

II. Zur Dokumentation und maschinellen Bearbeitung
von Zeitungstexten in der Außenstelle Bonn
von Manfred W. Hellmann

Übersicht

Allgemeiner Teil: Zur Dokumentation bei Zeitungstexten (42)

1. Ausgangslage (42)
2. Allgemeines zum Medium "Tageszeitungen" (42)
 - 2.1. Heterogenität im einzelnen (43)
 - 2.2. Ambivalenz und Homogenität im ganzen (45)
 - 2.2.1. Führungs- und Beeinflussungsmittel einer Gruppe (45)
 - 2.2.2. Kommunikationsmittel zur Leserschaft (45)
3. Wissenschaftliche Fragestellungen: (46)
Einzelanalyse - Dokumentation
4. Dokumentation bei Zeitungstexten (47)
 - 4.1. Allgemeine Bedingungen (47)
 - 4.2. Zeitliche Einheiten und Einteilungen (48)
 - 4.3. Mengenproblem und Kapazität (49)
 - 4.4. Begründbarkeit einer Textauswahl (51)
 - 4.5. Bedingungen einer zureichenden Auswahl (52)
 - 4.5.1. Allgemeine Feststellungen (52)
 - 4.5.2. Qualitative Analyse (54)
 - 4.5.2.1. Feststellung spezifischer Dominanzen (54)
 - 4.5.3. Quantitative Analyse (57)
 - 4.5.3.1. Definition der Mengeneinheiten (57)
 - 4.5.4. Feststellung der Seitenhäufigkeiten (59)
 - 4.5.4.1. Seitenhäufigkeiten in der Zielmenge (59)
 - 4.5.4.2. Seitenhäufigkeiten in der Auswahlmenge (61)
 - 4.6. Ermittlung der Modellmenge (62)
 - 4.6.1. Aufnahmeeinheit (62)
 - 4.6.2. Stichprobeneinheit (62)
 - 4.6.3. Aufnahmeintervall, Aufnahmedichte (63)
 - 4.6.4. Zahl der Stichproben (63)
 - 4.6.5. Wahrscheinlichkeitsquotient (64)
 - 4.6.6. Aufnahmezeitraum (64)
 - 4.7. Zusammenfassung der notwendigen Schritte (64)

Spezieller Teil: Dokumentation und maschinelle Verarbeitung der Bonner
Zeitungsjahrgänge (66)

5. Allgemeines (66)
6. Zur Kapazität (66)
7. Bestimmung der Gesamtmengen (67)
 - 7.1. Begründung zur Wahl der Ostjahrgänge (67)
 - 7.2. Begründung zur Wahl der Westjahrgänge (68)
 - 7.3. Vergleichbarkeit der ausgewählten Zeitungen (69)
8. Bestimmung der Zielmengen (70)
 - 8.1. Abgrenzung gegenüber Gesamtmengen (70)
 - 8.2. Quantitative Beschreibung der Jahrgänge (71)
9. Dominanzen (72)
10. Auswahlmodus (72)
 - 10.1. Auswahlmodus für ND 64 und 54 (72)
 - 10.2. Auswahlmodus für WELT 64 (74)
11. Modus der Stichprobenverteilung (74)
 - 11.1. Horizontaler Turnus (74)
 - 11.2. Vertikaler Turnus (75)
 - 11.3. Verschiebungen und Abweichungen (75)
12. Zusätze zur Modellmenge (76)
13. Besondere Gesichtspunkte bei der Übertragung der Texte auf Datenträger (77)
 - 13.1. Artikel (77)
 - 13.2. Schreibkonventionen (77)
 - 13.2.1. Informationskonstanz bei einzelnen Wörtern (Ersatzzeichen und Sonderzeichen) (78)
 - 13.2.2. Informationskonstanz bei Wortgruppen (Transkriptionen) (79)
 - 13.2.3. Informationskonstanz des Artikel-Charakters (Informationskarten) (80)
14. Arbeitsgang (Schema) (82)
 - 14.1. Erläuterungen zum Arbeitsablauf (84)
15. Übersicht über den Stand der Arbeiten (87)
 - 15.1. Texte (87)
 - 15.2. Auswertung (89)

Anhänge :

- Anhang Ia) Seitenberechnung für ND 64, Welt 64, ND 54 (91)
b) Berechnung der Häufigkeit jeder Seite in den Auswahlmengen (94)
- Anhang IIa) Liste der für die Modellmengen aufgenommenen Seiten (95)
b) Liste der zusätzlich aufgenommenen Seiten (97)
- Anhang III Auszug aus den Erläuterungen zum Ausfüllen der Informations-
karten (98)
- Anhang IV Übersicht über die in der Außenstelle vorhandenen wichtigeren
Programme (108)
- Anmerkungen (115)

Allgemeiner Teil : Zur Dokumentation bei Zeitungstexten

1. Ausgangslage

Als die Außenstelle Bonn im August 1964 gegründet wurde, erhielt sie die Aufgabe zugewiesen, die seit der politischen Teilung Deutschlands eingetretenen Veränderungen der deutschen Sprache in den beiden Teilen Deutschlands zu beobachten, zu registrieren und zu analysieren. Es bestand Klarheit darüber, daß diese Aufgabe nur Teil der größeren, die deutsche Sprache der Gegenwart überhaupt wissenschaftlich zu erforschen, sein kann.

Als Materialgrundlage für alle Untersuchungen kamen von vorn herein nur gedruckte, veröffentlichte Texte in Betracht, da uns die Beschaffung von Material gesprochener Sprache aus der DDR, vor allem von spontan gesprochener Umgangssprache, jedenfalls in methodisch ausreichender Breite aus bekannten Gründen nicht möglich ist. ¹⁾

Die Untersuchungen sollten sich vor allem, wenn auch nicht ausschließlich, auf den Wortschatz richten, und zwar auf den Wortschatz allgemein, sofern er überhaupt für sprachliche Veränderungen in Betracht kommt. Es sollte jedenfalls nicht ein bestimmtes Sachgebiet von vorn herein bevorzugt werden.

Das Textmaterial sollte aktuell sein und weder zu sehr von einigen oder einem bestimmten Verfasser stammen noch ausgeprägt fachsprachlichen Charakter haben.

Schließlich sollte das Material so gewählt werden, daß es ost-west-vergleichende Studien zuläßt, d.h. es sollte eine nach Zweck und Ziel seiner Entstehung, Aufbau, Erscheinungs- und Verbreitungsart möglichst vergleichbare Struktur besitzen. ²⁾ Diesen Bedingungen zufolge entschied sich die Außenstelle für die Aufnahme von Zeitungstexten, und zwar von Texten aus Tageszeitungen großer Verbreitung und überregionaler Bedeutung.

2. Allgemeines zum Medium "Tageszeitung"

Tageszeitungen der bezeichneten Art weisen eine Reihe gemeinsamer charakteristischer Züge auf, die den vorher genannten Bedingungen entsprechen :

2.1. Heterogenität im einzelnen

1. Vielheit von Verfassern

An jeder Zeitungsausgabe arbeitet eine Vielzahl von Redaktionsmitgliedern, Korrespondenten und anderen Autoren mit; theoretisch sind ebenso viele individuelle Schreibgewohnheiten zu erwarten. Ausgleichend wirken dagegen: der Einfluß der Redaktionsleitung oder einer ihr übergeordneten Instanz, die normierende Kraft des "teamworks" ("Kollektivs") in der Redaktion selbst, der Einfluß der Agenturen (in der DDR ADN, in der BRD vor allem dpa, ap und upi).

2. Vielfalt von Sachgebieten

Theoretisch ist die Zahl der Sachgebiete, die in einer Zeitung Berücksichtigung finden können, nahezu unbegrenzt groß. Allerdings wird jede Zeitung bestimmte Sachgebiete mehr pflegen als andere; allgemeine Tageszeitungen also überwiegend Sachgebiete von allgemeinerem Interesse. Die Hauptsachgebiete, die in nahezu jeder Ausgabe jeder großen Tageszeitung vertreten sind, werden in "Sparten" oder Ressorts zusammengefaßt; die Gliederung und Bezeichnung dieser Ressorts (am häufigsten Politik, Wirtschaft, Sport, Kulturelles, Unterhaltung, Lokales o.ä.) ist wie ihr Umfang von Zeitung zu Zeitung verschieden. Auch innerhalb einer Zeitung können die Sparten nach Rang und Breite der Behandlung wechseln³⁾. Die Hauptsachgebiete fächern sich in eine systematisch nur schwer zu fassende Fülle von engeren und engsten Sachgebieten auf⁴⁾. Überschneidungen mehrerer Sachgebiete sind häufig. Sie treten besonders dann auf, wenn Zeitungen eine Sparte besonders pflegen: eine stark politisch orientierte Zeitung (wie das ND) behandelt auch kulturelle oder wirtschaftliche Zusammenhänge unter einem politischen Aspekt, bei Wirtschaftszeitungen verhält es sich entsprechend anders.

3. Vielfalt an Themen⁵⁾

Bedingt durch das ihr wesensgemäße Streben nach Aktualität, durch ihre Bindung an den Strom der Zeit ist die Zeitung einem ständigen Wechsel der Themen unterworfen bzw. auf diesen angewiesen. Nur sehr wenige Themen können sich

über längere Zeit hin halten⁶⁾. Die Zahl der Themen muß als beliebig groß angesetzt werden; sie wird begrenzt ausschließlich durch den unterschiedlichen Informations- (Propaganda-, Werbe-, Sensations- usw.) -wert, den die Redaktion den Ereignissen, Sachverhalten oder Gegenständen, die ihr zur Kenntnis kommen, zumißt.

4. Mehrzahl an Zielen oder Zwecken des Schreibens

In den verschiedenen "Artikeln"⁷⁾ einer Zeitung werden ersichtlich verschiedene, wechselnde Ziele oder Zwecke verfolgt. Neben den "klassischen" publizistischen Zielen der Unterrichtung (Information), Beeinflussung (Meinungs- und Verhaltensbildung), Belehrung und Unterhaltung, die allerdings im Einzelfall selten rein in Erscheinung treten, könnte man noch Werbung (Bedarfsweckung und Bedarfslenkung) und Aufruf (Aufforderung zum Handeln oder bestimmten Verhaltensweise) als publizistische Ziele eigenen Charakters definieren⁸⁾.

5. Vielfalt an Schreib- bzw. Mitteilungsformen⁹⁾

Schon der im engeren Sinne redaktionelle Teil einer Zeitung weist eine Fülle verschiedener "Artikelformen" auf (Leitartikel, Kommentare, Berichte, Meldungen, Glossen usw.), daneben finden sich zahlreiche Sonderformen, die teils von der Redaktion, teils von privaten Auftraggebern zu verantworten sind, wie Romanabdrucke, Leserbriefe, Anzeigen, Tabellen, Wetterberichte, Abdrucke von Reden, Dokumenten, amtlichen Bekanntmachungen usw. Sie alle sind außerdem nach Umfang und Aufmachung (Schriftgrad der Überschrift, Spaltenzahl, Platzierung) unterschieden.

Die hier angedeutete außerordentliche Heterogenität, die für alle großen Tageszeitungen kennzeichnend ist, macht es wahrscheinlich, daß das Material auch sprachlich vielfältig, vielschichtig differenziert ist.¹⁰⁾ Sie stellt den Bearbeiter allerdings auch vor besondere Probleme, wenn er trotzdem allgemeingültige Aussagen machen will.

2.2. Ambivalenz und Homogenität im ganzen

Besondere Eigentümlichkeiten des Mediums "Tageszeitung" ergeben sich aus ihrer Eigenschaft als Massenkommunikationsmittel; sie haben, ebenso wie die Vielfalt in thematischer, inhaltlicher und formaler Hinsicht, große Bedeutung auch für die Sprache (Sprachgebrauch) der Zeitung.

Zwei Gesichtspunkte sind zu unterscheiden:

2.2.1. Zeitungen sind **O r g a n** für die (einheitliche oder nicht einheitliche) Meinung oder die Absichten **d e r R e d a k t i o n** bzw. für die Meinung oder Absichten derer, denen die Redaktion die Spalten ihrer Zeitung öffnet, etwa des Herausgebers, der Regierung, politischer, wirtschaftlicher, sozialer, weltanschaulicher (Interessen-) Gruppen bzw. **d e r d o m i n i e r e n d e n** oder allein herrschenden unter diesen **G r u p p e n** (wie der SED-Führungsgruppe in der DDR.) Jede Zeitung ist also Verbreitungsmittel für die Seh-, Denk- und damit auch Redeweise der Redaktion bzw. der sie beeinflussenden Gruppen oder Gruppe. Sie wird auch sprachlich deren Eigentümlichkeiten widerspiegeln. Der Grad der Beeinflussung und damit der gruppensprachlichen Eigentümlichkeiten ist allerdings sehr verschieden; er wird dort gering sein, wo in einer Zeitung bzw. in der Redaktion mehrere Gruppenmeinungen konkurrieren; er wird da am größten sein, wo eine Gruppe nicht nur eine bestimmte Zeitung, sondern die veröffentlichte Meinung überhaupt beherrscht, wo die Zeitung sich selbst als Verbreitungsmittel dieser einen Meinung versteht.

2.2.2. Zeitungen müssen aber selbst dann, wenn sie sich im bezeichneten extremen Sinn als Verbreitungsmittel eines Meinungsmonopols verstehen, Rücksicht nehmen auf die Aufnahmebereitschaft und Aufnahmefähigkeit ihrer Leserschaft, d.h. bei weit verbreiteten Tageszeitungen: einer breit gestreuten und differenzierten Leserschaft¹¹⁾. Im Normalfall müssen sie allgemeinverständlich bleiben und dürfen nur bis zu einem gewissen Grade enger fach-

oder gruppensprachlichen Einflüssen nachgeben. Der Grad der Rücksichtnahme und des Bemühens um Allgemeinverständlichkeit ist auch hier verschieden; er wechselt nicht nur von Zeitung zu Zeitung, sondern auch innerhalb der Sparten und Ausgaben derselben Zeitung; man läßt gelegentlich bewußt auch enger fachlich usw. gebundene Redeweisen zu, um den Ansprüchen einer differenzierten Leserschaft gerecht zu werden, meist jedoch ohne ein einmal gewonnenes mittleres Niveau zu sehr zu überschreiten.

Jede große Tageszeitung enthält also - auch sprachlich - beide Komponenten. Die jeweilige Position zwischen den beiden extremen Möglichkeiten bestimmt jedoch wesentlich den individuellen Charakter einer Zeitung.

Diese doppelte Eigenheit der Zeitung, verbunden mit der ihr ebenso eigentümlichen Verpflichtung zur Aktualität, läßt die Zeitung als besonders geeignetes Medium für Untersuchungen im Rahmen des der Außenstelle gestellten Arbeitsthemas erscheinen. Je nachdem, ob im Rahmen spezieller Untersuchungen einmal mehr Gewicht auf die Komponente der politisch-geistig werbenden und führenden Sprache oder mehr auf die der allgemein kommunikativen Sprache gelegt wird, müssen oder können dann noch Texte weniger ambivalenten Charakters hinzugezogen werden.

3. Wissenschaftliche Fragestellungen

(Einzelanalysen und Dokumentation)

Das der Außenstelle gestellte Rahmenthema verlangt sowohl die Untersuchung einer ganzen Anzahl verschiedenster, teils engerer, teils weiterer Spezialgebiete, als auch Sammlungen und Untersuchungen zum Wortschatz allgemein. Für alle diese, im einzelnen noch nicht übersehbaren, Untersuchungen mußte das Material erst beschafft und bereitgestellt werden. So weit wie möglich sollten für diese Zwecke die Mittel der elektronischen Datenverarbeitung eingesetzt werden.

Es war allerdings klar, daß die Außenstelle weder den Ehrgeiz noch auch je die Möglichkeit haben würde, die im Rahmen des Gesamtthemas anfallenden einzelnen Aufgaben selbst zu bearbeiten. Der Zielsetzung des Instituts entsprechend sollte daher damit begonnen werden, nicht nur für den Bedarf der Außenstelle selbst und des Instituts, sondern auch anderer interessierter Forscher Material bereitzustellen. Gerade im Hinblick auf die Benutzung durch andere Wissenschaftler mußte die Möglichkeit für eine möglichst große Zahl verschiedener Fragestellungen geschaffen werden, und zwar derart, daß, sofern sich die Fragestellung aus dem aufgenommenen Material nicht voll beantworten läßt, wenigstens ein Grundmaterial geliefert werden kann, das dann vom betreffenden Forscher durch spezieller ausgewähltes Material ergänzt werden kann. Damit trat neben und vor die Aufgabe, Einzelanalysen zu erstellen, die Aufgabe der Dokumentation¹²⁾. Entsprechend der vorher erläuterten Beschränkung auf das Medium "Zeitungen" handelt es sich vorerst um Dokumentation bei Zeitungstexten.

Die folgenden Ausführungen beziehen sich grundsätzlich auf diese.

4. Dokumentation bei Zeitungstexten

4.1. Allgemeine Bedingungen

An eine Dokumentation im bezeichneten Sinne werden bestimmte Anforderungen gestellt:

1. Die Texte müssen in ausreichender Vielfalt aufgenommen werden. Diese Bedingung wird durch die Eigenart des Mediums "Zeitung" erfüllt (siehe oben).
2. Die Texte müssen in ausreichend großer Menge aufgenommen werden, und zwar in so großer, daß
 - a) Zufälligkeiten in der Zusammenstellung der einzelnen Textteile (Teiltex-te) ausgeglichen werden,

- b) auch zur Bearbeitung speziellerer Fragestellungen noch genügend Material zur Verfügung gestellt werden kann,
 - c) Rückschlüsse auf größere Zusammenhänge sprachlicher Art möglich sind.
3. Es darf nichts von vorn herein ausgeschlossen werden, was sprachwissenschaftlich interessant werden könnte.
 4. Die Texte sollen aus möglichst in sich geschlossenen einzelnen Texteinheiten bestehen.
 5. Es darf bei der Aufnahme möglichst kein Informationsverlust eintreten; dazu ist u.a. eine umfassende, genaue Kennzeichnung der einzelnen Texteinheiten erforderlich.
 6. Das Material muß insgesamt und in seinen Teilen möglichst schnell und leicht zugänglich und reproduzierbar sein.

Schon im Hinblick auf Punkt 6 erweist sich das Mittel der maschinellen (elektronischen) Datenverarbeitung als unumgänglich.

4.2. Zeitliche Einheiten und Einteilungen

Zeitungen sind, im Gegensatz zu Büchern, kein in sich geschlossener, abgeschlossener Zusammenhang gestalteter Sprache, sondern ein kontinuierlicher, sich in bestimmten, regelmäßigen Abständen (Tagen, Wochen usw.) erneuernder Strom von Sprache.

Markierungen, die zum Ziel haben, innerhalb dieses Stroms "Anfang" und "Ende" zu setzen, haben - abgesehen von der durch Drucktechnik und Erscheinungsweise gegebenen Unterteilung in einzelne Ausgaben - alle etwas Unverbindliches an sich; die gebräuchlichste Einteilung ist jedoch die nach Jahrgängen¹³⁾. In jedem Fall ist die Einheit des Jahrgangs die größte

Einteilung, die sich aus den Zeitungen selbst gewinnen läßt¹⁴⁾; es ist kein allgemeines Kriterium erkennbar, nach welchem man mehrere Jahrgänge zu einer Einheit zusammenfassen könnte.

Andere Einteilungen nach kleineren Einheiten, etwa nach Quartalen, Monaten oder Wochen, sind denkbar. Welche Einheit oder Einheiten man zur Grundlage der Dokumentation wählt, hängt ausschließlich von den Fragestellungen ab, die man an das Material zu stellen beabsichtigt, bzw. von dem Zweck, für den es bestimmt ist.

Grundsätzlich muß gelten: Je allgemeiner oder vielschichtiger die schon formulierten oder zu erwartenden Fragestellungen sein können, um so ausgedehnter muß das Material und der Zeitraum sein, aus dem es stammt; je enger und spezieller die zu erwartenden Fragestellungen, um so kleiner darf der Zeitraum gewählt werden, um so stärker werden dann aber auch zeitraumbedingte und damit auch themenbedingte Eigenheiten des Materials, besonders im Wortschatz, hervortreten. Der Bearbeiter muß diese Eigenheiten im voraus berücksichtigen; er hat sich darüber Rechenschaft abzulegen, warum er gerade das Material aus diesem Zeitraum (z.B. Woche) und nicht aus einem der vielen möglichen anderen gewählt hat.

