

# SPRACHKUNST

Beiträge zur Literaturwissenschaft

Jahrgang 1/1970

Heft 1/2

S O N D E R D R U C K

V E R L A G B Ö H L A U

## TRAKL

### *Versuche zur maschinellen Analyse von Dichtersprache* \*

Von Wolfgang Klein und Harald Zimmermann (Saarbrücken)

#### 1 Vorbemerkungen

An der Forschungsstelle ‚Automatische Syntaktische Analyse‘ des Germanistischen Instituts der Universität Saarbrücken wurde in den letzten Jahren ein Verfahren entwickelt, beliebige deutsche Sätze mit einem Elektronenrechner syntaktisch zu analysieren<sup>1)</sup>. Ausgearbeitet wurde es an Texten aus einer mittleren Sprachschicht, wie sie sich etwa in der (gehobenen) Journalistenprosa dokumentiert. Das heißt aber keineswegs, es sei in seiner Anwendung auf eben diese Sprachschicht beschränkt: Im Prinzip kann es jeden deutschen Satz analysieren, sofern er den gängigen Regeln der Grammatik entspricht. Die Frage ist allerdings, in welchem Ausmaß sich das hier mit Erfolg angewandte Verfahren auf Sprachschichten übertragen läßt, deren Syntax von der üblichen, als korrekt empfundenen mehr oder minder abweicht — etwa auf der einen Seite die (schriftlich fixierte) gesprochene Umgangssprache<sup>2)</sup> und — auf der anderen Seite — die in anderer Hinsicht, aber im gleichen Maße von der „Regel“ abweichende Sprache der lyrischen Dichtung. Das sind natürlich zwei Extreme, und es versteht sich, daß das Saarbrücker Analyseprogramm in seiner gegenwärtigen Form diesen beiden Sprachverwendungen nicht gerecht werden kann. Dafür ist es nicht konzipiert.

Aber es wäre verlockend, die Möglichkeiten des Elektronenrechners in anderer Hinsicht für die Untersuchung der Dichtersprache zu nutzen, bescheidener gesagt: zu untersuchen, welche Möglichkeiten es hier gibt. Die Verfasser dieses Artikels haben daher, von den Professoren Hans Eggers und Helmut Kreuzer (beide Saarbrücken) angeregt, begonnen, unabhängig von ihrer Arbeit an der Saarbrücker Forschungsstelle ein Arbeitsprogramm zur automatischen Analyse von Dichter-

\*) Veränderte Fassung des Arbeitsberichtes in: Linguistische Arbeiten des Germanistischen Instituts und des Instituts für Angewandte Mathematik der Universität des Saarlandes, Bd. 6 (1969), Bl. 22—46 (nicht im Buchhandel).

<sup>1)</sup> Dieses Verfahren, seine Grundlagen, seine Effektivität und seine Anwendungsmöglichkeiten sind ausführlich beschrieben in HANS EGGERS U. a., Elektronische Syntaxanalyse der deutschen Gegenwartssprache, Tübingen 1969.

<sup>2)</sup> Wie sie etwa Tucholsky in seinen Wendriner-Geschichten effektiv simuliert. Zum Problem der gesprochenen Umgangssprache vgl. etwa H. STEGER, Gesprochene Sprache, in: Satz und Wort im heutigen Deutsch, Düsseldorf 1967, S. 259—291.

sprache zu entwerfen und an einem Beispiel, dem Werk Georg Trakls, zu realisieren. Die Vorarbeiten erwiesen sich als relativ langwierig, mußten auch verschiedentlich unterbrochen werden. Immerhin kann nun ein Zwischenbericht gegeben werden.

## 2 „Syntaktische“ Untersuchungen

Alles, was an einem Text objektiv beschreibbar ist, muß theoretisch auch von einer Maschine erfaßt werden können. Das besagt nicht, eine vollständige Erfassung der formalen, besser gesagt: formalisierbaren Gegebenheiten eines Textes könne unbedingt auch seine poetische Wirkung erklären: die hängt auch von zahlreichen extratextuellen Determinanten ab, die zwar zumindest teilweise objektivierbar sind, aber die Reichweite linguistischer Textanalysen übersteigen.

Die Sprache ist ein System von Zeichen; einen Text in einer bestimmten Sprache bilden heißt von einem Zeichensystem einen bestimmten Gebrauch machen; poetische Texte sind wie alle anderen zunächst einmal ein spezieller Fall von Zeichenverwendung; es erscheint uns daher gerechtfertigt, die gesamte Untersuchung der Übersichtlichkeit halber in die bekannten drei „Dimensionen“ der Morrisschen Zeichentheorie<sup>3)</sup> — „Syntax“ (jedwede Beziehung von Zeichen, genauer: von Zeichenträgern als den materiellen Repräsentationen von Zeichen, untereinander), „Semantik“ (Beziehung des Zeichens zum Bezeichneten), „Pragmatik“ (Beziehung des Zeichens zum Interpreten, das heißt zu dem, der es als Zeichen deutet, für den es mithin etwas bezeichnet) — zu unterteilen. In diesem Sinne ist, was wir tun, primär syntaktisch orientiert<sup>4)</sup>. Untersucht wird vor allem die Konstellation der Phoneme, teilweise auch der Phonemgruppen, und der Wörter (das heißt das, was man als „Syntax“ im engeren Sinne auffaßt<sup>5)</sup>).

Wir führen nun exemplarisch einige Punkte aus der Phonemanalyse an. Es ist nicht schwer, sich weitere Möglichkeiten auszumalen, und es läßt sich absehen, daß die meisten zu nichts führen werden. Aber ein gewisses Herumprobieren wird sich wohl nicht vermeiden lassen bei der Suche nach jenen Prinzipien, die, dem Verfertiger vielleicht unbewußt, die Lautstruktur von Texten determinieren<sup>6)</sup>. (Im folgenden stehen C für „Konsonant“, V für „Vokal“ und P für „Phonem“.)

<sup>3)</sup> Vgl. dazu CHARLES W. MORRIS, *Foundations of the Theory of Signs*, Chicago 1938, 12. Aufl. 1966. Daß wir uns dieser Einteilung bedienen, heißt keineswegs, daß wir Morris' eigene Ästhetiktheorie übernehmen, etwa so, wie er sie in seinem Buch *Signification and Significance*, Cambridge, Mass., 1964, dargelegt hat.

<sup>4)</sup> Vgl. zur syntaktischen Analyse im weiteren Sinne DIETER KRALLMANN, *Statistische Methoden in der stilistischen Textanalyse*, Diss., Bonn 1965; ders., *Stilistische Textbeschreibung mit statistischen Methoden*, in: *Nicht-numerische Informationsverarbeitung*, hrsg. von R. Gunzenhäuser, Wien 1968, S. 330—345.

<sup>5)</sup> Wenn man will, mag man darin den Reflex einer phonologischen bzw. einer syntaktischen Grammatikkomponente sehen, die hier in der „performance“ untersucht werden. Es muß ausdrücklich betont werden, daß es hier nicht um das Sprachsystem, sondern um seine Verwendung geht.

<sup>6)</sup> Vgl. dazu etwa KARL KNAUER, *Die Analyse von Feinstrukturen im sprachlichen Zeitkunstwerk*, in: *Mathematik und Dichtung*, hrsg. von H. Kreuzer und R. Gunzenhäuser, München 1969, 3. Aufl., S. 193—210.

1. Häufigkeiten der Phoneme, prozentualer Anteil am Gesamtbestand, Verhältnis  $V : C^7$ , Verhältnis Labiale : Dentale usw., jeweils für den gesamten Text und für bestimmte Segmente, etwa Zyklen.

