
Ursula Christmann

Methoden zur Erfassung von Literalität

1 Einführung und Überblick

Es dürfte inzwischen Allgemeingut sein, dass Lesen und Leseverstehen kein einheitlicher Prozess ist, sondern üblicherweise nach der Komplexität der beteiligten Teilprozesse in mehrere Ebenen aufgliedert wird, die von der Buchstaben- und Wortidentifikation über die semantische und syntaktische Analyse von Wortfolgen und Sätzen bis hin zu Prozessen des Aufbaus einer kohärenten Textbedeutungsstruktur, dem Erkennen der Autorintention und der Bewertung des Gelesenen reichen. Zur Erforschung dieser Teilprozesse steht heute eine Fülle unterschiedlichster Methoden zur Verfügung. Welche Methode zum Einsatz kommt, hängt in erster Linie von dem Erkenntnisinteresse und der Korngröße der Analyse ab. Auf hohem Abstraktionsniveau lassen sich die verschiedenen Verfahren danach unterscheiden, ob der Verstehensprozesses oder das Verstehensprodukt beleuchtet werden soll.

Die prozessorientierten Methoden (Abschnitt 2) werden primär in der Grundlagenforschung zum Sprach- und Leseverstehen eingesetzt. Dazu gehört insbesondere die Erfassung von Lesezeiten, Blickbewegungen sowie Reaktionszeiten. Im Mittelpunkt des Beitrags steht jedoch die breite Palette der produktorientierten Methoden (vom einfachen Lückentest bis hin zu sophistizierten Wiedergabeverfahren), die für die Anwendungsforschung interessanter sein dürften. Dabei unterscheide ich nach dem Grad der Komplexität zwischen der Erhebung textnaher (Abschnitt 3) und textferner Verstehensprodukte (Abschnitt 4). Abschließend werde ich Methoden zur Erfassung von empirisch gesicherten Korrelaten des Lese- und Textverstehens (Abschnitt 5) ansprechen.

2 Erfassung des Verstehensprozesses

Bei der Erforschung des Verstehensprozesses werden bevorzugt chronometrische Methoden eingesetzt (historischer Überblick: Mitchell 2004). Sie basieren auf dem Gedanken, dass Verstehen ein Vorgang ist, der sich in der Zeit entfaltet. Lange Verarbeitungszeiten sind dementsprechend immer ein Indikator für einen hohen Verarbeitungsaufwand. Der Verarbeitungsaufwand wiederum hängt von der Zugänglichkeit und damit dem Aktivierungsgrad der mentalen Repräsentation ab, die für das Verstehen relevant ist. Es wird also von der gemessenen Zeit auf die für das Verstehen notwendige Aktivierung von Konzepten geschlossen (vgl. Haberlandt 1994). Die verfügbaren Online-Methoden werden häufig verwendet, wenn es um Fra-

181

gen der Worterkennung und des lexikalischen Zugriffs geht, sie können und werden jedoch auch auf Satz- und zunehmend auf Textebene eingesetzt.

Lesezeiten

Die Messung von Lesezeiten ist eine weit verbreitete Methode zur Online-Erfassung des Leseverstehens. Sie beruht auf der Annahme, dass die gemessene Zeit ein Maß für den kognitiven Aufwand beim Lesen darstellt. Als valideste Methode zur Erfassung von Lesezeiten gilt die Moving-Window-Technique (vgl. Hemforth 2006). Dabei wird den Versuchspersonen der zu lesende Text portionsweise auf dem Bildschirm geboten. Auf Tastendruck verschwindet der gelesene Satz und der nächste erscheint auf dem Bildschirm.

Im Prinzip handelt es sich dabei um ein einfaches und ökonomisches Verfahren, das bei vielen Textverarbeitungsexperimenten allein oder in Kombination mit anderen Methoden eingesetzt wird. Allerdings ist zu beachten, dass Lesezeiten nicht notwendigerweise Auskunft darüber geben können, ob und in welcher Weise der dargebotene Text verstanden wurde. Zur Kontrolle des Verstehens ist der Einsatz zusätzlicher Techniken, wie z. B. die Online-Beantwortung von Fragen zum Text notwendig.

Auch kann es eine Reihe von Störeffekten geben, die die Interpretation der erhaltenen Daten erschweren oder sogar verfälschen können. So kommt es beispielsweise entscheidend darauf an, an welchem Punkt des Verstehensprozesses die Lesezeiten erfasst werden, nach einem Wort, einer Phrase, einem Satz. So lassen sich Unterschiede in der Lesezeit eines Ausdrucks in einem wörtlichen gegenüber einem metaphorischen Kontext nur dann nachweisen, wenn die Lesezeiten phrasenweise und nicht nach dem gesamten metaphorischen Ausdruck erfasst werden (Janus, Bever 1985)

Blickbewegungsmessung

Eine genauere Erfassung des Leseprozesses erlaubt die Blickbewegungsmessung. Im Prinzip handelt es sich dabei um eine relativ alte Methode, die bereits zu Anfang des 20. Jahrhunderts eingesetzt wurde (vgl. Radach 1996). In den letzten 20 Jahren hat sie, auch auf Grund erheblicher Verbesserungen von Aufnahme- und Auswertungstechniken, einen immensen Aufschwung erlebt. Allerdings erfordern Blickbewegungsmessungen eine technisch aufwändige (und nicht ganz billige) Laborausstattung.

Wie funktioniert das Ganze? Die Blickbewegungsforschung nutzt die Tatsache, dass das Auge während des Lesens nicht kontinuierlich Zeile für Zeile über den Text gleitet, sondern für Bruchteile von Sekunden verharret, bevor es zum nächsten Haltepunkt springt. Die Haltepunkte werden als *Fixationen*, die Sprünge als *Sakkaden* bezeichnet. Wird im Text zurückgegangen, treten Rücksprünge oder sog. *Regressionen* auf. Die Verarbeitung visueller Zeichen erfolgt nur während der Fixationen.

Fixationen, Sakkaden und Regressionen sind die zentralen Größen mit denen die Blickbewegungsmessung arbeitet (Überblick: Günther 1989; Rayner 1996; Radach, Kennedy 2004). Aus ihrer Abfolge werden verschiedene Maße abgeleitet. Prominente Maße sind die Fixationsdauer (die Summe aller Fixationen auf ein Wort gilt als Indikator für den kognitiven Verarbeitungsaufwand), die Sakkadenlänge, die Relation zwischen Sakkadenlänge und Fixationsdauer, Regressionen bzw. regressive Sakkaden (als Indikator für Verstehensprobleme; Überblick und operationale Definitionen gebräuchlicher räumlicher und zeitlicher Maße: Radach, Kennedy 2004, 6f.).

Zur Veranschaulichung ein paar Werte: Die Dauer einer Fixation beträgt im Durchschnitt zwischen 200 und 250 ms, die mittlere Sakkadenlänge liegt bei 20 bis 40 ms, pro Sekunde erfolgen 3 bis 4 Sakkaden und 10 bis 15 % der Fixationen bestehen aus Regressionen, die zur Reanalyse problematischer Textstellen genutzt werden. Sämtliche Maße können in Abhängigkeit von der Wortlänge und der -häufigkeit (z. B. lange Wörter werden länger fixiert als kurze, seltene länger als häufige; Rayner, Duffy 1986), dem Satzkontext, der *Textschwierigkeit* (z. B. längere Fixationen, kürzere Sakkaden, mehr Regressionen und kürzere Wahrnehmungsspanne bei schweren Texten; vgl. Rayner et al. 1989), dem *Vorwissen* und den *Zielsetzungen* der Leserinnen und Leser beträchtlich variieren. (Überblick über empirische Befunde: Inhoff, Rayner 1996).

Bei einem typischen Blickbewegungsexperiment werden Wörter oder kurze Textpassagen vorgegeben, die sog. Zielwörter enthalten

Beispielsatz nach Irmen (2007):

„Oft werden Autoren um ihren Beruf beneidet. Nur selten wird dabei berücksichtigt, wie schwer es *diese Männer* haben, ihren Lebensunterhalt zu verdienen.“

„Oft werden Autoren um ihren Beruf beneidet. Nur selten wird dabei berücksichtigt, wie schwer es *diese Frauen* haben, ihren Lebensunterhalt zu verdienen“.

