

URSULA CHRISTMANN

## **Verstehens- und Verständlichkeitsmessung Methodische Ansätze in der Anwendungsforschung**

### **1. Verstehen – Verständlichkeit: Begriffsklärung und Teilkompetenzen**

Entgegen der Alltagsintuition lässt sich die Verständlichkeit eines Textes nicht allein durch bestimmte objektiv feststellbare Textmerkmale wie etwa Wortschwierigkeit, Wortlänge, Satzlänge oder Satzkomplexität bestimmen, sondern sie erfordert neben der Berücksichtigung der semantischen Struktur und der Organisation von Textinhalten immer auch den Rückgriff auf das konkrete Verstehen eines Textes durch einen Leser bzw. eine Leserin. Das erscheint unmittelbar einsichtig, wenn man bedenkt, dass beispielsweise ein und derselbe juristische Text, selbst wenn er in syntaktisch-stilistischer Hinsicht optimiert ist, für juristische Experten/-innen leicht verständlich, für juristische Laien hingegen nahezu unverständlich sein kann. Dies ist im Rahmen der kritischen Auseinandersetzung mit der so genannten klassischen Lesbarkeitsforschung<sup>1</sup> (einschließlich der daraus abgeleiteten Lesbarkeitsformeln) auch immer wieder empirisch bestätigt worden: Die syntaktisch-stilistischen Oberflächenmerkmale von Texten erlauben keine Vorhersagen darüber, ob ein Text verstanden wird oder nicht. Will man solche Vorhersagen treffen, muss man die Merkmale der Rezipienten/-innen in das Konzept der Verständlichkeit einbeziehen. Texte sind demnach nicht an sich verständlich oder unverständlich, sondern sie sind dies immer nur für konkrete Leser/-innen mit ihren individuell unterschiedlichen kognitiven und motivationalen Voraussetzungen.

Entsprechend geht die Forschung zum Textverstehen heute übereinstimmend davon aus, dass der Verstehensprozess immer als eine Wechselwirkung zwischen vorgegebenem Text und der Kognitionsstruktur

---

<sup>1</sup> Standardwerk: GEORGE R. KLARE, *The measurement of readability*. Ames 1963.

des Lesers aufzufassen ist.<sup>2</sup> Was von einem Text verstanden wird, hängt demnach sowohl von den objektiven Merkmalen der Textinhaltsstruktur als auch von der Kognitionsstruktur der Rezipienten/-innen, d. h. dem Vorwissen, den Zielsetzungen sowie den Erwartungen und Einstellungen ab. Analog dazu kann man die Forschung zum Textverstehen danach unterscheiden, ob eher die Leser- oder eher die Textseite dieses Interaktionsprozesses im Vordergrund steht.

Die leserseitig orientierte Forschung konzentriert sich dabei auf die kognitiven Aktivitäten und Fähigkeiten des Rezipienten und lässt sich unter dem Begriff ‚Textverständnis‘ subsumieren.<sup>3</sup> In einer Fülle von empirischen Überprüfungen (zur Identifikation von Einzelfertigkeiten) ließen sich dabei faktorenanalytisch vier Teilkompetenzen sichern, die beim Verstehen eines Textes beteiligt sind: (1) Kenntnis der Wortbedeutung; (2) Schlussfolgerungen des Lesers qua Sinnverstehen; (3) Nachvollzug der Textstruktur und Textgliederung; (4) Identifizierung der Intention des Textes bzw. des Autors.<sup>4</sup> Diese Teilfähigkeiten sind im Übrigen kompatibel mit den Prozessen des Textverstehens, die von der neueren kognitionspsychologischen Forschung gesichert wurden. Sie verdeutlichen darüber hinaus, dass das Textverständnis sich primär auf den semantischen, nicht auf den grammatischen Bereich bezieht.<sup>5</sup>

Bei der textseitig orientierten Forschung steht der Text als solcher im Mittelpunkt; gefragt wird danach, welche Textmerkmale einen Einfluss auf das Verstehen haben und wie Texte gestaltet werden müssen, damit sie möglichst effizient und reibungslos verarbeitet werden können. Diese Forschung wird üblicherweise unter dem Begriff ‚Textverständlichkeit‘ zusammengefasst. Dabei wurden – ebenfalls wieder empirisch – vier Dimensionen der Textverständlichkeit herausgearbeitet:<sup>6</sup>

(1) Sprachliche Einfachheit, z. B.: geläufige, kurze, anschauliche Wörter, grammatikalisch einfache Sätze; keine Satzschachtelungen, keine Nominalisierungen; (2) kognitive Gliederung/Ordnung, z. B.: Vorstrukturierungen, kohärente Inhaltsorganisation, sequentielle Struktu-

<sup>2</sup> URSULA CHRISTMANN/NORBERT GROEBEN, Textverstehen, Textverständlichkeit – Ein Forschungsüberblick unter Anwendungsperspektive, in: HANS-PETER KRINGS (Hrsg.), *Wissenschaftliche Grundlagen der Technischen Kommunikation*. Tübingen 1996, 129–189; URSULA CHRISTMANN/NORBERT GROEBEN, Psychologie des Lesens, in: BODO FRANZMANN/KLAUS HASEMANN/DIETRICH LÖFFLER/ERICH SCHÖN (Hrsg.), *Handbuch Lesen*. München 1999, 145–223.

<sup>3</sup> NORBERT GROEBEN, *Leserpsychologie I: Textverständnis – Textverständlichkeit*. Münster 1982.

<sup>4</sup> DONALD SPEARRITT, Identification of subskills of reading comprehension by maximum-likelihood factor analysis, in: *Reading Research Quarterly* 8 (1972), 92–111.

<sup>5</sup> Ausführlich: GROEBEN, *Leserpsychologie* (Fn. 3); CHRISTMANN/GROEBEN, *Psychologie des Lesens* (Fn. 2).

<sup>6</sup> NORBERT GROEBEN, *Die Verständlichkeit von Unterrichtstexten*. 2. Aufl., Münster 1978.

rierung, Hervorhebungen, Analogien; (3) Kürze/Prägnanz; (4) motivationale Stimulanz: z. B. konfliktgenerierende Fragen. Von diesen Dimensionen ist die Dimension der kognitiven Gliederung/Ordnung am gewichtigsten für die Verständlichkeit eines Textes.

## 2. Schwerpunkte der Verstehenserhebung

In beiden Forschungsbereichen, sowohl der leser- als auch der textorientierten Forschung, wird eine Fülle von empirischen Methoden zur Erfassung von Verstehen/Verständlichkeit eingesetzt. Das Spektrum reicht von Blickbewegungserfassung über die Beantwortung von Textfragen bis hin zur Überprüfung des Verstehens auf der Handlungsebene. Die vorhandenen Methoden lassen sich u. a. danach klassifizieren, ob sie eher in der Grundlagen- oder eher in der Anwendungsforschung eingesetzt werden. Dabei ist die Grundlagenforschung primär an einer Rekonstruktion des *Verstehensprozesses* interessiert. Dazu werden bevorzugt folgende Methoden eingesetzt:

- Messung von Augenbewegungen wie Fixationsdauer (als Indikator für den kognitiven Verarbeitungsaufwand), Sakkadenlänge, Relation zwischen Sakkadenlänge und Fixationsdauer, Regressionen bzw. regressive Sakkaden (als Indikator für Verstehensprobleme) etc.;<sup>7</sup>
- Lese- und Verstehenszeiten für Sätze, Texte, Implikationen aus Sätzen etc. (Zeitraum zwischen Textdarbietung und Tastendruck; Annahme: je mehr Zeit, desto höher der kognitive Aufwand).
- Priming-Technik: Kurz hintereinander werden zwei Wörter geboten, von denen das zweite Wort (Target) beurteilt werden muss (z. B. Wort vs. Nicht-Wort). Die Reaktionszeit zwischen dem ersten (Prime) und dem zweiten Element (Target) gilt als Indikator für den Verarbeitungsaufwand. Die Reaktionszeit ist beispielsweise verkürzt, wenn zwischen den beiden Elementen eine semantische Beziehung besteht.
- Lexikalische Entscheidungsaufgaben: Messung der Latenzzeit zwischen dem Einblenden eines Stimulus-Wortes und der Entscheidung, ob es vorher gelesen wurde.

---

<sup>7</sup> Überblick: ULRICH GÜNTHER, Lesen im Experiment, in: *Linguistische Berichte* 122 (1989), 283–320.

Bei all diesen Methoden<sup>8</sup> geht es im Prinzip immer um eine Bestimmung des über die Zeit erfassten Verarbeitungsaufwandes. Dabei ist zum einen zu beachten, dass der Verarbeitungsaufwand nicht nur eine Funktion von Textmerkmalen darstellt, sondern immer auch von den Lesermerkmalen (Interesse, Konzentriertheit, Wissen, Interpretationsfähigkeit etc.) und vom jeweiligen sprachlichen Kontext mitbestimmt wird. Aus diesem Grund ist der Rückschluss von der gemessenen Zeit auf kognitive Aktivitäten nicht immer unproblematisch. Unabhängig davon sind diese Methoden für die Anwendungsforschung wenig produktiv, weil sie primär an der Rekonstruktion von Verstehensprozessen in einem mittlerweile sehr hohen Auflösungsgrad interessiert sind. Ich lege demgegenüber den Schwerpunkt mehr auf die Methoden der Anwendungsforschung, die durch die Erhebung von Verstehensprodukten das Ausmaß und die Qualität der jeweiligen Textverarbeitung aufklären sollen. Diese Verfahren werden nach aufsteigender Komplexität der beteiligten kognitiven Prozesse im Folgenden besprochen.