Für allgemeine Erörterungen und die sehr vielschichtigen Fragestellungen, mit denen wir rechnen müssen, ist zweckmäßigerweise von der Großeinheit des Jahrgangs auszugehen.

4.3. Mengenproblem und Kapazität

Vor Beginn der Textaufnahme muß Klarheit bestehen über die Dimensionen, in denen wir uns bei der Aufnahme von Zeitungstexten bewegen. Die WELT hat im Jahrgang 1964 (ohne Beilagen!) in ca. 310 erschienenen Nummern ca. 5.600 Zeitungsseiten mit einer Textmenge von rund 15 Millionen laufen-

den Wörtern produziert. Das entspricht einem Buchwerk von über 30.000 Seiten. Eine geübte Schreibkraft brauchte etwa 5.500 Arbeitstage entspr. 24 Jahren (bei rund 230 Arbeitstagen p.a.), um diese Menge auf Datenträger zu übertragen; sie brauchte etwa die gleiche Zeit für die Korrektur¹⁵⁾. Selbst die von einer weniger umfangreichen Zeitung, etwa dem "Neuen Deutschland", in nur einem Monat produzierte Textmenge entspricht noch immer dem Umfang eines Buches von über 1.200 Seiten¹⁶⁾.

Es ist offensichtlich, daß solche Mengen die Schreibkapazität der weitaus meisten Institute überschreiten. Solange die elektronische Industrie nicht ein auch für Zeitungen brauchbares optisches Lesegerät auf den Markt bringt, wird die Aufnahme vollständiger Zeitungsjahrgänge also nicht möglich sein. Auch die Aufnahme ganzer Monats-Mengen ist noch sehr aufwendig. Hier und um so mehr bei noch kleineren Einheiten, die vollständig zu bewältigen wären, stellen sich jedoch die unter Absatz 4.2. erwähnten Bedenken ein.

Theoretisch besteht noch die Möglichkeit, sich die Tatsache zunutze zu machen, daß die meisten Zeitungsdruckereien ihre Setzmaschinen mittels Lochstreifen (TTS-Streifen) steuern. Sofern man sich regelmäßig diese Lochstreifen beschaffen kann, wäre es möglich, auch sehr große Textmengen ohne eigene Schreibebeit elektronisch zu speichern. Leider war diese Möglichkeit wegen einer Reihe praktischer Schwierigkeiten bisher nicht zu verwirklichen¹⁷⁾, abgesehen davon, daß mittels TTS-Streifen ohnehin kein älteres Material zu gewinnen ist.

Aus diesen Gründen erweist es sich als unvermeidlich, eine Auswahl aus dem jeweils vorgesehenen Gesamtmaterial zu treffen.

4.4. Begründbarkeit einer Textauswahl

Eine vollständige Aufnahme eines umfangreichen Textkorpus wäre aber zudem nur dann erforderlich, wenn es darum ginge, die Eigenheiten einzelner Teile dieses Korpus (etwa eines Romanwerkes, Zeitungsjahrgangs usw.) vollständig zu erforschen, d.h. wenn Fragestellungen zu erwarten sind, die sich auf dieses eine Werk in seiner Einmaligkeit richten. Dies ist aber im Rahmen der Dokumentation, die das Institut einschließlich der Außenstelle Bonn treibt bzw. zu treiben beabsichtigt, durchweg nicht der Fall. Die aufgenommenen Texte sollen nur Beispiele sein für die deutsche Sprache (östliche, westliche Zeitungssprache, Literatursprache, Trivialliteratursprache usw.) der Gegenwart. Ein "repräsentatives" Gesamtkorpus wird sich also erst aus einer gut zusammengestellten Mischung verschiedener solcher Texte zusammensetzen. Dann aber erfüllt eine Auswahl aus vielen solcher Texte ihren Zweck besser als wenige vollständig aufgenommene Texte, auch und gerade dann, wenn neben synchronischen auch diachronische Fragestellungen zu erwarten sind. Dies gilt ganz besonders für Untersuchungen zum Wortschatz (jedoch nicht nur für diese), da erwiesen ist, daß bestimmte Eigentümlichkeiten des Wortschatzes großer Textkorpora schon in Teilmengen gut erkennbar sind (Homogenität vorausgesetzt)¹⁸⁾.

Damit kann das Prinzip der Auswahl nicht nur aus praktischen Gründen notwendig, sondern auch als sinnvoll, im Hinblick auf die höhere Arbeitseffektivität im Vergleich zum Prinzip der vollständigen Textaufnahme sogar als besser bezeichnet werden.

Es ist somit zu unterscheiden zwischen der Gesamtmenge, d.h. derjenigen Menge, über die letztlich Aussagen gemacht werden sollen, und der tatsächlich aufgenommenen Auswahlmenge (Modellmenge), d.h. derjenigen Menge, mittels derer Aussagen über die Gesamtmenge gemacht werden sollen.

Ein wesentlicher Nachteil liegt zweifellos darin, daß es nicht möglich ist, experimentell zu beweisen, daß die Auswahl tatsächlich Abbild der Gesamtmenge ist, da diese ja für genaue Textanalysen gerade nicht zur Verfügung steht. Über den Grad der Entsprechung sind nur Wahrscheinlichkeitsaussagen möglich. Um so sorgfältiger muß die Aufnahme der Auswahl vorbereitet werden.

4.5. Bedingungen einer zureichenden Auswahl ¹⁹⁾

4.5.1. Allgemeine Feststellungen

Um eine Auswahl treffen zu können, sind folgende Schritte erforderlich:

1. Es ist zunächst die Gesamtmenge zu bestimmen und zu sichten, d.i. die Großeinheit von Texten, auf die sich die Dokumentation beziehen soll.
2. Innerhalb dieser (Gesamt-) Bezugsmenge ist die Zielmenge zu bestimmen, d.h. die Menge, aus der die aufzunehmende Auswahl gewonnen wird.
3. Die Zielmenge muß in allen ihren Eigentümlichkeiten so genau wie möglich bestimmt werden.
4. Aus dieser Zielmenge muß eine Auswahl gewonnen werden, die
 - a) alle Eigenheiten der Zielmenge möglichst maßstabgetreu widerspiegelt,
 - b) ihre Menge nach ausreicht, um statistische Untersuchungen sowie Rückschlüsse auf die Zielmenge zu erlauben;
 - c) alle die auf Seite 47/48 aufgeführten sechs Anforderungen, die für jede Dokumentation gelten, erfüllt,

d.h. es muß ein qualitativ und quantitativ zureichendes Modell erstellt werden²⁰⁾.

Erläuterungen

Zu 1:

Es ist für die zu schaffende Auswahl von großer Bedeutung, ob sie sich etwa auf ein einzelnes Werk, das Gesamtwerk eines Dichters, eine literarische Gattung oder Epoche usw. beziehen soll. Je komplexer oder ausgedehnter die bezogene Gesamtmenge ist, um so komplexer oder ausgedehnter muß auch die Auswahl insgesamt und in ihren Teilen sein.

Zu 2:

Die Unterscheidung von Gesamtmenge und Zielmenge ist erforderlich, weil nicht immer alle Teile eines großen Textkorpus, etwa des Gesamtwerkes eines Schriftstellers, eines gesamten Zeitungsjahrgangs samt allen Beilagen usw., in die Dokumentation einbezogen werden sollen. Es kann sinnvoll sein, bestimmte Teile, wie die Briefe, Tagebücher und Entwürfe aus dem Gesamtwerk eines Schriftstellers bzw. die literarische Beilage aus einem Zeitungsjahrgang, zu übergehen und eine entsprechend eingeschränkte Zielmenge zu definieren²¹⁾.

Zu 3:

Zu den Eigentümlichkeiten der Zeitung gehört zum Beispiel, daß ihr Textmaterial nicht nur eine "vertikale" Ausdehnung hat - meßbar an der jeweils zu einem bestimmten Zeitpunkt erschienenen Zahl von Druckseiten -, sondern auch eine "horizontale" - meßbar an der Zahl der in einem bestimmten Zeitraum in einem bestimmten Rhythmus erschienenen Ausgaben²²⁾. Verbunden mit den auf S.42/43 erwähnten grundsätzlichen Eigentümlichkeiten, besonders der Verfasser-, Themen- und Formenvielfalt, erweist sich die Zeitung also als ein äußerst komplexes Gebilde. Verfahren zur Dokumentation etwa bei Büchern oder auch weniger kompliziert aufgebauten Periodika können wahrscheinlich im wesentlichen als Vereinfachungen des bei Zeitungstexten anzuwendenden Verfahrens betrachtet werden.

4.5.2. Qualitative Analyse

4.5.2.1. Feststellung spezifischer Dominanzen

Wenn angenommen werden könnte, daß die verschiedenen Eigenheiten (wie Einfluß bestimmter Verfassergruppen, Themenwahl, Sachgebiete, Berichtsformen usw.) gleichmäßig über die ganze Jahrgangsmenge in vertikaler und horizontaler Richtung verteilt wären, wäre die Gewinnung einer Modellmenge kein Problem; man könnte sich darauf beschränken, in regelmäßigen Abständen etwa gleich große Textmengen an beliebiger Stelle zu entnehmen und sie zu einer Auswahlmenge zu vereinigen. Leider verhält es sich nicht so; vielmehr zeigen alle Zeitungen in jeder Ausgabe gewisse regelmäßige Dominanzen, die es unmöglich machen, die Auswahl dem Zufall zu überlassen, sofern Verzerrungen und Einseitigkeiten vermieden werden sollen.

a) Dominanzen in vertikaler Richtung:

Die meisten Zeitungen gliedern ihre Ausgaben in Sparten. Im Regelfalle sind die ersten 2 - 3 Seiten der Politik und aktuellen Nachrichten vorbehalten, darauf folgt "Wirtschaft" und "Kulturelles" (Feuilleton). Die letzte Seite enthält zumeist Vermischtes, dazwischen können Lokalseiten oder Seiten für Sport, Reise usw. eingeschaltet sein. Gelegentlich können bestimmte Sparten zu Lasten anderer stark erweitert werden. Sofern eine Auswahl nicht absichtlich bestimmte Sparten und damit Sachgebiete und Themen bevorzugen und andere vernachlässigen will, muß auf eine gleichmäßige Berücksichtigung aller Seiten geachtet werden, d.h. alle Seiten von der ersten bis zur letzten müssen entsprechend ihrer Verteilung in der Zielmenge auch in der Auswahlmenge enthalten sein.

b) Dominanzen in horizontaler Richtung:

(Wochenturnus)

Es läßt sich feststellen, daß bestimmte Themen und Sparten sich an bestimmten Tagen häufen. Montags nimmt durchweg der Sport einen breiten Raum ein, wogegen Wirtschaft, vor allem Börsenberichte (in westlichen Zeitungen), weitgehend fehlen. Zum Wochende hin verstärkt sich der Anteil von Artikeln zum Thema "Reise, Erholung, Freizeit" usw., samstags, oft auch mittwochs, häufen sich Kleinanzeigen, auch Feuilleton und Unterhaltung. Eine bestimmte Wochentagsausgabe enthält meist Berichte über Wissenschaft, Literatur usw. Auch Leserbriefe werden oft nur in bestimmten Wochentagsausgaben gebracht.

Es besteht ferner Grund zu der Annahme, daß bestimmte Verfasser bevorzugt an bestimmten Wochentagen Artikel veröffentlichen.

Auch hier kann es also nicht gleichgültig sein, ob bestimmte Wochentagsausgaben häufiger als andere in der Modellmenge enthalten sind. Es ist wiederum auf gleichmäßige Berücksichtigung aller Wochentagsausgaben zu achten.

(Monatsturnus)

Regelmäßigkeiten im Ablauf der einzelnen Monate ließen sich bisher nicht erkennen.

(Quartalsturnus)

Regelmäßigkeiten innerhalb der Quartale sind nur sehr schwach und nur im Wirtschaftsteil zu erkennen: Jeweils beim Übergang zu einem neuen Quartal finden sich statistische Übersichten über Wirtschaftsentwicklungen häufiger als sonst.

(Jahresturnus)

Innerhalb eines Jahrgangs sind deutlich ausgeprägte Dominanzen zu erkennen. Der Wechsel von Saat und Ernte prägt sich besonders in östlichen Zeitungen im Bereich der Wirtschaft außerordentlich stark aus; Jahresende und -beginn bringen eine ungewöhnliche Häufung von Übersichten, Entwicklungsberichten, Prognosen, Planungen usw., desgleichen die Zeit der Beratung und Verabschiedung des Haushalts (Frühjahr bzw. Herbst). Die Sommerpause wirkt sich auf den Sportteil (bes. Fußball) ebenso negativ wie auf die Sparte "Reise, Urlaub" usw. positiv aus. Die saisonbedingte Dominanz bestimmter Sportarten (Wintersport, Wassersport) ist ebenso spürbar wie der Einfluß der großen Feste (Weihnachten, Karneval, Ostern), und zwar nicht nur im Werbe-Teil. Unabhängig von diesen jährlich wiederkehrenden Dominanzen können selbstverständlich bestimmte Ereignisse (Olympiade, Nahostkrieg u.ä.) die Berichte der entsprechenden Sparte über längere Zeit hin nachhaltig beeinflussen.

Jede Abweichung von einer gleichmäßig über den ganzen Jahrgang hin verteilten Auswahl kann daher zu erheblichen Verschiebungen vor allem im Wortschatz führen. Die Bevorzugung etwa eines bestimmten Monats in einer Auswahl macht diese jedenfalls als Modellmenge für den Jahrgang unbrauchbar. Gleiches gilt für die Bevorzugung bestimmter Seiten, Sparten oder Artikelformen.

Die hier aufgezeigten Regelmäßigkeiten gelten nur für eine bestimmte Art von Tageszeitungen und auch für diese nicht ohne Einschränkung. Abweichungen sind jederzeit möglich. Für andere Arten von Tageszeitungen, etwa für Boulevardblätter²³⁾, müssen neue Untersuchungen angestellt werden. Es ist außerdem möglich, daß wesentliche Eigenheiten übersehen werden, die sich dann unkontrolliert verzerrend auf die in der Modellmenge herr-

schenden Verhältnisse auswirken²⁴⁾. Andererseits ist es möglich, Kriterien aufzugreifen und bei der Auswahl zu berücksichtigen, die sich als irrelevant erweisen können²⁵⁾.

4.5.3. Quantitative Analyse

4.5.3.1. Definition der Mengeneinheiten

Bevor man nach der Feststellung der qualitativen Eigenheiten der Zielmenge zur Ermittlung der Modellmenge übergehen kann, ist es notwendig, die Mengeneinheiten in der Zielmenge festzustellen bzw. zu definieren. Als Meßeinheit²⁶⁾ bietet sich die Seite als einwandfrei formal definierbare Einheit an, die im übrigen den Vorteil hat, daß sie sowohl für Texte rein vertikaler Ausdehnung (Bücher) als auch für Texte vertikaler und horizontaler Ausdehnung (Periodika) gilt²⁷⁾.

a) Vertikale Mengeneinheiten

Zeitungen erscheinen jedoch nicht in "Seiten", sondern in "Ausgaben". Der Begriff "Ausgabe" ist jedoch unklar. Es fragt sich, ob etwa Beilagen dazu gehören oder nur bestimmte Beilagen²⁸⁾. Tatsächlich kann sich die Menge, mit der eine Zeitung erscheint, zusammensetzen aus

1. der Grundmenge (gewöhnlich der einheitlich durchnummerierte Teil der Zeitung)
2. den regelmäßigen Beilagen (meist gesondert numeriert)
3. den unregelmäßigen Beilagen, den Sonderbeilagen zu einmaligen Ereignissen, Extrablättern usw. (gesondert oder nicht numeriert).

Die unter Einschluß aller Erweiterungen und Beilagen an einem Tag erschienene Menge nenne ich Erscheinungsmenge. Der Bearbeiter hat sich also zunächst zu entscheiden, ob er die gesamte Erscheinungsmenge oder nur einen Teil von ihr - und welchen - zur Grundlage

seiner weiteren Berechnungen machen will. Die Mengeneinheit, für die er sich entscheidet, nenne ich Ausgabe ²⁹⁾. Erscheinungsmenge $S^{(e)}$ und Ausgabe $S^{(a)}$ sind in Zahl der Seiten (S) meßbar, wobei S_e eine aus mehreren Teilen (G=Grundmenge, E1 und E2 = Erweiterungen 1 und 2) zusammengesetzte Größe ist :

$$S_G + S_{E1} + S_{E2} = S_e ; \text{ dabei ist } S_e \equiv S_a$$

Umfang der Ausgabe wie der Erscheinungsmenge können täglich wechseln. Die Erscheinungsmenge unterscheidet sich von der Ausgabe also in gleicher Weise wie die Gesamtmenge von der Zielmenge.

b) Horizontale Mengeneinheiten

Im folgenden wird vorausgesetzt

1. daß die Zielmenge 1 Jahrgang ist,
2. daß die Ausgabe die gesamte Erscheinungsmenge umfaßt (also $S_a = S_e$),
3. daß sich innerhalb des Jahrgangs die Erscheinungsweise der Zeitung nicht grundlegend ändert ³⁰⁾.

Zur Ermittlung der horizontalen Mengenverteilung hat der Bearbeiter eine Reihe von Feststellungen zu treffen :

1. Feststellung der Zahl der Ausgaben (A_j) pro Jahrgang (J), wobei die Zahl der Ausgaben gleich der Nummer (n) der letzten Ausgabe des Jahres ist : $A_j = n$
2. Feststellung der Gesamtzahl der im Jahrgang erschienenen Seiten :

$$S_{a1} + S_{a2} + S_{a3} + \dots + S_{an} = S_{aJ}$$

(Daraus ergibt sich die durchschnittliche Seitenzahl der Ausgaben

$$s_a = \frac{S_{aJ}}{n})$$

3. Feststellung des Erscheinungsintervalls (des in Tagen zu messenden Abstands zwischen den Ausgaben)

- a) Das durchschnittliche Erscheinungsintervall läßt sich errechnen als $\frac{365}{n}$, wobei sich bei täglich erscheinenden Ausgaben ein Wert nahe 1, bei Tageszeitungen ohne Sonntagsausgabe ein Wert nahe 1,2 und bei Wochenzeitungen nahe 7 ergeben muß ³¹⁾.

Geringe Abweichungen nach oben sind normal, da immer einige Zeitungsausgaben, abweichend vom regelmäßigen Erscheinungsintervall, ausfallen (wegen Betriebsstörungen, an Festtagen usw.) ³²⁾.

4.5.4. Feststellen der Seitenhäufigkeiten

Da die Ausgaben einer Zeitung verschieden lang sind, erscheinen nicht alle Seiten gleich häufig in der Zielmenge. Da aber bestimmte Sparten und damit Themen an bestimmte Seiten oder Seitengruppen mehr oder weniger eng gebunden sind, ist es notwendig, die Verteilung der einzelnen Seiten in der Zielmenge zu ermitteln, um sie in der Modellmenge berücksichtigen zu können. Bei der Aufnahme einer Zeitung, deren Umfang zwischen 8 und 20 Seiten schwankt, muß die Auswahl mehr Seiten der Seitennummern 1 - 8 als der Seitennummern 12 - 20 enthalten.

4.5.4.1. Seitenhäufigkeiten in der Zielmenge

- a) Durch eine teilweise ³³⁾ Auszählung des Jahrgangs ist der Umfang der einzelnen Ausgaben festzustellen.

Mindestumfang der Ausgaben: y Seiten

Maximalumfang der Ausgaben: $y + m$ Seiten

(m = Differenz zwischen der kleinsten und der größten Ausgabe in Seiten)

b) Verteilung der Seiten nach Ausgaben verschiedener Länge (Klassenbildung):

p	Ausgaben	enden mit Seite	y	, enthalten also	p	mal	y	Seiten
p_1	"	"	"	"	$y + 2$,	"	"	p_1 " ($y + 2$) "
p_2	"	"	"	"	$y + 4$,	"	"	p_2 " ($y + 4$) "
p_3	"	"	"	"	$y + 6$,	"	"	p_3 " ($y + 6$) "
.					.			
.					.			
$\frac{p_i}{n}$	"	"	"	"	$y + m$,	"	"	$\frac{p_i}{S_{aJ}}$ " ($y + m$) "

c) Ermittlung der Häufigkeit jeder Seite pro Klasse in der Zielmenge:

Es kommen im Material vor:

die Seiten 1 bis y in jeder Ausgabe, also je n -mal; die Seiten $y + 1$ und $y + 2$ je $n - p$ mal usw.:

S. 1	n	}	zus. y mal n
S. 2	n		
S. 3	n		
. .	.		
. .	.		
. .	.		
S. y	n		
S. $y + 1$	$n - p$		
S. $y + 2$	$n - p$		
S. $y + 3$	$n - (p + p_1)$		
S. $y + 4$	$n - (p + p_1)$		
S. $y + 5$	$n - (p + p_1 + p_2)$		
S. $y + 6$	$n - (p + p_1 + p_2)$		
. .	. .		
. .	. .		
. .	. .		
. .	. .		

$$\begin{array}{ll} S. y + m - 1 & n - (p + p_1 + p_2 + \dots + p_i) \\ S. y + m & n - (p + p_1 + p_2 + \dots + p_i) \end{array}$$

Den jeweils gefundenen Wert nenne ich F.