2. Übergangswahrscheinlichkeiten

a) Mit welcher Wahrscheinlichkeit folgt ein Phonem  $P_j$  auf ein Phonem  $P_i$  beziehungsweise allgemeiner: auf eine Phonemgruppe  $P_{i1} \dots P_{ir}$ , wobei (beispielsweise)  $r = 0, \dots, 5$ .

b) Mit welcher Wahrscheinlichkeit folgt ein Konsonant  $C_j$  auf einen Konsonanten  $C_i$ , wobei auch  $C_i = C_j$  sein kann? Die Vokale werden dabei gleichsam ausgeblendet, das heißt, sie werden einfach übergangen.

c) Mit welcher Wahrscheinlichkeit folgt ein Vokal  $V_j$  auf einen Vokal  $V_i$ ? Dabei werden also umgekehrt die Konsonanten nicht berücksichtigt. Selbstverständlich kann auch in diesem Fall  $V_i = V_j$  sein. Hier wie in b ist der gesamte fortlaufende Text Grundlage der Berechnung.

d) Mit welcher Wahrscheinlichkeit folgt auf einen bestimmten im Anlaut stehenden Konsonanten  $C_i$  wieder ein Anlautkonsonant  $C_i$ . Ausgeblendet werden also alle Phoneme mit Ausnahme der im Anlaut stehenden. Damit wird die Grundlage für eine Untersuchung der Alliteration geschaffen<sup>5</sup>). Statt eines bestimmten Konsonanten Q kann auch eine Konsonantengruppe (etwa „tr“) oder die Menge der stimmlosen, labialen, ... Konsonanten herangezogen werden.

3. Filter: Bestimmte Phoneme werden ausgeblendet, zum Beispiel alle außer einem bestimmten Phonem  $P_i$  oder der Folge  $P_i \dots P_j$ , alle Konsonanten, alle vorderen Vokale usw. Das würde uns auf Untersuchungen führen, wie sie etwa Karl Knauer für die Sonette Baudelaires unternommen hat (siehe Anm. 6).

4. Reim: Die Lautstruktur vom letzten betonten Vokal im Vers bis zum Versende wird analysiert. Es versteht sich, daß in analoger Weise auch der Binnenreim untersucht werden kann.

Bei der Untersuchung der „Syntax im engeren Sinne“ (des Satzbaus) soll versucht werden, das Saarbrücker Analyseprogramm — mit einigen technischen Modifikationen — auf das neue Material anzusetzen. Was dabei herauskommen wird, läßt sich kaum vorhersagen. Wir wollen hier daher nur kurz darauf hinweisen, welche Art von Ergebnissen man — vielleicht — zu erwarten hat.

Erstens ist unser Analyseprogramm für ein Material konzipiert, das durchgehend grammatikalisch sein muß, wenn die Analyse nicht fehlgehen soll; daher können uns alle Fälle, in denen es bei poetischer Sprache versagt, zumindest einen

<sup>7</sup>) Es erscheint vielleicht banal, etwas Derartiges wie das Verhältnis von Vokalen zu Konsonanten zu beredinen; daß es das keineswegs ist, zeigt etwa der Umstand, daß in manchen literarischen Kanons (späte Troubadourlyrik, russische Futuristen) konsonantische Abundanz als Vorzug in der Lyrik galt, während man umgekehrt meist geneigt ist, gerade in der „Klangfülle“, das heißt in der vokalischen Abundanz eines ihrer Charakteristika zu sehen.

<sup>8</sup>) Das entscheidende Problem in diesem Zusammenhang ist, ob es gelingt, statistische Signifikanztests für minimale Textsegmente, etwa Verszeilen oder allenfalls Strophen, zu entwickeln. Wir können auf diese Frage aber hier nicht weiter eingehen.

Hinweis darauf geben, daß hier möglicherweise eine Abweichung von der Grammatikalität vorliegt — wohlgemerkt: einen Hinweis; wir sind natürlich nicht so vermessen anzunehmen, daß das Analyseprogramm nie versagen könnte, sofern der analysierte Satz bloß grammatikalisch ist.

Zweitens fällt es relativ leicht, mit dem Arbeitsaufwand, den die Maschine zur Analyse aufbringen muß, ein — allerdings oberflächliches — Maß der Komplexität zu verbinden. Oberflächlich ist dieses Maß sowohl in einem technischen Sinn — es berücksichtigt nur die Oberflächenstruktur — wie in einem allgemeineren, insofern die maschinelle Analyse nur teilweise die strukturelle Dekodierung eines Satzes durch den Menschen simuliert, sein Arbeitsaufwand also nicht unbedingt dem der Maschine gleichzusetzen ist.

Drittens hat sich bereits bei der bisherigen Arbeit mit gewöhnlichem Material gezeigt, daß ein streng formaler Analysemechanismus zwangsläufig syntaktische Ambiguitäten aufdeckt, die der vom Sinn des Gesagten beeinflusste Leser zunächst zu übersehen geneigt ist<sup>9)</sup>.

Viertens endlich lassen sich syntaktische Äquivalenzen (das heißt Parallelitäten im Satzbau benachbarter Sätze) durch ein derartiges Analyseverfahren mit Leichtigkeit auffinden — allerdings auch hier nur in der Oberflächenstruktur. Man könnte also alle vier Faktoren, die nach Klaus Baumgärtner<sup>10)</sup> die Poetizität eines Textes konstituieren, durch ein syntaktisches Analyseprogramm in erster Annäherung bestimmen.

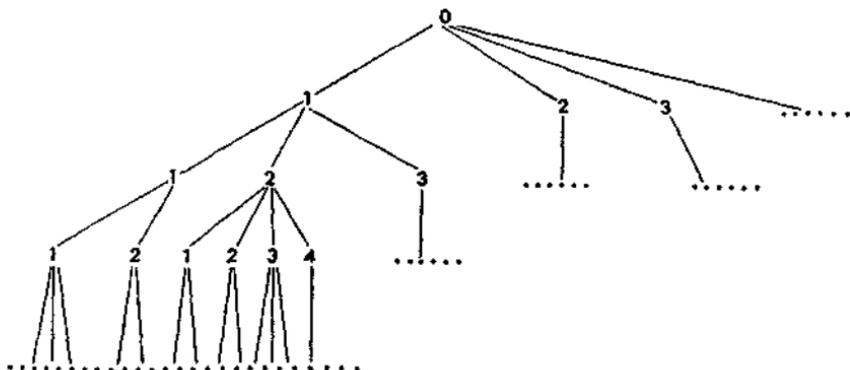
### 3 Semantik

Syntaktische Untersuchungen (im weiteren Sinne), zumal so einfache wie die angeführten, lassen sich relativ leicht mit einer Maschine durchführen, viel leichter jedenfalls als semantische. Wir sehen derzeit, trotz aller Anstrengungen auf diesem Gebiet, keinen Ansatz, der die Semantik computerreif zu formalisieren erlaubte. Einer der Gründe liegt sicher darin, daß bei der Kommunikation von Inhalten der rein sprachliche Bereich verlassen wird. Nun wäre es zwar nicht befriedigend, aber doch ein gewisser Fortschritt, wenn es gelänge, ein formales Verfahren zu entwickeln, das, den Referenzbereich zunächst noch außer acht lassend, den Vergleich textinterner semantischer Strukturen erlaubte. Ein solcher Vergleich läßt sich, so glauben wir, über ein sogenanntes „Begriffswörterbuch“, etwa das bekannte von

<sup>9)</sup> Man vergleiche etwa folgenden, in unseren Korpora belegten Satz: Man wollte das Schreckliche, Seltsame, das in Raum und Zeit Heimatlose einfangen, um es zu bannen. Das Wort einfangen ist dabei sinngemäß ein von wollte abhängiger Infinitiv, aber formal könnte es genauso gut Prädikat eines Relativsatzes mit dem Subjekt Heimatlose sein. Es fragt sich natürlich, ob derart fernliegende Mehrdeutigkeiten noch poetisch relevant sind.