### 3. Erhebung von Verstehensprodukten

#### 3.1 Beurteilungs- und Einschätzverfahren

##### *Semantisches Differenzial*

Bei dem von Osgood, Suci und Tannenbaum entwickelten semantischen Differential dürfte es sich um das bekannteste Verfahren zur Messung des konnotativen Bedeutungsgehalts von Begriffen bzw. der affektiven Qualität komplexerer Meinungsgegenstände handeln.<sup>9</sup> In der Psychologie wird es bevorzugt zur Beurteilung von Objekten, Personen des öffentlichen Lebens, ethnischen Gruppen, Nationalitäten, Werbeprodukten etc. sowie zur Erfassung des konnotativen Raums von Ähnlichkeiten und Unterschieden zwischen Begriffen wie ‚Gastarbeiter‘ vs. ‚Fremdarbeiter‘, ‚Frau‘ vs. ‚Mann‘, ‚Wissenschaftlerin‘ vs. ‚Wissenschaftler‘ oder ‚Einsamkeit‘ bei verschiedenen Nationalitäten etc. eingesetzt.<sup>10</sup> Darüber hinaus wurde es u. a. auch zur Bewertung

---

<sup>8</sup> Überblick bei GERT RICKHEIT/HANS STROHNER, *Grundlagen der kognitiven Sprachverarbeitung*. Tübingen 1993.

<sup>9</sup> CHARLES E. OSGOOD/GEORGE J. SUCI/PERCY H. TANNENBAUM, *The measurement of meaning*. Urbana (Illinois) 1957.

<sup>10</sup> Überblick: REINHOLD BERGLER (Hrsg.), *Das Eindrucksdifferential. Theorie und Technik*. Bern 1975.

ganz unterschiedlicher Textsorten, von literarischen Texten bis hin zu Verwaltungs- und Gesetzestexten genutzt.

Die Grundidee des semantischen Differentials besteht in der Annahme, dass Wörter bzw. Begriffe Punkte im semantischen Raum darstellen, der durch bestimmte Bedeutungsdimensionen definiert ist (auf denen die einzelnen Wörter lokalisiert sind). Entsprechend kommt es darauf an festzustellen, durch welche Bedeutungsdimensionen ein Konzept oder ein Meinungsgegenstand gekennzeichnet ist. Dazu wird ein Satz von siebenstufigen bipolaren Adjektivskalen vorgegeben, auf denen das fragliche Konzept einzuschätzen ist. Bei den Adjektiven handelt es sich um Gegensatzpaare wie ‚warm‘ vs. ‚kalt‘, ‚hart‘ vs. ‚weich‘, ‚heiter‘ vs. ‚traurig‘ etc. Urteilsgrundlage ist die „gefühlsmäßige Affinität des Urteilsgegenstands zu den Urteilsskalen“<sup>11</sup>, nicht der sachliche Zusammenhang (vgl. Abb. 1). Im Standarddifferential von Osgood sind diese Polaritäten sehr allgemein gehalten und haben für spezifische Konzepte teils nur eine metaphorische Aussagekraft.

Fasst man die Beurteilungen aller Beurteiler zusammen, so gelangt man zu einem über alle Personen gemittelten Polaritätenprofil für ein Konzept. Die Ähnlichkeit der Profile verschiedener Urteilsgegenstände (z. B. ‚männlich‘ vs. ‚weiblich‘) kann im nächsten Schritt bestimmt werden, indem man die Profile miteinander korreliert.

Darüber hinaus kann man die Zusammenhänge zwischen den einzelnen Adjektivskalen (Korrelationen über alle Versuchspersonen) in einer Faktorenanalyse zusammenfassen und gelangt so zu Dimensionen des semantischen Raums, auf denen die Urteilsgegenstände positioniert sind. Osgood hat in seinen (sprachvergleichenden) Untersuchungen mit dem Standarddifferential immer drei Dimensionen des semantischen Raums herausarbeiten können: Valenz (z. B. ‚angenehm‘ vs. ‚unangenehm‘); Potenz (z. B. ‚stark‘ vs. ‚schwach‘) und Aktivität (z. B. ‚erregend‘ vs. ‚beruhigend‘).

Während Osgood der Auffassung war, dass es universelle semantische Polaritäten gibt, die für alle Begriffe geeignet sind, hat die Empirie zwischenzeitlich gezeigt, dass die Ergebnisse von den verwendeten Beurteilungsmerkmalen abhängig sind. Darüber hinaus besteht bei der Verwendung universeller Eigenschaftslisten die Gefahr, dass methodische Artefakte produziert werden. Es kann nämlich durchaus sein, dass nur deshalb drei Dimensionen gefunden werden, weil immer mit den gleichen Adjektivskalen gearbeitet wurde. Hinzu kommt, dass die Versuchspartner/-innen auf Dimensionen reagieren müssen, die sie eigentlich für die Beurteilung des Gegenstands als völlig irrelevant erachten;

---

<sup>11</sup> JUERGEN BORTZ, *Lehrbuch der empirischen Forschung*. Berlin 1984, 129.

der ‚Zwang‘ zum Reagieren führt dann in der Regel zu Ankreuzungen im mittleren Bereich der Skala oder zu ‚missings‘.<sup>12</sup>

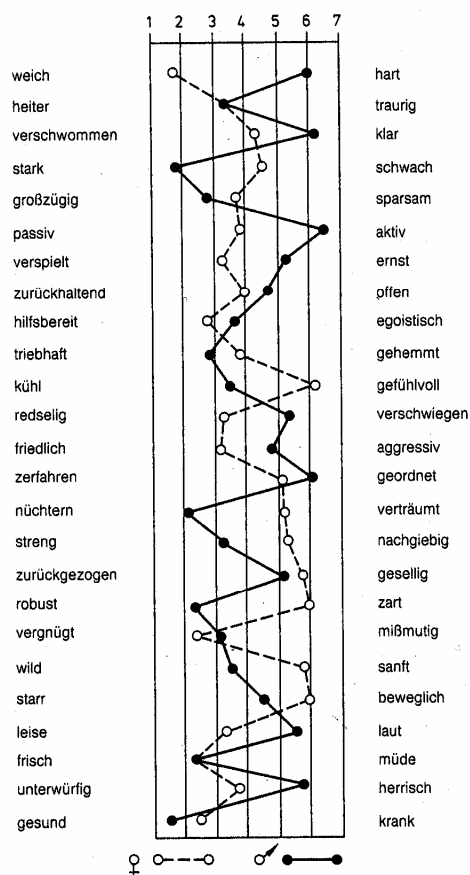


Abb. 7a

Abb. 1: Polaritätsprofil der Begriffe „männlich“ und „weiblich“ (nach Bortz, 1984, 129)

Auf Grund solcher Probleme werden heute häufig auch konzeptspezifische, d. h. dem jeweiligen Untersuchungsgegenstand angepasste Merkmale verwendet.<sup>13</sup> Im Unterschied zum universellen semantischen Differential erfasst das kontextspezifische Differential eher „die denota-

<sup>12</sup> BERGLER, Eindrucksdifferential (Fn. 10).

<sup>13</sup> Vgl. BERGLER, Eindrucksdifferential (Fn. 10).

tiven, direkten Beziehungen der Urteilsobjekte zu den Urteilsskalen“.<sup>14</sup> Für die Zusammenstellung solcher konzeptspezifischen Merkmale wird gefordert, dass sie für den Beurteilungsgegenstand charakteristisch und repräsentativ sowie auf den Sprachgebrauch der Zielgruppe zugeschnitten sind, bzw. dass die Adjektive eine repräsentative Stichprobe von Beurteilungsmerkmalen darstellen sollen, die die Befragten selbst zur Beschreibung verwenden.<sup>15</sup> Daher muss der Erstellung der Adjektivskalen eine explorative Analyse der assoziativen Bedeutungsaspekte eines Meinungsgegenstands sowie des Sprachgebrauchs der Zielgruppe vorausgehen, und die Repräsentativität der Bedeutungsmerkmale muss durch ein Experten-Rating gesichert werden. (Eine weitere Modifikation stellen Differentiale dar, bei denen statt einzelner Adjektive mehrere Merkmale oder Statements vorgegeben werden.)

