

Über die Erstellung eines Synoptischen Affinitäts-Lexikons (SAL) Kurzbericht

Im folgenden soll über ein Projekt berichtet werden, das z.Z. in der Forschungsgruppe für mathematisch-statistische Linguistik am 1. Physikalischen Institut der RWTH Aachen von Dipl.-Phys. H.-M. Dannhauer, Dr. phil. B. Rieger und Dr. rer. nat. D. Wickmann bearbeitet wird.

Das Vorhaben geht auf eine umfänglichere literar-historische Untersuchung deutscher studentischer Lyrik des 19. und 20. Jahrhunderts ¹ zurück, in der u.a. versucht worden war, die in den Gedichten sprachmaterialiter feststellbaren und semantisch bestimmten Textelemente (Wort- bzw. Motivkreise) quantitativ zu beschreiben und zu analysieren. Neben den Fragen nach *Häufigkeiten* und *Verteilungen* der bevorzugten Sprachmaterialien dieser Gedichte aus rund 150 Jahren trat die Untersuchung häufigster *Zuordnungsbeziehungen* verschiedener Wort- bzw. Motivkreise innerhalb dieser Texte in den Vordergrund: die Frage also nach Anzahl und Art regelhafter Kopplungen von Textelementen.

Dieser Ansatz wurde inzwischen unter Verwendung des gleichen Textmaterials weiterverfolgt, um einmal auf breiterer Grundlage ausgearbeitet zu werden. Die Zielvorstellung des ganzen Projekts besteht dabei darin, die in großen Textmassen generell vorfindlichen und konkret feststellbaren Zuordnungsbeziehungen exakt zu erfassen, die aus ihnen analysierbaren mittleren Bindungsintensitäten zwischen Wörtern, Begriffen, Motiven etc. zu ermitteln und diese kontext-abhängigen Werte dazu zu benutzen, sogenannte (semantische) Umgebungsfelder zu erstellen. Diese auch aus dem Schrifttum historisch zurückliegender Epochen rekonstruierbaren Umgebungsfelder bestimmter Wörter lexikographisch zugänglich zu machen und zwar so, daß sich Wandel oder Konstanz im Umfeld eines Begriffs durch die diachrone Gegenüberstellung synchroner Analyseergebnisse leicht überblicken läßt, ist der Sinn eines 'Synoptischen Affinitäts-Lexikons'.

Der zu dieser Tagung gerade noch rechtzeitig fertiggestellte Probe-Ausdruck ist freilich in der hier vorliegenden Form nicht zur Veröffentlichung bestimmt; dieses 650 Seiten umfassende Lexikon stellt vielmehr ein Zwischenergebnis der bisherigen Arbeit dar und möchte als Versuch verstanden sein, für einen vorerst noch relativ eng begrenzten Teilbereich des Schrifttums einer Zeit (studentische Lyrik) durch Gruppenmittelbildung die je zeitabhängig sich verändernden Normen des durchschnittlich Erwartbaren auf der Ebene semantischer Zuordnungen in lyrischen Texten zu bestimmen und überschaubar zu machen.

Das Ausgangsmaterial bilden rund 3000 Gedichte, deren entstehungszeitliche Häufungen die Aufteilung in drei sogenannte Publikations-Gruppen (*Publ-Gruppen*) nahelegte. Ein sprachmaterialer Merkmals-Katalog *K*, der rund 300 Wort- bzw. Motivkreise (*Womos*)*k* umfaßt, stellt gleichsam das auf häufigste, semantisch bestimmte Textelemente reduzierte Alphabet dar, aufgrund dessen jeder Gedichttext in jeder *Publ-Gruppe* verschlüsselt wurde. Der so erstellte Datensatz, der damit die an den Gedichttexten als Protokollbefund gewonnenen Informationen für die Weiterverarbeitung liefert, umfaßt demnach drei, den

¹B. Rieger: *Poetae Studiosi. Analysen studentischer Lyrik des 19. und 20. Jahrhunderts*. Ein Beitrag zur exaktwissenschaftlichen Erforschung literarischer Massenphänomene. Frankfurt/M.: Thesen 1970.

Publ-Gruppen entsprechende Teile.