<sup>10)</sup> Formale Erklärung poetischer Texte, in: *Mathematik und Dichtung*, hrsg. von H. Kreuzer und R. Gunzenhäuser, München 1969, 3. Aufl., S. 67—84.

Hallig-Wartburg<sup>11)</sup>, realisieren, ein Wörterbuch also, in dem die Wörter nicht nach einem formalen Prinzip wie dem Alphabet, sondern semantisch angeordnet sind: von gewissen Oberbegriffen ausgehend zu immer feineren Subgruppierungen. Sie bilden also eine semantische Hierarchie, die man sich in einem Stemma veranschaulichen kann:



0 bedeutet dabei „Wort überhaupt“, an jedem weiteren Knoten stehen bestimmte Wörter (genauer: Graphemfolgen zwischen zwei Leerzeichen), deren semantische Stellung im System — nicht etwa deren Bedeutung — sich durch eine den „Pfad“ reflektierende Ziffernfolge, also etwa 011241, 0362154, 021 usw., darstellen läßt. Jede Einheit in der semantisch zu analysierenden Textsequenz läßt sich demnach gemäß ihrer Position in diesem System charakterisieren; dabei wird ganz davon abgesehen, ob es „richtig“ ist, also in der Tat die semantischen Strukturen der betreffenden Sprache reflektiert. Nicht als ob dies belanglos wäre, aber zunächst wird ja nicht mehr gesucht als ein gemeinsames Bezugssystem. — Die semantische Beschreibung eines Textes sähe dann etwa so aus: 014120113240243102435643-21032..., wobei 0 jeweils eine Wortgrenze bezeichnet; bloß funktionale Morpheme („Synsemantica“) könnte man ausklammern oder mit 000 oder ähnlichem kennzeichnen. Aber das sind Detailprobleme, hier soll nur der Grundgedanke angedeutet werden. — Ein derartiges Verfahren ermöglicht es beispielsweise, die Jakobsonische Äquivalenztheorie („the poetic function projects the principle of equivalence from the axis of selection into the axis of combination“)<sup>12)</sup> auch auf semantischer Ebene einer formalen Bearbeitung zugänglich zu machen. Die Similarität oder Dissimilarität äußert sich in Form kleinerer oder größerer „semantischer Differenzen“, die bei nichtpoetischen Texten keine besondere Strukturierung erkennen lassen, bei poetischen aber gewisse Regelmäßigkeiten aufweisen.

<sup>11)</sup> R. HALLIG und W. VON WARTBURG, Begriffssystem als Grundlage für die Lexikographie, Berlin 1963, 2. Aufl.; für den Anfang wird es sicherlich zweckmäßiger sein, von einem einfacheren, nicht so detaillierten System auszugehen.

<sup>12)</sup> R. JAKOBSON, Poetics and Linguistics, in: Style in Language, hrsg. von TL A. Sebeok, Cambridge, Mass., 1960, S. 350—377, Zitat auf S. 358.

Wir können diesen Gedanken hier nicht weiter ausführen. Es ist klar, daß ein derartiges Verfahren nichts als ein primitiver Ansatz ist. Aber eine semantische Theorie, die angetan wäre, die Bedeutungsstrukturen einer natürlichen Sprache angemessen zu erfassen, vor allem in einer so hoch entwickelten Anwendungsform wie der lyrischen Dichtung, ist Utopia.

#### 4 P r a g m a t i k

Das gilt in noch höherem Maße für eine Theorie der Pragmatik, die ja die bloße Textanalyse transzendiert und per definitionem die rezipierenden Individuen in ihre Betrachtung einbeziehen muß; aus eben diesem Grunde gehört sie aber auch nicht mehr zu unserem primären Interessenbereich. Wir sind allerdings der Ansicht, daß eine Literaturtheorie, die den pragmatischen Aspekt ausklammerte, also textimmanent bliebe, ihrem Gegenstand nicht gerecht werden könnte. Erreichbar wäre in diesem Fall etwas, das sich gleichsam als Ideal so formulieren ließe: Texte werden so beschrieben, daß aus der Beschreibung neue Texte mit den gleichen Eigenschaften konstruiert werden können. Aber dieses Ziel kann natürlich nicht mehr als eine Art leitende Idee sein.

Sicherlich sind für die Poetizität eines Textes, oder allgemeiner noch: für das, was ein Betrachter für das Eigentümliche an einem bestimmten Text ansieht, nicht alle seine Merkmale gleichermaßen relevant. Die Frage, auf welche Merkmale es nun ankommt, läßt sich nur durch pragmatische Untersuchungen klären; und eben dabei ließe sich ein Computer wirkungsvoll einsetzen. Man könnte ihn etwa eine Serie von Textsequenzen produzieren lassen, die jeweils in einem bestimmten Merkmal abgewandelt werden, in den übrigen gleichbleiben, und diese Textsequenzen dann durch Versuchspersonen auf einer Skala der Poetizität einordnen lassen. Wahrscheinlich wird man dabei feststellen, daß die Bewertung nicht von einzelnen Merkmalen, sondern sehr komplexen Merkmalkonstellationen abhängt — ganz abgesehen davon natürlich, daß nicht nur textimmanente Faktoren den Grad der Poetizität bei der Rezeption bestimmen. Aber über den Nutzen eines solchen Vorgehens läßt sich ohne empirische Untersuchungen nicht allzuviel sagen.

#### 5 A u t o m a t i s c h e T r a n s k r i p t i o n

Es wurde bereits erwähnt, daß einen wesentlichen Teil unseres Programms Untersuchungen zur Lautstruktur ausmachen. Da die zu untersuchenden Texte im allgemeinen in gedruckter Form vorliegen und da sich im Deutschen wie in den meisten europäischen Sprachen die Lautstruktur nicht eindeutig aus der Orthographie ableiten läßt, müssen die Texte zunächst phonematisiert, das heißt in eine (sich auf Phoneme beschränkende) Lautschrift übertragen werden. Das von uns benützte Verfahren, das im übrigen auch einige Allophone berücksichtigt, wird weiter unten beschrieben. Hier soll nur auf einen speziellen Punkt eingegangen werden.

Die ganze bisherige Transkription war manuell. Aber der auf diese Weise phonematisierte Text kann als Basis für eine künftige maschinelle Umschrift dienen. Da der gleiche Text auch orthographisch aufgenommen worden ist, läßt sich jedem orthographischen Wort seine phonematische Entsprechung zuordnen. Dieses „Wörterbuch“\* müßte in zweierlei Weise ergänzt werden:

1. Es müßten bestimmte „Ausspracheregeln“ programmiert werden, etwa jene, die Max Mangold dem ‚Aussprache-Duden‘ vorangestellt hat<sup>15</sup>). Was sich so nicht erfassen ließe, müßte

2. direkt ins „Wörterbuch“ aufgenommen werden. Das gilt etwa für zahlreiche Fremdwörter wie auch für „Polyphone“ — Graphemfolgen, die verschieden ausgesprochen werden können, wie etwa „ödem“ [CE:DCM oder CEDE:M]<sup>14</sup>), „Wachstube“ [WAXSCHTU:BC oder WAKSTU:BC] usw.

