI_C-Test_Konzeption_10_2002.html
- SENDER, R. J. / Smith, E. A. (1967): *Automated readability index*. Tech. Rep. Cincinnati Univ., OH.
- SIGOTT, Günther / Köberl, Johann (1996): Deletion patterns and C-Test difficulty across languages. In: Grotjahn, Rüdiger (Hrsg.). *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bd. 3. Bochum: Brockmeyer, 159-172. <http://www.c-test.de/deutsch/index.php?lang=de§ion=originalia>
- SUMBLING, Mick / Viladrich, Carme / Doval, Eduardo / Riera, Laura (2014): C-test as an indicator of general language proficiency in the context of a CBT (SIMTEST). In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test: Aktuelle Tendenzen / The C-Test: Current trends*. Frankfurt/M.: Lang, 55-110.
- SVETACHOVA, Yulia (2015): *C-test item difficulty prediction: Exploring the linguistic characteristics of C-tests using machine learning*. Unveröffentlichte Masterarbeit, Universität Tübingen, Seminar für Sprachwissenschaft.
- VOSS, Bernd (2009): 25 Jahre *Language Testing* – 25 Jahre C-Test. [Rezension: Sonderausgabe *Language Testing*, (3) 25/2008 und Rüdiger Grotjahn (Hrsg. 2006): *Der C-Test: Theorie, Empirie, Anwendungen / The C-Test: Theory, Empirical Research, Applications*. Frankfurt/M.: Lang]. *Fremdsprachen und Hochschule* 81, 143-152.
- WOCKENFUSS, Verena / Raatz, Ulrich (2006): Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests. In: Grotjahn, Rüdiger (Hrsg.): *Der C-Test: Theorie, Empirie, Anwendungen*. Frankfurt/M.: Lang, 211-242.

Hebah Ahmed hat ihren Masterabschluss in Computerlinguistik an der Universität Tübingen im Jahr 2022 erworben. Sie arbeitet als freie Mitarbeiterin mit LEAD Graduate School & Research Network der Universität Tübingen zusammen, um zur Verbesserung und Digitalisierung von Bewertungsinstrumenten für Zweitsprachen, insbesondere für das Arabische beizutragen. Außerdem arbeitet sie in der Sprachverarbeitungsindustrie. Ihr Forschungsschwerpunkt ist hier Sprachverarbeitung des Arabischen (NLP) und Linguistische Komplexität.

E-Mail: hebah.ahmed@uni-tuebingen.de

Hala Attoura lehrt seit 2001 als Lektorin und Koordinatorin für Arabisch am Asien-Orient-Institut der Universität Tübingen. Sie ist staatlich geprüfte Übersetzerin und Dolmetscherin (Universität Mainz) und auf diesem Feld seit längerer Zeit tätig. Darüber hinaus forscht sie in der Kulturwissenschaft zu Kulturtransfer und Akkulturation sowie zum Transfer von Sprichwörtern. Hala Attoura war nach einem Abschluss in Visueller Kommunikation an der Universität Damaskus und dem Studium der Vorderasiatischen Altertumskunde, Altorientalistik und Klassischen Archäologie an der FU Berlin im Rahmen zahlreicher archäologischer Ausgrabungsprojekte tätig.

E-Mail: hala.attoura@uni-tuebingen.de

Claudia Duttlinger leitete von 2002-2019 das Fremdsprachenzentrum der Universität Tübingen. Davor war sie wissenschaftliche Mitarbeiterin der Romanistik der Universität Trier. Nach ihrem Studium der Romanistik und Anglistik an der Universität Siegen promovierte sie 1999 an der Universität Trier zu Komplimenten im Spanischen. Zwischen 2020 und 2023 koordiniert sie bei LEAD Graduate School & Research Network die Erstellung einer digitalen Plattform für verschiedene Sprachen zu C-Tests und Übungen mit individualisiertem Feedback und Scaffolding. Seit August 2023 ist das Projekt dem Hector-Institut für Empirische Bildungsforschung angeliebert. Claudia Duttlinger betreut außerdem C-Test-AGs bestehend aus Mitgliedern verschiedener Hochschulen in konzeptionellen Fragen. Ihre Forschungs-/Arbeitsschwerpunkte seit 2002 sind Testen und Qualitätssicherung beim Fremdsprachenerwerb. Sie ist Mitglied der Wissenschaftlichen UNICert®-Kommission.

E-Mail: claudia.duttlinger@uni-tuebingen.de