



Ein Prüfungssimulator für DaF

Hans Jürgen Heringer, Herrsching am Ammersee

ISSN 1470 – 9570

Ein Prüfungssimulator für DaF

Hans Jürgen Heringer, Herrsching am Ammersee

In diesem Beitrag wird eine etwas außergewöhnliche App vorgestellt. Sie simuliert eine DaF-Prüfung. Lerner können sich damit auf die echte Prüfung und die Prüfungssituation vorbereiten. Die mündliche Prüfung ist natürlich ausgenommen.

0. Vorbemerkung

Ein Prüfungssimulator ist eine interaktive elektronische Prüfung, mit der Lerner Prüfungen simulieren können als Training für die echte Prüfung. Der Simulator muss also die echte Prüfung authentisch abbilden und dem Testling eine authentische Bewertung seiner Leistung geben. Die Bewertung kann am Ende stehen, sie kann aber auch zum Training und zur Kontrolle Schritt für Schritt oder Item für Item oder Abteilung für Abteilung gegeben werden.

Karriererelevante Prüfungen sind standardisiert. Bei ihnen ist auch die Bewertung weitgehend standardisiert. Standardisierte Prüfungen im DaF-Bereich sind vor allem die DSH, TestDaF, Goethe/ Telc Zertifikat, Deutsch Test für Zuwanderer. Für alle habe ich Simulatoren entwickelt. Hier stelle ich den Zertifikat-Simulator vor.¹ Ich folge dabei dem Prüfungsablauf, wähle aber nach Itemtyp aus.

Alle diese Prüfungen testen die sprachlichen Grundfähigkeiten. Bisweilen mit leicht unterschiedlicher Benennung. Vor allem vor Grammatik scheuen sich manche. Klingt wohl zu bedrohlich.



Abbildung 1: sprachliche Grundfertigkeiten, die in der Zertifikatsprüfung und entsprechend im Simulator getestet werden.

¹ Der Simulator zur Zertifikatsprüfung ist für Android-Geräte unter <https://play.google.com/store/search?q=goethe%20zertifikat%20simulator> verfügbar.

1. Aufbau und Design

1.1. Itemtyp 1: Multiple Choice (Textverständnis)

The screenshot shows a mobile application interface for a reading comprehension test. At the top, there is a status bar with 'Carrier', signal strength, '1:44 PM', and '100%' battery. Below this is a navigation bar with four tabs: 'Leseverstehen' (selected), 'Sprachbausteine', 'Hörverstehen', and 'Schriftlicher Ausdruck'. On the right of the navigation bar, the time '148:59' is displayed. Below the navigation bar, the test title 'Atompilz im Riesengebirge' is shown. The main content area is divided into two columns. The left column contains the text of the article, which discusses a simulated nuclear attack in the Riesengebirge region. The right column contains four multiple-choice questions, each with three radio button options. The questions are: 'Die Gruppe Ztohoven', 'Die Atombombe', 'Nur wenige Zuschauer', and 'Der junge Tscheche Roman Tyc'. At the bottom of the screen, there is a progress bar with buttons for 'LV1', 'LV2' (selected), 'LV3', 'SB1', 'SB2', 'HV1', 'HV2', 'HV3', 'SA', 'Zurücksetzen', 'Bewerten', 'Weiter', and 'Ende'.

Abbildung 2: Itemtyp 1: Multiple Choice (Textverständnis)

Leseverstehen wird je nach Prüfung mit unterschiedlichen Textsorten in unterschiedlicher Textlänge getestet. Verbreitet ist dabei die multiple choice- (manchmal auch fill in-) Aufgabenstellung. In dieser Prüfung wird nach Goethe-Vorgabe ein Dreier-MC (eine Multiple-Choice Aufgabe mit 3 Antwortoptionen) gewählt, der allerdings methodisch nicht als der beste gesehen wird. Die Wahrscheinlichkeit der Zufallstreffer ist zu hoch. Darum sind allgemein Vierer-MCs besser.