Bei der automatischen Phonematisierung eines orthographisch eingegebenen Textes wird zunächst abgefragt, ob das jeweils bearbeitete Wort im „Wörterbuch“ vorhanden ist; wenn ja, wird das entsprechende phonematische Korrelat dafür eingesetzt, wenn nein, wird das Wort an die Regeln verwiesen und ihnen gemäß umgeschrieben. Polyphone werden in einer Liste ausgedruckt und müssen manuell nachgetragen werden. Durch diese Teilautomatisierung — denn mehr ist dieses Verfahren ja nicht — läßt sich die Transkription zumindest erheblich beschleunigen. Schon mit einem Mikroglossar, das lediglich die zweihundert häufigsten Wörter umfaßt, läßt sich jedes zweite Textwort transkribieren. Sicher kann ein derartiges Verfahren den Theoretiker nicht befriedigen, aber für die Praxis scheint es brauchbar. Die Fehlerzahl dürfte kaum höher liegen als bei manueller Umschrift.

## 6 Wahl des Materials

Als Untersuchungsmaterial haben wir Trakls ‚Dichtungen‘ gewählt, und zwar hauptsächlich aus folgenden Gründen:

1. Sie sind nicht allzu umfangreich; unter Ausschluß der Prosateile enthält der Text etwa 60.000 Phoneme; das entspricht rund 12.000 Wortformen (siehe unten, S. 136). Der Wortschatz ist begrenzt und daher gut zu überschauen (ca. 1500 verschiedene Wortformen).

2. Bei Trakl dürfte eine Lautanalyse besonders aufschlußreich sein. Walther Killy hat die These vertreten, Trakl sei bei der häufigen Umarbeitung seiner Gedichte oft nicht nach inhaltlichen, sondern klanglichen Kriterien verfahren<sup>15</sup>). Diese These ließe sich durch eine maschinelle Lautanalyse leicht nachprüfen; aber

<sup>15</sup>) Der Große Duden 6, Aussprachewörterbuch, Mannheim 1962, S. 46—80; für die in Vorbereitung befindliche 2. Aufl. hat Prof. Mangold diese Regeln teilweise umgearbeitet und neu formuliert.

<sup>14</sup>) Erklärung der Lautschrift unten, S. 130; die Betonungsunterschiede sind hier nicht angegeben.

<sup>15</sup>) Siehe etwa in W. KILLY, Wandlungen des lyrischen Bildes, Göttingen 1964, 4. Aufl., S. 122.

ganz gleich, ob sie stichhaltig ist: Es kann kein Zweifel bestehen, daß die Lautstruktur in ganz entscheidendem Maße den eigentümlichen Trakl-Ton mitbestimmt.

3. Die syntaktische Struktur der "Traklschen Gedichte weicht zwar von der der Alltagsprosa deutlich ab, aber doch wiederum nicht so stark, daß die Anwendung eines auf durchschnittliche Prosa konzipierten Parsers von vornherein sinnlos erschiene. Man stelle sich etwa vor, Hölderlins späte Hymnen oder Jandl-Gedichte sollten syntaktisch analysiert werden!

4. Schließlich bieten sich Iraklis Gedichte für semantische Untersuchungen wie die in Abschnitt 3 skizzierten an: Bestimmte semantische Konfigurationen, die von den in der Alltagssprache üblichen stark abweichen (z. B. Farbadjektiv+Abstraktum) tauchen bei Trakl überaus häufig auf und sind markant genug, auch bei einem relativ undifferenzierten Beschreibungsverfahren wie dem geschilderten nennenswerte Aufschlüsse zu erbringen.

## 7 Vorarbeiten

### 7.1 Das Lautsystem

Im Zusammenhang mit der Transkription sind zwei grundsätzliche Probleme zu klären, nämlich

- a) nach welchen Prinzipien soll transkribiert werden, und
- b) wie ist die Umschrift technisch zu verwirklichen?

Maßgeblich für die Lautung ist für uns nicht die Aussprache des Dichters, hier also Trakls, sondern die des durchschnittlichen Rezipienten. Wir wählen also die hochsprachliche Lautung, wie sie im Aussprachewörterbuch des Duden<sup>16)</sup> verzeichnet ist.

Als Eingabemedien stehen bei der benutzten Rechenanlage<sup>17)</sup> Lochkarten und Lochstreifen zur Verfügung. Wir haben (5-K.anal-)Lochstreifen gewählt, da sie uns auf Grund der Maschinenkonzeption geeigneter scheinen. Die Zeichen des Internationalen Phonetischen Alphabets (IPA) können als solche weder bei der Eingabe (Lochstreifen) noch der Ausgabe (über Schnelldrucker) dargestellt werden, da keine entsprechenden Schrifttypen vorhanden sind. Nach einigen Vorüberlegungen haben wir uns für die folgende Umschrift entschieden. Die Reihenfolge entspricht dabei zugleich der alphabetischen Folge der Phoneme.

Ein- und Ausgabealphabet sind als den technischen Gegebenheiten des Fernschreibers und des Schnelldruckers angepaßte Codierungen zu verstehen. Wie zu ersehen ist, wird bei der Eingabe die Dehnung eines Vokals durch Doppelschreibung gekennzeichnet. Ausgenommen davon sind einmal [e:] und [o:], da sie im Deutschen häufig sind, während die kurzen Vokale [e] und [o] nur in Fremdwörtern

<sup>16)</sup> Duden 6 (zit, Anm. 13). In einigen Zweifelsfällen stand uns Prof. Dr. Max Mangold, der Bearbeiter des Aussprache-Duden, beratend bei. Die Verantwortung für die Umschrift liegt natürlich bei uns selbst.

<sup>17)</sup> Näheres unter 8.1.

aufzutreten. Anstelle einer Verdoppelung wurde bei [e:] statt AEAE die Zeichenkombination AV verwendet. Der Doppelpunkt tritt als Längenzeichen nur bei der Ausgabe auf. Aus ökonomischen Gründen wurde dieses Zeichen bei der Eingabe nicht verwendet, da auf Grund der benutzten 5-Kanal-Lochstreifen vor jedem

Lauf-Nr.	IPA <sup>18)</sup>	Ausgabe (Drucker)	Eingabe (Lochstreifen)
0	[a]	A	A
1	[a:]	A:	AA
2	[b]	B	B
3	[d]	D	D
4	[e]	E	EE
5	[e:]	E:	E
6	[ɛ]	AE	AE
7	[ɛ:]	AE:	AV
8	[ɔ]	C	C
9	[f]	F	F
10	[g]	G	G
11	[ʒ]	J	ZZ
12	[h]	H	H
13	[i]	I	I
14	[i:]	I:	II
15	[k]	K	K
16	[l]	L	L
17	[m]	M	M
18	[n]	N	N
19	[ŋ]	NJ	NJ
20	[o]	O	OO
21	[o:]	O:	O
22	[ɔ]	Q	Q
23	[œ]	QE	QE
24	[ø:]	OE:	OE
25	[p]	P	P
26	[r]	R	R
27	[s]	S	S
28	[ʃ]	SCH	SCH
29	[z]	Z	Z
30	[t]	T	T
31	[u]	U	U
32	[u:]	U:	UU
33	[y]	Y	Y
34	[y:]	Y:	YY
35	[v]	W	W
36	[x]	X	X
37	[ç]	CH	CH

18) Zur Erläuterung der IPA-Symbole vgl. die Einleitung des Aussprache-Duden.

Längenzeichen eine Umschaltung auf Ziffernzeichen und danach eine Rückschaltung auf Buchstabenzeichen erforderlich gewesen wäre.

Einige Schwierigkeiten ergeben sich aus dem Begriff des Phonems selbst. So stehen [x] und [ç] in komplementärer Distribution und wären demnach als Allophone eines einzigen Phonems anzusehen. Dagegen spricht nun wieder ihre phonetische Verschiedenheit, die in unserem Fall den Ausschlag gab. Ebenso haben wir das [ŋ] als eigenes Phonem angesetzt, obwohl es in keiner Position distinktiv ist. An Phonemen, die nur in sogenannten „Lehnwörtern“ vorkommen, sind bisher lediglich [e], [o] und [ʒ] vorgesehen. [ʒ] ist nicht belegt. Es ist darüber hinaus phonetisch gesehen unbefriedigend, daß nicht zwischen [i] und [I] unterschieden wird, doch schien uns hier ein Kompromiß zwischen Exaktheit und Ökonomie angebracht <sup>19)</sup>.

