IQ-Tests für Texte : Computergestützte Verständlichkeitsmessung

Autor(en): **Kercher, Jan**

Objekttyp: **Article**

Zeitschrift: **Sprachspiegel : Zweimonatsschrift**

Band (Jahr): **71 (2015)**

Heft 1

PDF erstellt am: **24.05.2024**

Persistenter Link: https://doi.org/10.5169/seals-587129

Nutzungsbedingungen

Die ETH-Bibliothek ist Anbieterin der digitalisierten Zeitschriften. Sie besitzt keine Urheberrechte an den Inhalten der Zeitschriften. Die Rechte liegen in der Regel bei den Herausgebern.
Die auf der Plattform e-periodica veröffentlichten Dokumente stehen für nicht-kommerzielle Zwecke in Lehre und Forschung sowie für die private Nutzung frei zur Verfügung. Einzelne Dateien oder Ausdrucke aus diesem Angebot können zusammen mit diesen Nutzungsbedingungen und den korrekten Herkunftsbezeichnungen weitergegeben werden.
Das Veröffentlichen von Bildern in Print- und Online-Publikationen ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Die systematische Speicherung von Teilen des elektronischen Angebots auf anderen Servern bedarf ebenfalls des schriftlichen Einverständnisses der Rechteinhaber.

Haftungsausschluss

Alle Angaben erfolgen ohne Gewähr für Vollständigkeit oder Richtigkeit. Es wird keine Haftung übernommen für Schäden durch die Verwendung von Informationen aus diesem Online-Angebot oder durch das Fehlen von Informationen. Dies gilt auch für Inhalte Dritter, die über dieses Angebot zugänglich sind.

Ein Dienst der *ETH-Bibliothek*
ETH Zürich, Rämistrasse 101, 8092 Zürich, Schweiz, www.library.ethz.ch

http://www.e-periodica.ch

IQ-Tests für Texte

Computergestützte Verständlichkeitsmessung

Von Jan Kercher[1]

Immer wieder werde ich gefragt, ob man mit Computerprogrammen wirklich wissenschaftlich und objektiv die Verständlichkeit von Texten messen kann. Die Antwort darauf ist relativ einfach: Ja, das kann man – wenn man bereit ist, Textverständlichkeit ähnlich zu behandeln, wie man Intelligenz bei IQ-Tests behandelt. Was bedeutet das?

Sowohl Verständlichkeit als auch Intelligenz sind äusserst komplexe Phänomene: Sie werden von einer Vielzahl unterschiedlichster Faktoren beeinflusst, was letztlich zu einer fast unüberschaubaren Vielzahl von Ausprägungen der Verständlichkeit bzw. der Intelligenz führt. Oder anders ausgedrückt: Streng genommen gibt es so viele «Intelligenzen», wie es Lebewesen gibt, und so viele «Verständlichkeiten», wie es Texte gibt.

Wenn man sich mit solchen Phänomenen näher befasst, stellt man deshalb schnell fest, dass es kaum möglich ist, sie einerseits in ihrer gesamten Komplexität und Individualität zu erfassen und andererseits objektiv und praktikabel zu messen. Denn es handelt sich um einen klassischen Zielkonflikt: Je eher die Komplexität des Phänomens erfasst wird, desto schwieriger wird eine objektive und praktikable Messung. Und je objektiver und praktikabler das Messinstrument, desto schwieriger wird eine Erfassung der Komplexität des Phänomens.

Um diesen Konflikt aufzulösen, müssen demnach zwangsläufig bei der Erfassung der Komplexität oder bei der Objektivität und

1 Dr. Jan Kercher, Lehrbeauftragter an der Universität Hohenheim (D), Dissertation zum Thema «Verstehen und Verständlichkeit von Politikersprache» sowie zahlreiche wissenschaftliche Aufsätze in diesem Bereich – *jankercher@gmail.com*
Der vorliegende Text erschien zuerst im Blog «Sprachlupe» des Communication Lab, Institut für Verständlichkeit, Ulm (*http://blog.comlab-ulm.de* – direkt: *http://tinyurl.com/kercher-iq*). Die Zwischentitel hat der «Sprachspiegel» eingefügt.

Praktikabilität der Messung Abstriche gemacht werden. Sowohl die Intelligenzmessung durch IQ-Tests als auch die computergestützte Verständlichkeitsmessung durch Instrumente wie den Hohenheimer Verständlichkeitsindex setzen auf Abstriche bei der Komplexitätserfassung zugunsten einer möglichst maximalen Messobjektivität und einer hohen Praktikabilität. Dadurch unterscheiden sie sich von alternativen Messansätzen wie beispielsweise dem Hamburger Verständlichkeitskonzept oder dem Verständlichkeitsmodell von Groeben, die die Komplexität des Verständlichkeitsphänomens angemessener erfassen wollen, dafür aber zwangsläufig Abstriche bei der Objektivität und Praktikabilität hinnehmen müssen.

Lieber messbar als vielschichtig

Der Grund für diese Prioritätensetzung ist der praktische Anwendungsbezug von IQ-Tests und computergestützter Verständlichkeitsmessung: Nur eine maximale Messobjektivität und Messökonomie ermöglichen ein Messinstrument, dessen Einsatz einen vertretbaren Aufwand mit sich bringt und dessen Ergebnisse nicht schon allein aufgrund von Zweifeln an seiner Objektivität abgelehnt werden können. Denn genau dies sind die Schwächen von alternativen Messinstrumenten wie dem Hamburger Verständlichkeitskonzept: Der Messaufwand ist ungleich höher und das Ergebnis ist stets angreifbar, da es stark von der Subjektivität der jeweiligen Tester beeinflusst wird. Hinzu kommt: Solche Verfahren können – im Gegensatz zu computergestützten Messverfahren – nur von geschulten Experten durchgeführt werden.