Untersuchungsfrage: Führt die im Text verwendete maskuline Form („Autoren“) zu einer Vorstellung männlicher Personen? Wenn ja, sollte der Ausdruck „diese Frauen“ länger fixiert werden als „diese Männer“.

Ergänzend sei angemerkt, dass Blickbewegungsmessungen auch bei auditiver Sprachverarbeitung möglich sind. In diesem Fall wird die Bedeutung des Satzes in einer bildlichen oder realen Szene dargeboten und es wird geprüft, in welcher Weise das Auge der gesprochenen Sprache folgt (erstmal: Tanenhaus et al. 1995).

Die klassische Blickbewegungsforschung beruht auf zwei Annahmen, der sog. „*immediacy*“- und der „*eye-mind*“-Annahme (Just, Carpenter 1980). Danach wird ein Wort während der Fixation vollständig und unmittelbar verarbeitet und die Blickdauer (Summe der Fixationen auf das Wort) ist ein In-

dikator für die Verarbeitungszeit des betreffenden Wortes. Entsprechend wird angenommen, dass Fixationen und die Verarbeitung des fixierten Wortes im Gehirn eng verknüpft sind. Mittlerweile ist allerdings deutlich geworden, dass diese Annahmen nicht ganz zutreffend sind.

Die Fixationszeit ist nämlich nicht notwendigerweise ein Indikator für die Zeit, die zum Erkennen eines Wortes benötigt wird, wie dies in prominenten Modellen des Lesens (z. B. Just, Carpenter 1980) unterstellt wird. Es ist durchaus möglich, dass das Auge zum nächsten Fixationspunkt springt, bevor die Verarbeitung eines Wortes abgeschlossen ist. Dafür spricht einerseits das Auftreten von Regressionen, andererseits der sog. „*Spillover-Effekt*“, nach dem das auf ein seltenes (lange fixiertes) Wort folgende Wort länger fixiert wird, als wenn es einem geläufigen Wort folgt (Rayner, Duffy 1986). Außerdem weiß man inzwischen, dass entgegen früheren Annahmen die Dauer von Fixationen nicht nur von der Wortidentifikation, sondern auch von der syntaktischen und semantischen Analyse des gesamten Satzes bestimmt wird (einen neueren Überblick über empirische Studien gibt ein Sonderheft des *European Journal of Cognitive Psychology* aus dem Jahre 2004). Demzufolge ist es keineswegs klar, welche kognitiven Prozesse während der Fixation ablaufen. Wie bei allen Reaktionszeitmessungen in der Psychologie besteht auch hier das grundsätzliche Problem, dass nicht sicher von der gemessenen Zeit auf den Ablauf bestimmter kognitiver Prozesse zurückgeschlossen werden kann.

Blickbewegungsmessungen sind bevorzugt im Rahmen von Studien zur Silbenverarbeitung, Wortverarbeitung, Verarbeitung syntaktischer Strukturen, Inferenzen und lexikalischen Ambiguitäten eingesetzt worden. Dabei werden in der Regel nur Sätze oder ganz kurze Texte als Sprachmaterial verwendet. Aktuell beginnt man jedoch auch damit, Blickbewegungsmuster beim Lesen längerer Texte zu untersuchen, und zwar in Abhängigkeit von verschiedenen Leser- resp. Leserinnen-Merkmalen wie Leseperspektive, Lesestrategien, Vorwissen und Arbeitsgedächtniskapazität (z. B. Kaakinen et al. 2003) sowie in Kombination mit anderen Onlineverfahren wie lautes Denken (Kaakinen, Hyönä 2005). So werden z. B. perspektiven-relevante Textinhalte länger fixiert und zugleich besser behalten als perspektiven-irrelevante Inhalte. Der Effekt wird dabei durch die Arbeitsgedächtniskapazität (AGK) moderiert: Personen mit hoher AGK zeigen beim ersten Lesen längere Fixationen, Personen mit niedriger AGK zeigen diese bei den Rückwärtssprüngen (Diskussion der Befunde im Rahmen der Kapazitätstheorie). Bemerkenswert ist darüber hinaus, dass in einer ersten Untersuchung Blickbewegungen beim Lesen zusammen mit Protokollen des lauten Denkens kombiniert erhoben wurden (Kaakinen, Hyönä 2005). Dabei zeigte es sich, dass Denkvorgänge, die eine tiefere Verarbeitung indizieren (Selbst-Erklärungen und Fragen) mit längeren Fixationszeiten einhergehen als Denkvorgänge, die mit einer flachen Verarbeitung verbunden sind (Assoziationen und Paraphrasen).

Die Mehrzahl der Blickbewegungsuntersuchungen wurde mit guten Leserinnen und Lesern durchgeführt. Blickbewegungsstudien mit Kindern und älteren Menschen sind selten. Auch gibt es vergleichsweise wenige Studien, die sich mit interindividuellen Unterschieden beim Lesen befassen (Überblick: Richter, Christmann 2002). Dabei lassen sich zwar durchaus Unterschiede in der Blickbewegung zwischen guten und schlechten Lesern finden (z. B. mehr Sakkaden, mehr Regressionen und längere Fixationen bei schlechten Lesern), die allerdings nicht die Ursache schlechter Leseleistung darstellen.

Priming-Technik

Priming (auch Bahnung oder Voraktivierung) bedeutet, dass die Verarbeitung eines Wortes zu einer Voraktivierung eines anderen nachfolgenden Wortes (Target oder Zielreiz) führt, sofern die beiden Wörter z. B. semantisch aufeinander bezogen, phonetisch ähnlich oder hoch assoziiert sind. In einem typischen priming-Experiment werden kurz hintereinander zwei Wörter geboten (z. B. *Baum—Blüte*), von denen das zweite Wort, also der Zielreiz reaktionsrelevant ist und beurteilt werden muss (z. B. Wort vs. Nicht-Wort, semantische Nähe zum Prime-Wort). In der Kontrollbedingung wird das Prime-Wort durch ein anderes Wort ersetzt (z. B. *Fuchs—Blüte*). Gemessen wird die Reaktionszeit in der Experimental- und der Kontrollbedingung. Die Zeitdifferenz zwischen den beiden Bedingungen stellt den Priming-Effekt dar, der in der Regel umso größer ist, je enger die Beziehung zwischen Prime und Target ist. Die Reaktionszeit gilt dabei als Indikator für den Verarbeitungsaufwand. Variiert wird dabei häufig der zeitliche Abstand zwischen der Darbietung von Prime und Target (= SOA: Stimulus Onset Asynchrony; liegt zwischen 50 und 1300 ms).

So hat man beispielsweise festgestellt, dass bei polysemen Wörtern wie *Bank, Kapelle, Maus* etc. bei allen Bedeutungsvarianten nach kurzen SOAs (bis ca. 500ms) ein Priming Effekt auftritt, dass nach längeren SOAs jedoch nur noch die Bedeutung geprimed wird, die vom Satz-Kontext unterstützt wird (Neely 1977).

Beispiel:

Maria ging im Park spazieren und setzte sich auf eine B a n k ;
Zielreiz: *Geld* vs. *Tisch*

Die übliche Erklärung für den Priming-Effekt ist, dass bei der Verarbeitung des Prime-Wortes nicht nur seine Repräsentation aktiviert wird, sondern auch diejenigen Konzepte, die mit dem Prime-Wort verbunden sind. Priming-Untersuchungen werden bevorzugt auf Wortebene und Satzebene, aber auch auf Textebene durchgeführt.