Auf Grund der Codierung sind Doppeldeutigkeiten möglich. Die gleiche Zeichenkombination, die für ein einziges Phonem steht, kann auch dann auftreten, wenn jedes Symbol für sich einen Lautwert anzeigt. So wird für [ç] (wie in ICH) die Zeichenkombination CH eingesetzt, aber auch die — seltene — Phonemkombination [ə] + [h] (wie in *GEHÖRT*) wird durch die Zeichenfolge CH dargestellt. Um Fehler zu vermeiden, muß in den Fällen, wo einzelne Phoneme transkribiert werden, bei Mehrdeutigkeit ein Schrägstrich als Trennzeichen zwischen den Symbolen stehen, also zwar ICH, aber GC/HOERT).

## 7.2 Zusatzinformationen

Die Gedichte wurden über Fernschreiber in der Reihenfolge der verwendeten Trakl-Ausgabe in Lautschrift auf Lochstreifen übertragen. Dabei waren zusätzlich folgende Konventionen zu beachten:

a) Jede Wortform sollte durch eine genaue Stellenangabe identifizierbar sein. Daher wird die Zugehörigkeit zu einem Gedicht festgehalten, das seinerseits intern durch eine (fortlaufende) Nummer gekennzeichnet ist.

b) Daneben sind Seite und Verszeile der Ausgabe, Hinweise auf die Strophe und den Abschnitt sowie die absolute Nummer einer Wortform innerhalb des Gedichts vermerkt. Dies wird dadurch erreicht, daß zu Beginn einer neuen Seite die Seitennummer — in eckigen Klammern — geschrieben wird.

c) Bei Gedichtanfang wird der Titel — ebenfalls in eckigen Klammern eingeschlossen — angegeben <sup>20)</sup>.

d) Hinweise auf Zeile, Strophe und Abschnitt eines Gedichts gibt die Anzahl der Zeilentransporte (= Zwischenräume zwischen den Verszeilen):

<sup>19)</sup> Es kommt uns nicht darauf an, ob dieses System streng phonematisch aufgebaut ist, ob es sich als „phonematisch mit Berücksichtigung einiger Allophone“ charakterisieren läßt oder dergleichen.

<sup>20)</sup> Dies dient zur Kennzeichnung eines Gedichtanfangs. Der Wortlaut des Titels wird gesondert aufbewahrt. Er ist bei der Bearbeitung des Textmaterials vorläufig nicht berücksichtigt, doch kann jederzeit darauf zurückgegriffen werden.

- 1 Transport = noch zur Verszeile gehörig (bei überlangen Zeilen)
- 2 Transporte = neue Verszeile
- 3 Transporte = neue Verszeile, neue Strophe
- 4 Transporte = neue Verszeile, neue Strophe, neuer Abschnitt

e) Weiterhin werden die einem Wort folgenden Satzzeichen (maximal drei aufeinanderfolgende) festgehalten:

Zeichen	Internwert (und Ausgabe)	Lochstreifen (Symbol)
Punkt	1	.
Komma	2	,
Semikolon	3	;
Doppelpunkt	4	:
Anführungszeichen	5	'
Fragezeichen	6	×
Ausrufezeichen	7	!
Klammer	8	), (
Gedankenstrich	9	—

f) Im Deutschen liegt im allgemeinen der Hauptton auf dem ersten Vokal einer Wortform. Abweichungen von dieser „Regel“ sind bei den Eingabedaten dadurch gekennzeichnet, daß vor dem Trägervokal des Haupttons ein Gleichheitszeichen (=) eingefügt ist (Beispiel: Narziß — NARTS=IS). Ausgenommen von dieser Codierungsvorschrift ist der Vokal [9] (in den Vorsilben „be-“ und „ge-“), da [9] niemals betont ist. Nebentöne sind nicht gekennzeichnet.

Es schien uns vorerst nicht angebracht, weitere Informationen zu den Wortformen anzugeben, etwa eine Differenzierung nach syntaktischen Funktionen vorzunehmen. Die ohnehin recht langwierigen Vorarbeiten wären noch umfangreicher und die erforderlichen Textkontrollen und -korrekturen zudem erschwert worden.

## 8 Maschinelle Verarbeitung

### 8.1 Zur Verarbeitung sprachlicher Daten mit Hilfe eines Elektronenrechners

Ehe wir unsere Vorgehensweise bei der maschinellen Verarbeitung schildern, wollen wir einige grundlegende Bemerkungen vorausschicken, die erste Vorstellungen von den später verwendeten Begriffen vermitteln sollen: Bei der Bearbeitung von Sprache — in unserem Falle von Dichtersprache — durch den Elektronenrechner muß man davon ausgehen, daß dazu ein genau festgelegtes und für die Rechenanlage verständliches „Programm“ (=Folge von Bearbeitungsanweisungen oder Instruktionen in einer „Maschinensprache“) gehört, das von dem Linguisten oder Programmierer in allen Einzelheiten ausgearbeitet ist. Der Computer „kann“ also nur das, was ihm in Programmform vorgeschrieben wurde. Sein „Gedächtnis“ ist der Datenspeicher, der die zur Verarbeitung benötigten Informationen enthält. Er ist aus einer großen Zahl von Zellen (=Speicherplätzen, -worten) zusammen-

gesetzt. Eine Zelle wiederum besteht aus einzelnen, kleinsten Informationseinheiten, den „Bits“.

Wesentliche Vorteile des Computers sind die hohe Geschwindigkeit, in der Daten verarbeitet werden können, und der große Genauigkeitsgrad der Datenerfassung. Zum Abschluß der kurzen Vorbemerkungen noch ein Hinweis zur Kapazität der Saarbrücker Rechananlage: Die *Electrologica X 1* (Aufstellungsjahr 1963) besitzt einen Kernspeicher mit 32.000 (davon 20.000 frei verfügbaren) Speicherzellen zu je 27 Bits, einen Lochstreifenleser (maximale Einlesegeschwindigkeit: 1000 Symbole/sec) und einen Lochstreifenstanzer, einen Lochkartenleser und -Stanzer (zwei Karten/sec), drei Magnetbaineinheiten (Kapazität: 3,6 Millionen X 1-Worte), eine elektrische Schreibmaschine (zehn Zeichen/sec) und einen Schnelldrucker (10 Druckzeilen/sec; eine Druckzeile = 120 Symbole).

Ein Nachteil dieser Rechananlage ist die fehlende Kompatibilität, das heißt fehlende Austauschbarkeit mit anderen Rechnersystemen. Es ist also nicht möglich, Programme von oder auf anderen Rechananlagen zu verwenden. Darüber hinaus sind auch die Magnetbänder speziell für die *X 1* entwickelt und die darauf befindlichen Informationen (Text, Wortindices) anderen Maschinen ebenfalls nicht direkt zugänglich.

Bei der Frage der zu verwendenden Programmiersprache (es stehen die problemorientierte Sprache *Algol* und eine Assemblersprache, der *Kieler Code*, zur Verfügung) haben wir uns für den *Kieler Code*, eine speziell für die *EL X 1* entwickelte Maschinensprache, entschieden, da *Algol* im wesentlichen für die Bearbeitung numerischer Probleme entworfen wurde und uns für die linguistische Datenverarbeitung als wenig geeignet erscheint.