4.5.4.2. Seitenhäufigkeiten in der Auswahlmenge

(Umrechnung auf die Anteile der einzelnen Seiten in der Auswahlmenge)

Die gesuchte Häufigkeit f jeder Seite pro Klasse in der Auswahlmenge (Modellmenge) M verhält sich zur Gesamtzahl der Seiten (S_M) in der Auswahlmenge wie die Häufigkeit F jeder Seite pro Klasse in der Zielmenge zur Gesamtzahl der Seiten (S_{aJ}) in der Zielmenge :

$$\frac{f}{S_M} = \frac{F}{S_{aJ}} \quad ; \quad f = F \cdot \frac{S_M}{S_{aJ}} \quad , \text{ wobei } F \text{ jeweils einsetzbar,}$$

$$\frac{S_M}{S_{aJ}} \quad \text{ein konstanter Faktor ist.}$$

Sollten nicht für die gesamte Zielmenge, sondern nur für einige Teile von ihr exakt ausgezählte Werte vorliegen, so sind zweckmäßigerweise sowohl bei der Ermittlung von F als auch von f die tatsächlich ausgezählten Werte zugrunde³⁴⁾zulegen.

Sofern die Größe der Auswahlmenge bekannt ist, kann damit der Anteil aller Seiten in der Auswahl bestimmt werden.

Die Größe der Auswahlmenge kann vorgegeben sein (etwa durch Kapazitätsgrenzen), sie kann auch abhängen von der Auswahlquote (in Prozent der Zielmenge), vom Wahrscheinlichkeitsquotienten oder von Zahl und Umfang vorgegebener Stichproben.

4.6. Ermittlung der Modellmenge

Es gibt eine ganze Reihe von Möglichkeiten, eine Modellmenge sinnvoll zusammenzustellen. Welche dieser Möglichkeiten gewählt wird, hängt in erster Linie von dem Zweck ab, welcher der Materialaufnahme zugrundeliegt, in zweiter Linie von den Ansprüchen, die an das Material gestellt werden.

Nach diesen Gesichtspunkten sind dann einige Vorentscheidungen zu treffen:

4.6.1. Aufnahmeeinheit

Als erstes ist die *Aufnahmeeinheit*, die der Modellmenge zugrundeliegt, zu bestimmen. Wenn bisher vorausgesetzt wurde, daß die Seite als formal eindeutige Grund- und Zählseinheit der Zielmenge auch als Aufnahmeeinheit dient, so geschah das aus praktischen Gründen und weil ihr auch eine thematisch-inhaltlich gliedernde Funktion im Aufbau einer Ausgabe zukommt. Außer der Seite ist auch etwa der "Artikel" oder die "Ausgabe" als Aufnahmeeinheit denkbar. Der Artikel bietet den Vorteil sehr fein differenzierbarer Streuung und thematischer Einheitlichkeit, den Nachteil schwieriger Definition und sehr unterschiedlichen Umfangs; die Seite bietet den Vorteil mittlerer Streuungsmöglichkeit, formaler Eindeutigkeit und begrenzter thematischer Gebundenheit, den Nachteil unterschiedlichen Umfangs und den Nachteil, daß Textzusammenhänge zerrissen werden können, da gelegentlich Artikel auf einer Seite begonnen und auf einer anderen fortgesetzt werden. Die Ausgabe hat den Vorteil relativer Geschlossenheit, den Nachteil zu großen Umfangs, der dazu führen kann, daß nur sehr wenige Ausgaben aufgenommen werden können.

4.6.2. Stichprobeneinheit

Wird aus diesem Grund die Ausgabe als Aufnahmeeinheit abgelehnt und eine der beiden kleineren Einheiten gewählt, ist noch zu entscheiden, wie viele von diesen Aufnahmeeinheiten bei jeder Stichprobe jeweils aufgenom-

men werden sollen. Neben der Aufnahmeeinheit ist damit der weitere Begriff der Stichprobeneinheit einzuführen. Sie hat mindestens die Größe der Aufnahmeeinheit oder eines Mehrfachen von dieser ³⁵⁾.

4.6.3. Aufnahmeintervall, Aufnahmedichte

Diese Überlegung führt zur Frage des Aufnahmeintervalls. Das Aufnahmeintervall ist der zeitliche Abstand zwischen zwei Stichproben, und zwar, dem Grundsatz der gleichmäßigen Auswahl entsprechend, ein möglichst gleich großer Abstand zwischen allen Stichproben. Da immer davon ausgegangen werden muß, daß für die Herstellung der Modellmengen nur eine bestimmte, begrenzte Arbeitskapazität zur Verfügung steht, muß das Aufnahmeintervall um so größer, d.h. die Aufnahmedichte um so geringer werden, je größer die einzelnen Stichprobeneinheiten sind. Es gibt jedoch eine obere Grenze ³⁶⁾ für die Größe des Aufnahmeintervalls (= Untergrenze der Aufnahmedichte): sie wird dann überschritten, wenn mit hoher Wahrscheinlichkeit zu befürchten ist, daß sogar wesentliche sprachliche Erscheinungen, soweit sie von horizontalen Dominanzen her beeinflußt sind, von den Stichproben nicht mehr berührt werden ³⁷⁾.

4.6.4. Zahl der Stichproben

Die Zahl der Stichproben in der gesamten Modellmenge ist abhängig vom Aufnahmeintervall oder umgekehrt. Allerdings läßt sich, ebenso wie für das Aufnahmeintervall, eine Grenze angeben:

Die Zahl der Stichproben (meist identisch mit der Zahl der Aufnahmeeinheiten) darf nicht kleiner sein als die Zahl der für den Beobachter wichtigen Kriterien, die in der Auswahl zu berücksichtigen sind, bzw. der Rhythmus des Auftretens dieses wichtigen Kriteriums. Hat etwa eine Zeitung einen durchschnittlichen Umfang von 20 Seiten und soll jede Seite mindestens 1 mal in der Modellmenge enthalten sein, sind mindestens 20 Stichproben (=Seiten) erforderlich. Ist der Wochenrhythmus ein wich-

tiges Kriterium, sind 52 Stichproben erforderlich; sind 100 verschiedene Sachgebiete in einer Zeitung festgestellt und soll jedes fünfmal vertreten sein, benötige ich mindestens 500 Artikel usw. Eine Mindestmenge läßt sich also meist schon von der Fragestellung her bestimmen.

4.6.5. Wahrscheinlichkeitsquotient

Ob diese Menge tatsächlich ausreicht, d.h. ob eine qualitativ zureichend ermittelte Auswahl auch rein quantitativ die Bedingungen einer aussagekräftigen Modellmenge erfüllt, ist eine mathematisch-statistische Frage, die im Teil III des Forschungsberichts mit entsprechenden Methoden untersucht wird. Das Ergebnis dieser Berechnungen ist der Wahrscheinlichkeitsquotient, der ausdrückt, mit welcher Wahrscheinlichkeit eine in der Zielmenge vorhandene seltene Erscheinung (Wortform) auch in der Modellmenge erwartet werden kann und umgekehrt³⁸⁾.

4.6.6. Aufnahmezeitraum

Der Zeitraum, aus dem die Stichproben genommen werden, muß genau dem Zeitraum entsprechen, für den die Modellmenge aussagekräftig sein soll. Jede Auswahl kann Modell sein nur für den Zeitraum, der durch die erste und die letzte Aufnahmeeinheit - plus je einem halben Aufnahmeintervall vor der ersten und nach der letzten Aufnahmeeinheit - begrenzt wird³⁹⁾.

4.7. Zusammenfassung

der erforderlichen Schritte zur Ermittlung einer Modellmenge

Unter der Voraussetzung, daß

- a) die Rahmen-Zielsetzung der Dokumentation,
- b) der Rahmen der Kapazität (Arbeitskräfte, Hilfsmittel, Zeit)

festliegen, sind folgende Schritte notwendig:

1. Festlegung der Gesamtmenge und Erscheinungsmenge
daraus: Festlegung der Zielmenge und der Ausgabe
2. Bestimmung der Eigentümlichkeiten der Zielmenge
 - a) Dominanzen in vertikaler Richtung
 - b) Dominanzen in horizontaler Richtung
3. Messung (Auszählung) der Zielmenge,
der Ausgaben und
des Erscheinungsintervalls
in Seiten bzw. Tagen
4. Feststellung der Seitenverteilung in der Zielmenge
daraus: Ermittlung der Seitenanteile (F)
5. Bestimmung der Aufnahmeeinheit für die Auswahlmenge
6. Bestimmung des Aufnahmeintervalls (\rightarrow Aufnahmedichte)
daraus: Ermittlung der Zahl der Stichproben
7. Bestimmung der Aufnahmequote nach dem gewünschten Wahrscheinlichkeitsquotienten (bzw. umgekehrt)
8. Errechnung der insgesamt aufzunehmenden Auswahlmenge in Aufnahmeeinheiten (sofern nicht vorgegeben), damit:
Bestimmung des Umfangs der Stichproben
9. Errechnung der Seitenverteilung f in der Auswahlmenge (unter Berücksichtigung von Punkt 4)
10. Verteilung der Aufnahmeeinheiten (Seiten) auf die Stichproben und der Stichproben über den ganzen Aufnahmezeitraum anhand des unter 9 ermittelten Seitenschlüssels.
(Bei diesem Arbeitsvorgang ist, zur Vermeidung des Vorwurfs willkürlicher Verteilung, ein möglichst fester Verteilungsturnus anzustreben.)

Je nachdem, ob die Aufnahmequote, Umfang und Verteilung der Auswahlmenge oder der Wahrscheinlichkeitsquotient vorgegeben oder abgeleitet sind, muß die Reihenfolge der Schritte 7 - 10 entsprechend vertauscht werden.

Spezieller Teil: Dokumentation und maschinelle Verarbeitung der
Bonner Zeitungsjahrgänge

5. Allgemeines

Zu den Zielen und Methoden der in der Außenstelle betriebenen Dokumentation bei Zeitungstexten siehe die allgemeinen Bemerkungen unter Punkt 1 bis 3 und 4.1. bis 4.2. Entsprechend den dort zusammengestellten grundsätzlichen Bemerkungen sollte sich die Dokumentation auf mehrere ganze Jahrgänge überregionaler Tageszeitungen politischen Charakters und "repräsentativer" Art aus West und Ost beziehen. Da östliches Zeitungsmaterial schwerer zugänglich ist als westliches und da sich zudem das wissenschaftliche Interesse zunächst mehr auf das östliche Material richtete, sollte mit einem Zeitungsjahrgang aus der DDR begonnen werden.

6. Zur Kapazität

Für die Übertragung der Texte auf Datenträger standen der Außenstelle zwei bis vier, durchschnittlich 3,5 Halbtagskräfte zur Verfügung, von denen jeweils zwei mit dem Abschreiben, die übrigen mit der Korrektur und dem Lesen der abgeschrieben Texte beschäftigt waren⁴⁰⁾. Es mußte also von einer Schreibkapazität von 2 mal 20 Wochenstunden, also (bei durchschnittlich 46 Arbeitswochen pro Jahr) von 1840 Arbeitsstunden pro Jahr ausgegangen werden.

Da zum Abschreiben einer Zeitungsseite, sofern alle Schreibkonventionen⁴¹⁾ beachtet werden, ungefähr 9 Arbeitsstunden benötigt werden, ergibt sich eine theoretische Jahreskapazität von etwas über 200 Seiten. Durch den gerade bei Zeitungstexten sehr hohen Aufwand an Korrekturarbeiten sinkt die Schreibkapazität um rund ein Viertel, da etwa dieser Anteil der Seiten nach beendetem Korrekturen-Lesen neu geschrieben werden muß. Die obere Grenze der Kapazität lag also bei 150 Seiten pro Jahr⁴²⁾.

Wir gingen bei der Arbeitsplanung davon aus, daß grundsätzlich mehr als ein Jahrgang, also mindestens zwei, pro Jahr abgeschrieben und zur Verfügung gestellt werden sollten. Daraus ergab sich, daß die aus jedem Jahrgang aufgenommene Menge im Durchschnitt nicht mehr als 70 bis 80 Seiten umfassen durfte.

7. Bestimmung der Gesamtmengen

Als Grundlage für die Dokumentation von Zeitungstexten aus der DDR wurde

1. der Jahrgang 1964 des "Neuen Deutschland" (Organ des Zentralkomitees der SED), Berliner Ausgabe⁴³⁾,

für Zeitungstexte aus der Bundesrepublik

2. der Jahrgang 1964 der WELT, Hamburg, Berliner Ausgabe, gewählt.

Für vergleichende Untersuchungen diachronischer Art wurde als zweite Stufe der Textaufnahme

3. der Jahrgang 1954 des "Neuen Deutschland"⁴⁴⁾,

und

4. der Jahrgang 1954 der WELT, Berliner Ausgabe, ausgewählt.

7.1. Begründung zur Wahl der Ost-Jahrgänge

Das "Neue Deutschland" ist im Rahmen der von Staat und Partei kontrollierten Presse der DDR die einzige überregionale, in der ganzen DDR verbreitete Tageszeitung (im Jahre 1964 täglich, auch sonntags, erscheinend). Sie ist als Zentralorgan der SED in besonderer Weise repräsentativ für die

Meinung (und Sprache) der in der DDR herrschenden Gruppe und maßgebend für den größten Teil der DDR-Presse. Das "Neue Deutschland" (ND) ist nicht in erster Linie Verkaufsobjekt, sondern bewußt Erziehungs-, Propaganda- und Führungsmittel der Partei.

In der Erfüllung dieser Funktion ist es extrem "politisch" orientiert, d.h. es schränkt die Vielfalt und den Umfang der aus westlichen Zeitungen gewohnten Themen und Sparten erheblich ein bzw. bezieht diese mit in den Bereich des Politischen ein⁴⁵⁾. Mit dem Jahrgang 1964 wurde deshalb begonnen, weil dieser der letzte war, der bei Beginn der Arbeiten (Anfang 1965) abgeschlossen vorlag. In der zweiten Stufe (Jahrgang 1954) wurde ein Zehnjahresabstand gewählt, da angenommen wird, daß ein relativ großer Abstand nötig ist, um wesentliche Veränderungen im Wortschatz in West und Ost sichtbar zu machen. Es ist jedoch geplant, den Abstand später zu halbieren und noch je einen Jahrgang 1959 aufzunehmen.

7.2. Begründung zur Wahl der West-Jahrgänge

Für den Bereich der Bundesrepublik läßt sich keine Tageszeitung benennen, die in entsprechender Weise repräsentativ genannt werden kann. Die auflagenstärkste Zeitung, die Bild-Zeitung, ist ein überwiegend unpolitisches Boulevard-Blatt; unter den überregionalen "seriösen" Zeitungen, der "Frankfurter Allgemeinen Zeitung", der "WELT", vielleicht noch der "Süddeutschen Zeitung", "Stuttgarter Zeitung" und der "Frankfurter Rundschau", eine im gleichen Sinne "repräsentative" zu benennen, ist kaum möglich. Keine von ihnen gibt in dem Maße die Meinung einer Gruppe wieder, wie dies bei dem ND der Fall ist, keine ist auch in dem Maße "regierungsuffiziös". Am ehesten hat die Wahl zwischen FAZ und WELT zu fallen.

Die Wahl fiel schließlich, trotz etwas geringerer Auflage⁴⁶⁾, auf die WELT, da sie die wichtigste politische Zeitung des größten deutschen Zeitungskonzerns ist. Auch hier wurde die Berliner Ausgabe herangezogen. Es ist möglich

und geplant, auf die in Mannheim aufgenommenen ⁴⁷⁾ Texte aus der FAZ zurückzugreifen und sie mit den WELT-Texten zu vergleichen. Sollten sich dabei erhebliche Abweichungen im Wortschatz ergeben, könnten auch FAZ-Texte laufend aufgenommen werden.

7.3. Vergleichbarkeit der ausgewählten Zeitungen

Es leuchtet ein, daß die beiden ausgewählten Zeitungen nach Zielsetzung, Herkunft, Themenwahl, Darstellungsweise usw. nur bedingt miteinander vergleichbar sind. Im strengen Sinne ist aber keine einzelne westdeutsche Zeitung mit dem ND voll vergleichbar. Als Organ der SED entsprächen dem ND am ehesten die Zeitungen der großen Parteien in der Bundesrepublik, in anderer Hinsicht wäre es vielleicht mit einem Gewerkschaftsorgan vergleichbar, in wieder anderer, z.B. in stilistischer Hinsicht eher mit der Bild-Zeitung oder der Deutschen National-Zeitung oder auch, was die Themenwahl betrifft, mit manchen kleineren Regionalblättern.

Es kann aber nicht Aufgabe unserer Dokumentation sein, die erheblichen Unterschiede zwischen einer repräsentativen Zeitung der DDR und der Bundesrepublik künstlich zu überdecken, etwa indem man aus verschiedenen westdeutschen Zeitungen einen Mischtext herstellt, der der Eigenart des ND möglichst nahe käme ⁴⁸⁾. Die Eigenart der Textgrundlage ist zu wahren. Zwar dient die Dokumentation nicht letztlich dem Ziel, die Eigenart zweier Zeitungen miteinander vergleichbar zu machen (das wäre eine zeitungswissenschaftliche Fragestellung), aber sie bedient sich der Texte zunächst zweier, später mehrerer Zeitungen mit jeweils besonderen Eigenheiten, um daran die Eigenheiten der publizierten Sprache in beiden Teilen Deutschlands und ihre Entwicklung vergleichend erkennbar zu machen.

8. Bestimmung der Zielmengen

8.1. Abgrenzung gegenüber Gesamtmengen

a) Neues Deutschland 1964:

Als Zielmenge für den gesamten Jahrgang wurde festgelegt:

die zwischen dem 1. Januar und 31. Dezember 1964 produzierte Menge,
mit Ausnahme der Sonntagsausgabe
und der wöchentlich erscheinenden Beilage
"Die gebildete Nation".

b) DIE WELT 1964:

Als Zielmenge wurde festgelegt:

die zwischen dem 1. Januar und 31. Dezember 1964 produzierte Menge,
mit Ausnahme aller gesondert oder römisch numerierten Beilagen,
nämlich:

"Die Welt der Literatur" (14-tägig)

"Die Reise-Welt" (14-tägig)

"Die geistige Welt" (wöchentlich)

sowie der Stellenanzeigen-Beilage in der Samstagsausgabe (wöchentlich).

c) Neues Deutschland 1954:

Als Zielmenge wurde festgelegt:

die zwischen dem 1. Januar und 31. Dezember 1954 produzierte Menge,
und zwar einschließlich der Sonntagsausgabe (anstelle der fehlenden Montagsausgabe),
mit Ausnahme der Unterhaltungsbeilage (14-tägig).

d) DIE WELT 1954:

Als Zielmenge wurde die zwischen dem 1. Januar und 31. Dezember 1954 produzierte Menge festgelegt,

mit Ausnahme der gesondert oder römisch numerierten Beilage (näheres noch nicht ermittelt).