Lexikalische Entscheidungsaufgaben

Bei lexikalischen Entscheidungsaufgaben liest die Versuchsperson zunächst einen Satz oder einen kurzen Text und muss dann so schnell wie möglich entscheiden, ob ein nachfolgendes Wort ein „Wort“ oder ein „Nicht-Wort“ ist. Gemessen wird die Latenzzeit zwischen dem Einblenden eines Stimulus-Wortes und der Reaktion. Typischerweise sind die Latenzzeiten für vertraute Wörter sowie für Wörter, die semantisch ähnlich zu vorher gelesenen Wörtern sind oder einen Bezug zum Kontext aufweisen, geringer (zusammenfassend: Haberlandt 1994, 17). Das nachfolgende Beispiel verdeutlicht, dass man diese Methode durchaus auch auf Textebene einsetzen kann (Sharkey, Mitchell 1985), z. B. um zu prüfen, ob durch die Beschreibung einer Szene ein Skript aktiviert wird.

„The children’s birthday party was going quite well. They all sat around the table and prepared to sing.“

Entscheidungswort: *candles*.

Die Ergebnisse zeigen, dass skriptrelevante Targets zu schnelleren Entscheidungszeiten führen als skriptirrelevante Targets.

Wiedererkennen (item-recognition)

Beim Wiedererkennungsparadigma wird ein Zielwort während oder kurz nach dem Lesen einer Textpassage eingeblendet. Dabei können positive Targets verwendet werden, die im Text vorkamen, oder negative Targets, die nicht im Text enthalten waren. Geprüft wird die Wiedererkennungslatenz (ja/nein) bei korrekten Antworten, korrekten Zurückweisungen und falschen Alarmen (Haberlandt 1994)

Ein Beispiel für die Verwendung positiver Targets stellt eine Untersuchung zur psychologischen Realität mentaler Modelle von Glenberg et al. (1987) dar, bei der es um die Verarbeitung räumlicher Gegebenheiten ging.

Die Versuchspersonen lasen folgenden Text:

John was preparing for a marathon in August. After doing a few warm-up exercises he put on / took off his sweatshirt and went jogging. He jogged halfway around the lake without too much difficulty. Further along his route, however, John’s muscles began to ache.

Zielwort: *sweatshirt*

Die Reaktionszeiten zeigen, dass auf Objekte, die räumlich mit der Hauptfigur assoziiert sind (hier: *put on his sweatshirt*) schneller reagiert wird als auf solche, die räumlich getrennt sind (hier: *took off his sweatshirt*). Dies wird damit erklärt, dass der Leser den Handlungen der Hauptperson in einem räumlichen Situationsmodell folgt, was als Beleg für den Aufbau eines mentalen Modells gewertet wird.

3 Erhebung textnaher Verstehensprodukte

Für die Anwendung interessanter dürfte die Erfassung von Verstehensprodukten sein. Textnahe Erhebungsmethoden überprüfen das Verstehen in enger Anlehnung an den Text, wie das z. B. bei Reproduktions-, Wiedererkennungs- und Einsetzverfahren der Fall ist. Sie erfassen dabei eher das unmittelbare Textverständnis im Sinne einer Vorstellung über den Sinngehalt, wobei allerdings schwer entscheidbar ist, in welchem Ausmaß Gedächtnis- und Schlussfolgerungsprozesse beteiligt sind und ob eine korrekte Antwort auf dem Verstehen von Inhalten oder auf bloßer Merkfähigkeit beruht.

Lückentest (cloze procedure)

Der klassische Lückentest oder der sog. Cloze-Test basiert auf Einsetz- bzw. Ergänzungsprozessen (ausführlich: Groeben 1982). Dabei wird nach dem klassischen Verfahren von Taylor (1953, 1956) jedes fünfte Wort in einem Text ausgelassen, das dann von den Versuchspersonen einzusetzen ist.

Beim erstmaligen Lesen gilt die Anzahl der richtig eingesetzten Wörter als Maß für die Lesbarkeit des Textes. Gibt man den Lückentest nach dem Lesen eines Textes, erhält man ein Maß für das Textverständnis; mit zunehmender Zeitspanne zwischen Lesen und Testvorgabe wird der Cloze-Test schließlich zu einem Behaltenstest. Beim Einsatz des Verfahrens ist zu berücksichtigen, dass mindestens 20 % der Wörter ausgelassen sein sollten und dass der Text mindestens 250 Wörter lang sein muss. Der Cloze-Test ist methodisch intensiv überprüft worden und weist in der Regel hohe Reliabilitäten und eine hohe Übereinstimmung mit Multiple-Choice-Tests auf. Allerdings ist die Inhaltsvalidität des Verfahrens, d. h. wie gut der Test tatsächlich das Textverständnis misst, nicht ganz unproblematisch. Während Taylor beim Vergleich der Testwerte vor und nach dem Lesen des Textes einen validen Informationszuwachs feststellte, haben Nachuntersuchungen (Coleman, Miller 1968) z. T. keinen signifikanten Unterschied zwischen Cloze-Tests mit und ohne Textkenntnis feststellen können. Vermutlich ist die Kontextdetermination durch den umgebenden Text so hoch, dass auch ohne Textkenntnis durch Schlussfolgerungen hohe Werte im Cloze-Test erreichbar sind. Die Cloze-Werte sind somit kein reines Maß für das Textverständnis, da sie Merkmale und Fähigkeiten der Leserin mit abbilden.

Die Validität der klassischen cloze-procedure lässt sich allerdings erhöhen, wenn man nicht systematisch jedes 5. Wort auslässt, sondern die Auslassungen auf Inhaltswörter beschränkt (empirisch: Tuinman et al. 1975) oder auch statt Wörtern Satzteile bzw. ganze Sätze auslässt.

Das Auslassen von Satzteilen oder ganzen Sätzen, die z. B. strukturelle Information tragen oder eine inhaltliche Schlüsselstellung im Text einnehmen, stellt dabei m. E. eine gute Möglichkeit dar zu prüfen, ob das Wesentliche eines Textes verstanden wurde. Die Auswertung ist allerdings aufwändiger

als beim klassischen Cloze-Test, weil der semantische Gehalt der „Einsetzungen“ auf alle Fälle inhaltsanalytisch auszuwerten ist.

Die Wahl der Auslassungen sollte natürlich immer an der Textsorte, der Textkomplexität und den Lernzielen orientiert sein. Bei Sachtexten ist das Auslassen von sog. Kohärenzindikatoren eine gute Möglichkeit, das Textverständnis zu überprüfen. Geeignet erscheinen mir dabei insbesondere Konnektiva (z. B. kausal, temporal, adversativ, additiv), die Sätze und Teilthemen miteinander verbinden und die sich auch empirisch als verarbeitungswirksam erwiesen haben (z. B. Magliano et al. 1999; Millis et al. 1995; Millis, Magliano 1999). Gelingt es, fehlende Konnektiva in einem Text zu rekonstruieren, dann dürfte dieser auch tiefer verarbeitet worden sein. Geeignet erscheint auch das Auslassen konzeptuell-inhaltlicher Relationen, die zur Verdeutlichung von Zusammenhängen auf globaler Textebene eingesetzt werden (z. B. Konzept—Evidenz; Konzept—Spezifikation, Konzept—Beispiel, Konzept—Begründung, Problem—Lösung, These—Antithese, Frage—Antwort, Grund—Folge etc.; vgl. Dressler 1972; Taxonomie bei Sanders et al. 1992), oder von Topic-Indikatoren bzw. Signalen, mit denen die Stellung eines Themas im Gesamttext verdeutlicht werden soll. Hier bestünde dann die Aufgabe der Leserinnen und Leser darin, Evidenzen, Spezifikationen, Beispiele, Begründungen und Problemlösungen zu rekonstruieren.

C-Tests

Eine Weiterentwicklung der klassischen Cloze-Tests stellen die sog. C-Tests dar, die zuerst von Raatz und Klein-Braley (1981) entwickelt wurden. Im Unterschied zum klassischen Lückentest werden hier nicht ganze Wörter, sondern Teile von Wörtern getilgt, und zwar jeweils die erste oder zweite Hälfte jedes zweiten oder dritten Wortes. Die Tilgung erfolgt ab dem zweiten Satz. In der Regel werden kurze Texte mit ca. 70 Wörtern verwendet, die 20–25 Lücken aufweisen.