Bisher sind drei größere Programme zur maschinellen Verarbeitung des Phonetextes erstellt. Sie haben jeweils verschiedene Funktionen und sind nach ihrer wichtigsten bezeichnet als ‚Texteingabe‘-, ‚Lautwörterbuch‘- und ‚Lautindex‘-(=Textausgabe-)Programm <sup>21)</sup>.

## 8.2 Das Texteingabe-Programm

Das Texteingabe-Programm liest die Lochstreifendaten in den Kernspeicher ein und speichert sie in umkodierter Form auf Magnetband, das der weiteren Textbearbeitung auf Grund seiner schnelleren Verarbeitungsmöglichkeit als eigentlicher Informationsträger zugrunde liegt. Für jede Wortform und die zugehörigen Informationen ist dabei eine gleich große („feste“) Anzahl von Speicherzellen verwendet. Diese Einheit gliedert sich folgendermaßen auf:

<sup>21)</sup> Außer diesen Programmen zu einem deutschen Phonetindex wurden in Saarbrücken inzwischen Indexprogramme für folgende Alphabete ausgearbeitet:

Lateinisches Alphabet (Graphemindex)

Mittelbulgarisches Alphabet (Index zu Pseudo-Kaisarios)

Bulgarisch-Russisches (kyrillisches) Alphabet (in Technotran-Umschrift)

Esperanto-Alphabet.

Zelle (= 27 Bits)	Inhalt	Bitzahl	Maximal speicherfähige Dezimalzahl
0—6	Phoneme	je 4 x 6	= 28 Phonemwerte, maximaler dezimaler Phonemwert = 63
7	Gedicht-Nr.	10	1023
	Strophe	6	63
	absolute Wort-Nr.	11	2047
8	Vers-Nr.	9	511
	Seite (Text)	11	2047
	abweichende Betonungs-Nr. des betonten Phonems im Wort	5	31
	Trennungszeichen	1	1
	Abkürzung	1	1
	Abschnitt	5	31
9	Satzzeichen	3 x 5	je 31

Das Eingabeprogramm dient zugleich der Kontrolle der Informationen, da die Daten während des Einlesens über Drucker ausgegeben werden können. Dieser Ausdruck ist verhältnismäßig leicht auf fehlerhafte Angaben hin zu überprüfen<sup>22)</sup>.

### 8.3 Das Lautwörterbuch-Programm

Hauptaufgabe dieses Programms ist die Sortierung der Wortformen in der Reihenfolge des Lautalphabets (siehe oben, S. 130). Auf Grund der begrenzten Kapazität der Rechanlage können maximal 1300 Wortformen in einem Vorgang bearbeitet werden. Sortiert wird dabei im Kernspeicher ohne Zuhilfenahme von Magnetbändern, da ein geeignetes Bandsortierprogramm nicht zur Verfügung steht<sup>23)</sup>. Die Informationen zu jeder Wortform bleiben dabei erhalten. Bei einer größeren Gesamtsortiermenge werden die zu sortierenden Wörter über einen Filter (Vergleich des ersten oder der ersten beiden Phoneme) für den jeweiligen Sortiervorgang ausgewählt. Die Ergebnisse der Sortierläufe werden anschließend wieder auf Magnetband gespeichert. Weitere Auswahlmöglichkeiten — auch in Kombination untereinander — sind:

- a) Beschränkung auf Wortformen, die bestimmte (in einer Tabelle anzugebende) Phoneme enthalten.
- b) Bearbeitung des letzten Wortes einer Verszeile.
- c) Auswahl von Gedichten, deren (interne) Laufnummern auf Lochstreifen angegeben werden.

<sup>22)</sup> Die Daten selbst können auf beliebig vielen Lochstreifen eingegeben werden, da ein besonderes Zeichen das Ende einer Eingabemenge zeigt. Als „Stopzeichen“ ist die Basiszahl "10" verwendet.

<sup>23)</sup> Es ist damit zu rechnen, daß Ende dieses Jahres der Universität des Saarlandes eine neue Rechanlage mit ungleich größerer Leistungsfähigkeit zur Verfügung stehen wird, so daß dann alle technischen Probleme zufriedenstellend gelöst werden können.

Darüber hinaus kann dieses Programm zur Herstellung eines Häufigkeitswörterbuchs (der Wortformen) verwendet werden. Zunächst wird dabei die Häufigkeit des Vorkommens gleicher Wortformen festgehalten, anschließend diese in alphabetischer Reihenfolge stehenden Wörter nach der Häufigkeit sortiert. Schließlich ist es möglich, die Wortformen vom letzten zum ersten Phonem hin, also „rückläufig“, zu sortieren. Darauf werden wir bei der Betrachtung des Reimwörterbuchs eingehen. Parallel zu diesen Verarbeitungsmöglichkeiten kann zusätzlich eine statistische Untersuchung durchgeführt werden. Gezählt werden dabei die bearbeiteten Wortformen, die Phoneme insgesamt und die Anzahl der einzelnen Phoneme. Das Ergebnis wird auf Lochstreifen ausgestanzt.

#### 8.4 Das „Lautindex“-(Textausgabe-)Programm

Neben der Übernahme einer Reihe von Funktionen des Lautwörterbuch-Programms — vor allem der Auswahlmöglichkeiten — bestimmt sich die Hauptaufgabe dieses Programms in der Ausgabe der einzelnen Ergebnisse. Unter dem eigentlichen „Index“ wird dabei die Ausgabe der alphabetisch sortierten Wortformen mit einfachem Belegstellennachweis (Angabe von Seite und Verszeile), dem Hinweis auf Betonungswechsel und der Angabe der Beleganzahl zu einer Wortform verstanden. Daneben wird dieses Programm zur Ausgabe des Häufigkeitswörterbuchs (nur Angaben zum Wortlaut und zur Häufigkeit der Wortformen) verwendet. Wie bei den Programmen „Texteingabe“ und „Lautwörterbuch“ kann auch hier eine Text- oder Wörterbuchausgabe mit sämtlichen gespeicherten Informationen zum Wort (Gedichtnummer, Verszeile, Strophe, Wortnummer, Abschnitt, Betonungswechsel, Satzzeichen nach dem Wort) erfolgen. Darüber hinaus dient es zur Ausgabe des rückläufigen Wörterbuchs (einschließlich aller Informationen).

## 9 Erste Ergebnisse

Die Möglichkeiten, die sich bereits für eine Auswertung des Textmaterials auf der Grundlage der beschriebenen Programme ergeben, sind recht vielfältig. Wir haben nach Abschluß der ersten Kontrollen bereits einige Untersuchungen durchgeführt, deren erste Ergebnisse hier in Auszügen mitgeteilt werden sollen.