Ich beschreibe kurz den oberen Teil des Screens. Sie sehen hier die Grundnavigation nach den vier Prüfungsteilen. Es sind Buttons, mit denen Sie an den Beginn des jeweiligen Teils springen können. Rechts sehen Sie die Zeit, die Ihnen in der Prüfung noch zur Verfügung steht.

Darunter sehen Sie einen Vorschlag, wie viel Zeit der Prüfling dieser Aufgabe widmen sollte, und dazu, wie viele Punkte er bestenfalls gewinnen kann. Rechts bekommt der

Prüfling gezeigt, wie viel Punkte er insgesamt von der Höchstzahl der erreichbaren Punkte gewonnen hat.

1.2. Itemtyp 2: Fill-in (Strukturen oder Grammatik)

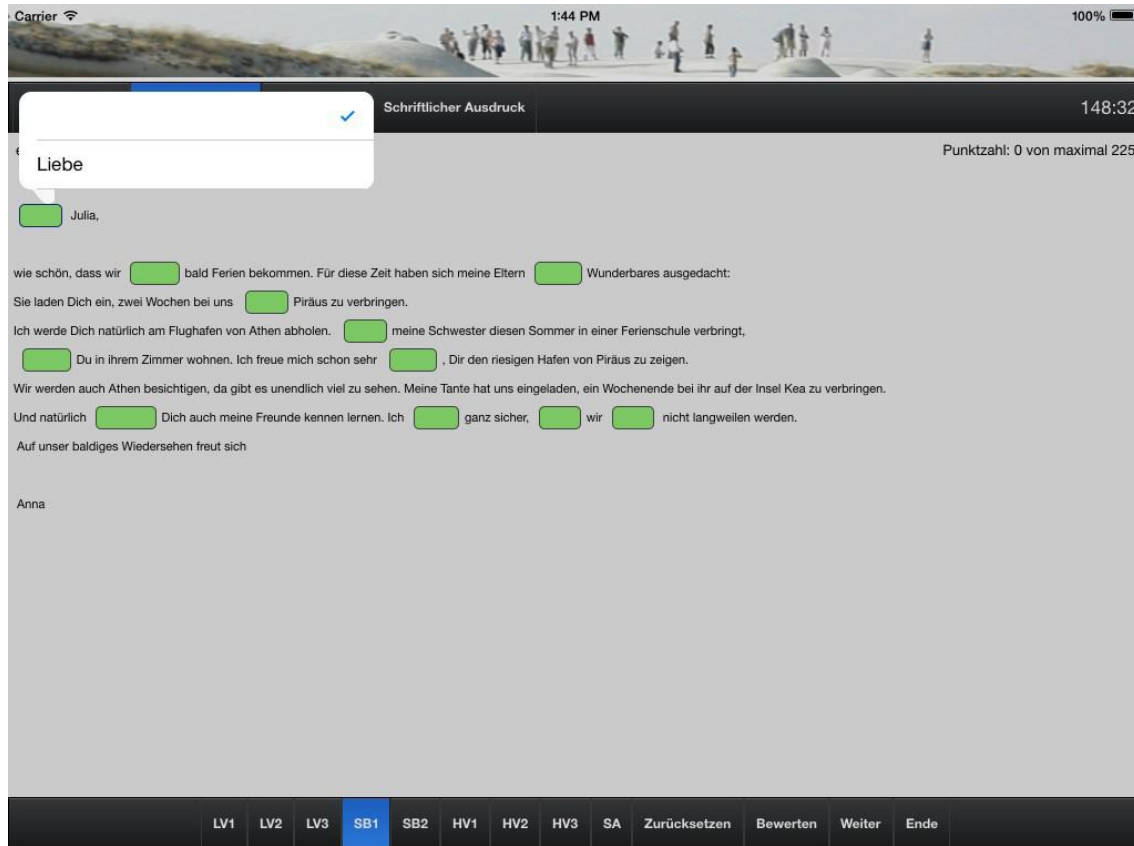


Abbildung 3: Itemtyp 2: Fill-in (Strukturen oder Grammatik)

Dieser Itemtyp wird hier exemplifiziert am Beispiel eines Grammatiktests, für den der Typus beliebt ist. Dabei sind diverse Varianten möglich. Je nach gewünschtem Schwierigkeitsgrad kann man freie Einsetzung verlangen oder die Schwierigkeit reduzieren durch Vorgaben und darunter wieder durch die Anzahl der Vorgaben den Grad der Schwierigkeit variieren.