8.2. Quantitative Beschreibung der Jahrgänge

a) für ND 1964:

Zahl der Erscheinungsmengen (e): 360

Gesamt-Seitenzahl (S_{eJ}): 2820

Zahl der Ausgaben (a): 308

Gesamt-Seitenzahl (S_{aJ}): 2125

Durchschnittliche Seitenzahl (s_a): 6.9

Durchschnittliches Erscheinungsintervall:

(bezogen auf Ausgaben):

1,1883

(bezogen auf Erscheinungsmengen):

1,0167

b) für WELT 1964:

Zahl der Erscheinungsmengen (e): 304

Gesamt-Seitenzahl (S_{eJ}): 8550

Zahl der Ausgaben (a): 304

Gesamt-Seitenzahl (S_{aJ}): 5635

Durchschnittliche Seitenzahl (s_a): 18.54

Durchschnittliches Erscheinungsintervall:

(bezogen auf Ausgaben):

1,2038

(bezogen auf Erscheinungsmengen):

1,2038

c) für ND 1954 :

Zahl der Erscheinungsmengen (e) : 304

Gesamt-Seitenzahl (S_{eJ}) : 2163

Zahl der Ausgaben (a) : 304

Gesamt-Seitenzahl (S_{aJ}) : 1976

Durchschnittliche Seitenzahl (s_a) : 6.5

Durchschnittliches Erscheinungsintervall :

(bezogen auf Ausgaben) : 1,2006

(bezogen auf Erscheinungsmengen) : 1,2006

d) für WELT 1954 :

noch nicht ermittelt.

9. Dominanzen

Zu den Dominanzen vgl. das unter Punkt 4.5.2.1. Gesagte. Für das ND wäre noch hervorzuheben, daß die jahreszeitlich bedingten Dominanzen in der Themewahl dort besonders ausgeprägt sind. Themen zur Saatzeit, zur Getreide-, Kartoffel- und Rübenernte, zur Planerfüllung (um die Zeit des Jahreswechsels) beherrschen oft wochenlang sogar die erste Seite; sie nehmen den Charakter von Propagandakampagnen an. Westliche Zeitungen wie die WELT vermeiden derartige Häufungen selbstverständlich mit Rücksicht auf die Leser.

10. Auswahlmodus

10.1. Auswahlmodus für ND 64 und 54 :

a) Aufnahmeeinheit :

Entsprechend den unter Punkt 4.6.1. dargelegten Überlegungen wurde die Seite als Aufnahmeeinheit gewählt. Der Nachteil, daß Textzusammenhänge (Artikel) gelegentlich auseinandergerissen werden, wurde dabei in Kauf ge-

nommen. Sie lassen sich außerdem ggf. ergänzen⁴⁹⁾.

b) Aufnahmeintervall:

Mit Rücksicht auf die begrenzte Kapazität konnte das Aufnahmeintervall nur in der Größenordnung zwischen drei Tagen (= 120 Stichproben) und zwei Wochen (= 26 Stichproben) gewählt werden. Wir entschieden uns für ein durchschnittliches Aufnahmeintervall von einer Woche (= 52 Stichproben).

c) Umfang der Stichproben:

Bei dem gewählten Aufnahmeintervall bestand nur die Wahl zwischen Stichprobeneinheiten von einer oder zwei Seiten Umfang. Wir entschieden uns für einen Umfang von einer Seite pro Stichprobe, da uns die damit erreichte Auswahlmenge von 52 Seiten groß genug für den beabsichtigten Zweck schien und da, sollte sich diese Annahme als falsch erweisen, eine Verdoppelung dieser Menge innerhalb des einmal gewählten Auswahlmodus ohne Schwierigkeiten möglich ist.

Für die beiden ND-Jahrgänge ergeben sich damit folgende Werte⁵⁰⁾:

	ND 1964	ND 1954
Umfang der Zielmenge:	2125 Seiten	1976 Seiten
Aufnahmeintervall:	1 Woche	1 Woche
Zahl der berücks. Ausg.: (= Stichproben)	52	52
Zahl der Seiten: (= 1 Seite p. Stichprobe)	52	52
Auswahlquote:	2,447; abgerundet 2,4%	2,632; abg. 2,6%

In diesen beiden Fällen ist die Aufnahmequote also abhängig vom Aufnahmeintervall und dem Umfang der Stichproben. Dieser Weg mußte eingeschlagen werden, da bisher noch keinerlei Erfahrungswerte für die Ermittlung von Modellmengen dieses Umfangs, insbesondere nicht aus Zeitungstexten, vorlagen.

10.2. Auswahlmodus für WELT 64

Nachdem für die beiden Jahrgänge des ND die Aufnahmequote einmal festlag, mußte bei der Aufnahme des westlichen Vergleichsjahrgangs von dieser, nicht vom Umfang der Stichproben ausgegangen werden; dabei hatte das Aufnahmeintervall gleich zu bleiben. Für WELT 1964 ergibt sich somit folgende Berechnung⁵¹⁾:

Umfang der Zielmenge	:	5.635 Seiten
Auswahlquote : 2,4%	=	135 Seiten
Aufnahmeintervall	: durchschn.	1 Woche
Zahl der Stichproben	:	52
Umfang der Stichproben	: durchschn.	2,6 Seiten

Da nur ganze Seiten aufgenommen werden, mußte der Umfang der Stichproben entweder 2 oder 3 Seiten betragen⁵²⁾. Um die Vergleichbarkeit weiter zu steigern, wurden die gleichen Tagesausgaben wie bei ND 64 zur Aufnahme bestimmt.

11. Modus der Stichprobenverteilung

Die Bestimmung der Ausgaben, aus denen die Stichproben genommen werden, geschah also bei den beiden zuerst aufgenommenen Ost-Jahrgängen nach folgendem Modus:

11.1. Horizontaler Turnus:

Jede Wochentagsausgabe soll berücksichtigt werden; daraus ergibt sich:

1. Woche: Montagsausgabe; 2. Woche: Dienstagsausgabe; 3. Woche: Mittwochsausgabe; usw. (- je 8 Tage Abstand); 6. Woche: Samstagsausgabe; 7. Woche: Montagsausgabe (= 2 Tage Abstand)

11.2. Vertikaler Turnus

Jede Seite soll berücksichtigt werden:

1. Ausgabe: Seite 1; 2. Ausgabe: Seite 2; 3. Ausgabe: Seite 3;
4. Ausgabe: Seite 4 usf. bis alle vorhandenen Seiten durchlaufen sind; danach Sprung auf Seite 1.

11.3. Verschiebungen und Abweichungen

a) Verschiebung des Turnus (11.1.) gegenüber Turnus (11.2.):

Unter den gegebenen Bedingungen unseres Materials zeigte es sich, daß beim ND meist eine bestimmte Seite auf die jeweiligen Wochentagsausgaben entfiel, also etwa auf die Montagsausgaben meist die Seite 1, auf die Samstagsausgaben meist die Seite 6 usw. Um dies zu vermeiden, wurde nach Durchlaufen je eines Turnus (also nach 6 Stichproben) der Beginn des neuen Turnus um eine Seite verschoben. Also:

6. Ausgabe: Seite 6; 7. Ausgabe nicht Seite 1, sondern Seite 2;
8. Ausgabe S. 3 usf.

b) Ausgleich entsprechend Seitenhäufigkeit:

Gemäß dem unter Punkt 4.5.4.1. dargelegten Verfahren wird für jede Seite ihre Häufigkeit in der Zielmenge absolut und anteilig ermittelt; daraus wird der Anteil jeder Seite an der Auswahlmenge von 52 Seiten (bei ND 64 und ND 54) bzw. 135 Seiten (bei WELT 64) errechnet. Vgl. dazu die Berechnung des Seitenschlüssels im Anhang Ib.

Entsprechend diesen Anteilen wird die nach dem Turnus 11.1, 11.2 und 11.3a ermittelte Verteilung korrigiert⁵³⁾.

c) Materialbedingte Abweichungen:

Das Archivmaterial, auf das wir uns stützen⁵⁴⁾, war leider gelegentlich lückenhaft. So war nicht immer die Berliner Ausgabe des ND vorhanden;

in diesen Fällen wurde auf die Republikausgabe ausgewichen. Fehlte auch diese, wurde die auf die fehlende folgende Ausgabe genommen.

Entsprechendes gilt auch für die WELT Berliner Ausgabe. Hier mußte in einigen Fällen die Hamburger Ausgabe genommen werden. Außerdem enthielt der uns übersandte Mikrofilm in 2 Fällen irrtümlich aufgenommene Seiten, nämlich solche aus der Beilage "Reise-Welt". Wir hielten den Fehler jedoch für so geringfügig, daß wir die entsprechenden Seiten in unserem Material belassen haben.

Die nach diesem System aus jedem Jahrgang tatsächlich aufgenommenen Seiten sind im Anhang IIa zusammengestellt.

12. Zusätze zur Modellmenge:

Zum Zwecke statistischer Einzeluntersuchungen (vgl. Teil III) wurde jeweils aus dem Monat Februar jedes der drei Jahrgänge eine bestimmte Zahl von Seiten zusätzlich aufgenommen, um eine höhere Auswahlquote zu erreichen. Aus ND 64-Februar wurden zu den 4 Seiten der Modellmenge weitere 14 Seiten aufgenommen, um eine Auswahlquote von 10% zu erreichen (Gesamtmenge Februar: 184 Seiten⁵⁵⁾; aus WELT 64-Februar zu den vorhandenen 10 Seiten weitere 12; aus ND 54-Februar zu den vorhandenen 4 Seiten weitere 5.

Ferner wurden zum Zwecke der vergleichenden Untersuchung des Wortschatzes bestimmter prominenter Redner einige Seiten mit offiziellen Reden von W. Ulbricht und einigen anderen Mitgliedern des Politbüros aus ND 64 aufgenommen. Der Modellmenge gehören nur jeweils 1 - 2 Seiten dieser Reden an; sie sind durch den Anhang jetzt jedoch komplett vorhanden.

Schließlich wurden aus verschiedenen Gründen noch einige weitere Seiten aufgenommen, vor allem bei ND 54. Sie sind in einem besonderen Anhang zusammengefaßt.

Umfang und Seitenverteilung der Zusätze siehe Anhang IIb.

13. Besondere Gesichtspunkte bei der Übertragung der Texte auf Datenträger

13.1. Artikel

Die Grundlage der statistischen Textermittlung und -berechnung bildet, wie erwähnt, die Zeitungseite.

Diese Einheit ist jedoch überwiegend formaler Natur. Darüberhinaus ist das Material in Sinneinheiten gegliedert, die wir "Artikel" nennen, wobei dieser Begriff hier eine erweiterte Bedeutung hat. Unter einem Artikel verstehen wir jede inhaltlich-thematisch zusammengehaltene und (bedingt) formal erkennbare Texteinheit, also auch etwa Wetterbericht, Börsentabellen, Kleinanzeigen, die unter einer Überschrift zusammengefaßt sind, Leserbriefe zum gleichen Thema, Romanfortsetzungen usw. Im einzelnen sind die Grenzen fließend. Sämtliche Artikel, die in einer Modellmenge enthalten sind, erhalten eine fortlaufende Numerierung⁵⁶⁾. Die maschinelle Verarbeitung des Materials orientiert sich vor allem an diesen Artikelnummern, nicht an den Seiten.

13.2. Schreibkonventionen⁵⁷⁾

Abgeschrieben, d.h. auf Lochkarten übertragen⁵⁸⁾ werden alle auf den Seiten vorhandenen Texte mit Ausnahme

- a) des Zeitungskopfes
- b) des Impressums
- c) reiner Zahlentabellen
- d) von Bildern, Zeichnungen usw.

In den Fällen c) und d) wird ein Vermerk aufgenommen, was an diesen Stellen übergangen wurde⁵⁹⁾.

13.2.1. Informationskonstanz bei einzelnen Wörtern und Zeichen

Texte bestehen aus Buchstabenfolgen, Zeichen (Satzzeichen) und Zahlen (die letzteren sind hier irrelevant) sowie aus Kombinationen dieser Gattungen.

Zahlreiche Buchstabenfolgen ("Wörter") und Zeichen sind nur im Textzusammenhang eindeutig, isoliert jedoch mehrdeutig.

Da unsere Datenverarbeitung zunächst auf die Isolierung der einzelnen "Wörter" und Zeichen abzielt, wird mittels bestimmter Vorschriften versucht, einen Teil der entstehenden Informationsverluste aufzufangen, soweit sie für die maschinelle Verarbeitung von Bedeutung sind.

Satzzeichen und Sonderzeichen:

Da eine Reihe von Zeichen auf den Lochkartenschreibern fehlen, müssen Ersatzzeichen⁶⁰⁾ verwendet werden:

Doppelpunkt	= / 0
Fragezeichen	= / 1
Anführungszeichen	= / 2
Ausrufezeichen	= / 3
Semikolon	= / 4
einf. Anführungszeichen	= / 5
Prozentzeichen	= / 6
Paragraphenzeichen	= / 7
öffnende eckige Klammer	= / 8
schließende eckige Klammer	= / 9
Abkürzungspunkt	= ' .
Auslassungspunkte	= ' ' ' .

Im übrigen gelten die schon in den "Schreibanweisungen für die Übertragung von Texten auf Lochstreifen" - Mannheimer Bericht - unter Punkt 5a, 7, 9 - 11, 13 - 17 erwähnten Vorschriften.

Abweichend gilt⁶¹⁾: Großschreibung (nur bei Substantiven und großgeschriebenen Anreden) wird durch das Zeichen * (Sternchen) vor dem Wort markiert; großgeschriebene Adjektive (vor allem als Namensbestandteile) werden durch ' * (Apostroph-Stern) vor dem Wort gekennzeichnet; Personen-Familiennamen erhalten zu dem Großschreibungszeichen noch das Kennzeichen Stern-Stern-blank vor dem Wort.

Unberücksichtigt bleiben semantische oder inhaltliche sowie grammatische Mehrdeutigkeiten.

Ebenfalls unberücksichtigt bleiben Schriftartunterschiede (Kursiv-, Sperr-, Fett- oder Petitdruck usw.), diakritische Zeichen, der Unterschied zwischen römischen und arabischen Zahlen (römische werden in arabische überführt⁶²⁾ sowie der Unterschied zwischen "Anführungsstriche unten" und "Anführungsstriche oben".

13.2.2. Informationskonstanz bei Wortgruppen

Bei der maschinellen Zerlegung unserer Texte in einzelne Wörter und Zeichen treten weitgehende Informationsverluste auf, die teils die Bedeutung bestimmter Gruppen und Wörter, teils charakteristische Eigenheiten größerer Textteile betreffen.

Diese Informationsverluste werden zum Teil durch ein System von Transkriptionszeichen aufgefangen⁶³⁾. Es entspricht im wesentlichen dem in Mannheim verwendeten System (vgl. "Schreibanweisungen" Punkt. 6a-f)⁶⁴⁾.

Zusätzlich werden transkribiert: Personeneigennamen samt ihren Titeln und Anreden, Sach-Eigennamen (nur mehrgliedrige)⁶⁵⁾, übersetzte Textstellen, Abkürzungen (soweit nicht mit Punkt abgekürzt)⁶⁶⁾ sowie Textstellen, die nicht schreibbar sind bzw. wegen Irrelevanz (Zahlenkolonnen) nicht geschrieben werden sollen⁶⁷⁾.

Die durch die Transkription dem Text beigefügten Informationen bleiben bei der Zerlegung den einzelnen Wörtern erhalten.

Ebenso können die zu einer transkribierten Textstelle gehörigen Wörter wieder zusammengefügt werden.

13.2.3. Informationskonstanz des Artikel-Charakters

Der Zeitungsleser entnimmt jedem Zeitungsartikel als ganzem schon bei flüchtigem Anschauen eine Reihe von Informationen, die ebenfalls bei der Zerlegung verlorengehen.

Ein Teil dieser Informationen, die sich auf den ganzen Artikel beziehen, wird auf "Informationskarten" gespeichert, d.h. chiffrierten Lochkarten, die jedem Artikel vorangestellt werden⁶⁸⁾. Die Informationskarte (IK) enthält (außer einer Chiffre für die Kartenart und der Nummer des Artikels, auf den sich die IK bezieht) folgende Angaben:

1. Chiffre der Zeitung mit Jahrgang, Monat, Tag, Seite
2. Herkunft der Zeitung (West oder Ost)
3. Aufnahmeprinzip (Statistisches = Modellmenge)
4. Art der Zeitung (Tages-, Wochen- usw.)
5. Angabe, ob es sich um ein Artikelfragment handelt⁶⁹⁾
6. Verfasser des Artikels (falls angegeben)
7. Agenturen (bis zu drei)
8. Sachgebiete (bis zu drei gleichzeitig)⁷⁰⁾
9. Innere Form
 - a) Publizistisches Ziel (bis zu zwei gleichzeitig)⁷¹⁾
 - b) Artikelart bzw. -gattung mit Sonderformen
 - c) Sparte, der der Artikel eingeordnet ist
10. Äußere Form (Aufmachung)
 - a) Breite (in Spalten)
 - b) Länge (in Zeilen)
11. Angaben über stilistische Kriterien (nicht ausgefüllt).

Auf einer weiteren Karte, der "Beikarte" (BK) oder "Themakarte" wird, außer der Artikelnummer, das Thema, das in dem Artikel behandelt wird, angegeben. Die Karte ist gegliedert in einen "Kopfteil", der das "Land" oder die "Länder" angibt, auf das sich das behandelte Thema bezieht, und einen "Textteil", der das Thema in Stichworten wiedergibt.

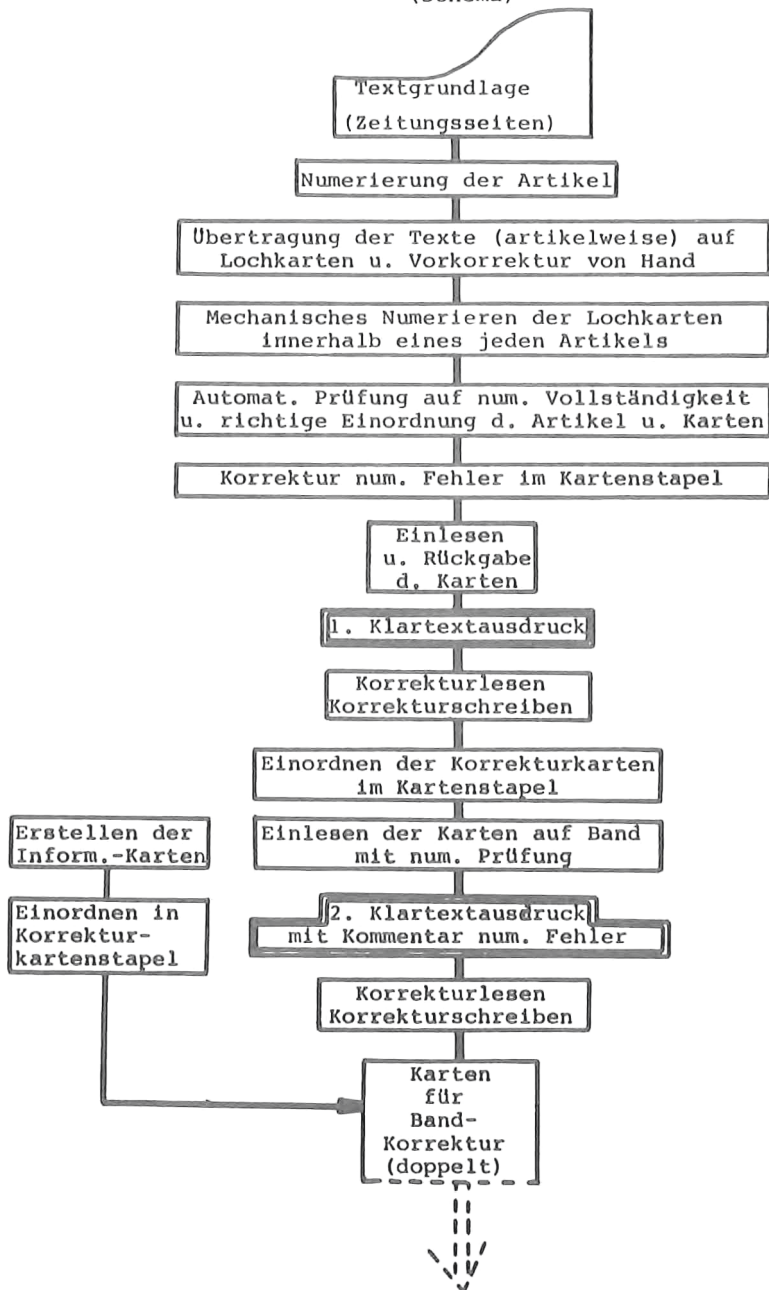
Die Angaben dieser beiden Karten werden in nicht chiffrierter, dafür begrenzter Form zusätzlich auf einer Handkartei (Artikelkartei) festgehalten.