C-Tests nutzen die natürliche Redundanz von Sprache auf Wort-, Satz- und Textebene und reduzieren diese; je besser die Sprachkompetenz, desto besser können Rezipienteninnen und Rezipienten solche Redundanzen nutzen, um Wortlücken zu rekonstruieren. Ein C-Test besteht aus 4–5 inhaltlich möglichst unterschiedlichen Texten, die von den Versuchspersonen zu rekonstruieren sind (Bearbeitungszeit: 5 Minuten pro Subtest). Der Gesamtpunktwert ergibt sich aus der Anzahl der richtig rekonstruierten Lücken. C-Tests messen die allgemeine Sprachkompetenz in einer Zielsprache im Vergleich zum Gruppendurchschnitt oder im Vergleich zu einem Muttersprachler bzw. einer Muttersprachlerin. Man kann sie sowohl normorientiert als auch kriteriumsorientiert einsetzen. Sie gelten als reliabel, valide, objektiv und einfach in der Entwicklung, Durchführung und Auswertung. Außerdem korrelieren sie in der Regel hoch mit dem Lese- und Hörverstehen. Sie werden in der Grundschule ebenso eingesetzt wie auf fortgeschrittenem Niveau, wie z.

B. zur Testung der Sprachkompetenz ausländischer Studierender an Universitäten (TestDaf = Test Deutsch als Fremdsprache; Grotjahn 2002; 2006).

Eine Variante der C-Tests sind die sog. TF-Tests (Teilfähigkeitstest; Baur, Spettmann 2005). Während C-Tests die allgemeine Sprachfähigkeit messen, kann mit TF-Tests die Kompetenz in Einzelbereichen von Lexik und Grammatik geprüft werden. Bei einer Prüfung der lexikalischen Kompetenz werden die Stamm-Morpheme der thematischen Wörter getilgt, bei einer Überprüfung der grammatikalischen Kompetenz morphologische Indikatoren wie Artikel, Pronomina, Verbformen etc.

C-Tests werden meist zur Erfassung der Sprachbeherrschung in einer Fremdsprache eingesetzt, man kann sie aber auch als muttersprachliche Tests konzipieren (Wockenfuß, Ratz 2006). In diesem Fall werden die Tests mit zunehmendem Alter immer leichter, bis alle Lücken richtig ergänzt werden. Möglicherweise könnte man dies dadurch kompensieren, dass zunehmend restriktivere Zeitbeschränkungen bei der Testbearbeitung eingeführt werden. Entsprechende Studien fehlen allerdings noch.

Gebundene Textreproduktion (Cued Recall)

Bei der gebundenen Textreproduktion (cued recall) wird nach der Textrezeption ein Wort vorgegeben mit der Aufgabe, denjenigen Satz des Textes zu reproduzieren, in dem das Wort vorgekommen ist. Die Wörter stellen sozusagen Gedächtnishilfen dar, die den Abruf der verarbeiteten Informationen erleichtern sollen. Das Verfahren ist eher für kurze Texte geeignet und misst das Textverständnis nur sehr punktuell. Andererseits stellt es ein sehr ökonomisches Verfahren dar, weil die Reproduktion auf diejenige Textstelle beschränkt werden kann, die besonders interessant ist.

Rekognitions- bzw. Wiedererkennungsverfahren

Bei Wiedererkennungsverfahren erhalten die Versuchspersonen die Aufgabe, zu entscheiden, ob Sätze, die dem fraglichen Text in veränderter oder unveränderter Form entnommen wurden, in dem vorher gelesenen Text wörtlich enthalten sind. Problematisch dabei ist, dass die Versuchspersonen zum einen schlicht raten können, zum anderen, dass Wiedererkennen auch ohne Verstehen möglich ist. D. h. die Verstehensleistung wird hier u. U. überschätzt. Hinzu kommt, dass auch dieses Verfahren das Textverständnis nur sehr punktuell misst und vor allem semantische Bezüge zwischen Sätzen unberücksichtigt lässt. Allerdings kann man das Verfahren auch etwas sophistizierter gestalten, indem man Wiedererkennenaufgaben bzw. Testsätze nach den jeweils thematischen Ebenen des Verarbeitungsprozesses (Oberflächenebene, propositionale Ebene und Situationsmodellebene) differenziert (Beispiel aus der Grundlagenforschung: Fletcher, Chrysler 1990). Die Ebenenunterscheidung kann man m. E. nutzen, um gezielt ein oberflächliches Verstehen von einem tieferen Verstehen (auf Situationsmodellebene) abzugrenzen.

Multiple-Choice-Aufgaben

Multiple-Choice-Aufgaben sind dadurch gekennzeichnet, dass in Form von Antwortmöglichkeiten Fragen zum Textinhalt gestellt werden, wobei die richtige Antwort wiederzuerkennen bzw. anzukreuzen ist. Bei der Konzipierung solcher Tests, die weit verbreitet sind, werden verschiedene Teilfähigkeiten des Textverstehens (mindestens: Kenntnis der Wortbedeutungen, Textgliederung, Schlussfolgerungen und Bewertungen) durch mehrere Aufgaben abgedeckt. Je nach Forschungsinteresse sind grundsätzlich auch Tests konzipierbar, bei denen nur eine Teilfähigkeit (wie wörtliches vs. schlussfolgerndes Textverstehen) erfasst wird. Klassische Leseverständnistests, die mit Multiple-Choice-Fragen operieren, sind z. B. „Verständiges Lesen“ (LV 5–6 und LV 7–9; Anger et al. 1971; 1972) oder der Hamburger Lesetest für 3. und 4. Klassen (Lehmann et al. 2006). Die Vorteile von multiple-choice-Tests bestehen vor allem in der ökonomischen Durchführung und der objektiven Auswertung. Problematisch ist aber die Inhaltsvalidität, die Frage also, ob Leseverständnistests tatsächlich das Leseverständnis messen. Dies kann man einfach durch den Vergleich der Beantwortung von Multiple Choice-Fragen mit und ohne Textrezeption prüfen. In einer entsprechend angelegten Untersuchung von Weaver und Bickley (1967) erreichten die Versuchspersonen ohne vorherige Textrezeption 67 % der richtigen Antworten der Textleser. Auf diese Problematik der Inhaltsvalidität ist in mehreren amerikanischen Studien schon sehr früh hingewiesen worden. Im deutschsprachigen Raum ist die Frage der Konstruktvalidität von Multiple-Choice-Leseverständnis-Aufgaben erst im letzten Jahr systematisch empirisch untersucht worden, und zwar an Schülern der 7.–9. Klassenstufe (Rost, Sparfeldt 2007). Den Schülern und Schülerinnen wurden verschiedene Versionen eines Leseverständnistests vorgelegt, und zwar in den Varianten „Original“, „ohne Text“, „ohne Fragen“ und „ohne Text und ohne Fragen“. Außerdem wurde die verbale Intelligenz miterhoben. Die Ergebnisse zeigen, dass in allen reduzierten Versionen deutlich mehr Fragen richtig beantwortet werden, als es bei einer rein zufälligen Beantwortung zu erwarten gewesen wäre. Beim Vergleich der Versionen „Original“ vs. „ohne Text“ wurden sogar eineinhalb Mal so viele Items überzufällig richtig angekreuzt. Mit abnehmender Informationshaltigkeit der Versionen nimmt die Anzahl der korrekten Ankreuzungen ab, liegt aber bei der reduziertesten Variante (ohne Text und ohne Fragen) immer noch über dem Zufall. Die verbale Intelligenz korreliert mit der Originalversion zu $r = .6$, mit der informationsreduziertesten Variante nur noch mit $.19$. Aus diesen Befunden folgt zumindest, dass Multiple-Choice-Tests nicht nur Leseverständnis erfassen.