Jede der drei *Publ-Gruppen* besteht aus einer bestimmten Anzahl I Texten, deren laufende Nummer mit i bezeichnet sei:

$$1 \leq i \leq I$$

Jedem einzelnen Gedicht i entspricht dabei eine längere oder kürzere Kette n der in der Reihenfolge ihres Auftretens erfaßten *Womos*. Danach ergibt sich der Umfang (die Länge) N einer *Publ-Gruppe* als Summe aller Gedichtlängen n , gemessen in der Anzahl sämtlicher in ihnen auftretender *Womos*:

$$N = \sum_{i=1}^I n_i$$

Den Umfang und die Breite des gesamten so aufbereiteten und hier herangezogenen Ausgangsmaterials zeigt folgende Tabelle. Sie läßt erkennen, wie die insgesamt $N_{Ges} = 96000$ Belegungen der rund $K = 300$ *Womos* in sämtlichen $I_{Ges} = 3000$ Texten sich auf die *Publ-Gruppen* 1, 2 und 3 verteilen, wobei $\bar{n}$ die mittlere Anzahl der in einer *Publ-Gruppe* erfaßten *Womos* pro Gedichttext angibt:

Publ-Gruppe	N	I	$\bar{n}$
1. (1822-1843)	21000	550	38
2. (1863-1909)	50000	1300	38
3. (1921-1963)	25000	1100	23
Gesamt:	96000	2950	33

Die Frage nach den in dieser Menge von Texten vorherrschenden Zuordnungsbeziehungen versucht nun zweierlei zu erhellen: einmal sollen regelhafte, d.h. über ein mit dem Zufall erklärbares Maß hinausgehende gleichzeitige Belegungen verschiedener *Womos* innerhalb der durch die einzelnen Gedichte gegebenen Texteinheiten jeder *Publ-Gruppe* aufgefunden und beschrieben werden; zum anderen sollen die daraus ermittelbaren Vorkommenswahrscheinlichkeiten zur Berechnung eines Maßes dienen, welches die Intensität einer möglicherweise vorhandenen *Affinität* (Bindung) bzw. *Repugnanz* (Abstoßung) zwischen *Womos* zu bestimmen gestattet.

Da letzteres die Grundlage und Voraussetzung bildet für die Erstellung von Umgebungsfeldern einzelner Wörter, Begriffe, Motive etc., mag folgendes Modell den methodischen Angang veranschaulichen:

Man stelle sich eine Fläche vor, auf der durch die Aufteilung in einzelne Fächer insgesamt I Kästchen oder Zellen entstanden sind, wobei jede dieser Zellen einen leeren Gedichttext repräsentiere. Über diese Zellen werden nun Kugeln verschiedener (nämlich $K = 300$) Typen regellos verstreut, wobei jede Kugel einem Womo eines bestimmten Typs ($k, k', k'', \text{etc.}$) entspreche. – Setzt man einmal voraus, daß tatsächlich zwischen den Kugeln der Typen k und k' eine Bindung oder Anziehung irgendwelchen Art besteht, dann ist die Wahrscheinlichkeit dafür, dass beide jeweils zusammen in eine Zelle fallen, sicherlich größer als wären sie nicht gekoppelt. In Umkehrung dieser Überlegung würde eine Überprüfung der Zelleninhalte demnach aufgrund der so feststellbaren Häufigkeiten gleichzeitiger Vorkommen

von k und k' die Grundlage bilden können für die Berechnung ihrer Wahrscheinlichkeit. Erst daraus kann auf die Intensität einer möglicherweise vorhandenen Bindung (Affinität) oder Abstoßung (Repugnanz) zwischen k und k' formal geschlossen werden. – Genau dies vermag ein Koeffizient zu leisten, der in Anlehnung an einen in der Statistik gebräuchlichen Korrelations-Koeffizienten für die hier vorliegende besondere Fragestellung entwickelt wurde:

Der sogenannte *A f f i n i t ä t s - K o e f f i z i e n t* (Affko) $s_{kk'}$ gibt dabei an, mit welchen Intensitätsgraden welche *Womos* einander in den analysierten Texten vornehmlich zugeordnet sind. Dabei errechnet sich die Affinität bzw. Repugnanz zwischen zwei *Womos* k und k' nach folgender Formel:

$$s_{kk'} = \frac{\sum_{i=1}^I (w_{ki} - w_{ki*}) \cdot (w_{k'i} - w_{k'i*})}{\sqrt{\sum_{i=1}^I (w_{ki} - w_{ki*})^2 \cdot \sum_{i=1}^I (w_{k'i} - w_{k'i*})^2}}; -1 \leq s_{kk'}, \leq +1$$

Darin bedeuten:

w_{ki} bzw. $w_{k'i}$ die absoluten Häufigkeiten des *Womos* k bzw. k' im i -tem Text einer *Publ-Gruppe*;

w_{ki*} bzw. $w_{k'i*}$ errechnen sich wie folgt:

$$w_{ki*} = \frac{N_k}{N} \cdot n_j \text{ bzw. } w_{k'i*} = \frac{N_{k'}}{N} \cdot n_i,$$

wobei

N_k bzw. $N_{k'}$, die absoluten Häufigkeiten der *Womos* k bzw. k' in einer *Publ-Gruppe* sind.