Zur Beurteilung der Verständlichkeit eines Textes sind kontextspezifische Differentiale (s. u.), sicherlich besser geeignet als das universelle Differential, das eher dann zum Einsatz gelangen sollte, wenn es um den Vergleich ganz unterschiedlicher Meinungsgegenstände und deren konnotativer Bedeutung geht. Man sollte sich auch bewusst sein, dass beim semantischen Differential der Assoziationsraum zweifach reduziert wird: Einmal durch die Vorgabe von Adjektiv-Skalen und zum zweiten durch die Zusammenfassung der Adjektivskalen zu Faktoren. Dadurch können relevante Unterschiede zwischen Begriffen verwischt werden. Man sollte sich also vorher überlegen, ob man mit möglichst wenigen generellen Faktoren operieren will, oder ob man wissen möchte, wie viele Bedeutungsdimensionen unterscheidbar sind.

#### *Verständlichkeitsrating*

Ein konzeptspezifisches semantisches Differential wurde von Langer et al. zur Einschätzung der Verständlichkeit von Texten eingesetzt, um darauf aufbauend Dimensionen der Textverständlichkeit zu bestimmen, die dann sehr ökonomisch zur Verständlichkeitsmessung eingesetzt werden können.<sup>16</sup> Als Merkmale für die Skalen des semantischen Differentials wurden konkrete Text- und Stilmerkmale verwendet, die sich in der Forschung als relevant erwiesen hatten (z. B. ‚folgerichtig‘ vs. ‚zusammenhangslos‘; ‚weitschweifig‘ vs. ‚aufs Wesentliche reduziert‘).

---

<sup>14</sup> BORTZ, Empirische Forschung (Fn. 11), 129.

<sup>15</sup> Vgl. BERGLER, Eindrucksdifferential (Fn. 10).

<sup>16</sup> INGHARD LANGER/FRIEDEMANN SCHULZ v. THUN/REINHARD TAUSCH, *Verständlichkeit in Schule, Verwaltung, Politik und Wissenschaft*. München 1974.

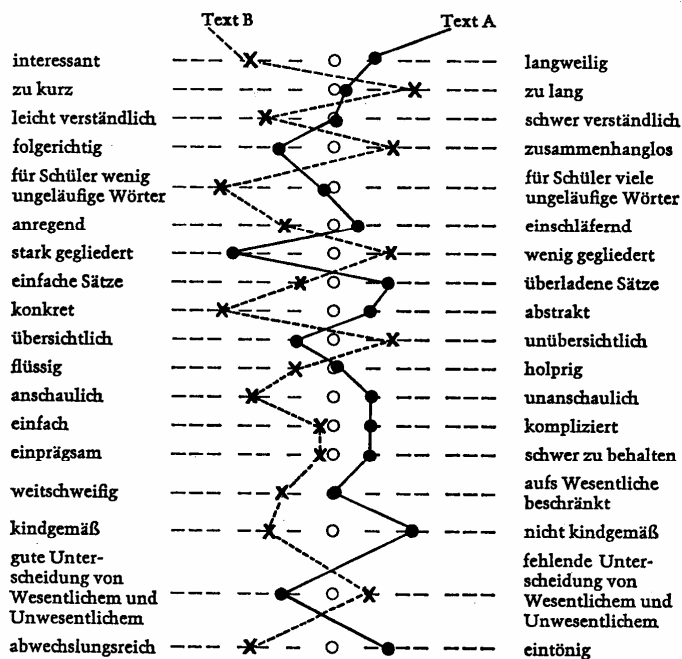


Abb. 2: Einschätzung zweier Texte hinsichtlich 18 relevanter Text- und Stilmerkmale  
(nach Langer et al. 1974, 50)

Die Abbildung zeigt die verschiedenen siebenstufigen Adjektivskalen sowie die Einschätzung zweier Texte auf diesen Skalen. Insgesamt wurde eine Fülle von Texten aus unterschiedlichen Bereichen von mindestens zehn Experten auf diesen Skalen eingeschätzt. Die Zusammenhänge zwischen den Skalen wurden korrelativ ermittelt und anschließend im Rahmen einer Faktorenanalyse zu Dimensionen der Verständlichkeit zusammengefasst (die Dimensionen werden durch die am höchsten auf den Faktoren ladenden Einzelmerkmalen definiert). Auf diese Weise gelangten die Autoren zu vier Dimensionen der Verständlichkeit, die als Dimensionsbilder dargestellt werden und die Merkmale enthalten, durch die die Dimensionen definiert sind.



## Dimensionsbild: Einfachheit

<b>Einfachheit</b>	+2	+1	0	-1	-2	<b>Kompliziertheit</b>
einfache Darstellung	[ ]					komplizierte Darstellung
kurze, einfache Sätze						lange, verschachtelte Sätze
geläufige Wörter						ungeläufige Wörter
Fachwörter erklärt						Fachwörter nicht erklärt
konkret						abstrakt
anschaulich						unanschaulich

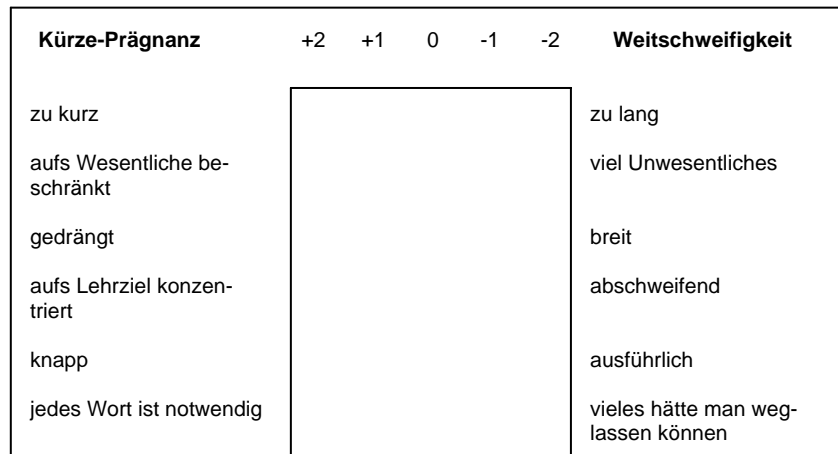
## Dimensionsbild: Gliederung-Ordnung

<b>Gliederung-Ordnung</b>	+2	+1	0	-1	-2	<b>Ungegliedertheit Zusammenhanglosigkeit</b>
gegliedert	[ ]					ungegliedert
folgerichtig						zusammenhanglos, wirr
übersichtlich						unübersichtlich
gute Unterscheidung von Wesentlichem und Un- wesentlichem						schlechte Unterschei- dung von Wesentlichem und Unwesentlichem
der rote Faden bleibt sichtbar						man verliert oft den roten Faden
alles kommt schön der Reihe nach						alles geht durcheinander

Abb. 3: Dimensionsbilder für vier zentrale Verständlichkeitsdimensionen  
(nach Langer et al., 1974, 50), Teil 1/2.

Diese vier Dimensionen (vgl. Abb. 3, Teil 1 u. 2), die als weitgehend unabhängig voneinander angesehen werden, können nun direkt zur Messung der Textverständlichkeit eingesetzt werden. Zur Beurteilung ist eine fünfstufige Rating-Skala vorgesehen. Die Verständlichkeit eines Textes wird dann durch die vier Kennwerte auf den Dimensionen ‚Einfachheit‘, ‚Gliederung-Ordnung‘, ‚Kürze-Prägnanz‘ und ‚zusätzliche

## Dimensionsbild: Kürze-Prägnanz



## Dimensionsbild: Zusätzliche Stimulanz

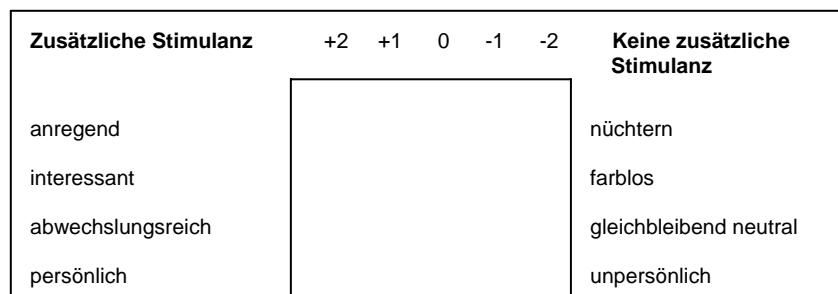


Abb. 3: Dimensionsbilder für vier zentrale Verständlichkeitsdimensionen  
(nach Langer et al., 1974, 50), Teil 2/2.

Stimulanz' angegeben. Damit ist allerdings noch nicht geklärt, ob Texte, die auf den Dimensionen als hoch verständlich eingeschätzt werden, tatsächlich auch besser verstanden und behalten werden, d. h. ob die Dimensionen valide sind. Dies konnte an Hand einer Fülle von Texten (Gesetzestexte, wissenschaftliche Texte, Texte aus Politik und Wirtschaft, Vertragstexte, Schulbuchtexte etc.) mit unterschiedlichen Probandengruppen (Schüler, Studierende, Erwachsene unterschiedlicher sozialer Herkunft) empirisch gesichert werden. Dabei wurde stets das Verstehen/Behalten eines Originaltextes mit dem Verstehen/Behalten einer verständlichkeitsoptimierten Fassung verglichen. Das Verstehen

wurde dabei mittels Cloze-Tests, Multiple-Choice-Aufgaben, freier Wiedergabe und offenen Fragen erfasst (s. u.).