In der technischen Implementierung gibt es dafür mehrere Möglichkeiten. Während in der Paper-and-Pencil-Version das gesamte Einsetzmaterial irgendwo beisammen dasteht und in den Text einzusetzen ist, kann man elektronisch bessere Realisierungen implementieren. Eine anspruchsvolle Realisierung ist das Drag'n Drop, das für Phones und Tablets nur schwer zu realisieren ist (in der WebVersion des Zertifikats ist es aber verwendet). Einfacher ist ein sog. Listenfeld, das an Ort und Stelle aufgeht und die Alternativen anbietet, aus denen per Click ausgewählt wird. Auf dem Schirm oben enthalten die Listenfelder jeweils drei Optionen gemäß Vorlage. Eine simplere und

anspruchslösere Realisierung macht man über Buttons, wo aus der Lösungsmenge per Mausclick zu wählen ist. Diese Version wird oft bevorzugt.

1.3. Itemtyp 3: Multiple Choice (Hörverstehen)

The screenshot displays a mobile application interface for a listening comprehension test. At the top, there is a navigation bar with four tabs: 'Leseverstehen', 'Sprachbausteine', 'Hörverstehen' (which is highlighted in blue), and 'Schriftlicher Ausdruck'. The time is 1:48 PM and the battery is at 100%. Below the navigation bar, the test title is 'etwa 10 Minuten, maximal 25 Punkte' and the current score is 'Punktzahl: 12.5 von maximal 225'. The instructions state: 'Mit Starten über den blauen Abspielknopf können Sie ein Gespräch hören. Dazu sollen Sie 10 Aufgaben lösen. Sie hören diesen Text zweimal. Bei jeder Aufgabe sollen Sie feststellen: Habe ich das im Text gehört oder nicht? Markieren Sie beim ersten Hören oder danach: R = richtig, F = falsch. Lesen Sie zuerst die Aufgaben.' The test content is organized into two columns. The left column contains five statements, each with a 'R' (right) or 'F' (false) radio button. The right column contains five statements, each with a 'R' or 'F' radio button. At the bottom, there is a navigation bar with buttons for 'LV1', 'LV2', 'LV3', 'SB1', 'SB2', 'HV1', 'HV2' (highlighted in red with '12.5' above it), 'HV3', 'SA', 'Zurücksetzen', 'Bewerten', 'Weiter', and 'Ende'.

Abbildung 4: Itemtyp 3: Multiple Choice (Hörverstehen)

Dieser Hörverstehenstest ist in multiple choice realisiert, entspricht also eigentlich dem Item Typ 1. Für ihn gelten darum die dortigen Kriterien. Der Unterschied besteht in der Prüfungstechnik darin, dass es nur um Zweier-MCs geht, die bezüglich der Validität gering eingeschätzt werden. Der Prüfling hat ja jeweils eine fifty-fifty-Chance. Darum muss in der Auswertung hier ein besonders strenger Maßstab angelegt werden, vielleicht auch ein Malus für falsche Lösungen vorgesehen sein.

Der Zweier-MC könnte hier gerechtfertigt sein, weil man den Schwerpunkt auf dem Hörverständnis sieht, das ja sozusagen zwei Fähigkeiten zugleich testet: Einerseits das Verstehen des gehörten Texts und andererseits das Verstehen des Lösungstexts.

Auf diesem Screen habe ich auch den Bewertungsmechanismus gezeigt (vgl. Abb.4): Bei jedem Click können Sie über den Bewerten-Button Ihr Ergebnis einspielen.