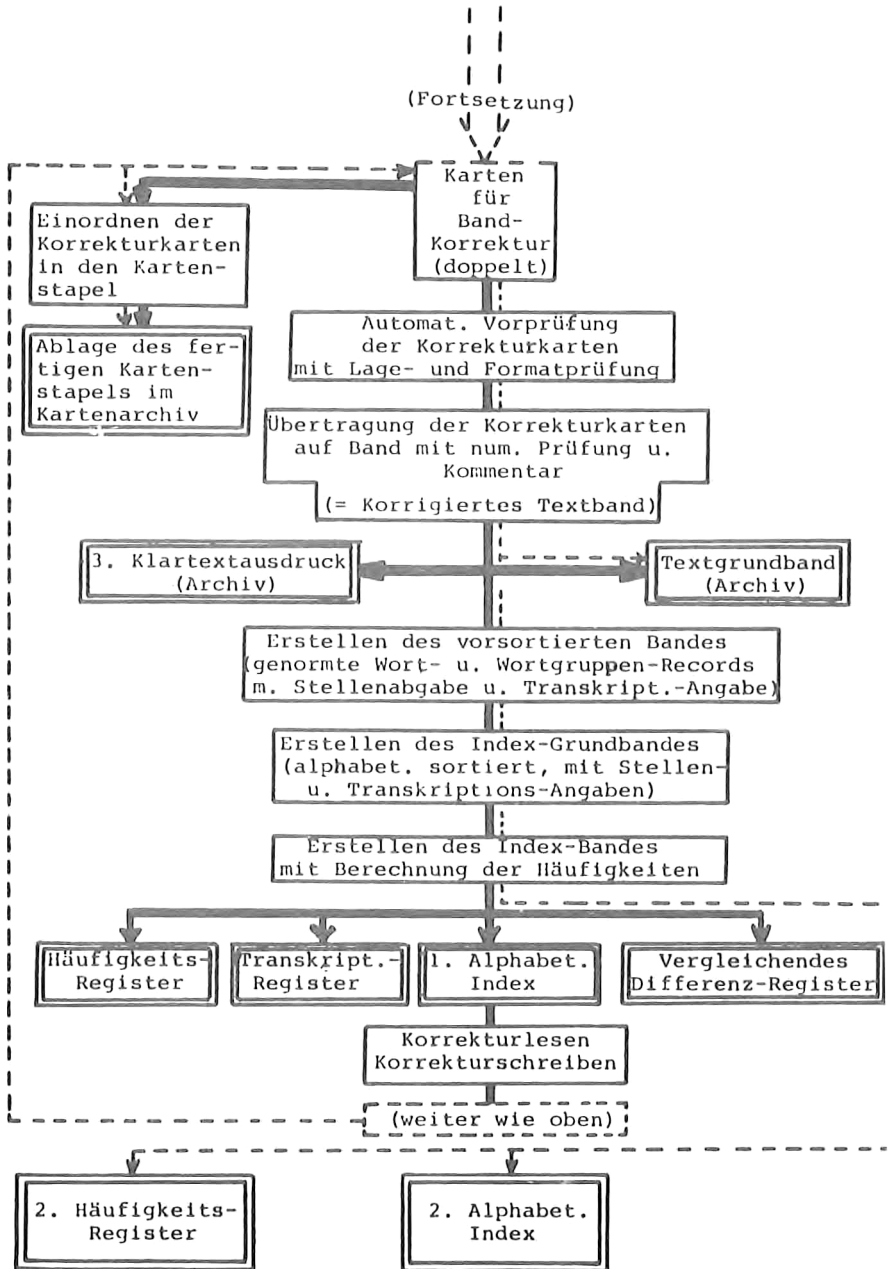
Auf einer dritten Lochkarte ("Zusatz"- oder "Verweiskarte") (ZK) können Artikel angegeben werden, die mit dem behandelten inhaltlich oder thematisch korrespondieren oder ihn ergänzen (falls es sich um ein Fragment gehandelt hat)⁷².

Eine Übersicht über den Aufbau der Informationskarte siehe im "Auszug aus den Erläuterungen zur Informationskarte" im Anhang III.

A_r_b_e_i_t_s_g_a_n_g

(Schema)





14.1. Erläuterungen und technische Angaben

zum Arbeitsablauf und zu den verwendeten Geräten

1. Ablochen. Wir benutzen die IBM-80-Spalten-Standard-Lochkarte in zwei Farben (für Ost- und Westtexte)⁷³⁾.

Aufbau einer Textkarte: Sp. 1-4 = Artikelnummer

Sp. 5-7 = Kartenummer

Sp. 8 = blank

Sp. 9-80 = fortlaufender Text, ohne Rücksicht auf Zeitungs-Zeilenlänge.

Geräte⁷⁴⁾: zwei IBM-Schreibblocher 026 (mit Textzeile am oberen Rand der Karten)⁷⁵⁾.

Das Gerät verfügt über 46 Zeichen (Symbolausführung H).

2. Mechanisches Numerieren. Die Artikelnummer wird schon beim Ablochen in die Karten dupliziert, die laufende Kartenummer im Doppler IBM Typ 519 artikelweise hinzugefügt (gestanzt und aufgedruckt).
3. Automatische numerische Prüfung. Programm PRUEF des IDS mit Unterprogrammen (vgl. Anhang IV, Programmübersicht Nr. 1).
4. Klartextausdrucke. Der Klartextausdruck gibt den Text kartenweise wieder (jede Ausdruckzeile entspricht einer Karte). Jeder Artikel beginnt auf einer neuen Seite. Unregelmäßigkeiten in Numerierung und Kartenformat werden kommentiert. Der Klartextausdruck kann eine für das Korrigieren eingerichtete Form haben.
Am Ende eines jeden Artikels erscheint ein Schreibervermerk (Namenssigle und Datum des Schreibens).
5. Erste Korrekturphase. Die Texte werden nur durch Austausch der falschen gegen die neuen richtigen Karten im Kartenstapel korrigiert. Erst danach werden die Texte insgesamt auf Band gelesen und dabei nochmals numerisch geprüft.

6. Zweite Korrekturphase (auf Magnetband).

Vorprüfung: Programm VORKOR (Programmübersicht Nr. 2)

Korrektur: Programm KORR und NDNUM (Programmübersicht Nr. 3 und 4)

In der zweiten Korrekturphase werden meist auch die Informationskarten (sowie Bei- und Zusatzkarten) hinzugefügt (jeweils als erste Karte jedes Artikels).

7. Textband

Band: IBM MC1.800, Länge 1000 feet

Schreibdichte: 556 b.p.i.

Recordlänge: 14 á 6

Blockung: variabel, gewöhnlich 1400

Schreibmodus: wählbar, gewöhnlich BCD

Allgemein: Siehe Programmbeschreibung EINAUS des IPK Bonn (Jahresbericht des IPK 1966/7, S. 39)

Dieses Textband ist Datenträger für die Suchwort-Programmläufe.

8. Erstellen des Index. Siehe "Programmübersicht" Nr. 16-20

8a. Der alphabetische Wortformen-Index enthält alle Wortformen mit Stellenangaben (Artikel- und Zeilennummer), ferner Angaben zur etwaigen Transkription des einzelnen Wortes zu jeder Fundstelle sowie relative und absolute Häufigkeit. Bestimmte oder alle Transkriptionen können dabei ausgefiltert werden. Bindestrich-Komposita werden unter allen ihren Teilwörtern aufgeführt, jedoch nur einmal gezählt.

8b. Index nach Kriterien der Informationskarten. Es können Indices aus Artikeln erstellt werden, auf die ein oder mehrere gewünschte Kriterien der Informationskarte zutreffen. Gleiches gilt für Indices nach Beikarten (Indices zu Artikeln über bestimmte Themen). (Programmübersicht Nr. 5)

9. Transkriptionsregister. Vom Index-Grundband ausgehend können alphabetische Spezial-Verzeichnisse zu allen Transkriptionen in Form alphabetischer Wortformen-Indices oder in Form von Wortgruppen-Registern erstellt werden (Programmübersicht Nr. 19).

10. Kombinierbarkeit. Die Auswahl- bzw. Spezialisierungsverfahren nach Kriterien der Informationskarten und nach Transkriptionen können kombiniert werden⁷⁶⁾. Die Zahl der sich aus der Kombination der vielfachen Abfragemöglichkeiten der Informations- und Beikarte mit den Möglichkeiten des Transkriptionssystems ergebenden sinnvollen und realisierbaren speziellen Verzeichnisse liegt bei mehreren tausend (die Zahl der theoretischen Möglichkeiten ist nicht genau angebar, da einige Chiffregruppen laufend erweiterbar sind; sie liegt aber über 2 Milliarden).

11. Vergleichende Differenzregister zur Ermittlung des abweichenden oder gemeinsamen Wortschatzes zweier Texte liegen noch nicht vor; das Programm ist in Arbeit. (Programm-Übersicht Nr. 21)

12. Dritte Korrekturphase. Erfahrungsgemäß zeigen sich manche Schreibfehler erst im Index⁷⁷⁾. Nochmalige Korrektur über alle Stufen der Verarbeitung hinweg ist also notwendig. Nach diesem Korrekturgang liegt die Fehlerquote in unseren Texten⁷⁸⁾ unter 0,5 Promille (wahrscheinlich bei 0,2 - 0,1 Promille). Dies dürfte ein bei großen Textmengen in der Praxis äußerst hoher Grad von Fehlerfreiheit sind. Er liegt selbstverständlich weit über der Fehlerfreiheit der Zeitungstexte selbst (Druckfehler im Zeitungstext werden beim Ablochen berichtigt).

15. Übersicht über den Stand der Arbeiten

(Stand von Frühjahr 1968)

15.1. Texte

1. Neues Deutschland 1964

a) Modellmenge:

52 Seiten, 680 Artikel, ca. 180.000 lfd. Wörter

Dieser Text lag einschließlich Informationskarten während des Jahres 1967 fertig zur Benutzung vor; es bestehen ein alphabetischer Index sowie Teilindices, ein Häufigkeitsregister und einige Transkriptionsregister.

b) Zusatzmenge 1 ("Februarmenge")

14 Seiten, 140 Artikel ca. 50.000 lfd. Wörter

Benutzbar seit Sommer 1967, Informationskarten teilweise ausgefüllt.

c) Zusatzmenge 2 (offizielle Reden)

13 Seiten, 25 Artikel, ca. 46.000 lfd. Wörter

Benutzbar seit Frühjahr 67. Informationskarten teilweise ausgefüllt.

d) Zusatzmenge 3 (sonstige Seiten)

3 Seiten, 30 Artikel, ca. 10.000 lfd. Wörter

Benutzbar seit Sommer 67. Informationskarten fertig ausgefüllt.

Das Gesamtkorpus aus ND 64 umfaßt somit 82 Seiten mit über 280.000 lfd. Wörtern.

Ein Gesamtindex und Häufigkeitsregister zu allen vier Einzelmengen gemeinsam werden vorbereitet.

2. DIE WELT 1964

a) Modellmenge:

135 Seiten, 1640 Artikel, ca. 410.000 lfd Wörter

Der Text liegt seit Ende 1967 korrigiert zur Benutzung vor. Informationskarten sind zu ca. 1/3 erstellt. Es sind ein Teilindex sowie Teilregister von Transkriptionen vorhanden.

- b) Zusatzmenge 1 ("Februarmenge")
12 Seiten, ca. 125 Artikel, ca. 35.000 lfd. Wörter
Korrigiert benutzbar seit Anfang 1968 (noch ohne Informationskarten).
- c) Zusatzmenge 2 (offizielle Reden)
entfällt.
- d) Zusatzmenge 3 (sonstige Seiten)
5 Seiten, 40 Artikel, ca. 14.000 lfd. Wörter
Korrigiert benutzbar seit Frühjahr 68 (noch ohne Informationskarten).

Das Gesamtkorpus umfaßt somit 152 Seiten mit ca. 460.000 lfd. Wörtern.

3. Neues Deutschland 1954

Dieser Text wurde noch vor dem Jahrgang WELT 64 auf Lochkarten übertragen, jedoch erst nach der WELT 64 korrigiert.

- a) Modellmenge
52 Seiten, 610 Artikel, ca. 165.000 lfd. Wörter
Der Text lag im Frühjahr 1968 einmal korrigiert und somit begrenzt zur Benutzung vor (noch ohne Informationskarten). Indices sind im Herbst 1968 zu erwarten.
- b) Zusatzmenge 1 ("Februarmenge")
5 Seiten, 62 Artikel, ca. 16.000 lfd. Wörter
Benutzbarkeit wie unter a)
- c) Zusatzmenge 2 (offizielle Reden)
entfällt
- d) Zusatzmenge 3 (sonstige Seiten)
11 Seiten, ca. 130 Artikel, ca. 35.000 lfd. Wörter
Benutzbarkeit wie unter a).

Das Gesamtkorpus aus ND 54 umfaßt also 68 Seiten mit ca. 216.000 Wörtern. Es wird voraussichtlich im Herbst 1968 einschließlich Informationskarten voll zur Verfügung stehen.

4. DIE WELT 1954

Mit der Aufnahme der Texte aus WELT 54 wird im Spätsommer 1968 begonnen werden. Die Auswahl der Texte wird voraussichtlich 90 Seiten umfassen.

Erweiterungēn

Es ist zu erwarten, daß in absehbarer Zeit westliche Zeitungstexte aus der laufenden Produktion über Zeitungslochstreifen (TTS-Streifen) unmittelbar auf Band genommen werden können. Ein Umwandlungsprogramm, das die Streifentexte in eine unseren Schreibkonventionen stark angenäherte Form bringt, ist in der Außenstelle erstellt worden ⁷⁹⁾.

15.2. Auswertung

Die Außenstelle hat sich, ihrer Aufgabe entsprechend, auf solche Auswertungsprogramme konzentriert, die für Wortschatzuntersuchungen erforderlich sind. Um die Bonner Texte jedoch auch für Mannheimer syntaktische Untersuchungen und umgekehrt verwertbar zu machen, ist in der Außenstelle ein Umwandlungsprogramm erstellt worden, das die Bonner Textbänder auf die für die Mannheimer Bänder vom Rechenzentrum Darmstadt geforderten Konventionen umstellt bzw. Mannheimer Bänder auf die in Bonn gültigen Konventionen umstellt (vgl. Programm-Übersicht Nr. 12 und 13).

Zur Auswertung der Texte in Form von Indices oder Transkriptionsregistern, die mit Abfragen nach Kriterien der Informationskarte gekoppelt sein können, vgl. Punkt 14.1., Abschnitt 8-10. Solche Register aufgrund kombinierter Abfragen wurden auf den Tagungen des Instituts im Herbst 1965 und Frühjahr 1968 vorgelegt (vgl. Anmerkung 76).

Als wesentlich wichtiger für Arbeiten zum Wortschatz hat sich das unter Nr. 6 und 7 der Programm-Übersicht kurz beschriebene System "MOSU" zur Suche von Wörtern mit Kontext erwiesen. Es können aufgrund vorzuziehender Buchstabenkombinationen ganze Wörter oder Wortteile, d.h. Wort"stämme", -anfänge, -endungen, gesucht werden⁸⁰⁾. Einschränkung auf Suche nach Substantiven oder Nicht-Substantiven ist möglich. Alle gefundenen Wörter werden gesondert sowie in ihrem Kontext (variabel zwischen 1 und 9 Zeilen Kontext oder 2 - 7 Wörter Kontext oder 1 Satz Kontext) ausgedruckt. Jedem Befehl können Angaben aus der Informationskarte (dechiffriert) beigegeben werden⁸¹⁾. Der Kontext kann von allen störenden Zusatz- oder Transkriptionszeichen gereinigt werden.

Mit Programmläufen dieser Art hat die Außenstelle in den Jahren 1966 und 1967 18 verschiedene wissenschaftliche Arbeitsvorhaben unterstützt. Der Umfang der ausgegebenen Menge beträgt über 300.000 Zeilen Kontext. Unter den 12 im Jahre 1967 geförderten Arbeitsvorhaben befanden sich 4 von ehemaligen oder jetzigen Mitarbeitern des Instituts in Mannheim, Bonn und Innsbruck, 4 betrafen Staatsexamensarbeiten oder Dissertationen der Bonner Universität, 4 weitere betrafen Arbeitsvorhaben von anderen Wissenschaftlern. Ein Teil der Arbeiten ist abgeschlossen.

Das Programmsystem wird weiter ausgebaut und verfeinert werden.

Anhang I a / 1

Seitenberechnung für ND 1964:

Ausgegangen wurde, wie unter Punkt 10.1. beschrieben, vom Umfang der Auswahlmenge, d.h. von einer Seite pro Zeitungswoche = 52 Seiten.

Ausgezählt wurden 153 Erscheinungsmengen aus 5 Monaten (Januar, Februar, April, Juli, Oktober) mit 1200 Seiten, bzw. 132 Ausgaben (d.h. ohne Sonntagsausgaben und Beilagen) mit 912 Seiten.

Auf 360 Jahres-Erscheinungsmengen umgerechnet ergibt sich daraus eine Gesamt-Bezugsmenge S_{eJ} von 2820 Seiten; auf 308 Jahrgangs-Ausgaben umgerechnet eine Zielmenge S_{aJ} von 2125 (vgl. Mengenbeschreibung unter Punkt 8.2.)⁸⁾ 52 Seiten von 2125 sind 2,45 %, \sim 2,4 %.

Aufstellung der ausgezählten Ausgaben verschiedener Länge in Klassen:

Klasse	Länge d. Ausgaben in Seiten (y bis y+m)	Zahl der Ausgaben (p bis p_i)	Zahl der Seiten pro Klasse (p mal y)	Häufigkeit (F) jeder Seite in 912 S. ausgez. Menge
1	4	17	68	4 x 132
2	5/6	46	276	2 x 115
3	7/8	65	520	2 x 69
4	9/10	--	--	2 x 4
5	11/12	4	48	2 x 4
		n=132	912	

Entsprechend der unter 4.5.4.2. ermittelten Formel $f = F \cdot \frac{S_M}{S_{aJ}}$ wird die Anzahl

jeder Seite pro Klasse in der Auswahlmenge von 52 Seiten ermittelt; dabei ist $\frac{S_M}{S_{aJ}}$ hier = 0,057⁸³⁾.

Anhang 1 a / 2

Seitenberechnung für WELT 1964

Ausgezählt wurden 131 Ausgaben aus 5 Monaten (Januar, Februar, April, Juli, Oktober) mit insgesamt 2429 Seiten⁸⁴⁾. Auf das Jahr, d.h. auf 304 Ausgaben bezogen, ergibt sich eine Zielmenge von ~ 5635 Seiten (vgl. Mengenbeschreibung Punkt 8.2.).

Bei einer Auswahlquote von 2,4 % muß der Umfang der Auswahlmenge somit 135 Seiten betragen.

Aufstellung der ausgezählten Ausgaben verschiedener Länge in Klassen:

Klasse	Länge d. Ausgaben in Seiten (y bis y+m)	Zahl der Ausgaben (p bis p _i)	Zahl d. Seiten pro Klasse (p mal y)	Häufigkeit (F) jeder Seite in 2428 ausgez. Menge
1	12	10	120	12 x 131
2	13/14	26	364	2 x 121
3	15/16	35	560	2 x 95
4	17/18	21	378	2 x 60
5	19/20	8	160	2 x 39
6	21/22	10	220	2 x 31
7	23/24	0	0	2 x 21
8	25/26	7	182	2 x 14
9	27/28	4	112	2 x 14
10	29/30	1	30	2 x 10
11	31/32	3	96	2 x 9
12	33/34	5	170	2 x 6
13	35/36	1	36	2 x 1
		n=131	2428	

Entsprechend der Formel $f = F \cdot \frac{S_M}{S_{aJ}}$

wird die Anzahl jeder Seite pro Klasse in der Auswahlmenge von 135 Seiten ermittelt; dabei ist $\frac{S_M}{S_{aJ}}$ hier = 0,0556.

Anhang I a / 3

Seitenberechnung für ND 1954:

Ausgegangen wurde, wie bei ND 64, vom Umfang der Auswahlmenge, d.h. von einer Seite pro Zeitungswoche = 52 Seiten.

Ausgezählt wurden 104 Erscheinungsmengen aus 4 Monaten (Januar, April, Juli, Oktober) mit 740 Seiten bzw. ebenfalls 104 Ausgaben⁸⁵⁾ (ohne Beilagen) mit 676 Seiten.