Für die überraschenden Ergebnisse lassen sich mehrere Gründe anführen: Zum einen sind die Wahlantworten zu einer Frage u. U. nicht gleichwahrscheinlich, d. h. der Leser kann durch Kombinieren die richtige Antwort erraten. Zum anderen kann man die verschiedenen Fragen, die den gesamten

Textinhalt abdecken sollen, zueinander in Beziehung setzen und auf dieser Grundlage einen kohärenten Textsinn konstruieren. Daraus folgt, dass Multiple-Choice-Tests z. B. weniger das Textverständnis messen als übergreifende Faktoren der (verbalen) Intelligenz wie „schlussfolgerndes Denken“ (vgl. Groeben 1982). Nach den Ergebnissen von Rost und Sparfeldt spielt die verbale Intelligenz in der Tat eine gewichtige Rolle, sie erklärt aber nicht die überzufällig richtigen Ankreuzungen in der Version „ohne Text und ohne Fragen“, da sie hier nur noch wenig Kriteriumsvarianz aufklärt. Was als potentieller Erklärungsfaktor bleibt, ist das allgemeine Welt- und Hintergrundwissen, dessen Einfluss bislang aber noch nicht systematisch untersucht wurde und bei dem mehrere Inhaltsbereiche berücksichtigt werden müssten. Für die Konstruktion von Multiple-Choice-Items folgt daraus, dass gute Distraktoren wenig „von Faktoren wie Weltwissen, Intelligenz und Testerfahrung abhängen“ sollten (Rost, Sparfeldt 2007, 313). Auch sollten die Distraktoren von Multiple-Choice-Items, die meist spontan erstellt werden, hinsichtlich ihrer Plausibilität eingeschätzt werden.

Eine Weiterentwicklung der Multiple-Choice-Tests, die deren Nachteile zu vermeiden sucht, ist der chunked reading-Test (CRT; Carver 1970). Während beim Multiple-Choice-Test die richtige Antwort anzukreuzen ist, ist beim CRT ein falscher Satzteil aus vier richtigen herauszufinden. Dazu muss für den Verständnistest ein Satzteil so verändert werden, dass die Satzsequenz einen in sich kohärenten, aber vom ursprünglichen Text abweichenden Sinn ergibt. Das nachfolgende Beispiel (Carver, Darby 1971, 35, in dt. Übersetzung; zit. nach Groeben 1982, 71f.) verdeutlicht das Vorgehen.

Gelesener Text:

„Die Ansichten des Laien über Computer und deren Leistungsfähigkeit verleihen dieser Erfindung den Status eines unfehlbaren und unparteiischen Lehrers. Es ist durchaus möglich, dass die Zukunft diese Einschätzung bestätigt, aber zur Zeit ist es wichtig festzustellen, was Computer nicht können, auch wenn kein wichtigerer Grund vorliegt als das eine Ziel: eine realistische Bewertung zu erreichen. Computer können keine Entscheidungen treffen, die nicht vorher vom Menschen programmiert wurden.“

Verständnis-Test:

- | | |
|---|--|
| <p>1. (A) Die Ansichten des Laien
(B) über zukünftige Lehrmaschinen
(C) und deren Leistungsfähigkeit
(D) verleihen dieser Erfindung

(E) den Status</p> | <p>3. (A) aber zur Zeit
(B) ist es wichtig festzustellen,
(C) was Computer nicht können,
(D) denn der Lehrer als Person ist wichtiger
(E) als das eine Ziel:</p> |
| <p>2. (A) eines unfehlbaren

(B) und unparteiischen Lehrers.
(C) Es ist durchaus möglich,
(D) dass die Zukunft
(E) eine veränderte Lage bringt,</p> | <p>4. (A) immer neue Maschinen zu konstruieren.
(B) Computer können keine
(C) Entscheidungen treffen,
(D) die nicht vorher
(E) vom Menschen programmiert wurden.</p> |

Beim CRT kann die richtige Wahlantwort nicht mehr einfach durch Schlussfolgerungsprozesse herausgefunden werden. Empirisch zeigte sich, dass beim bei Nicht-Lesen des Textes im CRT ein Leistungsabfall von 54 % auftritt (Carver, Darby 1971), während es bei herkömmlichen Multiple-Choice-Tests nur 33 % sind (Weaver, Bickley 1967). Der CRT misst somit den verstandenen Textsinn valider als Multiple-Choice-Tests.

Verifikationsverfahren

Bei diesem Verfahren müssen die Versuchspersonen vorgegebene Sätze hinsichtlich ihrer Richtigkeit beurteilen, d. h. es ist zu entscheiden, ob eine Aussage zu einem vorher gelesenen Text passt oder nicht. Die Urteile sollen Aufschluss darüber geben, in welchem Ausmaß ein Text verstanden wurde und welche Inferenzen gebildet worden sind.

4 Erhebung textferner Verstehensprodukte

Alle bislang vorgestellten Verfahren sind primär darauf konzentriert, zu überprüfen, ob der Sinngehalt eines Textes erfasst wurde. Darüber hinaus bezieht sich das Verstehen jedoch auch auf das Ausmaß, in dem die neue Textinformation in das eigene Wissenssystem integriert und wie das erworbene Wissen umgesetzt wird. Dies setzt zwar im ersten Schritt die Erfassung des Sinngehalts voraus, impliziert aber zusätzlich in mehr oder minder großem Ausmaß auch eine Loslösung von der zugrundeliegenden Textoberflächeninformation, d. h. es werden textfernere Verstehensprodukte erhoben.

Beantwortung von Fragen zum Text

Zu den textfernen Verstehensprodukten gehört zunächst einmal die Beantwortung von Textfragen mit offener Antwortmöglichkeit. Mit solchen Fragen können sowohl erworbenes Faktenwissen, Überblickswissen und Zusammenhangswissen erhoben (wissensorientierte Fragen) als auch das tiefere Textverständnis geprüft werden (verständnisorientierte Fragen).

Unter Rückgriff auf die Ebenen des Textverstehens kann man Fragen grundsätzlich danach unterscheiden, ob sie sich auf die Mikrostruktur des Textes (z. B. Konzeptfragen, Wie-Fragen, Wer-Fragen), seine Makrostruktur (z. B. Zusammenfassungen, Fragen zum Textaufbau) oder auf das Situationsmodell (Erklärungsfragen, Inferenzfragen, Problemlösefragen, Anwendungsfragen, Bewertungsfragen etc.) beziehen. Grundsätzlich sollten sich Fragen zur Überprüfung eines tieferen Verstehens dabei eher auf Situationsmodell-Ebene bewegen. Zur Formulierung von Fragen kann man sich dabei auch von Fragetaxonomien anregen lassen (z. B. Graesser et al. 1993), die für die Konstruktion von Textfragen genutzt werden kann.

Zur Überprüfung des Verständnisses bieten sich insbesondere Fragen an, die die Versuchspersonen veranlassen, Inferenzen zu ziehen. Mit einer sol-

chen Erfassung der Inferenzfähigkeit werden das konstruktive Verstehen und der kognitive Umgang mit den aufgenommenen Inhalten überprüft. Typische Inferenzfragen sind z. B. „Warum“-Fragen, die zu Inferenzen über Ursachen und übergeordnete Ziele führen, oder „Wie“-Fragen, die Inferenzen zu Handlungen und untergeordneten Zielen auslösen sollen (vgl. Graesser, Hemphill 1991). Der Bereich der Inferenzziehung ist für die Überprüfung des tieferen Verstehens noch nicht vollständig systematisch erforscht und überprüft worden, bietet aber m. E. eine breite Palette von Möglichkeiten zur Fragenformulierung (Überblick über Inferenztypen: Singer 1994).