Übrigens hier bei wahllos „R2“ gewinnen Sie die Hälfte der möglichen Punkte, unten zugeordnet zu sehen und oben im Gesamtergebnis.

Zur unteren Navigationsleiste: Hier sind die Einzelteile anzusteuern, mit „weiter“ erscheinen jeweils fortlaufend die nächsten Aufgaben.

1.4. Itemtyp 4: Halbfreie Textproduktion (Schreiben)

Carrier 1:48 PM 100%

Leseverstehen Sprachbausteine Hörverstehen **Schriftlicher Ausdruck** 149:02

etwa 30 Minuten, maximal 45 Punkte Punktzahl: 12,5 von maximal 225

Sie können hier den Prüfungsteil Schriftlicher Ausdruck üben. Ihr Brief wird automatisch bewertet. Sie können nicht sicher sein, dass menschliche Korrektoren gleich bewerten.

Liebe(r) ...
Bei uns ist alles in Ordnung.
Ich hoffe, es geht dir genau so gut wie mir.
Ich habe meine Abschlussprüfung bestanden. Das ist ein tolles Gefühl!
Wenn du nach Deutschland kommst, würde ich dich gern zu uns in den Schwarzwald einladen.
Da können wir allerhand gemeinsam unternehmen. Überleg dir schon: Wozu hättest du Lust?
Hier gibt es viele Möglichkeiten. Wenn du möchtest, kannst du gern jemanden mitbringen. Wir haben viel Platz hier.
Ich würde mich sehr, sehr freuen, wenn wir uns sehen könnten.
Bis dahin mit herzlichen Grüßen

Deine Claudia

Anrede

Datum

Abschied

Unterschrift

Sie würden die Einladung gern annehmen. Beantworten Sie den Brief an Claudia.
Gehen Sie auf folgende Inhaltspunkte ein:

Wie ist das Wetter? Welche Kleidung brauchen Sie?
Was sie unternehmen könnten Ob Sie jemanden mitbringen können

< >

Abbildung 5: Typ 4: Halbfreie Textproduktion (Schreiben)

Dieser Itemtyp ist eine Besonderheit. Freie Textproduktion kann nicht durch vorgegebene Punktezuordnung bewertet werden. Darum werden hier menschliche Rater eingesetzt, in der Regel zwei, um eine gewisse Reliabilität zu erreichen. Wie weit das gelingt, ist eine andere Frage.

Eine Besonderheit dieses Simulators ist, dass die Textproduktion des Prüflings automatisch bewertet wird und das Ergebnis ihm direkt angezeigt wird. Das Verfahren werde ich im Folgenden genauer erklären, weil es eine Innovation ist und in anderen elektronischen Prüfungen bisher nicht realisiert wird.

Wenn der Prüfling die Prüfung beendet hat, bekommt er einen ausführlichen Bewertungsscreen mit den für jeden Teil erreichten Punktzahlen und eine Einschätzung seiner Leistung. Insbesondere natürlich, ob sie bestanden hätte oder nicht.

2. Grundlagen der automatischen Textbewertung

2.1. Ausgangslage und Zielsetzung

Die Bewertung schriftlicher Produktionen allgemein, wie die der Produktionen von DaF-Lernern ist ein ewiges Problem. Eigentlich müssten wir hier wie für alle Tests die Erfüllung der Kriterien verlangen:

- Validität
- Objektivität
- Reliabilität

Das Ziel unseres Projekts war, mittels textueller Parameter einen Score zu erzeugen, der möglichst hoch mit der Bewertung durch menschliche Rater korreliert. Um ein möglichst robustes Verfahren zu entwickeln, gingen wir rein statistisch vor.