Auf 304 Jahreserscheinungsmengen umgerechnet ergibt sich daraus eine Gesamtmenge S_{eJ} von 2163 Seiten; auf 304 Jahrgangsausgaben umgerechnet eine Zielmenge S_{aJ} von 1976 Seiten (vgl. Mengenbeschreibung, Punkt 8.2.). 52 Seiten von 1976 Seiten sind 2,63 %.

Aufstellung der ausgezählten Ausgaben verschiedener Länge in Klassen:

Klasse	Länge d. Ausgaben in Seiten (y bis y+m)	Zahl der Ausgaben (p bis p_i)	Zahl d. Seiten pro Klasse (p mal y)	Häufigkeit (F) jeder Seite in 676 S. ausgezählter Menge
1	4	10	40	4 x 104
2	5/6	59	354	2 x 94
3	7/8	34	272	2 x 35
4	9/10	1	10	2 x 1
n = 104			676	

Entsprechend der Formel $f = F \cdot \frac{S_M}{S_{aJ}}$ wird die Anzahl jeder Seite pro Klasse

in der Auswahlmenge von 52 Seiten ermittelt; dabei ist $\frac{S_M}{S_{aJ}}$ hier = 0.0769.

Anhang I b Berechnung der Häufigkeit jeder Seite in der Auswahlmenge

Neues Deutschland 64				DIE WELT 64			
Seite	Vork. in der ausgez. Menge (912 S.)	Vork. in der Auswahlmenge (52 S.)		Seite	Vork. in der ausgez. Menge (2429 S.)	Vork. in der Auswahlmenge (135 S.)	
F mal 0,057 = f				F mal 0,0556 = f			
			rechn. rund				rechn. rund
1	132	7,52	7	1	131	7,28	8
2	132	7,52	7	2	131	7,28	7
3	132	7,52	7	3	131	7,28	7
4	132	7,52	7	4	131	7,28	7
5	115	6,56	7	5	131	7,28	8
6	115	6,56	7	6	131	7,28	7
7	69	4,00	5	7	131	7,28	7
8	69	4,00	4	8	131	7,28	7
9	4	0,23	-	9	131	7,28	8
10	4	0,23	1	10	131	7,28	7
11	4	0,23	-	11	131	7,28	7
12	4	0,23	-	12	131	7,28	7
<hr/>				13	121	6,73	7
<hr/>				14	121	6,73	7
<hr/>				15	95	5,28	5
<hr/>				16	95	5,28	5
<hr/>				17	60	3,34	4
<hr/>				18	60	3,34	3
<hr/>				19	39	2,17	2
<hr/>				20	39	2,17	2
<hr/>				21	31	1,72	2
<hr/>				22	31	1,72	2
<hr/>				23	21	1,17	1
<hr/>				24	21	1,17	1
<hr/>				25	21	1,17	1
<hr/>				26	21	1,17	1
<hr/>				27	14	0,78	1
<hr/>				28	14	0,78	1
<hr/>				29	10	0,56	1
<hr/>				30	10	0,56	-
<hr/>				31	9	0,50	1
<hr/>				32	9	0,50	1
<hr/>				33	6	0,33	-
<hr/>				34	6	0,33	-
<hr/>				35	1	0,05	-
<hr/>				36	1	0,05	-
<hr/>				<hr/>			
<hr/>				2.428	<hr/>		135
<hr/>				<hr/>			

ANHANG II a

Liste der für die Modellmengen aufgenommenen Seiten

Nr.	N D 1964		DIE W E L T 1964		N D 1954		DIE W E L T 1954	
	Datum	Seite	Datum	Seiten	Datum	Seite	Datum	Seiten
1)	Do 2.	1.	Do 2.	1,	Fr 1.	1.	Fr 1.	1.
2)	Fr 10.	1.	Fr 10.	3,	So 10.	1.	Mo 11.	1.
3)	Sa 18.	1.	Sa 18.	5,	Di 19.	1.	Di 19.	1.
4)	Mo 20.	1.	Mo 20.	7,	Mi 27.	1.	Mi 27.	1.
5)	Di 28.	1.	Di 28.	10,11,12	Do 4.	2.	Do 4.	2.
6)	Mi 5.	2.	Mi 5.	13,14	Fr 12.	2.	Fr 12.	2.
7)	Do 13.	2.	Do 13.	15,16	Sa 20.	2.	Sa 20.	2.
8)	Fr 21.	2.	Fr 21.	1,	Mo 22.	2.	Mo 22.	2.
9)	Sa 29.	2.	Sa 29.	2,	Di 2.	3.	Di 2.	3.
10)	Mo 2.	3.	Mo 2.	3,	Mi 10.	3.	Mi 10.	3.
11)	Di 10.	3.	Di 10.	5,	Do 18.	3.	Do 18.	3.
12)	Mi 18.	3.	Mi 18.	7,	Fr 26.	3.	Fr 26.	3.
13)	Do 26.	3.	Do 26.	9,10	Sa 3.	4.	Sa 3.	4.
14)	Fr 3.	4.	Fr 3.	11,12,13	Mo 5.	4.	Mo 5.	4.
15)	Sa 11.	4.	Sa 11.	15,16	Di 13.	4.	Di 13.	4.
16)	Mo 13.	4.	Mo 13.	17,18,31,32	Mi 21.	4.	Mi 21.	4.
17)	Di 21.	4.	Di 21.	1,13,14	Do 29.	4.	Do 29.	4.
18)	Mi 29.	4.	Mi 29.	19,20	Fr 7.	5.	Fr 7.	5.
19)	Fr 15.	5.	Fr 15.	21,22	Sa 15.	5.	Sa 15.	5.
20)	Sa 23.	5.	Sa 23.	3,	Mo 17.	5.	Mo 17.	5.
21)	Mo 25.	5.	Mo 25.	5,	Di 25.	5.	Di 25.	5.
22)	Di 2.	6.	Di 2.	7,	Mi 2.	6.	Mi 2.	6.
23)	Mi 10.	6.	Mi 10.	8,	Do 10.	6.	Do 10.	6.
24)	Do 18.	6.	Do 18.	9,10	Fr 18.	6.	Fr 18.	6.
25)	Fr 26.	6.	Fr 26.	11,12	Sa 26.	6.	Sa 26.	6.
26)	Sa 4.	7.	Sa 4.	2,13,14	So 27.	6.	Mo 28.	6.
				17,18,26,27				

noch nicht ermittelt

ANHANG II b

Liste der zusätzlich aufgenommenen Seiten

a) Seiten zur Verdichtung der jeweiligen
Februarmenge

N D 64			DIE W E L T 64			N D 54		DIE W E L T 54	
Datum	Seiten		Datum	Seiten		Datum	Seite	Datum	Seiten noch nicht festgelegt
Fr 31. 1.	2,3		Sa 1.2.	1,2,4		So 31. 1.	1	Mo 1.2.	
Di 5. 2.	1		Mo 10.2.	5,6,7		Di 9. 2.	5	Mo 8.2.	
lo 10. 2.	2,3		Mo 17.2.	8,9,10		Di 16. 2.	6	Di 16.2.	
Do 13. 2.	6,8		Di 25.2.	11,12,13		Mi 24. 2.	8	Fr 26.2.	
lo 17. 2.	2,7				So 28. 2.	4			
Fr 21. 2.	4								
Di 25. 2.	4,5								
sa 29. 2.	1,8								

b) Seiten zur Ergänzung vorhandener Artikel
(offizielle Reden)

Di 5. 2.	(1),3 5,6,7			
Do 13. 2.	3,4			
Do 6. 2.	3,4,5			
sa 29. 2.	(1),4, 5,6,8			
so 1. 3.	5,6	(Einklammerung = teilweise)		

c) Sonstige zusätzliche Seiten

Di 22. 1.	2	Mi 22.1.	1,2	So 11. 4.	7	
Di 2. 9.	8	Do/Fr 7./8.5.		Sa 9. 1.	8	Sa 9.1.
Di 6.10.	3		1,2	So 17. 1.	3	
		Sa 11.4.	2	Do 29. 4.	8	
				So 23. 5.	6	
				So 4. 7.	6,7	
				So 15. 8.	2	
				So 26. 9.	3	
				So 7.11.	7	
				So 19.12.	1	

Anhang III

Auszug aus den Erläuterungen zur Informationskarte

1. Spaltenfeld (Sp. 1 - 4)
Kartenkennung
2. Spaltenfeld (Sp 5 - 9)
Jahrgangschiffre und Artikelnummer
3. Spaltenfeld (Sp. 10)
Angabe, ob vollständiger Artikel oder Fragment
4. Spaltenfeld (Sp. 11 - 13)
Publikationsform
5. Spaltenfeld (Sp. 14)
Herkunft des Textes (Ost oder West)
6. Spaltenfeld (Sp. 15)
Aufnahmeprinzip (1 = Modellmenge, 2ff = Zusätze)
7. Spaltenfeld (Sp. 16 - 26)
Stellenangabe: Sigle des Textes (der Zeitung), Jahr, Monat, Tag, Seite
8. Spaltenfeld (Sp. 27 - 48)
Angabe des Verfassers (sofern namentlich genannt) bzw. Ersatzformen solcher Angaben
9. Spaltenfeld (Sp. 49 - 57)
Sigle(n) der Agentur(en), soweit erwähnt (bis zu drei nebeneinander)
10. Spaltenfeld (Sp. 58 - 63)
Angabe des oder der Sachgebiete, in das (die) der Artikel gehört. Es können bis zu drei Sachgebiete angegeben werden. Die Reihenfolge soll dem Gewicht (Rang) entsprechen.
Hauptsachgebiete: 1 = Politik, 2 = Wirtschaft, 3 = Soziales, 4 = Sport, 5 = Kulturelles

Gliederung der Sachgebiete

I	Politik			
IA	Allgemeine Lage			
IB	Persönlichkeiten			
IC	Übernationale Zusammenschlüsse und Vereinbarungen			
ID	West-Ost-Politik (bestimmter Staaten)			
IE	Westliche Außenpolitik			
IF	Östliche Außenpolitik			
IG	Politik der Blockfreien und Neutralen (als Gruppen)			
IH	Politik einzelner blockfreier und neutraler Staaten			
II	Militär- und Verteidigungspolitik (internat. u. national)			
IJ	Europäische Politik	(auch einzelner Staaten)		
IK	Afrikanische Politik	"	"	"
IL	Asiatische Politik	"	"	"
IM	Arabisch-nahöstl. Politik	"	"	"
IN	Deutsche Außenpolitik			
IO	Deutsche West-Ost-Politik (gesamtdeutsch)			
IP	Innenpolitik auf Bundesebene bzw. der DDR			
IQ	Innenpolitik auf Länderebene			
IR	Innenpolitik auf Partei(en)ebene			
IS	Deutsche Wirtschaftspolitik			
IT	Innenpolitik best. Staaten			
IU	Deutsche Kulturpolitik			
IV	Deutsche Sozialpolitik			
IW	Verbände			
IX	Kommunalpolitik			
IY				
IZ	Sonstiges			

- 2 Wirtschaft
- 2A Allgemeine Lage
- 2B Persönlichkeiten
- 2C Technik
- 2D Übernationale Zusammenschlüsse (außer europäischen)
- 2E Europäische Zusammenschlüsse
- 2F Organisationen, Verbände
- 2G Planung
- 2H Arbeitskräfte, Personalstruktur
- 2I Industrie, Grundstoffindustrie
- 2J Industrie, Investitionsgüter -
- 2K Industrie, Konsumgüter-
- 2L Landwirtschaft, Forsten
- 2M Seefahrt, Fischerei
- 2N Handwerk
- 2O Handel, Außen-
- 2P Handel, Innen- (auch gesamtdeutscher)
- 2Q Betriebsorganisation, Rationalisierung
- 2R Finanzen, Börse, Versicherungen
- 2S Verkehr, Verkehrswege
- 2T Verkehrsmittel (Luft-, Land-, Seefahrzeuge) und - technik
- 2U Stadtplanung, Raumordnung
- 2V Wohnungsbau, Bauwesen
- 2W Dienstleistungsindustrie u. -gewerbe
 (Fremdenverkehr, Hotel-, Gaststätten u. dergl.)
- 2X Werbung
- 2Y Vergnügung, Unterhaltung, Massenmedien (nur wirtschaftlich)
- 2Z Sonstiges
- 2(Energiewirtschaft
- 2) Handel-Versorgung

- 3 Soziales
- 3A Allgemeines, Statistik
- 3B Persönlichkeiten
- 3C Bevölkerung
- 3D Verbrechen, Vergehen (auch Beschuldigungen)
- 3E Rechtswesen
- 3F Gesundheitswesen, Medizin
- 3G Verbände, Organisationen
- 3H Körperpflege
- 3I Versorgung (Renten usw.)
- 3J Rassenprobleme
- 3K Kommunikationsmittel (Presse, Rundfunk, Fernsehen, Film;
 aber nicht Artikel zum künstlerischen Rang!)
- 3L Wissenschaft und Forschung
- 3M Schul-, Erziehungs-, Bildungswesen
- 3N Beruf und Arbeitswelt
- 3O Haus und Garten
- 3P Familie (Ehe)
- 3Q Liebe (Ehe), Erotik (sog. Privat-oder Intimsphäre)
- 3R Hobbies, Freizeit
- 3S Geographisches
- 3T Mode
- 3U
- 3V Veranstaltungen
- 3W Naturereignisse, Wetter
- 3X Unglücksfälle, Todesfälle
- 3Y Streiks, Unruhen, Aufstände
- 3Z Sonstiges

4	Sport
4A	Allgemeines
4B	Persönlichkeiten, Verbände
4C	
4D	Großveranstaltungen mehrerer Sparten (Olympiade, allgemein)
4E	Leichtathletik
4F	Schwerathletik, Boxen, Ringen
4G	
4H	Turnen, Gymnastik, Tanz
4I	
4J	Ballspiele, Mannschafts- (Fußball, Handball, Hockey usw.)
4K	
4L	Ballspiele, Einzel- (Tennis, Tischtennis usw.); Golf, Cricket
4M	Fechten
4N	Pferdesport
4O	Wassersport
4P	Wintersport mit Eislauf
4Q	Schießsport
4R	Radsport
4S	Motor- und Flugsport
4T	
4U	Bergsteigen, Wandern
4V	
4W	Spiele und Rätsel
4X	
4Y	
4Z	Sonstiges

- 5 Kulturelles
- 5A Allgemeines
- 5B Persönlichkeiten
- 5C
- 5D Organisationen, Gruppen
- 5E kulturelle Entwicklung in einzelnen Ländern
- 5F Musik (Konzert)
- 5G Oper, Operette, Musical
- 5H Theater, Ballett
- 5I Malerei, Graphik
- 5J Bildhauerei, Plastik
- 5K Architektur
- 5L Literatur
- 5M Sprache
- 5N Wissenschaft, Forschung, Universitäten (nur zu deren kulturellen Aufgaben und Leistungen, nicht zur sozialen Funktion)
- 5O Massenmedien (Film, Rundfunk, Fernsehen)
- 5P Philosophie, Weltanschauung, Ideologie
- 5Q
- 5R Religion und Kirche
- 5S
- 5T Geschichte
- 5U
- 5V Volks- und Brauchtum
- 5W
- 5X
- 5Y
- 5Z Sonstiges

11. Spaltenfeld (Sp. 64-70)

Innere und äußere Form

a) Innere Form

1) Publizistische Ziele (Sp. 64-65)

Es können 2 Chiffren gleichzeitig angegeben werden.

1 = Unterrichtung

2 = Belehrung

3 = Beeinflussung

4 = Unterhaltung

5 = Wirtschaftswerbung

2) Mitteilungsform (Sp. 66)

A = Nachricht (Unterrichtung, Wirtschaftswerbung)

B = Bericht (Unterrichtung, Belehrung, Beeinflussung,
Wirtschaftswerbung)

C = Background-Bericht (Unterrichtung, Belehrung,
Beeinflussung)

D = Abhandlung (Belehrung, Unterrichtung)

E = Beitrag (Belehrung, Unterrichtung, Unterhaltung,
Wirtschaftswerbung)

F = Tips, Vorschlag, Anweisung (Belehrung, Unterrichtung
Unterhaltung, Wirtschaftswerbung)

G = Hauptkommentar (Beeinflussung, Belehrung, Unterrichtung)

H = Glosse (Beeinflussung, Unterhaltung)

I = "Spitzmarke" (Beeinflussung, Unterhaltung)

J = Geschichte (Unterhaltung, Belehrung)
(Kurzgeschichte, Erzählung)

- K = Plauderei (Unterhaltung, Beeinflussung, Belehrung)
L = "kleines Feuilleton" (Unterhaltung, Beeinflussung,
Belehrung)
M = Großanzeige (Wirtschaftswerbung, Beeinflussung,
Unterrichtung, Belehrung)
N = Anzeige (Wirtschaftswerbung, Beeinflussung,
Unterrichtung, Belehrung)

Sonderformen

- O = Text zu einem Bild, Schaubild, graph. Darstellung usw.
P = Leserbrief
Q = Interview
R = Abdruck eines literarischen Dokuments (auch Gedichtes)
S = Abdruck einer amtlichen Verlautbarung (Gesetz, Vertrag,
Bekanntmachung, Aufruf), Kommuniké
T = Abdruck einer Rede
U = Abdruck eines fremden Artikels
V = Klein- und Privatannonce
W = Tabelle, Wetterbericht
X = Fortsetzungsroman
Z = Sonstiges

Für diese Sonderformen braucht kein publizistisches Ziel angegeben zu werden; die dafür vorgesehenen Spalten (64-65) werden dann mit 00 ausgefüllt.

3) Angabe, ob es sich um eine Übersetzung handelt (Sp. 67)

4) Angabe der Sparte (Sp. 68)

- | | |
|---------------|-----|
| Politik | = 1 |
| Wirtschaft | = 2 |
| Sport | = 3 |
| Technik-Motor | = 4 |

Feuilleton	=	5
Lokales	=	6
Anzeigen	=	7
Vermischtes	=	8
Sonstiges	=	0

In Zeitungen, die selbst keine Sparten angeben (meist am oberen Rand der Seite), kann diese Spalte zunächst frei bleiben.

- b) Äußere Form (Sp. 69-70)
- 1) Länge des Artikels in Zeilen (nach Klassen gruppiert),
 - 2) Breite des Artikels in Spalten.

12. Spaltenfeld (Sp. 71-77)

Angaben über stilistische Kriterien.
(noch nicht entwickelt!)

13. Spaltenfeld (Sp. 78-80)

Länge des Artikels in Karten.

Erläuterungen zur Beikarte (Themakarte)

1. Spaltenfeld (Sp. 1-4)

Kartenkennung

2. Spaltenfeld (Sp. 5-9)

Jahrgangschiffre und Artikelnummer

3. Spaltenfeld (Sp. 10)

bleibt leer

4. Spaltenfeld (Sp. 11-80) Themaangabe

a) Kopfteil:

enthält Angabe des oder der im Artikel behandelten Landes (Länder).

Bundesrepublik = BRD, DDR = DDR, für alle anderen Staaten die Chiffren der amtlichen Autokennzeichen.

Es können bis zu vier Staaten im Kopfteil genannt werden.

Danach ist zweimal blank zu setzen. Bezieht sich der Artikel auf kein bestimmtes Land, ist nach Spalte 10 zweimal blank zu setzen.

b) Textteil: Stichwort-Angabe des behandelten Themas.