### 9.1 Lautindex

Die Ausgabe des Index, von dem wir eine Probeseite anfügen, umfaßt insgesamt 52 Seiten. Die letzten beiden Ziffern der Belegzahl bezeichnen die (absolute) Verszeile des Gedichts, die Ziffern davor beziehen sich auf die Seite der zugrunde gelegten Ausgabe. Ein \* nach der Belegzahl gibt an, daß dieses Wort nicht auf der ersten Silbe betont ist. Ausnahme: der Vokal der ersten Silbe ist ein [ə] (Schriftbild: C):

Normalfall:

AIGNC 7212 (= Seite 72 Zeile 12)

Beispiel für Betonungswechsel mit Kennzeichnung:

ALTAE:RC 19315\* (= Seite 193 Zeile 15)

Beispiel für Betonungswechsel ohne Kennzeichnung:

BCMA:LT 3612 (= Seite 36 Zeile 12)

## 9.2 Reimwörterbuch

Durch die Kombination einer Gedichtauswahl (=alle reimenden Gedichte) mit der Bearbeitung nur des letzten Wortes einer Verszeile (=Reimwort) und einer anschließenden rückläufigen Sortierung war es möglich, automatisch ein Reimwörterbuch zu Trakls Lyrik herzustellen; es umfaßt 28 Seiten, eine davon ist als Probe beigegeben. Die Abkürzungen der Überschrift bedeuten:

GED	Gedichtnummer	AB	Abschnitt
ST	Strophe	BE	Betonung (bei Wechsel)
AWN	Wortnummer	T	Trennungszeichen im Wort
VSN	Versnummer	A	Abkürzung, falls Wort länger als 28 Phoneme
SEIT	Seite der Ausgabe		
ZEICHEN	Satzzeichen (Art)		

## 9.3 Lautübersicht <sup>24)</sup>

Gesamtzahl der Wortformen: 12.179

Gesamtzahl der Phoneme: 59.216

Anzahl der Phoneme/Wortform: 4,86

Phonem	Belegzahl (absolut)	Beleganteil (%)	Meier (%)
[a]	3821	6,45	6,19
[a:]	725	1,22	1,74
[b]	1104	1,86	1,84
[d]	2750	4,64	4,71
[e]	20	0,03	0,11
[e:]	1075	1,82	2,39
[ɛ]	1527	2,58	2,61
[ɛ:]	120	0,20	0,27
[ə]	5975	10,09	9,64
[f]	1489	2,51	2,36
[g]	1269	2,14	2,20
[h]	612	1,03	1,08
[i]	3583	6,05	6,73

<sup>24)</sup> Zum Vergleich geben wir die durchschnittliche Häufigkeit des betreffenden Phonems nach Meier an. Vgl.: H. MEIER, Deutsche Sprachstatistik, Hildesheim 1967, 2. Aufl., S. 250/1; die hier auf zwei Stellen abgerundeten Werte beziehen sich auf Prosatexte mit insgesamt 50.000 Lauten.

Phonem	Belegzahl (absolut)	Beleganteil (%)	Meier (%)
[i:]	991	1,67	2,37
[k]	1112	1,88	1,59
[l]	2816	4,76	3,49
[m]	1639	2,77	2,52
[n]	6263	10,58	10,18
[ŋ]	540	0,91	0,69
[o]	13	0,02	0,11
[o:]	571	0,96	0,94
[ɔ]	900	1,52	1,52
[œ]	79	0,13	0,12
[ø]	162	0,27	0,24
[p]	515	0,87	0,76
[r]	4703	7,94	7,50
[s]	2310	3,90	4,63
[ʃ]	1411	2,38	1,36
[z]	1008	1,70	2,27
[t]	5113	8,63	8,17
[u]	1929	3,26	2,86
[u:]	288	0,49	0,95
[y]	425	0,72	0,64
[y:]	352	0,59	0,37
[v]	1085	1,83	1,94
[x]	281	0,46	0,66
[ç]	640	1,08	2,23

#### 9.4 Lautverteilung nach Gedichtzyklen

Einen ersten Einblick in die Untersuchungsmöglichkeiten gibt eine Übersicht über den unterschiedlichen prozentualen Anteil einzelner Phoneme am Lautbestand in verschiedenen Gedichtzyklen. Zu diesem Zweck wurde der Lyrikband in die einzelnen Zyklen zerlegt, die dann jeweils lautstatistisch untersucht wurden:

##### Abkürzungen:

W = Gesamtzahl der untersuchten Wörter im Zyklus; P = Gesamtzahl der Laute im Zyklus; Q = Durchschnittliche Anzahl der Phoneme je Wort im Zyklus.

Zyklus	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
W	652	1305	794	467	1380	777	653	1924	748	1648	1407	427
P	3103	6134	3536	2315	6961	3751	3114	9243	3535	8065	7306	2153
Q	4,76	4,70	4,45	4,96	5,04	4,83	4,77	4,80	4,73	4,89	5,19	5,04
a	7,09	7,17	6,67	5,96	6,92	6,96	6,36	6,51	6,96	5,93	5,54	5,25
a:	1,06	1,24	0,99	1,38	1,11	1,33	1,25	1,18	1,19	1,18	1,41	1,49
b	1,93	1,71	1,61	2,07	1,77	2,19	1,51	1,86	2,52	1,85	1,79	1,95
d	3,51	4,42	3,76	3,41	3,74	5,04	5,01	5,13	3,88	5,41	5,30	5,53
e	—	0,03	0,03	0,13	0,04	0,05	—	0,03	0,06	0,03	—	—

Zyklus I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	
e:	1,71	1,74	1,67	1,77	1,65	2,11	2,31	1,87	1,84	1,89	1,71	1,81
ε	2,47	2,30	1,90	2,59	2,67	2,48	2,76	2,60	2,43	2,82	2,83	2,69
ε:	0,16	0,23	0,37	0,30	0,26	0,11	0,23	0,16	0,17	0,20	0,15	0,19
ə	9,99	8,67	8,57	10,07	9,50	9,57	9,99	10,62	9,48	11,12	11,16	10,87
f	2,61	2,90	3,03	2,68	2,80	2,61	2,34	2,49	2,69	2,03	2,30	1,77
g	2,13	2,10	1,78	2,38	2,17	1,84	2,34	1,83	2,15	2,41	2,35	2,56
ɜ	—	—	—	—	—	—	—	—	—	—	—	—
h	1,03	1,03	1,41	0,86	1,12	0,88	1,06	1,09	0,93	0,82	1,18	0,74
i	6,48	6,65	6,25	5,96	6,49	5,89	6,49	6,33	5,80	5,82	5,27	4,46
ir	1,29	1,64	2,88	1,56	1,51	1,87	1,86	1,75	1,39	1,69	1,33	1,72
k	1,81	1,86	2,15	1,86	1,82	1,95	1,51	2,00	1,92	1,81	1,83	2,00
l	4,90	4,84	4,13	5,23	4,28	4,32	3,60	4,89	5,71	4,79	5,15	5,20
m	3,06	2,82	3,54	3,07	2,72	2,35	2,44	2,87	2,01	2,54	3,11	2,51
n	12,38	10,68	11,11	10,71	11,28	9,28	11,72	10,55	10,84	10,12	9,53	9,99
ŋ	0,90	0,77	0,76	0,82	0,78	0,85	1,03	1,06	0,88	0,98	0,94	1,12
o	0,03	0,02	—	—	0,03	—	0,10	0,05	—	—	—	—
o:	0,94	0,86	0,99	1,04	0,75	0,93	0,96	0,99	0,59	1,09	1,27	0,98
ɔ	1,93	1,63	1,64	1,73	1,60	1,44	1,32	1,33	1,81	1,18	1,60	1,72
œ	0,08	0,11	0,06	0,17	0,17	0,19	0,06	0,12	0,06	0,14	0,18	0,23
ø	0,26	0,29	0,34	0,17	0,17	0,40	0,35	0,20	0,37	0,29	0,32	0,23
p	0,74	0,67	0,54	0,91	1,02	1,07	0,96	1,04	0,71	0,86	0,84	0,84
r	7,67	7,55	7,72	7,56	8,10	8,50	7,55	7,98	7,78	7,91	8,21	8,45
s	3,22	3,52	2,63	4,10	3,70	3,92	4,59	4,25	3,73	4,17	4,20	4,23
f	2,26	2,02	2,46	1,94	2,63	2,67	2,15	2,13	2,49	2,48	2,61	2,74
z	1,26	1,74	1,98	1,90	1,60	1,60	2,47	1,98	1,47	1,76	1,40	0,93
t	8,96	9,42	9,73	8,90	8,92	8,58	7,77	7,56	9,31	8,47	8,54	8,64
u	3,19	3,31	3,87	3,28	3,49	3,41	2,95	3,08	3,90	3,08	2,82	3,53
u:	0,48	0,54	0,34	0,43	0,43	0,51	0,30	0,60	0,45	0,45	0,56	0,46
y	0,61	0,72	0,93	1,04	0,91	0,75	0,64	0,63	0,65	0,62	0,68	0,60
y:	0,52	0,62	0,48	0,52	0,42	0,45	0,45	0,52	0,74	0,78	0,75	0,79
v	1,81	2,12	2,01	1,86	1,64	1,95	1,99	1,59	1,87	1,90	1,68	2,18
x	0,52	0,75	0,40	0,30	0,43	0,64	0,51	0,40	0,37	0,31	0,48	0,79
ç	1,06	1,30	1,27	1,34	1,37	1,33	1,06	0,74	0,88	1,09	0,97	0,84