Insgesamt kann die Verständlichkeit mit Hilfe der vier Dimensionen äußerst ökonomisch eingeschätzt werden. Hinzu kommt, dass das Konzept auf unterschiedliche Texte anwendbar ist, Hinweise zur Verbesserung von Texten gibt, valide und trainierbar ist. Problematisch erscheint allerdings die Theorielosigkeit des Ansatzes sowie der Rekurs auf die relativ subjektiven und intuitiven Expertenratings, die nur eine indirekte Erfassung der Verständlichkeit erlauben.<sup>17</sup>

### 3.2 Erhebung textnaher Verstehensprodukte

Direktere Erhebungsmethoden, die ich im Folgenden darstellen werde, setzen unmittelbar an der Erfassung von Verstehensprodukten an. Dabei unterscheide ich zwischen eher textnahen Erhebungsmethoden, also Methoden, die das Verstehen in enger Anlehnung an den Text überprüfen, wie das bei Reproduktions-, Wiedererkennung- und Einsetzverfahren der Fall ist, und textfernen Erhebungsmethoden; zu diesen gehören vor allem die freie Wiedergabe, die Beantwortung von Textfragen oder die Überprüfung des Verstehens auf Handlungsebene. Während textnahe Methoden eher das unmittelbare Textverständnis im Sinne einer Vorstellung über den Sinngehalt erfassen, sind bei textfernen Methoden in hohem Maße auch Gedächtnis- und Schlussfolgerungsprozesse beteiligt. Generell besteht bei allen Verfahren das Problem, dass unter Umständen schwer entscheidbar ist, ob eine korrekte Antwort auf dem Verstehen von Inhalten oder auf bloßer Merkfähigkeit beruht.

#### *Subjektive Informationsmessung*

Das auf Shannon zurückgehende Verfahren der subjektiven Informationsmessung (auch: *progressive cloze procedure*)<sup>18</sup> ist ein typisches Ergänzungsverfahren, das die subjektiven Erwartungsstrukturen der Rezipienten/-innen berücksichtigt.<sup>19</sup> Dabei wird ein Text Buchstabe für Buchstabe von links nach rechts erraten. Die Anzahl der Wahlvorgänge

---

<sup>17</sup> MANFRED HOFER, Textverständlichkeit: Zwischen Theorie und Praxeologie, in: *Unterrichtswissenschaft* 2 (1976), 143–150.

<sup>18</sup> CLAUDE E. SHANNON, Prediction and entropy of printed English, in: *Bell System Technical Journal* 30 (1951), 50–64.

<sup>19</sup> GROEBEN, Leserpsychologie (Fn. 3); GROEBEN, Verständlichkeit von Unterrichtstexten (Fn. 6).

bei jedem Zeichen gibt dabei dessen Wahrscheinlichkeit innerhalb des Satzes/Textes an. Daraus lässt sich ein Entropieprofil des Satzes ableiten und die mittlere Information ermitteln. Die Erwartung für einen Buchstaben wird vom Inhalt und dem Bekanntheitsgrad des Textes für den Rezipienten bestimmt. Information tritt dann auf, wenn ein Buchstabe nicht der Erwartung des Rezipienten entspricht, also etwas Neues auftritt (Nach der Informationstheorie stellt nur das Neue, das dem Leser nicht bekannt ist, eine Information dar). Das Informationsmaß drückt also den subjektiven Überraschungswert einer Zeichenfolge aus und erlaubt damit eine flexible Berücksichtigung der Erwartungsstrukturen des Rezipienten. Die Anzahl richtig erratener Zeichen stellt einen Indikator für die Bekanntheit des Textinhalts dar. Wird ein Text ohne vorheriges Lesen erraten, dann ist dies ein Indikator dafür, dass der Rezipient schon von vorneherein praktisch alles im Text Mitgeteilte weiß. Der Prozentsatz der richtig geratenen Zeichen nach dem Lesen im Vergleich zu dem vorher ohne Lesen erratenen Text stellt ein Maß für die durch die Verarbeitung des Textes hinzugewonnene Information dar.

Das Verfahren wurde im deutschen Sprachraum von Weltner weiterentwickelt und speziell auf die Messung der Textschwierigkeit bzw. Textverständlichkeit zugeschnitten.<sup>20</sup> Der Grundgedanke besteht darin, dass das Informationsausmaß dem Überraschungswert entspricht, den ein Text für eine/-n Leser/-in hat. Je mehr Leser/-innen über den im Text behandelten Sachverhalt wissen und je mehr sie die Textform (stilistische, syntaktische, strukturelle Besonderheiten) kennen, desto geringer ist der Informationswert für sie und desto verständlicher ist der Text. Entsprechend kann man das sukzessive Erraten von Zeichen eines Textes, der vorher noch nicht gelesen wurde, als Maß für dessen Verständlichkeit für eine/-n bestimmten Leser/-in ansetzen. Weltner hat das ursprüngliche Rateverfahren von Shannon vereinfacht, indem die Versuchspersonen nur einmal raten muss und dann das richtige Zeichen mitgeteilt bekommt. Außerdem hat er die Rateeinheit von (ursprünglich) Buchstaben über Silben bis zu Worten vergrößert. Gemessen werden die richtig erratenen Zeichen (Redundanz erster Ordnung), die dann zu den Informationswerten der vollständigen Schätzung in Beziehung gesetzt werden; über eine Regressionsgleichung lässt sich der subjektive Informationswert für jede einzelne Versuchsperson berechnen.

Der Vorteil der Methode besteht darin, dass die Verständlichkeit eines Textes relativ direkt an der Verarbeitung des Textsinns ansetzt und Schlussfolgerungs- sowie Gedächtnisprozesse zwar nicht ausschließt,

---

<sup>20</sup> KLAUS WELTNER, *Informationstheorie und Erziehungswissenschaften*. Quickborn bei Hamburg 1970.

aber gegenüber anderen Verfahren doch in den Hintergrund drängt.<sup>21</sup> Der Nachteil liegt darin, dass der Erhebungsprozess doch sehr aufwendig und langwierig ist.

#### *Lückentest (cloze procedure)*

Demgegenüber sind Schlussfolgerungsprozesse beim Lückentest oder der so genannten *cloze procedure* explizit nicht ausgeschlossen. Der Lückentest basiert ebenfalls auf Einsetz- bzw. Ergänzungsprozessen.<sup>22</sup> Dabei wird nach dem klassischen Verfahren von Taylor jedes fünfte Wort in einem Text ausgelassen, das dann von den Versuchspersonen einzusetzen ist.<sup>23</sup> Beim erstmaligen Lesen gilt die Anzahl der richtig eingesetzten Wörter als Maß für die Lesbarkeit des Textes. Führt man den Lückentest nach dem Lesen eines Textes durch, erhält man ein Maß für das Textverständnis; mit zunehmender Zeitspanne zwischen Lesen und Testvorgabe wird der Cloze-Test schließlich zu einem Behaltentest. 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 keinen signifikanten Unterschied zwischen Cloze-Tests mit und ohne Textkenntnis feststellen können.<sup>24</sup> 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 die Textverständlichkeit, da sie Merkmale und Fähigkeiten der Leser/-innen abbilden. Aus diesem Grund ist die bereits dargestellte progressive *cloze procedure* (subjektive Informationsmessung), bei der die Kon-

---

<sup>21</sup> Zur Anwendung vgl. GROEBEN, Verständlichkeit von Unterrichtstexten (Fn. 6).

<sup>22</sup> Ausführlich: GROEBEN, Leserpsychologie (Fn. 3).

<sup>23</sup> WILSON L. TAYLOR, Cloze-procedure. A new tool for measuring readability, in: *Journalism Quarterly* 30 (1953), 415–433; DERS., Recent development in the use of cloze procedure, in: *Journalism Quarterly* 33 (1956), 42–48. Cloze-procedure. A new tool for measuring readability, in: *Journalism Quarterly* 30 (1953), 415–433; DERS., Recent development in the use of cloze procedure, in: *Journalism Quarterly* 33 (1956), 42–48.

<sup>24</sup> EDMUND B. COLEMAN/ GERALD R. MILLER, A measure of information gained during prose learning, in: *Reading Research Quarterly* 3/3 (1968), 369–386.

textdetermination nur von einer Seite aus erfolgt und die Textinhalte weniger leicht durch Schlussfolgerungen erraten werden können, das validere Verfahren.

Die Validität der klassischen *cloze procedure* lässt sich jedoch erhöhen, wenn man nicht systematisch jedes fünfte Wort auslässt, sondern die Auslassungen auf Inhaltswörter beschränkt oder auch statt Wörtern Satzteile bzw. unter Umständen ganze Sätze auslässt.<sup>25</sup>

Eine Weiterentwicklung des klassischen Lückentextes stellt die Removal of Information Procedure (RIP; Auslassmethode) dar, bei der die Versuchspersonen in einem Text diejenigen Inhaltswörter unterstreichen, die in einem Cloze-Test am schwierigsten auszufüllen wären.<sup>26</sup> Allerdings wird das Verfahren primär zur Überprüfung des Informationsgehalts bestimmter Wortarten in bestimmten Satzpositionen eingesetzt. Es misst daher weniger die Verständlichkeit des Textes, kann aber Hinweise zur Erhöhung der Verständlichkeit geben.