Analog zu $s_{kk'}$ lassen sich nun zu einem abfragbaren und jeweils festzuhaltenden *Womo* k , das dann als Start-Motiv (*Stamo*) bezeichnet wird, die Affko-Werte aller übrigen *Womos* k' , die dann als Ziel-Motive (*Zimos*) bezeichnet werden, berechnen. Aus den so ermittelten Index-Werten, die alle zwischen -1 und $+I$ liegen, wurden der Konzentration auf wichtigste Beziehungen halber aus den insgesamt jeweils 300 die 50 absolut größten Werte herausgegriffen, danach erst nach abnehmender Größe im positiven und negativen Bereich geordnet, mit Rangziffern versehen und in Listenform unter dem abgefragten *Stamo* ausgedruckt. Diese jeweils 50 *Zimos* biden das, was im folgenden das kontextuelle Umgebungsfeld eines *Stamos* genannt wird². Für jedes abgefragte *Stamo* lassen sich auf diese Weise drei, den drei *Publ-Gruppen* entsprechende Umgebungsfelder errechnen, die in synoptischem Parallel-Druck nebeneinander Konstanz und/oder Veränderung in der Zusammensetzung der Umgebungen über rund 150 Jahre hinweg leicht erkennen lassen.

In Abb. 1 und Abb. 2 sind zur Verdeutlichung zwei gegenüberliegende Seiten im Affinitäts-Lexikon abgedruckt, die beide das Umgebungsfeld des *Stamo* STRASSE wiedergeben. Abb. 1 führt in jeder der drei, den drei *Publ-Gruppen* entsprechenden Kolonnen die das Umgebungsfeld bildenden *Zimos* auf, jedoch in alphabetischer Reihenfolge. Dadurch wird

²Diese Anordnung der $s_{kk'}$ erlaubt in der Tendenz die Interpretation von abnehmender Affinität und zunehmender Repugnanz, wobei die ermittelten Index-Werte einzig zur Zuweisung der Rangziffern, bislang aber noch nicht zur Grundlage eines diachronen Vergleichs dienen können. Voraussetzung dafür wären Signifikanz-Tests, die bisher noch nicht vorgenommen werden konnten, obwohl die theoretische Verteilung von s inzwischen berechnet wurde. Eine eingehende Untersuchung der mathematisch-statistischen Problematik wird z.Z. von B.H.D. Dangerman ausgearbeitet und wird zusammen mit dem Lexikon (SAL) voraussichtlich 1972 erscheinen.