Wiedergabeverfahren

Will man wissen, in welchem Ausmaß und in welcher Intensität die Versuchspersonen den gesamten Textinhalt repräsentiert und verstanden haben, sind freie Wiedergabeverfahren die Methode der Wahl. Die vier gängigsten sind:

- *Freie Reproduktion:* Bei der freien Reproduktion werden die Versuchspersonen gebeten, den gesamten Text so genau und so vollständig wie möglich wiederzugeben.
- *Strukturierte freie Wiedergabe:* Eine möglichst vollständige Wiedergabe kann durch Vorgabe von Strukturierungshilfen unterstützt werden. Es wird quasi ein strukturelles Raster vorgegeben wie „Allgemeine Aussagen“, „Elaborationen“, „Details“, „Beispiele“, „Anwendungen“ etc., das möglichst umfassend und vollständig ausgefüllt werden soll. Mit dieser Methode soll sicher gestellt werden, dass die Versuchspersonen nicht einfach Aspekte, die ihnen vielleicht weniger wichtig erscheinen, weglassen. Statt eines strukturellen Rasters kann auch textspezifisch ein inhaltliches Raster vorgegeben werden (Anwendungsbeispiel: Christmann 1989).
- *Gelenkte Reproduktion:* Im Unterschied zur freien Reproduktion werden bei der gelenkten Reproduktion einzelne Wörter, z. B. der Satzanfang, vorgegeben; die Versuchspersonen werden gebeten, die betreffenden Satzanfänge zu ergänzen (daher auch: Satz-Ergänzungs-Methode).
- *Zusammenfassungen:* Gerade bei längeren Texten kann es je nach Erkenntnisinteresse sinnvoll sein, nicht den ganzen Text wiedergeben zu lassen, sondern Zusammenfassungen der Textinhalte zu erbitten. Zusammenfassungen repräsentieren die globale Textstruktur und werden nach van Dijk (1980) durch die Anwendung von Makroregeln wie „Auslassen“, „Selegieren“, „Generalisieren“ und „Konstruieren bzw. Integrieren“ gebildet. Die Analyse der erhaltenen Zusammenfassungen zeigt, welche Makroregeln im individuellen Fall angewendet wurden und wie und in welchem Ausmaß die Textinformation verdichtet wurde. „Generalisierungen“ und „integrative Sätze“ dürften dabei ein tieferes Textverständnis indizieren als „Auslassungen“ und „Selektionen“. Generell zeigen Zusammenfassungen, welche Informationen als besonders wichtig

erachtet und besonders intensiv verstanden und behalten wurden. Außerdem geben sie Aufschlüsse darüber, ob das Wesentliche eines Textes verstanden wurde. Eine Alternative zur Produktion von Zusammenfassungen durch die Probandeninnen und Probanden besteht darin, Zusammenfassungen mit Auslassungen vorzugeben und diese von den Leserinnen und Lesern ergänzen zu lassen (vgl. Alderson 2000).

Bei allen Wiedergabeverfahren ist die Auswertung zumindest bei längeren Texten in der Regel jedoch aufwändig. Damit die quantitative und auch die qualitative Verstehensleistung gemessen werden können, müssen sowohl der vorgegebenen Text als auch die Verstehensprotokolle in Einheiten zerlegt werden (Propositionen, Sätze, Idea Units). Nach dieser Vorarbeit kann zunächst unter quantitativer Perspektive ausgezählt werden, wieviel Einheiten behalten wurde. Unter qualitativer Perspektive muss zumindest bestimmt werden, wieviele der wiedergegebenen Einheiten korrekt oder inhaltlich angemessen im Sinne der Vorlage sind. Differenziertere Auswertungen verlangen daher die Entwicklung eines inhaltsanalytischen Kategoriensystems, das – immer in Abhängigkeit vom Erkenntnisinteresse – z. B. folgende Kategorien enthalten könnte: Oberflächliche vs. tiefe Verarbeitung (z. B.: Wurde Relevantes von Irrelevantem unterschieden? Wurden unterschiedliche Perspektiven unterschieden?), inhaltliche Angemessenheit, Stringenz, Kohärenz, Vernetztheit von Wissens-elementen etc. Für die inhaltsanalytische Auswertung gilt, dass die Kategorien präzise expliziert werden müssen und dass die Objektivität des Verfahrens durch Ermittlung der Interkodier-Übereinstimmung zwischen mindestens zwei Kodiererinnen bzw. Kodierern bestimmt werden muss (vgl. Rustemeyer 1992).

Ordnungsaufgaben

Ordnungsaufgaben sind Aufgaben, bei denen die Reihenfolge der Sätze oder (besser noch) Abschnitte des Ursprungstextes zerstört wird. Aufgabe der Rezipienteninnen resp. Rezipienten ist es, die korrekte Reihenfolge wiederherzustellen. Den Grundgedanken des Durcheinanderwürfeln von Sätzen und Textabschnitten kann man natürlich auch auf längere Texte anwenden (Beispiele bei Anderson 2002).

Sortieraufgaben dieser Art erscheinen besonders angemessen, wenn man das textstrukturelle Wissen erfassen will, also Wissen darüber, wie spezifische Textsorten (Erzähltexte, Märchen, wissenschaftliche Texte etc.) üblicherweise aufgebaut sind. Solche konventionalisierten Strukturen, die die globale Ordnung von Texten beschreiben, werden als Superstrukturen bezeichnet (van Dijk 1980), deren Kenntnis das Textverstehen nachweislich erleichtert (Überblick: Christmann, Groeben 1996). Das grundsätzliche Problem bei diesem Verfahren besteht darin, dass man sicherstellen muss, dass nicht auch andere Reihenfolgen einen Sinn machen.

5 Bedingungen und Korrelate des Leseverstehens

Die Qualität des Lesens wird von einer Reihe von Variablen beeinflusst, die zumindest in quasi-experimentellen Settings zur Überprüfung des Textverstehens miterhoben werden sollten. Dazu gehören insbesondere die Wortkenntnis, das Vorwissen und die Arbeitsgedächtniskapazität.

Wortkenntnis

Eine grundlegende Anforderung beim Lesen besteht darin, den im Text enthaltenen Wörtern Bedeutung zuzuordnen. Dies bezeichnet man als lexikalischen Zugriff. Die Schnelligkeit des lexikalischen Zugriffs hängt vom Ausmaß und der Flüssigkeit der Wortkenntnis ab und korreliert mit allgemeinen Lesefähigkeitsmaßen sowohl bei Kindern als auch bei Erwachsenen. Das Gleiche gilt für den Wortschatz. Wortschatz-Maße weisen in der Regel hohe Zusammenhänge mit Lesefähigkeiten auf (z. B. Graves 1989).

Zur Messung der Schnelligkeit des lexikalischen Zugriffs haben sich prinzipiell die bereits vorgestellten lexikalischen Entscheidungsaufgaben sowie Verifikationsaufgaben bewährt.

Zur Erfassung des Wortschatzes stehen sog. Wortschatztests für verschiedene Altersstufen (von 3 Jahren bis zum späten Erwachsenenalter) zur Verfügung.

Dazu gehört z. B. der Wortschatztest (WST von Schmidt und Metzler; Einsatzbereich: ab 16 Jahre). Er besteht aus 40 Aufgaben zur Wiedererkennung von Wörtern und ist als Einzel- und Gruppentest durchführbar. Eine Testaufgabe enthält je ein Zielwort und 5 Distraktoren. Der Proband hat die Aufgabe, das Zielwort in jeder Zeile herauszufinden und durchzustreichen.

Ein anderes Verfahren ist der Hawik R, der aus 44 Items besteht, die nach der Bedeutung von Testwörtern fragen („Was ist ein Hund“); er wird häufig zur Diagnostik des lexikalischen Niveaus eingesetzt.

Vorkenntnisse

Der Einfluss von Vorkenntnissen auf das Leseverstehen ist enorm. Zur Erfassung des Vorwissens werden in der Regel inhaltspezifische Multiple-Choice-Tests eingesetzt, die dann natürlich immer neu zu konstruieren sind.