#### *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. Das heißt, die Verstehensleistung wird hier unter Umständen überschätzt. Hinzu kommt, dass auch dieses Verfahren das Textverständnis nur sehr punk-

<sup>25</sup> Empirisch: JAAP TUINMAN, The removal of information procedure, in: *Journal of Reading Behavior* 3/2 (1971), 44–50.

<sup>26</sup> Vgl. TUINMAN, The removal of information procedure (Fn. 25).

tuell misst und vor allem semantische Bezüge zwischen Sätzen unberücksichtigt lässt.

### *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 wird jede der 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. Die Vorteile von Multiple-Choice-Tests bestehen vor allem in der ökonomischen Durchführung und der objektiven Auswertung. Problematisch ist aber die Inhaltsvalidität. Dies zeigt insbesondere der Vergleich der Beantwortung von Multiple-Choice-Fragen mit und ohne Textrezeption. In einer entsprechend angelegten Untersuchung von Weaver und Bickley erreichten die Versuchspartner/-innen ohne vorherige Textrezeption 67 % der richtigen Antworten der Textleser.<sup>27</sup> Für dieses überraschende Ergebnis lassen sich zwei Gründe anführen: Zum einen sind die Wahlantworten zu einer Frage unter Umständen nicht gleich wahrscheinlich, 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 zum Teil weniger das Textverständnis messen als übergreifende Faktoren der (verbalen) Intelligenz wie ‚schlussfolgerndes Denken‘.<sup>28</sup>

Eine Weiterentwicklung der Multiple-Choice-Tests, die deren Nachteile zu vermeiden sucht, ist der Chunked-Reading-Test (CRT).<sup>29</sup> Während beim Multiple-Choice-Test die richtige Antwort anzukreuzen ist, gilt es beim CRT einen falschen Satzteil aus vier richtigen herauszufinden. Dazu muss für den Verständnistest ein Satzteil so verändert werden, dass die Satzsequenz einen in sich kohärenten, aber vom ursprüng-

---

<sup>27</sup> WENDELL W. WEAVER/ALBERT C. BICKLEY, Sources of information for response to reading test items, in: *Proceedings of the 75<sup>th</sup> Annual Convention of the American Psychological Association* (1967), 293–294.

<sup>28</sup> GROEBEN, *Leserpsychologie* (Fn. 3).

<sup>29</sup> RONALD P. CARVER, Analysis of the ‚chunked‘ test items as measures of reading and listening comprehension, in: *Journal of Educational Measurement*, 7/3 (1970), 141–149.

lichen Text abweichenden Sinn ergibt. Das nachfolgende Beispiel verdeutlicht das Vorgehen:<sup>30</sup>

*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 wichtiger 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<br/>(B) über zukünftige Lehrmaschinen<br/>(C) und deren Leistungsfähigkeit<br/>(D) verleihen dieser Erfindung<br/>(E) den Status</p> | <p>3. (A) aber zur Zeit<br/>(B) ist es wichtig festzustellen,<br/>(C) was Computer nicht können,<br/>(D) denn der Lehrer als Person ist wichtiger<br/>(E) als das eine Ziel:</p>     |
| <p>2. (A) eines unfehlbaren<br/>(B) und unparteiischen Lehrers.<br/>(C) Es ist durchaus möglich,<br/>(D) dass die Zukunft<br/>(E) eine veränderte Lage bringt,</p>     | <p>4. (A) immer neue Maschinen zu konstruieren.<br/>(B) Computer können keine<br/>(C) Entscheidungen treffen,<br/>(D) die nicht vorher<br/>(E) vom Menschen programmiert wurden.</p> |

Beim CRT kann die richtige Wahlantwort nicht mehr einfach durch Schlussfolgerungsprozesse herausgefunden werden. Empirisch zeigte sich, dass beim Nicht-Lesen des Textes im CRT ein Leistungsabfall von 54 % erreicht wird,<sup>31</sup> während es bei herkömmlichen Multiple-Choice-

<sup>30</sup> Beispiel nach RONALD P. CARVER/CHARLES A. DARBY, Development and evaluation of a test of information storage during reading, in: *Journal of Educational Measurement*, 8/1 (1971), 33–44, 35; in dt. Übersetzung; zit. nach GROEBEN, *Leserpsychologie* (Fn. 3), 71 f.

<sup>31</sup> CARVER/DARBY, *Test of information storage* (Fn. 30).



Tests nur 33 % sind.<sup>32</sup> 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 werden und welche Inferenzen gebildet worden sind.

### 3.3 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 zugrunde liegenden 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). Zur Überprüfung des Verständnisses bieten sich z.B. Fragen an, die die Versuchspersonen veranlassen, Inferenzen zu ziehen. Mit einer solchen 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 unterge-

---

<sup>32</sup> WEAVER/BICKLEY, Sources of information (Fn. 27).

ordneten Zielen auslösen sollen.<sup>33</sup> Der Bereich der Inferenzziehung ist für die Überprüfung des tieferen Verstehens noch nicht vollständig systematisch erforscht und überprüft worden, bietet aber meines Erachtens eine breite Palette von Möglichkeiten zur Fragenformulierung. In der sprachpsychologischen Literatur werden eine Fülle von Inferenztypen unterschieden,<sup>34</sup> denen unterschiedliche Dimensionen des Schlussfolgerns zugrunde liegen: logisch-zwingende Inferenzen sind z. B. solche, die absolut notwendig sind, um einen kohärenten Textsinn aufzubauen, Brücken-Inferenzen verbinden Einzelinformationen miteinander und elaborative Inferenzen (auch: strategisch-bewusste Inferenzen) gehen weit über den vorgegebenen Text hinaus und verbinden den Textsinn mit dem Vorwissen der Rezipienten/-innen. Im Prinzip müssten all diese Inferenzdimensionen bei der Formulierung von Fragen zum Text berücksichtigt werden, was bislang aber nur unzureichend geschehen ist.

#### *Wiedergabeverfahren*

Will man wissen, in welchem Ausmaß und in welcher Intensität die Versuchspartner/-innen 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 Versuchspartner/-innen 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 der Methode soll sichergestellt werden, dass die Versuchspartner/-innen nicht einfach Aspekte, die ihnen vielleicht weniger wichtig erscheinen, weglassen. Statt eines strukturellen Rasters kann auch textspezifisch ein inhaltliches Raster vorgegeben werden.<sup>35</sup>

---

<sup>33</sup> ARTHUR C. GRAESSER/DAROLD HEMPHILL, Question answering in the context of scientific mechanisms, in: *Journal of Memory and Language* 30 (1991), 186-209.

<sup>34</sup> Überblick: MURRAY SINGER, Discourse inference processes, in: MORTON A. GERNSBACHER (Hrsg.), *Handbook of psycholinguistics*. San Diego (California) 1994, 479-515.

<sup>35</sup> Anwendungsbeispiel: URSULA CHRISTMANN, *Modelle der Textverarbeitung. Textbeschreibung als Textverstehen*. Münster 1989.

- *Gelenkte Reproduktion*: Im Unterschied zur freien Reproduktion werden bei der gelenkten Reproduktion einzelne Wörter, z. B. der Satz-anfang, vorgegeben; die Versuchspartner/-innen werden gebeten, die betreffenden Satzanfänge zu ergänzen (daher auch: Satz-Ergän-zungs-Methode).
- *Zusammenfassungen*: Gerade bei längeren Texten kann es je nach Er-kenntnisinteresse 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 durch die Anwendung von Makroregeln wie ‚Auslassen‘, ‚Selektieren‘, ‚Generalisieren‘ und ‚Konstruieren bzw. Integrieren‘ gebildet.<sup>36</sup> Diese Makroregeln sind rekursiv, d. h. sie können wieder auf bereits gebildete Makropropositionen angewen-det werden und führen dann zu Makropropositionen auf der nächst höheren (Abstraktions-)Ebene. Die Analyse der erhaltenen Zusam-menfassungen zeigt, welche Makroregeln im individuellen Fall an-gewendet wurden und wie und in welchem Ausmaß die Textinfor-mation verdichtet wurde. ‚Generalisierungen‘ und ‚integrative Sätze‘ dürften dabei ein tieferes Textverständnis indizieren als ‚Auslassun-gen‘ und ‚Selektionen‘. Generell zeigen Zusammenfassungen, wel-che Informationen als besonders wichtig erachtet und besonders in-tensiv verstanden und behalten wurden. Außerdem geben Sie Aufschlüsse darüber, ob das Wesentliche eines Textes verstanden wurde. Die Auswertung solcher Zusammenfassungen wird dabei deutlich erleichtert, wenn der zur Rezeption vorgelegte Text nach konventionalisierten Darstellungsregeln, -prozeduren und -katego-rien, sog. Superstrukturen<sup>37</sup>, aufgebaut ist, wie es beispielsweise in bestimmten Wissensbereichen wie Psychologie und Medizin, aber si-cherlich bisweilen auch im juristischen Bereich, der Fall ist.