STASSE

PUBL.-GRUPPE 1 (1822-1843)		PUBL.-GRUPPE 2 (1863-1909)		PUBL.-GRUPPE 3 (1921-1963)	
ABEND	.054	ABSCHIED	.063	ALL	.057
ALT ER	.063	AB GRUND	.058	ASPHALT	.177
BANDE	.054	ALT ER	.072	BAHN	.153
BANN	.052	BANN	.069	BETT	.175
BILD	.054	BERGGEBIRG	.059	BLUETE	.062
BLAU	.058	CHOR AL	.106	DACH	.093
BLICK EN	.056	DACH	.063	DUMPFG	.079
BORN	.054	DOM KATHD	.068	ERDE	.058
DACH	.094	DUNKEL N	.075	EWIG KEIT	.074
ELEND NOT	.107	ERDE	.080	FENSTER	.109
FENSTER	.166	EWIGKEIT	.056	FEUER	.095
FEST LICH	.077	FENSTER	.107	FLUSS	.059
FEUER	.085	FREMD E	.062	FRAU WEIB	.062
FRAGE	.208	FRIEDE N	.075	FUSS	.169
FREUND SCH-	.054	GELB	.089	GRAU	.125
FUSS	.105	GESICHT	.083	HAUCH	.065
GESICHT	.095	GLANZ	.057	HAUS	.188
GLOCKE	.059	GLOCKE	.063	HIMMEL	.089
GOLD	.064	GLUT	.060	HOFFNUNG	.072
GRABGRUFT	.064	HAUCH	.056	INNEN IG	.069
GRAU	.056	HAUS	.116	INSEL	.058
HEISS	.069	HIMMEL	.054	KIRCHE	.216
HORIZONT	.081	HIRTE	.065	KREUZ	.090
HORN TROMP	.577	HOHL	.056	KUEHL E	.071
INTERJ	.107	KIRCHE	.073	LEBEN	.104
JUBELJAUCH	.090	KLINGEN	.071	LIEBEMINNE	.070
KALT E	.059	KRANK H	.062	MAEDCHEN	.062
KIRCHE	.075	LERCHE	.058	MUED IGKT	.093
KLINGEN	.386	MILD LAU	.074	NASS E	.065
KOPFHAUPT	.096	NATUR	.063	NEON	.095
LEBEN	.059	NEBEL IG	.058	QUELLE N	.073
LUFT	.076	OED E	.155	REGEN NEN	.067
MOND	.058	RAUCH	.131	SCHATTEN	.062
MORGEN	.076	RAUM	.075	SCHMERZ EN	.059
MUED IGKT	.092	SCHLAF EN	.073	SCHREI EN	.089
NACHT	.061	SCHRITT	.133	SCHWEIGEM	.078
RAUSCHEN	.057	SEELE	.059	SEE	.062
ROTPURPUR	.063	SEE	.059	STADT	.097
SACHT	.058	SEHNSUCHT	.068	STEIN	.074
SCHMERZ EN	.070	SONNE	.059	STILL E	.067
SCHWEBEN	.059	STADT	.401	STROM EN	.068
SCHWER	.107	TRAUER N	.062	TIEF	.072
SEHNSUCHT	.057	TRAUM	.060	TRAUM	.070
STUMM	.171	WAGEN	.189	WAGEN	.145
TAU	.059	WEG PFAD	.080	WALD TANN	.068
UFER	.052	WIESE AUE	.083	WEICH	.071
WAGEN	.224	WILD	.080	WELT	.090
WANDER-SCH	.132	WOLKE	.072	WUND E	.078
WELLE	.073	WUNSCH	.083	ZIGARETTE	.061
WIND	.066	ZWEIG AST	.066	ZUNGE	.075

Abb. 1

STASSE

PUBL.-GRUPPE 1			PUBL.-GRUPPE 2			PUBL.-GRUPPE 3		
(1822-1843)			(1863-1909)			(1921-1963)		
1	HORN TROMP	.577	1	STADT	.401	1	KIRCHE	.216
2	KLINGEN	.386	2	WAGEN	.189	2	HAUS	.188
3	WAGEN	.224	3	OED E	.155	3	ASPHALT	.177
4	FRAGE	.208	4	SCHRITT	.133	4	BETT	.175
5	STUMM	.171	5	RAUCH	.131	5	FUSS	.169
6	FENSTER	.166	6	HAUS	.116	6	BAHN	.153
7	WANDER SCH	.132	7	FENSTER	.107	7	WAGEN	.145
8	ELEND NOT	.107	8	CHOR AL	.106	8	GRAU	.125
9	SCHWER	.107	9	GELB	.089	9	FENSTER	.109
10	INTERJ	.107	10	GESICHT	.083	10	STADT	.097
11	FUSS	.105	11	WUNSCH	.083	11	NEON	.095
12	KOPFHAUPT	.096	12	WEG PFAD	.080	12	MUED IGKET	.093
13	GESICHT	.095	13	DUNKEL N	.075	13	DACH	.093
14	DACH	.094	14	RAUM	.075	14	KREUZ	.090
15	MUED IGKT	.092	15	MILD LAU	.074	15	SCHREI EN	.089
16	JUBELJAUCH	.090	16	KIRCHE	.073	16	DUMPF	.079
17	FEUER	.085	17	ALT ER	.072	17	ZUNGE	.075
18	HORIZONT	.081	18	KLINGEN	.071	18	STEIN	.074
19	FEST LICH	.077	19	DOM KATHD	.068	19	QUELLE N	.073
20	KIRCHE	.075	20	HIRTE	.065	20	KUEHL E	.071
21	ALT ER	.063	21	ABSCHIED	.063	21	REGEN NEN	.067
22	TAU	.059	22	DACH	.063	22	NASS E	.065
23	GLOCKE	.059	23	GLOCKE	.063	23	FRAU WEIB	.062
24	BLICK EN	.056	24	TRAUER N	.062	24	MAEDCHEN	.062
25	ABEND	.054	25	FREMD E	.062	25	SEE	.062
26	BORN	.054	26	KRANK H	.062	26	ZIGARETTE	.060
27	BANN	.052	27	HOHL	.056	27	FLUSS	.059
1	UFER	.052	1	HIMMEL	.054	1	ALL	.057
2	BILD	.054	2	HAUCH	.056	2	ERDE	.058
3	FREUND SCH	.054	3	EWIG KEIT	.056	3	INSEL	.058
4	BANDE	.054	4	GLANZ	.057	4	SCHMERZ EN	.059
5	GRAU	.056	5	NEBEL IG	.058	5	SCHATTEN	.062
6	RAUSCHEN	.057	6	AB GRUND	.058	6	BLUETE	.062
7	SEHNSUCHT	.057	7	LERCHE	.058	7	HAUCH	.065
8	MOND	.058	8	BERGGEIRG	.059	8	STILL E	.067
9	SACHT	.058	9	SEELE	.059	9	WALD TANN	.068
10	BLAU	.058	10	SEE	.059	10	STROM EN	.068
11	SCHWEBEN	.059	11	SONNE	.059	11	INNEN IG	.069
12	KALT E	.059	12	TRAUM	.060	12	LIEBEMINNE	.070
13	LEBEN	.059	13	GLUT	.060	13	TRAUM	.070
14	NACHT	.061	14	NATUR	.063	14	WEICH	.071
15	ROTPURPUR	.063	15	ZWEIG AST	.066	15	HOFFNUNG	.072
16	GOLD	.064	16	SEHNSUCHT	.068	16	TIEF E	.072
17	GRABGRUFT	.064	17	BANN	.069	17	EWIGKEIT	.074
18	WIND	.066	18	WOLKE	.072	18	SCHWEIGEN	.078
19	HEISS	.069	19	SCHLAF EN	.073	19	WUND E	.078
20	SCHMERZ EN	.070	20	FRIEDE N	.075	20	HIMMEL	.089
21	WELLE	.073	21	WILD	.080	21	WELT	.090
22	MORGEN	.076	22	ERDE	.080	22	FEUER	.095
23	LUFT	.076	23	ROSE	.083	23	LEBEN	.104