Arbeitsgedächtniskapazität

Lesen stellt einen Vorgang dar, bei dem in hohem Maße das Arbeitsgedächtnis beansprucht wird. Dies betrifft auf Satzebene insbesondere semantische und syntaktische Integrationsprozesse, auf Textebene Prozesse der lokalen und globalen Kohärenzbildung. In all diesen Fällen muss im Arbeitsgedächtnis die aus dem Text eingelesene Information zu der aus dem Langzeitgedächtnis abgerufenen Information in Beziehung gesetzt, integriert und für weitere Verarbeitungsprozesse zur Verfügung gehalten werden (Carpenter et

al. 1994). Entsprechend überrascht es nicht, dass sich die Arbeitsgedächtniskapazität nicht nur als guter Prädiktor für die allgemeine Lesefähigkeit, sondern auch als Prädiktor für die Güte von Integrationsprozessen beim Lesen erwiesen hat (Überblick: Carpenter et al. 1994; zusammenfassend: Richter, Christmann 2002). Zur Messung der Arbeitsgedächtniskapazität haben Danemann und Carpenter (1980) ein Verfahren entwickelt, das als Lesespannentest bekannt geworden ist (deutsche Version: Hacker et al. 1994).

Dabei lesen die Probandinnen und Probanden eine Liste von zusammenhangslosen Sätze auf Verständnis und behalten gleichzeitig das jeweils letzte Wort dieser Sätze; die Anzahl der Sätze, bei denen dies gelingt, ist ein Indikator für die Lesespanne. Ein Satz wird dabei jeweils für 3 s auf dem Bildschirm präsentiert; die Versuchspersonen haben 1 Sekunde Zeit, per Tastendruck den Satz als richtig oder falsch zu beurteilen und sich das letzte Wort des Satzes zu merken (die Sätze sind schlicht und erfordern keine Vorwissensaktivierung; zwischen den Endwörtern bestehen keine semantischen Beziehungen). Nach der Darbietung der Sätze sind die Endwörter in der richtigen Reihenfolge wiederzugeben.

Beispiel aus Oberauer et al. (2006, 129):

Lesen (4s) und Beurteilen

„Alle Menschen haben einen Vater“	(richtig? / falsch?)
„Der Himmel hat eine Ecke“	(richtig? / falsch?)
„Ein Auto betreibt man mit Kohle“	(richtig? / falsch?)
„Ein Elefant ist größer als eine Maus“	(richtig? / falsch?)

Aufgabe:

Wiedergabe der letzten Wörter des Satzes in der richtigen Reihenfolge.
Hier: Vater, Ecke, Kohle, Maus.

Die Sätze werden in Gruppen von 2–6 Sätzen präsentiert, wobei jede Gruppe dreimal geboten wird. Die Präsentation erfolgt in aufsteigender Reihenfolge, und zwar so lange, bis die Versuchsperson keine zwei (der drei vorgegebenen) Satzlisten auf einer Stufe korrekt erinnert. Die Lesespanne ist die höchste Stufe, auf der 2 Satzlisten korrekt wiedergegeben werden können.

Fazit

Zur Erfassung des Lese- und Textverstehens liegt eine Fülle unterschiedlicher Verfahren vor, die sich sicherlich noch weiter ausdifferenzieren lassen. Welches dieser Verfahren im konkreten Fall zur Testung eingesetzt wird, hängt zum einen von dem jeweiligen Erkenntnisinteresse ab (welche Fertigkeit soll überprüft werden) zum anderen aber auch von dem zugrundeliegenden Text bzw. von der Aufgaben-Text-Relation. Nicht jeder Text eignet sich für jeden Testtypus.

Literatur

- Alderson, Charles (2000): *Assessing Reading*. Cambridge: Cambridge University Press.
- Anger, Hans; Bargmann, Rolf; Voigt, Martin (1971): *Verständiges Lesen (VL 5–6)*. Weinheim: Beltz.
- Baur, Rupprecht; Spettmann, Melanie (2005): *Arbeiten mit C-Tests und Lückentests im Rahmen des Gemeinsamen Europäischen Referenzrahmens für DaF*. <<http://www.daf-netzwerk.org/Arbeitsgruppen/AG4>> (Material für Lehrende – Lehren – Lernen – Beurteilen).
- Carpenter, Patricia; Miyake, Akida; Just, Marcel (1994): Working memory constraints in comprehension. Evidence from individual differences, aphasia, and aging. In: *Handbook of psycholinguistics*. Hrsg. von Morton Gernsbacher. San Diego, CA: Academic Press, S. 699–719.
- Carreiras, Manuel; Clifton Jr., Charles (Hrsg.) (2004): *The on-line study of sentence comprehension. Eyetracking, ERPs and beyond*. New York: Psychology Press.
- Carver, Ronald (1970): Analysis of the ‚chunked‘ test items as measures of reading and listening comprehension. In: *Journal of Educational Measurement*, Jg. 7, H. 3, S. 141–149.
- Carver, Ronald; Darby, Charles (1971): Development and evaluation of a test of information storage during reading. In: *Journal of Educational Measurement*, Jg. 8, H. 1, S. 33–44.
- Christmann, Ursula (1989): *Modelle der Textverarbeitung. Textbeschreibung als Textverstehen*. Münster: Aschendorff.
- Christmann, Ursula; Groeben, Norbert (1999): Psychologie des Lesens. In: *Handbuch Lesen*. Hrsg. von Bodo Franzmann, Klaus Hasemann, Dietrich Löffler und Erich Schön. München: Saur, S. 145–223.
- Coleman, Edmund; Miller, Gerald (1968): A measure of information gained during prose learning. In: *Reading Research Quarterly*, Jg. 3, H. 3, S. 369–386.
- Daneman, Meredyth; Carpenter, Patricia (1980): Individual differences in working memory and reading. In: *Journal of Verbal Learning and Verbal Behavior*, Jg. 19, S. 450–466.
- Dressler, Wolfgang (1972): *Einführung in die Textlinguistik*. Tübingen: Niemeyer.
- van Dijk, Teun (1980): *Textwissenschaft*. Tübingen: Niemeyer.
- Fletcher, Charles; Chrysler, Susan (1990): Surface forms, text bases, and situation models. Recognition memory for three types of textual information. In: *Discourse Processes*, Jg. 13, S. 175–190.
- Glenberg, Arthur; Meyer, Marion; Lindem, Karen (1987): Mental models contribute to foregrounding during text comprehension. In: *Journal of Memory and Language*, Jg. 26, S. 69–83.
- Graesser, Arthur; Hemphill, Darold (1991): Question answering in the context of scientific mechanisms. In: *Journal of Memory and Language*, Jg. 30, S. 186–209.
- Graesser, Arthur; Person, Natalie; Huber, John (1992): Mechanisms that generate questions. In: *Questions and information systems*. Hrsg. von Thomas Lauer, Eileen Peacock und Arthur Graesser. Hillsdale, NJ: Erlbaum, S. 303–360.
- Graves, Michael (1989): A quantitative and qualitative study of elementary school children’s vocabularies. In: *Journal of Educational Research*, Jg. 82, H. 4, S. 203–209.