Die Idee eines Scorings auf statistischer Basis ist verwandt mit der Konstruktion von Intelligenztests. Intelligenz wird hierbei als ein undefiniertes Konstrukt gesehen. Um Intelligenz zu testen, werden Items entwickelt, die keinen inhaltlichen Bezug zu overten Intelligenzleistungen haben müssen. Alles, was am Ende zählt, ist die hohe Korrelation mit dem Außenkriterium, in unserem Fall die Scorings menschlicher Rater. Selbstverständlich würde für andere Nutzungsarten die face validity eine größere Rolle spielen, etwa wenn ein System genutzt werden sollte zur Verbesserung der Schreibfähigkeiten.

Einfache textuelle Parameter zur Messung der Textqualität werden seit langem verwendet im Korrekturalltag und etwas präziser in der Verständlichkeitsforschung. Sie sind natürlich alle umstritten, wurden immer wieder variiert oder durch neue, in neuer theoretischer Tradition begründete ersetzt. In der Regel sind diese Parameter so ausgelegt, dass ihre hohe Ausprägung für Schwerverständlichkeit des Textes spricht. Dies kehrten wir tentativ um: Was schwerer verständlich ist, zeigt ein höheres sprachliches Niveau.

2.2. Beurteilungskriterien üblicher Praxis

Zur Bewertung von Texten wurden für menschliche Rater Kriterien und Dimensionen festgelegt und Explikationen in Skalen versucht.

Dieses Verfahren ist weithin üblich und beruht auf der verbreiteten Ansicht: Je differenzierter die Anleitung, umso besser. Es stößt aber an die Grenzen der intellektuellen Fähigkeiten von Ratern (Komplexität).

Als Beispiel die Kriterien für den Bereich der formalen Korrektheit.

Formale Richtigkeit	
Bewertet werden Syntax, Morphologie (und Orthografie/ Interpunktion). Zur Bewertung werden die Punkte wie folgt vergeben:	
5 Punkte	Keine bzw. nur vereinzelte Fehler in Syntax, Morphologie (und Orthografie/ Interpunktion)
4 Punkte	Einige Fehler in Syntax, Morphologie (und Orthografie/ Interpunktion), die jedoch das Verständnis nicht stören.
3 Punkte	Einige Fehler in Syntax, Morphologie (und Orthografie/ Interpunktion), die das Verständnis nur wenig beeinträchtigen.
2 Punkte	An mehreren Stellen beeinträchtigen die Fehler in Syntax, Morphologie (und Orthografie/ Interpunktion) das Verständnis erheblich.
1 Punkt	An vielen Stellen beeinträchtigen die Fehler in Syntax, Morphologie (und Orthografie/ Interpunktion) das Verständnis erheblich.
0 Punkte	So viele Fehler in Syntax, Morphologie (und Orthografie/ Interpunktion), dass der Brief kaum noch verständlich ist.

Abbildung 6: Kriterien für den Bereich der formalen Korrektheit

Skepsis ist vor allem angebracht, was die hohe Differenziertheit betrifft und die simplen Graduierungen: *Viele, mehrere, vereinzelte; wenig, kaum, erheblich*.

Menschliche Rater erscheinen erst einmal als Maß aller Dinge. Allerdings verdunkelt sich das Bild schnell. So kann man leicht überprüfen, dass menschliche Rater recht subjektiv urteilen, indem man den gleichen Text von mehreren Ratern beurteilen lässt.

Leider ist es mit der Validität aber auch nicht so weit her. Die folgende Grafik zeigt die Bewertung von 169 Brieftexten mit einer Höchstpunktzahl von 10.