Bei allen Wiedergabeverfahren ist die Auswertung zumindest bei län-geren Texten in der Regel jedoch aufwendig. Damit die quantitative und auch die qualitative Verstehensleistung gemessen werden können, müssen sowohl der vorgegebene Text als auch die Verstehensprotokol-le in Einheiten zerlegt werden (Propositionen, Sätze, *idea units*). Nach dieser Vorarbeit kann zunächst unter quantitativer Perspektive ausge-zählt werden, wie viel Einheiten behalten wurden. Unter qualitativer Perspektive muss zumindest bestimmt werden, wie viele der wiederge-gebenen Einheiten korrekt oder inhaltlich angemessen im Sinne der

---

<sup>36</sup> TEUN A. VAN DIJK, *Textwissenschaft*. München 1980.

<sup>37</sup> VAN DIJK, *Textwissenschaft* (Fn. 36).

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 Interkoder-Übereinstimmung zwischen mindestens zwei Kodierer/-innen bestimmt werden muss.<sup>38</sup>

#### *Diagnose von Wissensstrukturen*

Eine weitere Möglichkeit, das Textverstehen und den Wissenszuwachs nach der Textrezeption zu überprüfen, stellen sog. graphische Verfahren der Wissensdiagnose dar. Das Prinzip besteht darin, die Grundgedanken eines Textes und die zwischen ihnen bestehenden Relationen mit Hilfe von vorgegebenen Notationssystemen herauszuarbeiten und graphisch in Form von Netzwerken und zweidimensionalen Diagrammen (sog. *maps*) darzustellen. Die bis heute vorliegenden Verfahren, die auf der Grundlage von Theorien und Befunden zum semantischen Gedächtnis entwickelt wurden (hier insbesondere Netzwerk- und Schemansätze) unterscheiden sich vor allem hinsichtlich der Art und Anzahl der vorgegebenen Relationen sowie der Art der graphischen Veranschaulichung. Bekannt geworden ist vor allem die Netzwerktechnik von Dansereau et al., bei dem die Kerngedanken eines Textes mit Hilfe von drei Relationstypen graphisch veranschaulicht werden: hierarchische Relationen (z. B. Teil-Ganzes-Relation; Ist-Ein-Relation), Kettenrelationen (zeitliche Ordnung; kausale Sequenz) und Cluster-Relationen (Merkmale, Definitionen, Analogien).<sup>39</sup> Die Technik arbeitet primär mit innerbegrifflichen Relationen, zielt in erster Linie auf die Abbildung der vorgegebenen Textinhaltsstruktur und dürfte insbesondere für die Darstellung hierarchisch aufgebauter, fachwissenschaftlicher Texte geeignet sein.<sup>40</sup> Im Unterschied dazu legt die NAIT-Technik (*No-*

---

<sup>38</sup> Zur Inhaltsanalyse vgl. RUTH RUSTEMEYER, *Praktisch-methodische Schritte der Inhaltsanalyse. Eine Einführung am Beispiel der Analyse von Interviewtexten*. Münster 1992.

<sup>39</sup> DONALD F. DANSEREAU et al., Development and evaluation of a learning strategy training program, in: *Journal of Educational Psychology* 71/2 (1971), 64–73.

<sup>40</sup> Vgl. SIGMAR O. TERGAN, *Modelle der Wissensrepräsentation als Grundlage qualitativer Wissensdiagnostik*. Opladen 1986.

de Acquisition and Integration Technique)<sup>41</sup> ein stärkeres Gewicht auf die Abbildung von Wissensteilmengen, die über die unmittelbare Textinformation hinausgehen, indem sie beispielsweise eine Elaboration von Schlüsselkonzepten unter Rückgriff auf das eigene Vorwissen vorsieht.

Diese Verfahren<sup>42</sup> wurden zwar primär entwickelt, um die Textverarbeitung zu effektivieren, können aber auch eingesetzt werden, um den jeweiligen individuellen Wissenszuwachs zu erfassen. Dazu wird das Wissen vor und nach der Textrezeption mit Hilfe eines graphischen Verfahrens erhoben. Der Vergleich der beiden Strukturbilder ermöglicht eine Bestimmung des quantitativen und qualitativen Wissenszuwachses.

Als Beispiel für eine solche graphische Darstellung möchte ich die Heidelberger Struktur-lege-Technik<sup>43</sup> anführen, die ursprünglich zur Erhebung subjektiver Theorien entwickelt wurde, verschiedentlich aber auch (in modifizierter Form) zur Diagnose von Wissensstrukturen eingesetzt worden ist.<sup>44</sup> Danach werden die zentralen Konzepte eines rezipierten Textes auf Kärtchen geschrieben und mithilfe eines Regelsystems formaler Relationen in eine Struktur gebracht bzw. visualisiert. Die Versuchspartner/-innen erhalten dazu einen Leitfaden, in dem verschiedene Formalrelationen in alltagssprachlicher Formulierung in Abhängigkeit von Fragestellung, Problembereich und Untersuchungsstichprobe zusammengestellt und mit Beispielen erläutert sind.<sup>45</sup> Dabei werden folgende Kernrelationen unterschieden:

---

<sup>41</sup> GEORGE M. DIEKHOFF/PAMELA J. BROWN/DONALD F. DANSEREAU, A prose learning strategy training program based on network and depth-of-processing models, in: *Journal of Experimental Education* 50 (1981), 180-184.

<sup>42</sup> Überblick: CHARLES D. HOLLEY/DONALD F. DANSEREAU (Hrsg.), *Spatial learning strategies: Techniques, applications, and related issues*. London 1984; TERGAN, Wissensrepräsentation (Fn. 40); DAVID H. JONASSEN/KATHERINE BEISSNER/MICHAEL YACCI, *Structural knowledge. Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale (New Jersey) 1993.

<sup>43</sup> BRIGITTE SCHEELE/NORBERT GROEBEN, *Dialog-Konsens-Methoden*. Tübingen 1988.

<sup>44</sup> Vgl. z. B. STEFFEN-PETER BALLSTAEDT/HEINZ MANDL, *Diagnosis of knowledge structures in text learning*. Univ. Tübingen, Deutsches Institut für Fernstudien, Forschungsbericht Nr. 37. Tübingen 1985; MARTINA SCHEMANN, Diagnose von Wissensstrukturen: eine empirische Untersuchung, in: *Unterrichtswissenschaft* 23/1 (1995), 208-228.

<sup>45</sup> Zur alltagssprachlichen Flexibilisierungsversion der Dialog-Konsens-Methodik: BRIGITTE SCHEELE/NORBERT GROEBEN/URSULA CHRISTMANN, Ein alltagssprachliches Struktur-lege-Spiel als Flexibilisierungsversion der Dialog-Konsens-Methodik zum Einsatz auf verschiedenen Altersstufen, in: BRIGITTE SCHEELE (Hrsg.), *Struktur-lege-Verfahren als Dialog-Konsens-Methoden*. Münster 1992, 152-195; URSULA CHRISTMANN/NORBERT GROEBEN/MARGIT SCHREIER, Subjektive Theorien - Rekonstruktion und Dialog-Konsens, in: *SPIEL (Siegener Periodikum für internationale empirische Literaturwissenschaft)* 18/1 (1999), 138-153.

- (a) Relationen aus dem Bereich des Definierens im weiteren Sinn (SLT): z. B. das ist/das heißt; und; oder; damit/um zu; zum Beispiel/so wie;
- (b) Relationen aus dem Bereich der empirischen Hypothesen/Gesetzmäßigkeiten (SLT): z. B. führt zu; nur, wenn auch; nur, wenn nicht; je mehr, desto mehr/je weniger, desto weniger; je weniger, desto mehr/je mehr, desto weniger; je mehr;
- (c) Relationen zur Abbildung präskriptiv-deskriptiver gemischter Satzsysteme (ZMA): z. B. (für den Bereich der positiven Wertungen) *soll sein; deshalb; soll trotzdem nicht sein; führt allerdings auch zu; wird neutralisiert durch, letzte Möglichkeit;*
- (d) Relationen zur Beschreibung/Erklärung von Handlungsabfolgen (kommunikative Handlungsbeschreibung durch Flussdiagrammdarstellung): z. B. *indem; entweder ... oder ... oder; wenn, dann; und dann; und zugleich; Ziel erreicht.*

Mit solchen Relationen ist beispielsweise in einer Studie zur Rezeption des Dramas *Penthesilea* gearbeitet worden.<sup>46</sup> Aufgabe der Versuchspartner/-innen war es, nach der Rezeption des Dramas aus mehreren vorgegebenen Schaubildern eines auszuwählen, das dem eigenen Verständnis am ehesten entsprach und dieses Schaubild unter Rückgriff auf die oben angeführten Formalrelationen gemäß den eigenen Vorstellungen zu verändern (Konzepte streichen, Konzepte hinzufügen, Inhalte verändern, das Zueinander der Konzepte mit Hilfe der Relationen zu modifizieren und Belegstellen aus dem Text anzugeben). Abbildung 4 zeigt das psychoanalytische Interpretationsschaubild und verdeutlicht zugleich, wie Strukturbilder, die mithilfe der Heidelberger-Struktur-Legge-Technik erstellt werden, aussehen können.