Abb. 2

das Auffinden gleicher *Zimos* in den Umgebungen verschiedener *Publ-Gruppen* erleichtert. Abb. 2 führt diese *Zimos* ebenfalls, nun jedoch in der Reihenfolge ihrer abnehmenden Index-Werte geordnet, in den drei parallel vergleichbaren Kolonnen auf, wobei vor dem jeweiligen *Zimo* dessen Rangziffer erscheint, während der zugehörige Index-Wert (bei dem Nullen vor dem Komma nicht ausgedruckt wurden) nachfolgt.

Überblickt man die drei Umgebungsfelder (Abb. 2), so fällt zunächst auf, daß sich – klar abgesetzt – Entsprechungen jeweils im positiven und im negativen Bereich der Index-Werte ergeben. Für die *Publ-Gruppen 1* und *2* sind dies: KLINGEN, WAGEN, FENSTER, GESICHT, DACH, KIRCHE, ALT, GLOCKE; für die *Publ-Gruppen 2* und *3*: STADT, WAGEN, HAUS, FENSTER, KIRCHE, DACH; für die *Publ-Gruppen 1* und *3*: WAGEN, FENSTER, FUSS, MÜDIGKEIT, KIRCHE, DACH im positiven Bereich, so daß sich, als in den Umgebungen aller drei *Publ-Gruppen* vertreten, folgende *Zimos* ergeben: WAGEN, FENSTER, KIRCHE, DACH (mit allerdings recht unterschiedlichen Umgebungs-Rangziffern).

Ähnliche Entsprechungen lassen sich im negativen Bereich feststellen, wenngleich weniger zahlreich; Übergänge vom positiven in den negativen Bereich der Umgebungsfelder sind verhältnismäßig selten und meist nur beim Vergleich der *Publ-Gruppen 1* und *3* zu beobachten.

Zum Schluß noch eine kurze Bemerkung zur hohen Affinität der *Womos* HORN/TROMPETE und KLINGEN im Umgebungsfeld von STRASSE der *Publ-Gruppe 1*: dieser zunächst überraschende Sachverhalt wird leicht verständlich, wenn man bedenkt, daß die um 1822 bis 1843 entstandenen Gedichte weitgehend einem noch romantischen Straßenbild entsprachen, in dem die Postkutsche (WAGEN), das Posthorn (HORN/TROMPETE) und dessen KLINGEN nicht fehlten. In Analogie dazu erscheint – einem gewandelten Vorstellungsbereich STRASSE entsprechend – im Umgebungsfeld der *Publ-Gruppe 3* neben KIRCHE und HAUS das *Zimo* ASPHALT auf dem dritten Rang des positiven Bereichs.