Eine weitere Gemeinsamkeit verbindet die computergestützte Verständlichkeitsmessung und IQ-Tests: Beide haben ihre Ursprünge im frühen 20. Jahrhundert und können dadurch auf eine etwa 100-jährige Forschungsgeschichte zurückblicken. 1912 erfand William Stern den Intelligenzquotienten, der 1932 von David Wechsler zum heute gängigen IQ-Testverfahren weiterentwickelt wurde. Zu einer ähnlichen Zeit entwickelten sich in den USA die ersten Verständlichkeitsformeln, die als Vorläufer von Instrumenten wie dem Hohenheimer Verständlichkeitsindex angesehen werden können.

Redlichkeit wird vorausgesetzt

Was allerdings nicht verschwiegen werden sollte: Die beiden Messverfahren teilen nicht nur ihre wichtigsten Stärken (Objektivität, Praktikabilität), sondern auch ihre wichtigste Schwäche. Sowohl IQ-Tests als auch Verständlichkeitsindizes können durch gezielte Manipulation verzerrt bzw. «ausgetrickst» werden. So lassen sich die Ergebnisse bei IQ-Tests durch gezieltes Training verbessern, ohne dass dadurch zwangsläufig auch die tatsächliche Intelligenz steigt. Ebenso lassen sich Texte so manipulieren, dass Computerprogramme höhere Verständlichkeitswerte ermitteln, ohne dass die Verständlichkeit des Textes wirklich gesteigert wurde (z. B. indem an willkürlichen Stellen zusätzliche Punkte in die Sätze eingefügt werden).

Mit anderen Worten: Die Aussagekraft beider Verfahren beruht auf der Redlichkeit ihrer Anwender. Kann diese vorausgesetzt werden, stellt die Schwäche beider Ansätze kein entscheidendes Problem mehr dar. Noch etwas deutlicher wird dieser Befund, wenn man ihn mit einer Temperaturmessung per Thermometer vergleicht: Hält man während der Ermittlung der Raumtemperatur ein Feuerzeug unter das Thermometer, so wird die Messung zwangsläufig kein sinnvolles Ergebnis ermitteln. Dies ist allerdings nicht auf die mangelnde Messgenauigkeit des Thermometers zurückzuführen, sondern auf die gezielte Manipulation desjenigen, der das Feuerzeug hält.

Doch zurück zur Verständlichkeitsmessung per Computerprogramm: Wie funktioniert diese nun genau? Ganz ähnlich wie beim IQ-Test werden zunächst Merkmale des interessierenden Phänomens ermittelt, die sich objektiv messen bzw. in Zahlen ausdrücken (quantifizieren) lassen. Beispiele hierfür sind die durchschnittliche Satzlänge eines Textes in Wörtern, die durchschnittliche Wortlänge in Silben oder auch der Anteil an Fremdwörtern bzw. seltenen Wörtern. Zu Beginn der Verständlichkeitsforschung mussten diese Merkmale noch vergleichsweise langwierig von Hand ausgezählt werden, mittlerweile stehen hierfür Computerprogramme wie TextLab zur Verfügung, die diese lästige Zählarbeit in Bruchteilen von Sekunden erledigen. Die

quantifizierbaren Textmerkmale werden in einem zweiten Schritt in Beziehung zu Daten gesetzt, anhand derer man ihre Eignung für eine Verständlichkeitsmessung beurteilen kann. Beispiele hierfür sind die durchschnittliche Lesegeschwindigkeit von Probanden oder die Ergebnisse von Verständnisfragen zu einem Text. Auf dieser Grundlage lassen sich nun sogenannte Verständlichkeitsformeln oder Verständlichkeitsindizes konstruieren, die anhand eines bestimmten Massstabs oder «Benchmarks» einen Textverständlichkeitswert ermitteln.

Von «Bild-Zeitung» bis Dissertation

Nimmt man als Beispiel für solch ein Messinstrument den Hohenheimer Verständlichkeitsindex, so werden hier als Massstab zwei sehr unterschiedliche Arten von Texten herangezogen: Auf der einen Seite Artikel aus dem Politikteil der «Bild-Zeitung» (als Massstab für leicht verständliche Informationstexte), auf der anderen Seite Texte aus politikwissenschaftlichen Doktorarbeiten (als Massstab für schwer verständliche Informationstexte). Anhand eines Abgleichs mit den quantifizierbaren Textmerkmalen dieser Textgattungen lässt sich nun jeder beliebige andere Informationstext auf einer Skala von 0 (sehr unverständlich) bis 20 (sehr verständlich) einordnen. Ein Wert von 0 bis 5 Punkten entspricht hierbei der Verständlichkeit von politikwissenschaftlichen Doktorarbeiten, ein Wert von 15 bis 20 Punkten der Verständlichkeit von «Bild»-Artikeln. Texte auf einer Nachrichtenseite wie «Spiegel Online» liegen bei etwa 12 Punkten. Mit anderen Worten: Jeder Anwender kann sich selbst aussuchen, welche Textgattung er als Massstab für die Bewertung und Optimierung der eigenen Texte heranziehen möchte.

Auch hierin liegt demnach eine weitere Gemeinsamkeit von IQ-Tests und computergestützter Verständlichkeitsmessung: Man muss nicht zwangsläufig einen maximalen Wert anstreben. Häufig reicht auch ein niedrigerer bzw. mittlerer Wert, um die jeweils angestrebten (Kommunikations-)Ziele zu erreichen. Oder um es anders auszudrücken: Es muss nicht unbedingt immer der IQ von Leonardo da Vinci sein, der Schätzungen gemäss bei bis zu 220 Punkten lag. Häufig reicht auch der IQ von Bill Clinton (138 Punkte).