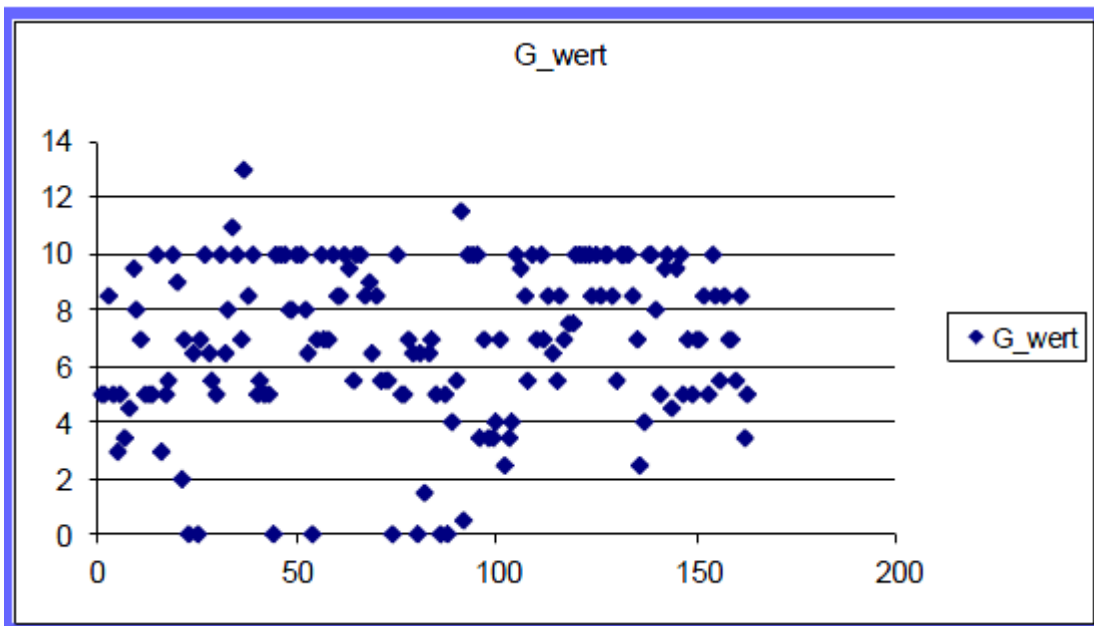


Abbildung 7: Bewertung von 169 Brieftexten

Die Ausreißer über die Höchstpunktzahl 10 hinaus wird man wohl als menschliche Irrtümer sehen müssen, aber immerhin drei an der Zahl und als Schnitt von zwei Ratern. Verblüffender erscheint die Massierung oben, besonders bei 10 Punkten und andererseits die dünne Besiedlung unten. Etwas mehr Streuung und wohl auch Trennschärfe wäre wohl doch nach der anzunehmenden Kompetenzverteilung der Schreiber zu erwarten.

Den Problemen menschlicher Rater sucht man Herr zu werden durch

- Einsatz mehrerer Rater,
- differenzierteres Rating in mehreren Dimensionen,
- Schulung der Rater.

2.3. Die untersuchten Parameter

In unserem Projekt haben wir viele Parameter getestet, die nach textlinguistischen Kriterien erfolgversprechend schienen. Die folgenden wurden am Ende verwendet. Man erkennt auf Anhieb, dass menschliche Rater zu einer so differenzierten und empirisch begründbaren Bewertung nicht in der Lage sein können:

- **Textlänge**

Anzahl aller Wortformenvorkommen des Textes

- **Wiederholrate**

Quotient aus Anzahl der Vorkommen aller Wortformen und Anzahl aller Wortformtypes (= token/ type ratio)

Menschliche Rater nutzen das Kriterium intuitiv, werden den Parameter wohl kaum verlässlich auswerten können. Da ist die Maschine wesentlich verlässlicher.

- **Fehler**

Zahl inkorrektur Wortformen(vorkommen)

Ein wichtiger Parameter ist die Ermittlung von Fehlern. Fehler sind für sich ein weites Feld. Schon die Definition, was ein Fehler ist, bleibt umstritten. Eine flexible und angemessene Fehlertypologie ist nach wie vor ein Desiderat.

Darum die Beschränkung: Falsche Wortformen sind zu ermitteln, wie dies ein Rechtschreibprüfprogramm tut.

- **W_Komplexität**

Durchschnittliche Buchstabenanzahl aller Wortformen

Die Verwendung komplexer Wörter wird hier rein an der Wortlänge gemessen. Dies ist ein sicherer und stabiler Parameter. Er ist in der Verständlichkeitsforschung immer wieder in unterschiedlichen Formeln, unterschiedlichem Gewand und unterschiedlicher Gewichtung im Gebrauch.