ANHANG IV

Übersicht über die in der Außenstelle vorhandenen
wichtigeren Programme

Nr.	Name Gebiet	Leistung
1	<u>PRUEF</u> Texterstellung	("Textprüfung") Übernimmt Daten von Karten oder Band auf Band, prüft Textfiles auf numerische Richtigkeit nach folgenden Kriterien: 1. Lückenlose, wiederholungslose, aufsteigende Folge von a) Artikeln, b) Karten (vergleicht beide Numerierungen), 2. Vorhandensein, Lage und Zugehörigkeit von Informationskarten, 3. Beginn und Ende der Artikel nach besonderen Bedingungen. Gibt 20 verschiedene Kommentare als Prüfprotokoll aus. Zusätzlich: Erstellt Textausdrucke in verschiedener Form mit Kommentierung.
	Unterprogr. Autor	EINAUS (IPK), 4 weitere für Ausdruck, 5 für Prüfung. W. Klutentreter, IDS
2	<u>VORKOR</u> Texterstellung	("Vorkorrektur") Übernimmt zur Bandkorrektur bestimmte Korrekturkarten auf Band, prüft auf Lage, Format und Erfüllung der von KORR (Nr.3) verlangten Eigenschaften. Gibt 13 Kommentare als Prüfprotokoll aus.
	Unterprogr. Autor	3 zum Prüfen W. Klutentreter, IDS
3	<u>KORR</u> Korrektur beliebiger files hier: Texterstellung	("Korrektur") korrigiert files (hier: Textfiles) anhand von Korrekturkarten durchErsatz, Löschen oder Neueinfügen einzelner Records (bei Löschen u. Neueinfügen s.Nr.4)
	Unterprogr. Autor	EINAUS, CHAR, BIST, ZELE T. Krumnack, IPK

Nr.	Name Gebiet	Leistung
4	<u>NDNUM</u> Texterstellung Unterprogr. Autor	("ND-Numerierung") Numeriert lückenhafte oder überbelegte files (oder Teile von files) neu durch - hier: von KORR korrigierte Textfiles. - H. E. Neuhaus, IPK
5	<u>INKASU</u> Textbereit- stellung Unterprogr. Autor	("Informationskarten-Suche") Sucht nach vorgegebenen Kriterien (auch Kom- binationen mehrerer Kriterien) Informations- und Beikarten ab, schreibt Nummern der betr. Artikel aus und/ oder stellt die betr. Artikeltexte auf Band zur Verfügung oder druckt aus. G. Billmeier, IDS
6	<u>MOSU 1</u> Wort-Suche Unterprogr. Autor	("Morphem-Suche" 1. Teil) Sucht auf Textbändern aufgrund vorgegebener Buchstabenkombinationen Belegwörter und übergibt sie mit Kontext und Stellenangaben sowie Informations- karten an Folgeprogramm. (Kann auch als Unterprogramm verwendet werden) Spezifikationen: a) Suche ist beschränkbar auf "Endungen", b) jede Stelle der Such-Kombination kann durch mehrere Zeichen belegt werden, c) Substantive oder Nicht-Substantive können ausgefiltert werden, d) reiner Zähllauf (ohne Belegausgabe) ist möglich, e) Kontext kann von 1-11 Karten variiert werden. EINAUS, SIEB, ZELE, ZAHL T. Krumnack, IPK
7	<u>MOSU 2</u> Wort-Suche	("Morphem-Suche" 2. Teil) Übernimmt von MOSU 1 gefundene Belegwörter mit Kontext, sortiert sie nach Such-Kombinationen (immer) und - je nach Spezifikation - innerhalb dieser alphabetisch, reinigt Kontext von nicht ge- wünschten Zeichen, gibt mit Hilfe von INFDEC (Nr.8) Informationen

Nr.	Name Gebiet	Leistung
		aus Informationskarte bei, beschneidet Kontext auf gewünschte Länge (Zahl von Karten, Zahl von Wörtern oder satzweise) und druckt aus. Liegt in 3 Standardvarianten, 2 nicht standardisierten Varianten und 1 erweiterter Variante vor:
a)	<u>SAMOS</u>	("Standard-Ausdruck für MOSU") Liefert: gefundenes Wort, Fundstellenangabe von Inform.-karte (falls gewünscht) und Kontext (mit Nummer) in folgenden drei Stufen: SAMOS 1: 3 Zeilen Kontext (= Belegzeile, Vorzeile, Folgezeile), sortiert alphabetisch bis 400 Belege je Suchkombination; SAMOS 2: 5 Zeilen Kontext; sortiert bis 250 Belege je Suchkombination; SAMOS 3: 7 Zeilen Kontext; (sortiert Belege nicht innerhalb der Suchkombination).
b)	<u>SAMOS W</u>	Wie SAMOS, aber Kontextbeschränkung nach angebbarer Zahl von Wörtern; Belege unbeschränkt sortierbar;
c)	<u>SAMOS S</u>	Wie oben, aber Kontextbeschränkung auf 1 Satz (von Satzschlußzeichen zu Satzschlußzeichen) (Belege nicht sortiert; begrenzte Kapazität);
d)	<u>SAMOS E</u>	Kombination von SAMOS W und INFDEC. Wie SAMOS W, decodiert zusätzlich bestimmte gewünschte Angaben aus Informationskarten und fügt sie den Belegen bei; erstellt nicht automatisch Druckspiegel.
	Unterprogr.	
	Autor	G. Billmeier, IDS
8	<u>INFDEC</u>	("Informationskarten-Dechiffrierung")
	Spezial- Unterprogramm	Dechiffriert angegebene Bereiche der Informations- karten in Klartext und übergibt an jeweiliges Hauptprogramm, vor allem: an PRUEF (Klartext der Inform.-karten am Beginn jede Artikels in Klartextausdrucken), an MOSU (SAMOS E) zur näheren Kennzeichnung der Belege. Mit eigenem Hauptprogramm: Erstellt dechiffrierte Wiedergabe der Inform.-karten in Karteikartenformat zur Kontrolle der Inform.-karten.

Nr.	Name Gebiet	Leistung
	Unterprogr.	-
	Autor	G. Billmeier, IDS
9	<u>SUSA</u>	("Suche von Stern-Abkürzungen")
	Wort-Suche	Sucht auf Textband Initial-Abkürzungen (= Wörter mit mehr als 1 "Stern" (Großschreibungszeichen)) mit Belegstellen. Auch als Unterprogr. zu SUBEL verwendbar.
	Unterprogr.	EINAUS, PACK
	Autor	H. D. Lutz, IDS
10	<u>SUBEL</u>	("Suche von Belegstellen")
	Wort-Suche	Sucht anhand vorgegebener (auch von SUSA übernommener) Belegstellenangaben auf Textband Kontext variablen Umfangs und druckt aus.
	Unterprogr.	EINAUS, PACK, (SUSA)
	Autor	H. D. Lutz, IDS
11	<u>SUBIN</u>	("Suche von Bindestrich-Komposita")
	Wort-Suche (Spezial-Progr.)	Sucht auf Textband Bindestrich-Komposita, klassifiziert sie nach 6 Gesichtspunkten, druckt Belege mit variablem Kontext aus.
	Unterprogr.	EINAUS, 5 interne
	Autor	G. Billmeier, IDS
12	<u>COSYMA</u>	("Convertierungs-System f. Mannheimer Bänder")
	Textbereit- stellung	Wandelt Band-Texte nach Mannheimer Schreibkonventionen in Band-Texte nach Bonner Konventionen um.
	Unterprogr.	EINAUS
	Autor	G. Billmeier, IDS

Nr.	Name Gebiet	Leistung
13	<u>COSYBO</u> Textbereit- stellung Unterprogr. Autor	("Convertierungs-System f. Bonner Bänder") Wandelt Band-Texte nach Bonner Schreibkonventionen in Band-Texte nach Mannheimer Konventionen um. EINAUS G. Billmeier, IDS
14	<u>DECOS</u> Textbereit- stellung (Lochstreifen) Unterprogr. Autor	("Decodierung von Streifentexten") Übernimmt Text von Lochstreifen (hier: TTS-Zeitungs- setzstreifen), codiert Zeichen um, stübert von unerwünschten Zeichen, gleicht Text z.T. an Bonner Schreibkonventionen an, bringt Text in Records bestimmter Länge, definiert und numeriert Artikel und Zeilen, erkennt Überschriften, druckt aus oder über- gibt an ESLOT. W. Klutentreter, IDS
15	<u>ESLOT</u> Textbereit- stellung (Lochstreifen) Unterprogr. Autor	("Erkennung v. Satzschlüssen in Lochstreifen-Texten") Übernimmt durch DECOS umcodierte und vorbereitete Streifentexte, unterscheidet Satzschlußpunkte von anderen Punkten, definiert Satzschlüsse, schreibt Satzanfänge klein (sofern nicht Substantiv am Satzanfang), kommentiert Entscheidungen und druckt aus. EINAUS, 6 - 7 interne G. Billmeier, IDS
16	<u>IDVS</u> Index- Erstellung Unterprogr. Autor	("IDS-Vorsortier-Programm") Normiert Texte im IDS-Code (Bonn) zu sortierbaren Einheiten, getrennt nach Einzelwörtern und Wort- gruppen (Transkriptionsgruppen und Bindestrich-Gruppen). EINAUS, IDWO, IDTA u. a. T. Krumnack, IPK
17	<u>IBSORT</u> Index- Erstellung	("IBSYS-Sortier-Programm") Sortiert nach IDVS genormte Band-files alphabetisch, getrennt nach Einzelwörtern und Wortgruppen. Oder: Sortiert von RELI (Nr. 19) erstellte files nach relativen Häufigkeiten.

Nr.	Name Gebiet	Leistung
	Unterprogr. Autor	s. IBSYS-Systemband T. Krumnack, IPK
18	<u>IDMI</u> Index- Erstellung Unterprogr. Autor	("IDS-Misch-Programm") Mischt von IBSORT getrennt sortierte Einzelwörter und Gruppen in alphabetischer Reihenfolge. EINAUS, BIST, ZELE, ZAHL T. Krumnack, IPK
19	<u>RELI</u> Index- Erstellung Unterprogr. Autor	("Listen mit relativen Häufigkeiten") Erzeugt aus von IDMI oder IBSORT sortierten files alphabetische Indices mit Angaben der Fund- stellen, der absoluten und relativen Häufigkeiten sowie der Zugehörigkeit zu einer Wortgruppe. Druckt aus oder übergibt an IBSORT. EINAUS, SIEB, BIST, ZELE, ZAHL T. Krumnack, IPK
20	<u>KRLI</u> Index- Erstellung Unterprogr. Autor	("Listen nach Klassen relativer Häufigkeit") Übernimmt files, die nach Häufigkeit sortiert sind (und innerhalb gleicher Häufigkeit alphabetisch), a) listet aus: Häufigkeitsklasse, Zähler pro Klasse, relative Häufigkeit, absolute Häufigkeit, Wort; b) gibt a) auf Band, zusätzlich mit Kennung, die den file bezeichnet, zwecks Weiterverarbeitung durch VERSYS (Nr. 21). T. Krumnack, IPK
21	<u>VERSYS</u> Vergleichs- register	("Vergleichssystem für Häufigkeitsfiles") a) Vergleicht zwei alphabetisch geordnete files (2 Bänder), erstellt daraus allgemeines Vergleichs- band (mit Angabe zu jeder Wortform, ob vorhanden und wie häufig (nach Klassen)). b) Erzeugt aus Vergleichsband spezielle Vergleichs- Listen (intern vorgesehen: 20 mögliche Listen). c) Vergleicht zwei alphabetisch geordnete files auf <u>einem</u> Band, erstellt daraus Listen nach b).

Nr.	Name Gebiete	Leistung
		<p>Es können Vergleichs-Register erzeugt werden nach:</p> <ul style="list-style-type: none">a) "linker" Differenzb) "rechter" Differenzc) "Durchschnitt" (Gemeinsames)d) Vereinigung (aber nach links und rechts getrennt).
		<p>Vergleich kann erfolgen nach</p> <ul style="list-style-type: none">1. vorhanden - nicht vorhanden2. gleich - ungleich3. häufig - selten (nach wählbarer Definition bzw. Gewichtung).
	Unterprogr.	
	Autor	H. E. Neuhaus, IPK

A n m e r k u n g e n

zum Forschungsbericht Hellmann (Außenstelle Bonn)

- 1) In begrenztem Umfang sind Untersuchungen zur gesprochenen Sprache möglich. In der Außenstelle wurde einige Monate lang von der damaligen wiss. Mitarbeiterin Frau Dr. Inge Kraft eine Reihe von Tonbandaufnahmen aus Landfunksendungen des Senders "Radio DDR", und zwar Interviews und Gespräche, untersucht und mit westlichen Landfunkinterviews sowie mit östlichen und westlichen Zeitungsartikeln über landwirtschaftliche Themen verglichen. Die in den Interviews von Radio DDR gesprochene Sprache zeigte, von einigen charakteristischen Fehlleistungen abgesehen, eine so weitgehende Übereinstimmung mit der Terminologie gedruckter Texte (Zeitungen), d.h. eine so starke Anpassung des Interviewten an die Redeweise des jeweiligen Interviewers, daß die Untersuchungen wieder eingestellt wurden. Frau Dr. Kraft hat während der Herbsttagung 1965 des Instituts in Mannheim über ihre Ergebnisse berichtet; ein Aufsatz wird in Kürze in "Deutsche Studien" Heft 23, Lüneburg 1968, erscheinen.
- 2) Zum Problem der Vergleichbarkeit unseres westlichen und östlichen Materials vgl. unten Punkt 7.3.
- 3) Dazu vgl. Anhang III, Abs. 11a (4).
- 4) Zur Berücksichtigung der in unserem Material vorgefundenen Sachgebiete, s. Anhang III, Abs. 10.
- 5) Der Begriff "Thema" ist nicht mit "Sachgebiet" zu verwechseln. Zur Kennzeichnung vgl. unten Punkt 13.2.3.
- 6) Gewisse Themen kehren allerdings in bestimmten Abständen wieder; dazu unten 4.5.2.1.b).
- 7) Zur Definition des Begriffs s.4.6.1. und 13.1.
- 8) E. Dovifat, Zeitungslehre Bd. 1, Berlin 1962 (Götschen H. 1039) unterscheidet nur "Nachrichtenstilform, Meinungsstilform, Unterhaltungstilform" (S. 120, Näheres S. 120 - 133).
W. Hagemann, Die Zeitung als Organismus, Heidelberg 1950, unterscheidet "Unterrichtung, Beeinflussung, Wirtschaftswerbung, Unterhaltung, Belehrung" (vgl. a.a.O. Kap. III, S.35 ff). Beide ordnen den publizistischen Zielen bestimmte Schreib- oder Stilformen zu. - Zur Anwendung dieser Kriterien auf unser Material vgl. Anhang III, Abs. 11a (1 und 2).

- 9) Vgl. dazu Anhang III, Abs. 11a (Innere Form). Unsere Einteilung lehnt sich an die Hagemanns an (a.a.O. S. 49 ff.).
- 10) Zwingend ist dieser Schluß freilich nicht; theoretisch ist es möglich, auch eine Vielzahl von Sachgebieten, Themen usw. mit relativ sparsamen sprachlichen Mitteln zu bewältigen.
- 11) Daraus den Schluß zu ziehen, man habe es bei Tageszeitungen sicher mit "Normal-" oder "Durchschnittssprache" zu tun, ist allerdings gefährlich; hier können zwei Fehlerquellen liegen:
 - a) Die Redaktion selbst kann sich in ihrer Leserschaft irren, sie kann sie über- und unterschätzen, wenn sie sich z.B. einer Redeweise bedient, die bei der Leserschaft oder einem Teil davon "nicht ankommt".
 - b) Der (wiss.) Betrachter kann sich irren, nämlich wenn eine Redaktion absichtlich, z.B. um eine propagandistische Wirkung nach außen zu erzielen, sich einer Redeweise bedient, die den Eindruck einer besonders fortschrittlichen (oder kämpferischen, sozialistischen, national gesinnten, christlichen, gebildeten, aufgeschlossenen usw.) Leserschaft erwecken soll. Dies dürfte aber wohl nur in einer Gesellschaft möglich sein, in der ein Meinungsmonopol herrscht. Im übrigen bleibt auch dann die Frage offen, inwiefern nicht auch eine solche in bestimmter Richtung forcierte Redeweise schließlich Wirkungen auf die Redeweise der Leser hat, gerade unter den Bedingungen eines Meinungsmonopols.
- 12) Unter "Dokumentation" wird hier nicht erst die (maschinelle) Aufbereitung von Texten, ihre Ausstattung mit ständig zu erweiternden Informationen zwecks (maschineller) Reproduktion sprachlicher Daten unter verschiedensten Fragestellungen verstanden, sondern auch schon die diesem Ziel dienende adäquate Ermittlung der entsprechenden Texte.
- 13) Ohne Zweifel hat diese und jede andere formale Einteilung keine unmittelbare qualitative Bedeutung. Eine Ausgabe vom 31. Dezember ist von der des 1. oder 2. Januar nicht grundsätzlich stärker unterschieden als von der des 30. Dezember.
- 14) Sie ist insofern auch berechtigt, als sich in jedem Jahrgang bestimmte Themen turnusmäßig, nämlich jahreszeitlich bedingt, wiederholen. Dazu vgl. unten 4.5.2.1.b).
- 15) Zum Vergleich die Zahlen für das "Neue Deutschland" des Jahrgangs 1964: Gesamtmenge ohne Sonntagsausgaben und Beilagen: ca. 2200 Seiten entspr. 7 Mill. lfd. Wörter (dabei ist berücksichtigt, daß 1 Seite ND durchschnittlich mehr Wörter enthält als 1 Seite WELT, da dort großräumige Anzeigen sehr selten sind). Diese Menge entspricht rund 14.000 Buchseiten.

- 16) ND Februar 1964: 184 Seiten ohne Beilagen und ohne Sonntagsausgaben; eine Zeitungsseite ND entspricht etwa 6-7 Buchseiten.
- 17) Die erste und trivialste dieser Schwierigkeiten besteht darin, daß die fast ausschließlich benutzten TTS-Streifen mit den in den Rechenzentren und wiss. Instituten vorhandenen Streifenlesegeräten nicht zu verarbeiten sind. Lesegeräte, die sowohl Standard- als auch TTS-Streifen lesen können, gibt es zwar, sie werden aber an den für uns erreichbaren Stellen nicht eingesetzt. Am Germanistischen Seminar der Universität Saarbrücken besteht zwar die Möglichkeit, TTS-Streifen in Standardstreifen umzuformen, jedoch ist dies ein umständlicher Weg, zumal bei regelmäßiger Umformung von Streifen in großen Mengen.
Eine weitere Schwierigkeit liegt darin, 1. daß die TTS-Streifen durchweg nicht korrigiert sind, 2. daß Überschriften, Anzeigen sowie andere Texte mit drucktechnischen Besonderheiten nicht über Streifen gesetzt werden und umgekehrt nicht alles, was auf den Streifen steht, auch in der Zeitung erscheint (größere Änderungen noch kurz vor dem Druck).
Ferner ist die Mehrdeutigkeit einiger Zeichen meist unerwünscht, wie z.B. Nichtunterscheidung von Abkürzungspunkt und Satzschlußpunkt, Großschreibung am Satzanfang und damit Nichtunterscheidung von der sonst willkommenen Großschreibung der Substantive usw. In der Außenstelle ist inzwischen ein Programm erarbeitet worden, das diese Schwierigkeiten weitgehend automatisch beseitigt. Es arbeitet z. Zt. mit einer Fehlerquote von knapp 5 Prozent. (Weiteres vgl. Anhang IV "Programmübersicht", Nr. 14 u. 15). Schließlich werden von elektronisch gespeichertem Material durchweg zusätzliche Informationen sprachlicher Art erwartet, die ein über Streifen eingeleiteter Text ohne Eingriffe von Hand natürlich nicht enthält. Dennoch erscheint es nach den bisher erreichten Ergebnissen als möglich, durch geeignete Aufbereitungsprogramme auch fremde Streifentexte so zu verbessern, daß sie sich als Textgrundlage für unsere Zwecke eignen, sofern man die Forderung nach Eindeutigkeit und Zusatzinformationen etwas reduziert.
- 18) Vgl. G. Herdan, The Advanced Theory of Language as Choice an Chance, Berlin-Heidelberg-New York 1966, S. 96 f. und 169 ff. .
- 19) Die Begriffe "repräsentative Auswahl bzw. Menge" und "Repräsentanz" bzw. "Repräsentativität" werden bewußt vermieden, da sie, wenigstens im allgemeinen Sprachgebrauch, die Vorstellung des Absoluten, fest Definierbaren auslösen können. Es gibt jedoch in diesem Sinne keine absolut repräsentative Auswahl.
- 20) Der Begriff "Modell" gibt das Gemeinte insofern adäquat wieder, als er die Vorstellung intendiert, daß ein Modell maßstabgetreu und funktionsgerecht zu sein hat, dies jedoch in Einzelheiten nur dann sein kann, wenn ein bestimmter Maßstab, der von den jeweils an das Modell zu stellenden Anforderungen abhängt, nicht unterschritten wird. Dabei ist klar, daß auch das beste Modell nur einen Teil der Eigenschaften der Bezugsgröße genau, viele andere nur annäherungsweise und einige gar nicht widerspiegelt.