- Groeben, Norbert (1982): *Leserpsychologie I. Textverständnis – Textverständlichkeit*. Münster: Aschendorff.
- Grotjahn, Rüdiger (2002): Konstruktion und Einsatz von C-Tests. Ein Leitfaden für die Praxis. In: *Der C-Test. Theoretische Grundlagen und praktische Anwendungen*. Bd. 4. Hrsg. von Rüdiger Grotjahn. Bochum: AKS-Verlag, S. 211–225.
- Grotjahn, Rüdiger (Hrsg.) (2006): *Der C-Test. Theorie, Empirie, Anwendungen / The C-Test. Theory, empirical research, applications*. Frankfurt a. M.: Lang.
- Günther, Ulrich (1989): Lesen im Experiment. In: *Linguistische Berichte*, Jg. 122, S. 238–320.
- Haberlandt, Karl (1994): Methods in reading research. In: *Handbook of psycholinguistics*. Hrsg. von Morton Gernsbacher. San Diego, CA: Academic Press, S. 1–31.
- Hacker, Winfried; Veres, Thora; Wollenberger, Esther (1994): Verarbeitungskapazität für Text. Ergebnisse der Entwicklung eines deutschsprachigen Prüfverfahrens des Arbeitsgedächtnisses. In: *Zeitschrift für Psychologie*, Bd. 202, S. 295–320.
- Hemforth, Barbara (2006): Psycholinguistische Methoden zur Untersuchung des Satz- und Textverstehens. In: *Text – Verstehen. Grammatik und darüber hinaus. Institut für Deutsche Sprache Jahrbuch 2005*. Hrsg. von Hardarik Blühdorn, Eva Breindl und Ulrich Waßner. Berlin, New York: de Gruyter, S. 205–221.
- Inhoff, Albrecht; Rayner, Keith (1986): Parafoveal word processing during eye fixations in reading. Effects of word frequency. In: *Perception and Psychophysics*, Jg. 34, S. 49–57.
- Irmen, Lisa (2007): What's in a (role) name? Formal and conceptual aspects of comprehending personal nouns. In: *Journal of Psycholinguistic Research*, Jg. 36, S. 431–456.
- Janus, Raizi; Bever, Thomas (1985): Processing metaphoric language. An investigation of the three stage model of metaphor comprehension. In: *Journal of Psycholinguistic Research*, Jg. 14, H. 5, S. 473–487.
- Just, Marcel; Carpenter, Patricia (1980): A theory of reading. From eye fixations to comprehension. In: *Psychological Review*, Jg. 87, H. 4, S. 329–354.
- Kaakinen, Johanna; Hyönä, Jukka; Keenan, Janice (2003): How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. In: *Journal of Experimental Psychology: Learning, Memory and Cognition*, Jg. 2, H. 3, S. 447–457.
- Kaakinen, Johanna; Hyönä, Jukka (2005): Perspective effects on expository text comprehension. Evidence from thinking aloud protocols, eyetracking, and recall. In: *Discourse Processes*, Jg. 40, H. 3, S. 239–257.
- Lehmann, Rainer; Peek, Rainer; Poerschke, Jan (2006): *HAMLET 3–4. Hamburger Lesetest für 3. und 4. Klassen*. Göttingen: Hogrefe. 2. überarbeitete Auflage.
- Magliano, Joseph; Trabasso, Tom; Graesser, Arthur (1999): Strategic processes during comprehension. In: *Journal of Educational Psychology*, Jg. 91, S. 615–629.
- Meyer, Martin (2003): Spezielle Verfahren III. Bildgebende Verfahren. In: *Psycholinguistik. Ein internationales Handbuch*. Hrsg. von Gert Rickheit, Theo Herrmann und Werner Deutsch. Berlin, New York: de Gruyter, S. 181–189.
- Millis, Keith; Magliano, Joseph (1999): The co-influence of grammatical markers and comprehender goals on the memory for short discourse. In: *Journal of Memory and Language*, Jg. 41, S. 183–198.
- Millis, Keith; Golding, Jonathan; Barker, Gregory (1995): Causal connectives increase inference generation. In: *Discourse Processes*, Jg. 20, S. 29–49.
- Mitchell, Don (2004): On-line methods in language processing. Introduction and

- historical review. In: *The on-line study of sentence comprehension. Eyetracking, ERPs and beyond*. Hrsg. von Manuel Carreiras und Charles Clifton Jr. New York: Psychology Press, S. 15–32.
- Neely, James (1977): Semantic priming and retrieval from lexical memory. Roles of inhibitionless spreading of activation and limited-capacity attention. In: *Journal of Experimental Psychology: General*, Jg. 106, S. 226–254.
- Oberauer, Klaus; Mayr, Ulrich; Kluwe, Rainer (2006): Gedächtnis und Wissen. In: *Lehrbuch Allgemeine Psychologie*. Hrsg. von Hans Spada. 3. Aufl. Bern: Huber, S. 115–197.
- Raatz, Ulrich; Klein-Braley, Christine (1981): The C-Test. A modification of the cloze procedure. In: *Practice and problems in language testing*. Hrsg. von Terry Culhane, Christine Klein-Braley und Douglas Stevenson. Essex: University of Essex Occasional Papers, S. 113–148.
- Radach, Ralph (1996): *Blickbewegungen beim Lesen. Psychologische Aspekte der Determination von Fixationspositionen*. Münster: Waxmann.
- Radach, Ralph; Kennedy, Alan (2004): Theoretical perspectives on eye movements in reading. Past controversies, current issues and an agenda for future research. In: *European Journal of Cognitive Psychology*, Jg. 16, S. 3–26.
- Rayner, Keith (1977): Visual attention in reading. Eye movements reflect cognitive processes. In: *Memory & Cognition*, Jg. 4, S. 443–448.
- Rayner, Keith; Duffy, Susan (1986): Lexical complexity and fixation times in reading. Effects of word frequency, verb complexity, and lexical ambiguity. In: *Memory & Cognition*, Jg. 14, S. 191–201.
- Rayner, Keith; Sereno, Sara; Morris, Robin; Schmauder, Rene; Clifton, Charles Jr. (1989): Eye movements and on-line comprehension processes. In: *Language and Cognitive Processes*, Jg. 4, SI 21–24.
- Richter, Tobias; Christmann, Ursula (2002): Lesekompetenz. Prozessebenen und interindividuelle Unterschiede. In: *Lesekompetenz. Bedingungen, Dimensionen, Funktionen*. Hrsg. von Norbert Groeben und Bettina Hurrelmann. Weinheim: Juventa, S. 25–58.
- Rost, Detlef; Sparfeldt, Jörn (2007): Leseverständnis ohne Lesen? Zur Konstruktvalidität von multiple-choice-Leseverständnisaufgaben. In: *Zeitschrift für Pädagogische Psychologie*, Jg. 21, H. 3/4, S. 305–314.
- Rustemeyer, Ruth (1992): *Praktisch-methodische Schritte der Inhaltsanalyse. Eine Einführung am Beispiel der Analyse von Interviewtexten*. Münster: Aschendorff.
- Sanders, Ted; Spooren, Wilbert; Noordman, Leo (1992): Toward a taxonomy of coherence relations. In: *Discourse Processes*, Jg. 15, S. 1–35.
- Sharkey, Noel; Mitchell, Don (1985): Word recognition in a functional context. The use of scripts in reading. In: *Journal of Memory and Language*, Jg. 24, S. 253–270.
- Singer, Murray (1994): Discourse inference processes. In: *Handbook of psycholinguistics*. Hrsg. von Morton Gernsbacher. San Diego, CA: Academic Press, S. 479–515.
- Streb, Judith (2003): Spezielle Verfahren II. Elektrophysiologische Methoden. In: *Psycholinguistik. Ein internationales Handbuch*. Hrsg. von Gert Rickheit, Theo Herrmann und Werner Deutsch. Berlin, New York: de Gruyter, S. 168–180.
- Tanenhaus, Michael; Spivey-Knowlton, Michael; Eberhard, Kathleen; Sedivy, Julie (1995): Integration of visual and linguistic information in spoken-language comprehension. In: *Science*, Bd. 268, S. 1632–1634.
- Taylor, Wilson (1956): Recent development in the use of cloze procedure. In: *Jour-*

- nalism Quarterly*, Jg. 33, S. 42–48.
- Taylor, Wilson (1953): Cloze-procedure. A new tool for measuring readability. In: *Journalism Quarterly*, Jg. 30, S. 415–433.
- Tuinman, Jaap (1971): The removal of information procedure. In: *Journal of Reading Behavior*, Jg. 3, H. 2, S. 44–50.
- Weaver, Wendell; Bickley, Albert (1967): Sources of information for response to reading test items. In: *Proceedings of the 75th Annual Convention of the American Psychological Association*. Washington, DC: APA, S. 293–294.
- Wockenfuß, Verena; Raatz, Ulrich (2006): Über den Zusammenhang zwischen Testleistung und Klassenstufe bei muttersprachlichen C-Tests. In: *Der C-Test. Theorie, Empirie, Anwendungen / The C-Test. Theory, empirical research, applications*. Hrsg. von Rüdiger Grotjahn. Frankfurt a. M.: Lang, S. 211–242.