- **Lexikalische Komplexität**

Quotient aus Anzahl der Inhaltswörterwortformenokkurrenzen mit Buchstabenanzahl > 8 und Anzahl aller Inhaltswörterwortformenokkurrenzen

Der Parameter ist eng verwandt mit der W_Komplexität, bewertet aber nur lange Wörter.

- **Morphologische Tiefe**

Durchschnittlicher Rang der 30 ranghöchsten korrekten Wortformen

Der Parameter gleicht alle Wortformen des Textes ab mit einer Tabelle, in der deutsche Wortformen nach Frequenz geordnet und invers mit Rangwerten versehen sind. Es geht also darum zu bewerten: Wie selten sind die Wortformen, die der Proband benutzt (und

also benutzen kann) Die Beschränkung auf die 30 seltensten ist willkürlich, könnte aber zur empirischen Anpassung variiert werden.

- **Lexikalische Tiefe**

Durchschnittlicher Rang der 20 ranghöchsten Lemmas

Der Parameter gleicht nach der Lemmatisierung alle Lemmas des Textes ab mit einer Tabelle, in der deutsche Wörter in der Grundform nach Frequenz geordnet und umgekehrt mit Rangwerten versehen sind. Es geht also darum zu bewerten: Wie selten sind die Wörter, die der Proband benutzt ((und) also benutzen kann)? Die Beschränkung auf die 20 seltensten ist willkürlich, könnte aber zur empirischen Anpassung variiert werden.

- **Lexikalische Elaboriertheit**

Durchschnitt alle Lemma-Ränge

Der Parameter gleicht alle Lemmas ab mit einer Tabelle, in der deutsche Wörter in der Grundform nach Frequenz geordnet und umgekehrt mit Rangwerten versehen sind. Es geht also darum zu bewerten: Wie selten sind die Wörter, die der Proband benutzt?

- **Lexikalische Varianz**

Quotient aus Anzahl Lemmas der Inhaltswörter und Anzahl Lemmas der Inhaltswörterokkurrenzen

Mit diesem Parameter soll überprüft werden, wie oft der Proband das gleiche Wort in unterschiedlichen Formen verwendet, landläufig gesprochen wiederholt.

- **Lexikalische Breite**

Quotient aus Anzahl Lemmas der Inhaltswörter und Anzahl der Inhaltswörterwortformen

- **Lexikalische Ladung**

Quotient aus Anzahl Lemmas der Inhaltswörterwortformenokkurrenzen und Anzahl der Funktionswörterokkurrenzen

Hiermit soll ermittelt werden, in welchem Verhältnis sozusagen inhaltsarme und somit auch hochfrequente Funktionswörter im Vergleich zu eher dicken Inhaltswörtern stehen. Dem liegt eine sicherlich diskutabile Kategorisierung zu Grunde.

- **Kohäsion**

Quotient aus Anzahl Konnektoren und Textlänge

Die Textlogik wird in Textkorrekturen oft als Kriterium ausgewertet. Sie äußert sich natürlich in Kohäsion und damit in vorkommenden Konnektoren.

- **K_rang**

Relative Häufigkeit von Kollokationspaaren (Bigrammen)

Der Probandentext wird fortlaufend in Paare der Wortformenokkurrenzen zerlegt. Die Paare werden abgeglichen mit einer Tabelle solcher Pärchen, die mit Frequenzen aus einem großen Korpus (Textwörter $n = 10^8$) versehen sind.

- **TR_rang**

Relative Häufigkeit von Kollokationstripeln (Trigrammen)

Der Probandentext wird fortlaufend in Tripel der Wortformenokkurrenzen zerlegt. Die Tripel werden abgeglichen mit einer Tabelle solcher Tripel, die mit Frequenzen versehen sind. Der Parameter kann unterschiedlich ausgewertet werden: Kommt das Trigramm in der Tabelle überhaupt vor? Wie frequent ist es?