Wie bei der freien Wiedergabe auch, können solche Strukturbilder nach quantitativen Gesichtspunkten (Anzahl der unterschiedenen Konzepte, Anzahl der Relationen) und inhaltsanalytisch ausgewertet werden. Gegenüber der freien Wiedergabe besteht der Nachteil des Verfahrens darin, dass die Versuchspartner/-innen vorab mit der graphischen Darstellung vertraut gemacht werden müssen. Der Vorteil ist darin zu sehen, dass in höherem Maße gerade auch überdauernde Wissensstrukturen erfasst werden, als dies bei anderen Verfahren der Fall ist.

---

<sup>46</sup> NORBERT GROEBEN, Interpretationsansätze zur *Penthesilea* und Gender-Sozialisation – eine dialog-hermeneutische Rezeptionsstudie, in: CARL PIETZCKER (Hrsg.), *Freiburger Literaturpsychologische Gespräche*. Freiburg 1998, 100–120.

Abb. 4: Schaubild eines psychoanalytischen Interpretationsansatzes  
(nach Groeben, 1998, S. 114 f.), Teil 1/2.

Abb. 4: Schaubild eines psychoanalytischen Interpretationsansatzes  
(nach Groeben, 1998, S. 114 f.), Teil 2/2.



### *Handlungsumsetzung*

Als letztes und vermutlich komplexestes Verfahren zur Erfassung des Textverstehens soll die Güte der Umsetzung des erworbenen Wissens in konkretes Handeln zumindest noch angesprochen werden. Dieses Verfahren bietet sich vor allem bei Sachtexten an, die die Durchführung bestimmter Handlungen anleiten wollen, und ist bisher vor allem im Umgang mit technischen Geräten eingesetzt worden. Erfasst wird hier z. B. die Anzahl und Abfolge der Handlungsschritte sowie die Genauigkeit und Schnelligkeit der Handlungsausführung.<sup>47</sup> Ob und in welchem Ausmaß dieses Verfahren auch zur Erhebung des Verstehens juristischer Texte brauchbar ist, hängt letztlich vom Inhalt des betreffenden Textes ab. Ein möglicher Anwendungsbereich könnten Texte zur Straßenverkehrsordnung sein, deren Verständnis beispielsweise im Rahmen von Computersimulationen überprüft werden könnte.

## 4. Notwendigkeit von Evaluationsmodellen

Mit Hilfe der beschriebenen Verfahren wird das Ergebnis des Verstehensprozesses für bestimmte Texte gemessen. Die resultierenden Daten kann man grundsätzlich in zwei Richtungen interpretieren: zum einen in Bezug auf die individuelle Verarbeitungskompetenz des/der jeweiligen Rezipienten/-in, zum anderen in Bezug auf das zugrunde liegende Ausgangsmaterial, den jeweiligen Text (s. o. Kap. 1.). Die in unserem Zusammenhang zentrale zweite Interpretationsrichtung setzt eine möglichst umfassende, ggf. repräsentative Stichprobe von Rezipienten/-innen voraus. Dies ist das quantitative Problem bei der empirischen Messung von Textverständlichkeit.

Daneben sind aber auch qualitative Problemdimensionen zu klären, wenn man eine solche empirische Verständlichkeitserhebung in der Praxis einsetzen will. Dazu gehört zunächst einmal die Frage, welches Ausmaß an Verständlichkeit man prinzipiell anstreben will. Denn ein beliebiger empirischer Messwert sagt bei aller Präzision nicht von selbst etwas über die verständlichkeitsbezogene Güte des untersuchten Textes aus. Dazu braucht es vielmehr eine normative Bestimmung des angestrebten Ausmaßes von Verständlichkeit, ein Evaluationsmodell. Die beiden wichtigsten Varianten einer solchen Modellierung bestehen in der Vorstellung einer ‚maximalen‘ vs. einer ‚optimalen‘ (im Sinne

---

<sup>47</sup> Überblick: URSULA CHRISTMANN/NORBERT GROEBEN, Anforderungen und Einflussfaktoren bei Sach- und Informationstexten, in: NORBERT GROEBEN/BETTINA HURRELMANN (Hrsg.), *Lesekompetenz in der Mediengesellschaft*. Weinheim 2002, 150–173.

einer mittleren) Verständlichkeit.<sup>48</sup> Das Konzept der maximalen Verständlichkeit ist – implizit – mit den (Hamburger) Einschätzskalen verbunden (s. o. Kap. 3.1), mit deren Hilfe z. B. bei der Verbesserung von Texten eine möglichst große Vereinfachung etc. erreicht werden soll, also eine Maximierung der Verständlichkeit. Dem hat Groeben unter der Verbindung von kognitions- und motivationspsychologischen Perspektiven das Konzept einer optimalen Verständlichkeit entgegengesetzt, d. h. die normative Vorstellung, dass Texte durchaus auch eine kognitive Herausforderung darstellen sollen, weil damit eine größere Verarbeitungsmotivation und -tiefe erreicht wird, die wiederum zu größeren Behaltenseffekten führt.<sup>49</sup>

Es sprechen also nicht nur praktische, sondern auch theoretische Gründe dafür, im Bereich der juristischen Texte keine maximale, sondern eine mittlere Verständlichkeit als Zielkriterium anzusetzen. Dabei bleibt allerdings das Problem, was für einzelne Leser/-innen bzw. Lesergruppen als mittlere Verständlichkeit gelten kann. Entsprechend der eingangs eingeführten Text-Leser-Interaktion als Grundstruktur des Verarbeitungsprozesses hängt das von den individuellen leserseitigen Voraussetzungen ab, vor allem in Bezug auf das Vorwissen, aber auch die sprachliche Kompetenz und Motivation.<sup>50</sup> Gerade im Bereich der Anwendungsforschung befasst sich mit diesem Problem der Theorie-Ansatz der ‚Aptitude-Treatment-Interaction‘ (Wechselwirkung zwischen Lehrmethode und Lernermerkmalen).<sup>51</sup> Innerhalb der evaluativen Modellierung von Textverständlichkeit für den juristischen Bereich müsste hier also von der Rechtswissenschaft elaboriert werden, welche rezipientenseitigen Merkmale als entscheidend für den Verarbeitungsprozess von juristischen Texten angesetzt werden können und sollen. Allerdings gibt es ein strukturelles Problem dabei. So können z. B. im Bereich der Schule aus den theoretischen Modellen zur Aptitude-Treatment-Interaction, die sich empirisch bewähren, Konsequenzen für die Individualisierung von Unterricht gezogen werden, d. h. es ist für un-

---

<sup>48</sup> Vgl. NORBERT GROEBEN, Verständlichkeitsforschung unter Integrationsperspektive: Ein Plädoyer, in: HEINZ MANDL (Hrsg.), *Zur Psychologie der Textverarbeitung*. München 1981; NORBERT GROEBEN/URSULA CHRISTMANN, Textoptimierung unter Verständlichkeitsperspektive, in: GERD ANTOS/HANS P. KRINGS (Hrsg.), *Textproduktion. Ein interdisziplinärer Forschungsüberblick*. Tübingen 1989, 165–196.

<sup>49</sup> GROEBEN, Verständlichkeit von Unterrichtstexten (Fn. 6).

<sup>50</sup> Vgl. GROEBEN, Leserpsychologie (Fn. 3).

<sup>51</sup> LEE CRONBACH/RICHARD SNOW, Aptitudes and instructional methods: A handbook of research in interactions. New York 1977; JOACHIM HASEBROOK, Aptitude-Treatment-Interaction, in: DETLEF H. ROST (Hrsg.), *Handwörterbuch Pädagogischer Psychologie*. Weinheim 1998, 10–13; RICHARD SNOW/JUDY SWANSON Instructional psychology: Aptitude, adaption, and assessment, in: *Annual Review of Psychology* 43 (1992), 583–626.

terschiedliche Gruppen von Schülern/-innen eine unterschiedliche Textschwierigkeit anzustreben, im Extremfall mit Abstufungen für jedes einzelne Individuum entsprechend seiner kognitiven und motivationalen Voraussetzungen. Im juristischen Bereich ist eine solche Variation von Schwierigkeit z. B. für Gesetzestexte, die allgemeingültig gelten und formuliert sein müssen, selbstverständlich nicht möglich. Das bedeutet, dass eine adressatenspezifische Formulierung nur für interpretative Kommentare etc. denkbar ist. Die Explikation von Evaluationsmodellen der Verständlichkeit für den juristischen Bereich wird daher vor allem auch die Ausarbeitung von Mehr-Stufen-Modellen der Informationsvermittlung umfassen müssen. Das derzeitige Kommentarwesen bezieht sich bislang primär auf die Zielgruppe der Richter und Anwälte. Kommentare für andere Zielgruppen mit geringer juristischer Vorerfahrung (man denke etwa an den in Verbänden organisierten Zusammenschluss von Geschädigten oder an Kleinunternehmer/-innen) liegen praktisch nicht vor. Eine Ausnahme stellen hier die vereinzelt ausgearbeiteten Studienkommentare (für Studierende der Rechtswissenschaft) dar. Sie verweisen nicht zuletzt darauf, dass eine adressatenspezifische Formulierung interpretativer Kommentare gerade auch für andere Zielgruppen, die nicht über juristisches Expertenwissen verfügen, notwendig und sinnvoll sein dürfte. Empirische Verständlichkeitserhebung kann erst dann praktisch wirksam werden, wenn sie mit solchen Modellen eines mehrstufigen Informationsflusses verbunden wird, in denen die Adressatenspezifität auch theoretisch zu begründen ist. Innerhalb einer solchen Modellierung ist mit den beschriebenen Methoden der Verständlichkeitsmessung dann zu überprüfen, ob die angezielten Grade einer optimalen (mittleren) Verständlichkeit erreicht werden oder nicht.