- 21) Zur Unterscheidung von Gesamtmenge und Zielmenge vgl. auch 4.5.3.1.a).
- 22) Vgl. 4.5.2.1. und 4.5.3.1.
- 23) E. Mittelberg, Wortschatz und Syntax der Bildzeitung, Marburg 1967, hat für seinen Untersuchungsgegenstand gezeigt, daß dort die herkömmliche Sparteneinteilung und Aufeinanderfolge der Sparten nicht gilt. Für die Organe bestimmter Institutionen oder Gruppen (Verbände usw.) ist ähnliches anzunehmen.
- 24) Die Dokumentation von Zeitungstexten steht hier vor ähnlichen Problemen wie die "repräsentativen Bevölkerungsumfragen": auch dort können alle nach fein gegliederten Kriterien wie Geschlecht, Alter, Familienstand, Schulbildung, Konfession, Beruf, Wohnort usw. vorgenommenen Befragungen zu falschen Ergebnissen führen, wenn tatsächlich nicht eines der berücksichtigten Kriterien, sondern ein nicht berücksichtigtes wie etwa die Programmwahl im Fernsehen den Ausschlag für das erwartete Verhalten der Bevölkerung gibt.
- 25) So kann etwa die scheinbar mögliche Unterscheidung von namentlich gekennzeichneten und anonymen Artikeln kein echtes Kriterium einer Auswahl sein, weil die Zeitungen hier sehr willkürlich verfahren.
- 26) Die Verwendung der Einheit "Seite" als Maß- oder Zählereinheit ist nicht zu verwechseln mit ihrer möglichen Verwendung als Aufnahmeeinheit. Dazu s. unten 4.6.1.
- 27) Sie gilt für Druckerzeugnisse jeglicher Art, also außer für Bücher und Periodika auch für Sonderdrucke, Flugschriften, Manuskripte, Plakate usw.
- 28) Beispiele für die Vielfalt der "Beilagen" allein in der WELT: Die Rubriken "Betrieb und Beruf" sowie "Schule und Hochschule" erscheinen nicht täglich, sind zwar im allgemeinen laufend durchnummeriert, nehmen aber doch nach Umfang und Aufmachung den Charakter von Beilagen an. Regelmäßig erscheinende Beilagen sind etwa die Stellenanzeigen in den Samstagausgaben (obwohl auch im übrigen durchnummerierten Teil der Zeitung noch Stellenanzeigen stehen können); ferner "Die geistige Welt" (ebenfalls samstags) und die "Reise-Welt" (freitags); diese Beilagen sind durchweg, jedoch nicht immer, gesondert (römisch) numeriert. Das schließt aber nicht aus, daß nicht auch in anderen Ausgaben mehrere Seiten für "Reise und Erholung" reserviert werden, ohne direkt Beilagen zu sein. - Die "Welt der Literatur" nimmt eine Sonderstellung schon im Format ein (halbes WELT-Format); sie erscheint 14-tägig donnerstags und ist gesondert abonnierbar, stellt also fast eine Zeitung in der Zeitung dar. Anlässlich der Olympischen Spiele haben mehrere Zeitungen Sonderbeilagen herausgebracht.

- 29) Nach dieser Definition kann die "Ausgabe" also auch nur etwa die Beilagen einer Zeitung enthalten. Die Summe aller Erscheinungsmengen ist also die Gesamtmenge, die Summe aller Ausgaben ist die Zielmenge.
- 30) Falls also eine Zeitung im Laufe eines Jahrgangs von wöchentlichem zu täglichem Erscheinen übergeht (oder umgekehrt), können die folgenden Überlegungen nicht angewandt werden.
- 31) Für ein 1-bändiges Buch ergibt sich der Wert 365, da n (Zahl der Ausgaben im Jahr) gleich 1 ist.
- 32) Im Jahrgang 1966 des "Neuen Deutschland", das täglich erschien, sind 359 Nummern herausgekommen, also 6 Nummern ausgefallen; die WELT hat 1966 305 Nummern herausgebracht, d.h. nach Abzug der 52 Sonntage sind 8 Nummern ausgefallen. Die ausgefallenen Nummern betreffen sämtlich die gesetzlichen Feiertage.
- 33) Erfahrungsgemäß reicht es aus, etwa jede vierte Woche oder jeden vierten Monat eines Jahrgangs, also ca. $1/4$ der Jahrgangsmenge, genau auszu-zählen. Größere Genauigkeit ist nicht erforderlich, da sie in der Auswahl ohnehin nicht berücksichtigt werden könnte.
- 34) Bei der Ermittlung unserer Auswahl wurden jeweils 4 bzw. 5 Monate genau ausgezählt (vgl. unten Punkt 10.1., Anm. 50 und 51, und Anhang Ia).
- 35) Wieviel Aufnahmeeinheiten (Seiten, Artikel) eine Stichprobeneinheit enthält, hängt von der Menge ab, die der Bearbeiter aufzunehmen für notwendig hält, d.h. von der gewünschten oder geforderten Aufnahmequote bzw. dem Wahrscheinlichkeitsquotienten.
- 36) Die untere Grenze für das Aufnahmeintervall ist selbstverständlich das Erscheinungsintervall.
- 37) Konkret bedeutet das: In einer auf einen ganzen Jahrgang bezogenen Modellmenge sind 10 ganze Ausgaben à 10 Seiten selbstverständlich besser als 10 einzelne Seiten, aber 100 gut verteilte einzelne Seiten sind wesentlich besser als 10 ganze Ausgaben.
- 38) Näheres s. Teil III, S 140 ff.
- 39) Läßt man also eine Auswahl am 10. März beginnen und führt sie mit einem Aufnahmeintervall von 6 Tagen bis zum 10. Juni fort, so kann diese Auswahl selbst bei genauester Beachtung aller Bedingungen einer zureichenden Auswahl genau für den Zeitraum vom 7. März bis 13. Juni Modell sein, jedoch niemals für den ganzen Jahrgang oder für einen anderen Zeitraum.

Damit ist nicht gesagt, daß nicht auch eine zeitlich sehr eng begrenzte Auswahl sinnvoll sein kann. Für eine Untersuchung etwa, für die ein diachronisch extrem tiefgestaffeltes Material benötigt wird (etwa bei eng begrenzten Wortschatzuntersuchungen über 10 bis 20 Jahrgänge hinweg) kann eine Auswahl aus nur einer Woche (allerdings immer der gleichen Woche) bei genügender Dichte eine sehr ergiebige Modellmenge darstellen; - freilich ist sie Modell für die Verhältnisse in dieser einen bestimmten Woche in diesen 10 bis 20 Jahrgängen. Bedingung ist dabei, daß man sich der zeitraumbedingten Besonderheiten gerade dieser kurzen Auswahlspanne genau bewußt ist. (Als Beispiel für sinnvolle Verwendung einer solchen zeitlich begrenzten Auswahl vgl. H. Bartholmes, Das Wort "Volk" im Sprachgebrauch der SED, Düsseldorf 1964). - Es sind also auch diachronisch gegliederte Modellmengen möglich, für die aber die aufgezeigten Grundsätze für die Ermittlung einer zureichenden Auswahl - entsprechend modifiziert angewandt - ebenfalls gelten.

- 40) Die noch verbleibende Korrekturarbeit - zu Beginn der Textaufnahme der Überwiegende Teil - wurde von einem wissenschaftlichen Mitarbeiter übernommen.
- 41) Zu den Schreibkonventionen vgl. unten Punkt 13.
- 42) Diese Kapazität ist jedoch nur mit eingearbeiteten Kräften bei fertig entwickeltem System zu erreichen. Die tatsächlich erreichte Leistung lag im ersten Jahr bei etwa einem Viertel, im zweiten Jahr bei etwa der Hälfte der angegebenen Kapazität. Im Jahre 1967 ist die Kapazität nahezu erreicht worden.
- 43) Es wurde die Berliner Ausgabe (B-Ausgabe) statt der Republik-Ausgabe (A) gewählt, einmal um auch Texte mit (mehr oder weniger) regionalem Kolorit zu erhalten, vor allem aber weil auch die WELT in einer Berliner Regionalausgabe erscheint; die beiden Zeitungen wurden dadurch besser vergleichbar.
- 44) Das ND 54 hatte noch keine Trennung zwischen Berliner und Republik-Ausgabe.
- 45) So ist z.B. im ND ein Feuilleton im Sinne von "Unterhaltung" so gut wie unbekannt; ein kulturell-literarischer Teil erscheint sporadisch und mit sehr wechselndem Umfang, sofern er nicht als Beilage erscheint (dazu siehe unten 8.1.). Einen Roman "unter dem Strich" enthielt der Jahrgang 1964 nicht, dagegen ein Teil des Jahrgangs 1954.
- 46) Auflagenhöhe der drei Zeitungen (nach Ausgaben des Bundespresse- und Informationsamtes) für 1967:
FAZ: 249.000, WELT: 247.000, ND: 600.000.
Die Angaben für die Auflagehöhe des ND differieren, es handelt sich um die wahrscheinlich annähernd zutreffende Zahl. Dabei muß noch berücksichtigt werden, daß bei einem überwiegend propagandistisch orientierten Organ wie dem ND der Anteil der nur verteilten, nicht gelesenen Exemplare verhältnismäßig hoch liegt.

- 47) Vgl. zu den Mannheimer Zeitungstexten Teil I, Textliste (1.1.e).
- 48) Die Herstellung solcher Mischtexte kann unter bestimmten Bedingungen sinnvoll sein, etwa dann, wenn es darum geht, individualstilistische Eigenheiten in mehreren Texten sonst großer Ähnlichkeit zu unterdrücken (in der linguistischen Forschungsstelle in Besançon ist man z.T. diesen Weg gegangen). In diesem Sinne stellen aber unsere Zeitungen selbst schon hochgradige Mischtexte dar. Eine Mischung höherer Stufe würde die Untersuchung jedenfalls nicht erleichtern.
- 49) Unvollständige Wörter am Ende eines Artikelfragments werden ergänzt. Einige Artikelfragmente aus Ulbricht-Reden wurden ergänzt; vgl. Punkt 12. Verweismöglichkeiten bei nachträglicher Aufnahme fehlender Artikelteile s. 13.2.3. ("Zusatzkarte").
- 50) Der Umfang der Zielmenge für ND 64 wurde ermittelt aufgrund der Auszählung von 5 Monaten (vgl. Anhang Ia/1), für ND 54 durch Auszählung von 4 Monaten (vgl. Anhang Ia/3).
- 51) Der Umfang der Zielmenge wurde ermittelt aufgrund der Auszählung von 5 Monaten (vgl. Anhang Ia,2).
- 52) Tatsächlich schwankt der Umfang der Stichproben zwischen 2 und 4 Seiten (s. Liste der aufgenommenen Seiten, Anhang IIa).
- 53) In der Praxis bedeutet das, daß Seiten aus verschiedenen Ausgaben gegeneinander ausgetauscht wurden, was besonders bei Seiten mit hohen Seitennummern, d.h. bei Seiten, die selten vorkommen, mehrfach nötig wurde.
- 54) Die aus dem ND 64 und 54 zur Aufnahme vorgesehenen Seiten erhielten wir freundlicherweise in Form von Mikrofilm-Rückvergrößerungen vom Archiv für gesamtdeutsche Fragen, Bonn; die aus WELT 64 vorgesehenen Seiten vom Archiv des Seminars für Publizistik an der Freien Universität Berlin, und zwar als Mikrofilm; auch hier nahm das Archiv für gesamtdeutsche Fragen die Rückvergrößerungen vor (s. auch Vorbemerkung S.8).
- 55) Diese Februar-Menge lag den Untersuchungen von G. Billmeier im Teil III zugrunde.
- 56) Die Artikel liegen in unserem Material und auf dem Magnetband in fortlaufender, aufsteigender Reihenfolge. Gelegentlich sind Nummern unbesetzt, da einige Male zunächst getrennt numerierte Artikel unter einer Nummer zusammengefaßt wurden; in anderen Fällen wurden Artikel nachträglich getrennt, dann erhielt einer der beiden Artikel eine der bisher freigeblichen Nummern. Die Numerierung entspricht also nicht vollständig der Artikelfolge auf den Zeitungsseiten. Die Zuordnung von materialinterner Artikelnummer zum Original-Zeitungsartikel ist mit Hilfe der Informationskarte dennoch gesichert (vgl. Anhang III, 1.Seite).

- 57) Vgl. dazu die Mannheimer Schreibanweisungen, Teil I, Punkt 2.4.
- 58) Näheres zu den einzelnen Stufen des Verfahrens unten in den "Erläuterungen zum Arbeitsablauf", Punkt 14.1.
- 59) Vgl. dazu unten 13.2.2. und Anm. 67.
- 60) Diese Ersatzzeichen sowie auch andere Bestandteile unserer Schreibkonventionen gehen auf die Schreibkonventionen des Instituts für Phonetik und Kommunikationsforschung der Universität Bonn (IPK) der Jahre 1964/65 zurück. Inzwischen hat das IPK seine Schreibkonventionen geändert (vgl. den Forschungsbericht 1966/2 des IPK, Teil II S. 25 ff.); bei uns wurden sie jedoch beibehalten. Eine wechselseitige Umformung ist, mit geringen Abweichungen, möglich.
- 61) Hier sind nur solche Abweichungen berücksichtigt, die entweder einen Zuwachs oder eine Minderung an Informationen beinhalten.
- 62) Im ND 64 wurde der Unterschied zunächst noch berücksichtigt, jedoch schon in den Zusätzen zum ND 64 und dann im ND 54 und WELT 64 fallengelassen.
- 63) Auch System und Form der Transkriptionszeichen wurden vom IPK übernommen, allerdings den differenzierten Ansprüchen unserer Zeitungstexte entsprechend ausgebaut und z.T. geändert. Es wurde dann mit einigen Änderungen auch von der Mannheimer Zentrale übernommen.
- 64) Abweichend von der Mannheimer Regelung werden Transkriptionen nur durch einfaches Plus-Zeichen geschlossen, das Symbol der Transkriptionsart also nicht wiederholt. Überlappende Transkriptionen, die allerdings sehr selten vorkommen, sind also nach unseren Regeln nicht möglich.
- 65) Im ND 64 wurden auch eingliedrige Sach-Eigennamen, z.B. Ortsnamen, noch transkribiert. In WELT 64 wurde diese Transkribierung bei eingliedrigen und eindeutigen Sach-Eigennamen jedoch aufgegeben, da sie zweifelsfrei auch aus dem Index zu entnehmen sind, also ein Informationsverlust nicht zu befürchten ist.
- 66) In WELT 64 sind Buchstabenabkürzungen auch ohne Transkription formal erkennbar, nämlich durch die mehrfache Großschreibung in einem Wort (vgl. Anhang IV, Programmübersicht Nr. 9).

- 67) Für nicht schreibbare Stellen in der Zeitung (Bilder, Zeichnungen u. dergl.) benutzen wir das Transkriptionszeichen +X...+, wobei die Transkription mit einer Stichwortangabe dessen, was an dieser Stelle in der Zeitung steht, gefüllt wird. Für nicht zu schreibende Stellen (Zahlenkolonnen u. dergl.) wird in entsprechender Weise die Transkription +Y verwandt.
- 68) Die Informationskarte mit ihren Erweiterungen wurde, aufgrund von Anregungen des IPK, im wesentlichen in der Außenstelle neu erarbeitet.
- 69) Falls ja, kann die Artikelnummer, unter der der fehlende Teil nachträglich aufgenommen wird, auf einer besonderen Karte vermerkt werden (siehe unten).
- 70) Unsere Gliederung in 5 Hauptsachgebiete mit je 20 bis 28 Untersachgebieten ist rein pragmatisch; sie erhebt keinen Anspruch auf systematische Folgerichtigkeit und Vollständigkeit. Sie entstand im übrigen in Anlehnung an das von Heinz Vater entwickelte und im Lüneburger Forschungsbericht des Jahres 1965 vorgelegte System von Sachgebieten, wurde allerdings erheblich abgeändert.
- 71) Vgl. oben 2.1., Abs. 4, Anm. 8.
- 72) Vgl. oben Punkt 12 (Ergänzungen zu offiziellen Reden). Die Möglichkeiten der Beikarte wurden bisher nur teilweise bei ND 64, die der Zusatzkarte noch gar nicht realisiert.
- 73) Die Verwendung von Lochstreifen in größeren Mengen ist in Bonn aus überwiegend organisatorischen Gründen leider nur unter großen Schwierigkeiten möglich. Die Datenerfassung im IIM ist ganz überwiegend auf Lochkarten eingestellt. Texterfassung über Streifen hat den Vorteil, daß Streifen erheblich einfacher aufzubewahren sind als Lochkarten, die teure und platzverbrauchende Schränke benötigen; demgegenüber bietet die Verwendung von Karten vor allem beim Korrigieren große Vorteile.
- 74) Die beiden Lochkarten-Schreibgeräte sind gemietet.
- 75) Zu Anfang wurde neben dem Schreibblocher ein Lochkartenprüfer IBM 056 verwendet. Das Kartenprüfen mit dem genannten Gerät erwies sich jedoch als unzweckmäßig, da mehr (scheinbare) Fehler angezeigt als wirkliche beseitigt wurden. Da eine Prüfung "von Hand" wegen unserer Transkriptionen ohnehin nötig ist, wurde Ende 1965 auf den Kartenprüfer zugunsten eines zweiten Schreibblockers verzichtet.

- 76) Auf der Mannheimer Herbsttagung 1965 des IDS wurden, allerdings auf der Grundlage begrenzten Textmaterials, folgende spezielle Register aus kombinierten Abfragen vorgelegt:
1. Alphab. Index der Artikel zum Sachgebiet "Wirtschaft" (=Abfrage nur nach Inform.-Karte)
 2. Verzeichnis der Abkürzungen in Reden W. Ulbrichts (=2 verschiedene Abfragen aus Inform.-Karte und Abfrage nach Abkürzungs-Transkription),
 3. Verzeichnis der Personennamen (mit Titeln) aus Artikeln über die Bundesrepublik (=Abfrage nach Beikarte und Abfrage nach Personennamen-Transkription).
- Auf der Jahrestagung März 1968 wurden u.a. als Probeläufe vorgelegt:
1. Spezial-Index der Punkt-Abkürzungen aus ND 64,
 2. Kontext-Register der Initial-Abkürzungen aus WELT 64.
- 77) Im Index zum ND 64 fanden sich noch ca. 40 Fehler.
- 78) Zunächst nur im ND 64.
- 79) Näheres s. oben 4.3., Anm. 17 und Programmübersicht Nr. 14 u. 15.
- 80) Auf der Jahrestagung März 1968 wurden zwei Probeläufe dieser Art vorgelegt, von denen der eine die Suche nach "Wörtern" (auch nach Wortteilen, sofern die gesuchte Zeichenkombination im Innern zu zusammengesetzter Wörter steht), der andere die Suche nach "Wortendungen" (in diesem Fall Substantiv-Suffixen) demonstrierte. Das Verfahren der Suche nach Endungen befriedigt noch nicht, da "Wortendung" zur Zeit noch definiert ist als Zeichenfolge, der mindestens zwei Konsonanten und ein Vokal vorausgehen und ein Leerzeichen folgen müssen.
- 81) Im Normalfall werden genaues Erscheinungsdatum und Seite des Artikels angegeben. Vgl. Programmübersicht, Nr. 7a.
- 82) Ungenauigkeit für die Werte von S_{eJ} und S_{aJ} : ± 20 Seiten. Diese mögliche Abweichung wirkt sich in der Berechnung der Seitenverteilung in der Modellmenge nicht mehr aus. - Die Abweichung gilt auch entsprechend für die Ermittlung der Werte für WELT 64 und ND 54.
- 83) Entsprechend der unter 4.5.4.2. vorgeschlagenen vereinfachenden Möglichkeit wird hier sowohl für die Berechnung von F als auch von f nicht der (ermittelte) Wert von S_{aJ} (=2125), sondern der (tatsächlich ausgezählte) Wert aus 5 Monaten (≈ 912) zugrundegelegt. Genauer müßte also statt S_{aJ} ein anderer Begriff (etwa S_{aT} , wenn T die ausgezählte Teilmenge ist) gesetzt werden. - Entsprechendes gilt für die Modellmengenberechnung der anderen beiden Jahrgänge.

- 84) Einschließlich unnumerierter Seiten, jedoch ausschließlich aller Beilagen (vgl. oben 8.1.b und 4.5.3.1.a, Anm. 28).
- 85) Zahl der Erscheinungsmengen und der Ausgaben stimmen im ND 54 (wie schon in WELT 64) überein, da zwar keine Montagsausgabe, jedoch eine Sonntagsausgabe erschienen ist, die mitberücksichtigt wurde.