Alles in allem ist zu unterscheiden:

- Parameter, die rein lokal auf dem Text operieren (Textlänge, Wiederholrate, W_Komplexität, Gliederung)
- Parameter, die auf Merkmale des Deutschen Bezug nehmen, die in Datenbanken gespeichert sind (Fehler, Lexikalische Komplexität, Morphologische Tiefe, Lexikalische Tiefe, Lexikalische Breite, Lexikalische Ladung, Subordinationstiefe, Kohäsion, K_rang usw.).

Die ersteren sind technisch attraktiver, weil unabhängiger von eventuell unsicherer empirischer Basis. Hingegen nehmen sie nicht direkt aufs Deutsche Bezug.

2.4. Durchführung und Ergebnis

Die folgenden Thesen bildeten unseren Ausgangspunkt:

- Je größer die Wiederholrate, umso schlechter der Text
- Je mehr Fehler, umso schlechter der Text
- Je größer W_Komplexität, umso elaborierter der Text

- Je größer Lexikalische Komplexität, umso elaborierter der Text
- Je größer Morphologische Tiefe, umso elaborierter der Text
- Je größer Lexikalische Tiefe, umso elaborierter der Text
- Je größer Lexikalische Elaboriertheit, umso elaborierter der Text
- Je größer Lexikalische Varianz, umso elaborierter der Text
- Je größer Lexikalische Breite, umso elaborierter der Text
- Je größer Lexikalische Ladung, umso elaborierter der Text
- Je größer Subordinationstiefe, umso elaborierter der Text
- Je größer Gliederung, umso elaborierter der Text
- Je größer Kohäsion, umso elaborierter der Text
- Je größer Konnektorvarianz, umso elaborierter der Text
- Je größer K_rang, umso elaborierter der Text
- Je größer TR_rang, umso elaborierter der Text

Wichtig war nun,

- welche Parameter überhaupt auf die Scores der menschlichen Rater laden,
- welche Parameter untereinander hoch korrelieren.

Untereinander hochkorrelierende Parameter sind reduzierbar, ausgewertet wurden die besonders ergiebigen Parameter:

- Wiederholrate
- Fehler
- Lexikalische Komplexität
- K_rang
- TR_rang

Mit unterschiedlichen Gewichtungen dieser Parameter wurde eine Korrelation von 0.68 zu den menschlichen Ratern erzielt. Menschliche Rater korrelieren untereinander gewöhnlich nur mit 0.66. Zuletzt gelang es sogar, zu geschulten Superratern eine Korrelation von 0.78 zu erreichen. Das scheint in jeder Hinsicht befriedigend.

Vielleicht ist aber die automatische Bewertung einfach besser als menschliche Rater. Auf jeden Fall ist sie transparent und empirisch reliabel. Diese Einsicht durchzusetzen ist aber eher undenkbar.

Quellenangaben



Heringer, Hans-Jürgen (2015): Goethe Zertifikat Simulator. Android-Version unter <https://play.google.com/store/search?q=goethe%20zertifikat%20simulator>



Maenner, Dieter (2008) Deutsch Prüfungstraining: Zertifikat Deutsch. Berlin: Cornelsen.



Heringer, Hans-Jürgen, Maenner, Dieter (2011) Deutsch Prüfungstraining: Deutsch Test für Zuwanderer. Berlin: Cornelsen.

Biographische Angaben

Hans Jürgen Heringer (hans-juergen@heringer.net) ist emeritierter Professor. Seit 1981 war er an der Universität Augsburg tätig, seine Arbeitsschwerpunkte sind formale Theorien der Sprache, vor allem Syntax und Semantik aber auch Didaktik der Deutschen Sprache sowie des Deutschen als Zweit- und Fremdsprache. Heringer lebt in Herrsching am Ammersee. Weitergehende Informationen zu Leben und Werk sind zu finden unter https://de.wikipedia.org/wiki/Hans_J%C3%BCrgen_Heringer.