## Literatur

- Ballstaedt, Steffen-Peter/Mandl, Heinz, Diagnosis of knowledge structures in text learning. Univ. Tübingen, Deutsches Institut für Fernstudien, Forschungsbericht Nr. 37. Tübingen 1985.
- Bergler, Reinhold (Hrsg.), Das Eindrucksdifferential. Theorie und Technik. Bern, Stuttgart, Wien 1975.
- Bortz, Juergen, Lehrbuch der empirischen Forschung. Berlin 1984.
- Carver, Ronald P./Darby, Charles A., Development and evaluation of a test of information storage during reading. Journal of Educational Measurement, 8/1 (1971), 33-44.

- Carver, Ronald P., Analysis of the ‚chunked‘ test items as measures of reading and listening comprehension. *Journal of Educational Measurement*, 7/3 (1970), 141-149.
- Christmann, Ursula, Modelle der Textverarbeitung. Textbeschreibung als Textverstehen. Münster 1989.
- Christmann, Ursula/Groeben, Norbert, Textverstehen, Textverständlichkeit – Ein Forschungsüberblick unter Anwendungsperspektive, in: Krings, Hans-Peter (Hrsg.), *Wissenschaftliche Grundlagen der Technischen Kommunikation*. Tübingen 1996, 129-189.
- Christmann, Ursula/Groeben, Norbert, Die Rezeption schriftlicher Texte, in: Günther, Hartmut/Ludwig, Otto (Hrsg.), *Schrift und Schriftlichkeit. Ein interdisziplinäres Handbuch internationaler Forschung*. Berlin, New York 1996, 1536-1545.
- Christmann, Ursula/Groeben, Norbert, Psychologie des Lesens, in: Franzmann, Bodo/Hasemann, Klaus/Löffler, Dietrich/Schön, Erich (Hrsg.), *Handbuch Lesen*. München 1999, 145-223.
- Christmann, Ursula/Groeben, Norbert, Anforderungen und Einflussfaktoren bei Sach- und Informationstexten, in: Groeben, Norbert/Hurrelmann, Bettina (Hrsg.), *Lesekompetenz in der Mediengesellschaft*. Weinheim 2002, 150-173.
- Christmann, Ursula/Scheele, Brigitte, Subjektive Theorien über (un-)redliches Argumentieren, in: König, Eckard/Zedler, Peter (Hrsg.), *Bilanz qualitativer Forschung*. Weinheim 1995, 63-99.
- Christmann, Ursula/Groeben, Norbert/Schreier, Margrit, Subjektive Theorien – Rekonstruktion und Dialog-Konsens, in: SPIEL (Siegener Periodikum für internationale empirische Literaturwissenschaft) 18/1 (1999), 138-153.
- Coleman, Edmund B./Miller, Gerald R., A measure of information gained during prose learning. *Reading Research Quarterly*, 3/3 (1968), 369-386.
- Cronbach, Lee/Snow, Richard, E., *Aptitudes and instructional methods: A handbook of research in interactions*. New York 1977.
- Dansereau, Donald F./Collins, Karen W./McDonald, Barbara A./Holley, Charles D./Garland, John/Drekhoff, George/Evans, Selby, H., Development and evaluation of a learning strategy training program. *Journal of Educational Psychology* 71/1 (1971), 64-73.
- Diekhoff, George M./Brown, Pamela J./Dansereau, Donald F., A prose learning strategy training program based on network and depth-of-processing models, in: *Journal of Experimental Education* 50 (1981), 180-184.
- Graesser, Arthur C./Hemphill, Darold, Question answering in the context of scientific mechanisms, in: *Journal of Memory and Language* 30 (1991), 186-209.
- Groeben, Norbert, *Die Verständlichkeit von Unterrichtstexten*. 2. Aufl., Münster 1978.

- Groeben, Norbert, Verständlichkeitsforschung unter Integrationsperspektive: Ein Plädoyer, in: Mandl, Heinz (Hrsg.), Zur Psychologie der Textverarbeitung. München 1981.
- Groeben, Norbert, Leserpsychologie I: Textverständnis - Textverständlichkeit. Münster 1982.
- Groeben, Norbert, Interpretationsansätze zur Penthesilea und Gender-Sozialisation - eine dialog-hermeneutische Rezeptionsstudie, in: Pietzcker, Carl (Hrsg.), Freiburger Literaturpsychologische Gespräche. Freiburg 1998, 100-120.
- Groeben, Norbert/Ursula, Christmann, Textoptimierung unter Verständlichkeitsperspektive, in: Antos, Gerd/Krings, Hans P. (Hrsg.), Textproduktion. Ein interdisziplinärer Forschungsüberblick. Tübingen 1989, 165-196.
- Günther, Ulrich, Lesen im Experiment. Linguistische Berichte 122 (1989), 283-320.
- Hasebrook, Joachim, Aptitude-Treatment-Interaction, in: Rost, Detlef H. (Hrsg.), Handwörterbuch Pädagogischer Psychologie. Weinheim 1998, 10-13.
- Hofer, Manfred, Textverständlichkeit: Zwischen Theorie und Praxeologie, in: Unterrichtswissenschaft 2 (1976), 143-150.
- Holley, Charles D./Dansereau, Donald F. (Hrsg.), Spatial learning strategies: Techniques, applications, and related issues. London 1984.
- Jonassen, David H./Beissner, Katherine/Yacci, Michael, Structural knowledge. Techniques for representing, conveying, and acquiring structural knowledge. Hillsdale (New Jersey) 1993.
- Klare, George R., The measurement of readability. Ames 1963.
- Langer, Inghard/Schulz v. Thun, Friedemann/Tausch, Reinhard, Verständlichkeit in Schule, Verwaltung, Politik und Wissenschaft. München 1974.
- Osgood, Charles E./Suci, George J./Tannenbaum, Percy H., The measurement of meaning. Urbana (Illinois) 1957.
- Rickheit, Gert/Strohner, Hans, Grundlagen der kognitiven Sprachverarbeitung. Tübingen 1993.
- Rustemeyer, Ruth, Praktisch-methodische Schritte der Inhaltsanalyse. Eine Einführung am Beispiel der Analyse von Interviewtexten. Münster 1992.
- Scheele, Brigitte/Groeben, Norbert, Dialog-Konsens-Methoden. Tübingen 1988.
- Scheele, Brigitte/Groeben, Norbert/Christmann, Ursula, Ein alltagssprachliches Struktur-lege-Spiel als Flexibilisierungsversion der Dialog-Konsens-Methodik zum Einsatz auf verschiedenen Altersstufen, in: Scheele, Brigitte (Hrsg.), Struktur-lege-Verfahren als Dialog-Konsens-Methoden. Münster 1992, 152-195).
- Schemann, Martina, Diagnose von Wissensstrukturen: eine empirische Untersuchung. Unterrichtswissenschaft 23/1 (1995), 208-228.
- Shannon, Claude E., Prediction and entropy of printed English. Bell System Technical Journal 30 (1951), 50-64.

- Singer, Murray, Discourse inference processes, in: Gernsbacher, Morton A. (Hrsg.), *Handbook of psycholinguistics*. San Diego (California) 1994, 479-515.
- Snow, Richard/Swanson, Judy, Instructional psychology: Aptitude, adaption, and assessment, in: *Annual Review of Psychology* 43 (1992), 583-626.
- Spearritt, Donald, Identification of subskills of reading comprehension by maximum-likelihood factor analysis, in: *Reading Research Quarterly* 8 (1972), 92-111.
- Taylor, Wilson L., Cloze-procedure. A new tool for measuring readability, in: *Journalism Quarterly* 30 (1953), 415-433.
- Taylor, Wilson L., Recent development in the use of cloze procedure, in: *Journalism Quarterly* 33 (1956), 42-48.
- Tergan, Sigmar O., *Modelle der Wissensrepräsentation als Grundlage qualitativer Wissensdiagnostik*. Opladen 1986.
- Tuinman, Jaap, The removal of information procedure. *Journal of Reading Behavior* 3/2 (1971), 44-50.
- van Dijk, Teun A., *Textwissenschaft*. München 1980.
- Weaver, Wendell W./Bickley, Albert C., Sources of information for response to reading test items. *Proceedings of the 75th Annual Convention of the American Psychological Association*, 2, 1967, 293-294.
- Weltner, Klaus, *Informationstheorie und Erziehungswissenschaften*. Quickborn bei Hamburg 1970.