Die Titel der Zyklen:

I = Die schöne Stadt; II = Kleines Konzert; III = Die Bauern; IV = Traum des Bösen; V = De Profundis; VI = Im Dorf; VII = Helian; VIII = Sebastian im Traum; IX = Der Herbst des Einsamen; X = Siebengesang des Todes<sup>25</sup>); XI = Gesang des Abgeschiedenen; XII = Offenbarung und Untergang<sup>26</sup>).

Wenn wir hier gleichsam als Modell eine erste Analysetabelle aufzeigen, so soll zunächst nur auf den Weg hingewiesen werden, auf dem wir zu konkreten Aussagen über unseren Test zu gelangen hoffen. Diese Tabelle kann derartige Aussagen allein nicht begründen, sondern liefert einige erste Ansätze und zeigt Probleme auf, die zu bearbeiten sind.

<sup>25</sup>) Ohne 'Die Verwandlung des Bösen' und 'Winternacht' (Prosa).

<sup>26</sup>) Ohne 'Offenbarung und Untergang' (Prosa).

Doch schon hierbei lassen sich einige Tendenzen feststellen. So kann man — um nur einige Beispiele zu nennen — beim Vokal [a] insgesamt gesehen eine Neigung zu zunehmend seltenerer Verwendung ablesen (Rückgang von einem Anteil von 7,09 Prozent auf 5,25 Prozent — eigentliche Ausnahme nur Zyklus IV). Einen ähnlichen Entwicklungsprozeß — wenn auch nicht durchgehend — scheint der Vokal [i] zu durchlaufen (Rückgang von 6,48 Prozent in Zyklus I auf 4,46 Prozent in Zyklus XII). Es wäre daneben auch zu untersuchen, welche Lautverteilungen von der Wahl der lexikalischen Elemente abhängen. So läßt sich die starke Zunahme des Phonems D und die gleichzeitige Abnahme von A sehr leicht aus der — gleichfalls statistisch zu belegenden — Neigung Trakls erklären, in seinen späten Gedichten zusehends den bestimmten Artikel (phonemisch: DE:R, DI:, DAS usw.) an Stelle des unbestimmten (phonemisch: AIN, AINC, AINCR usw.) zu setzen. Möglicherweise lassen sich auch die übrigen nach dem Augenschein signifikanten Tendenzen in der Lautentwicklung so erklären; wir wollen jedoch hier nichts präjudizieren, sondern abwarten, was sich aus den Untersuchungen, so wie sie in den einleitenden Abschnitten skizziert worden sind, ergibt.

## 10 Index

Neben dem Text der Gedichte in phonemischer Umschrift haben wir Trakls gesamte Werke, so wie sie vor der Publikation der kritischen Ausgabe vorlagen<sup>27)</sup>, in normaler Orthographie auf Lochstreifen aufgenommen. Von diesem Text wurde dann ein Index hergestellt, der ursprünglich in der von Hans Schwerte und Helmut Schanze herausgegebenen Reihe 'Indices zur neueren deutschen Literatur' (Frankfurt 1968 ff.) erscheinen sollte. Durch das Erscheinen der kritischen Ausgabe<sup>28)</sup> hat sich dieser Plan überholt. Wir haben statt dessen ein Verfahren ausgearbeitet, einen Index herzustellen, der nicht nur den Text, sondern auch sämtliche Varianten zu einer Textstelle erfaßt. Er wird nach Fertigstellung in der genannten Reihe erscheinen.

<sup>27)</sup> Das sind: Die Dichtungen, Salzburg (Müller) o. J., 12. Aufl., sowie Aus goldenem Kelch, Salzburg (Müller) o. J., 6. Aufl.

<sup>28)</sup> Georg Trakl, Dichtungen und Briefe, hrsg. von W. Killy und H. Szklenar, Salzburg 1969.

	1402*	1416*	1604*	1613*	HIN				1412
6	12009*					2116	2422	5827	8003
HAERAINBRICHT				17515*		8014	8127	8134	9209
HAERAP				5822*		9508	10024	10916	10920
2	17637*					17	13302	16626	17910
HAERAUF				3704*	HINAIN				1624*
2	5818*				HINANSCHTI:K				10027*
HAERBCN				1916	HINAP				2915*
HAERFOIR				5512*		6213*	8360*	10145*	12805*
3	7636*	16818*				9	12906*	13024*	13410*
HAERLICH				7105	HINAPSCHTI:K				9910*
2	7106					3	9920*	10901*	
HAERN				6727	HINAPZINJKT				10701*
HAERP				2223	HINAUFSCHTAIKT				10809*
HAERPST				4903	HINFLI:ST				6614
	7301	8012	9513	9610	HINGANJS				8378
	9812	10412	10801	10914	HINGEIT				17014
	11504	11512	11602	11701	HINGGSCHTRAECT				4049
14	16408				HINLAUSCHT				17219
HAERPSTC				7202	HINRQLT				19304
2	13312				HINSCHMILTST				8366
HAERPSTCS				6605	HINSCHTAERBCNT				13408
	10153	10206	12904	13105		3	13906	16523	
8	17117	17805	19314		HINZINJKT				7306
HAERPSTCSNAXT				17719		3	12811	13405	
HAERPSTGCBRQYNTCS				6102	HINTCR				4606

WORTLAUT (UMSCHRIFT)	GED	ST	AWN	VSN	SEIT	AB	BE	T	A	ZEICHEN
WAICRSCHPI:GCL	71	2	63	9	117	1				3
WAICRSCHPI:GCL	72	1	12	3	118	1				2
TSWILINJSSCHPI:GCL	40	1	20	3	66	1				
HY:GCL	3	2	32	6	11	1				1
HY:GCL	6	3	50	11	15	1				
HY:GCL	29	3	51	11	50	1				
HY:GCL	71	2	51	7	117	1				3
HY:GCL	72	1	8	2	118	1				
FAELZCNHY:GCL	40	1	15	2	66	1				
FLY:GCL	29	3	46	10	50	1				1
FLY:GCL	71	2	76	11	117	1				
DUNJKCL	23	6	109	21	39	2				
DUNJKCL	23	6	120	23	39	2				1
DUNJKCL	23	17	345	65	41	6				1
DUNJKCL	23	17	357	67	41	6				1
DUNJKCL	70	4	73	16	116	1				1
DUNJKCL	107	1	5	1	185	1				2
GCFUNJKCL	70	4	58	13	116	1				
SCHTAERNGCFUNJKCL	107	1	18	4	185	1				1
GBIMCL	23	14	300	56	40	5				1
HIMCL	23	14	290	54	40	5				1
TYMPCL	23	9	171	33	39	3				
TYMPCL	23	9	183	35	39	3				
SCHPI:L	107	2	32	7	185	1				2
GLQKNSCHPI:L	68	7	171	25	114	3				1
FQL	22	3	63	11	37	1				1
TQL	22	3	52	9	37	1				