

SSD
Box 5048
40221 Göteborg

890104

Att.: Birger Jerlehag

Hej!

Tyska institutionen i Lund har länge drivit ett projekt på data lagrade vid SSD, som under de senaste åren p.g.a tidsbrist har legat "slumrande". Ansvarig vid Tyska var och är Wolfgang Koch. Som timanställd programmerare vid institutionen har jag nu fått i uppdrag att åter få igång projektet, och tänkte börja med en översyn och dokumentation av existerande filer och program. Tyvärr har vi slarvat en del med dokumentationen här, varför jag nu (förhoppningsvis för sista gången!) skulle vilja be om ett antal upplysningar:

1. Vi har tänkt använda MacIntoshdatorer som terminaler. Företrädesvis skulle vi vilja gå över något nätverk för att slippa telefonkostnader etc. Vi är anslutna till en VAX-dator vid LDC, Gemini, som i sin tur är uppkopplad dels på DECnätet, dels på TCP/IP (med Telnet för terminalkörning och FTP för filöverföring). Dessutom finns som en tredje möjlighet DATAPAK, vilket dock skulle passa oss sämre. För att kunna kommunicera med värddatorn behöver jag veta vilket/vilka nät den är ansluten till, dess namn, samt projektets användaridentitet och lösenord.
2. Det skulle kunna vara praktiskt att kommunicera med er via elektronisk post, t.ex. via EARN/BITNET. Jag skulle därför behöva din/er adress.
3. Vi har här nere en serie utdrag ur de olika filer, som projektet arbetat med, men ingen information om hur de är upplagda. För att kunna komma åt dem i datorn skulle all information om filernas status vara nyttig: om de är utrullade, om deras storlek tillåter att man tar upp och tittar på dem med t.ex. en editor utan vidare, osv.

<u>DELMATERIAL</u>	<u>INNEHÅLL</u>	<u>LRECL</u>	<u>BLKSIZE</u>
1	Lemma-index Z	49	7240
2	Lemma	120	7200
3	Lemma-index	108	7560
4	Index Z	54	7992
5	Lemma-index W	49	7840
6	Index W	54	7992
7	Löpande Z art-kat 1	80	8000
8	2	80	8000
9	3	80	8000
10	4	80	8000
11	5	80	8000
12	0	80	8000
13	hela	80	8000
14	Löpande W art-kat 1	80	8000
15	2	80	8000
16	3	80	8000
17	4	80	8000
18	5	80	8000
19	0	80	8000
20	hela	80	8000
	Tabell art-kat. W	80	8000
	Tabell art-kat. Z	80	8000

DELMATERIALINNEHÅLLLRECLBLKSIZE

①	Lemma-index Z	49	7240
2	Lemma	120	7200
③	Lemma-index	108	7560
4	Index Z	54	7992
⑤	Lemma-index W	49	7840
6	Index W	54	7992
7	Löpande Z art-kat 1	80	8000
8	2	80	8000
9	3	80	8000
10	4	80	8000
11	5	80	8000
12	0	80	8000
⑬	hela	80	8000
14	Löpande W art-kat 1	80	8000
15	2	80	8000
16	3	80	8000
17	4	80	8000
18	5	80	8000
19	0	80	8000
20	hela	80	8000
	Tabell art-kat. W	80	8000
	Tabell art-kat. Z	80	8000

Följande delmaterial konverteras och medräknas i kostnadsuppskattningen

(Z = Süddeutsche Zeitung)
(W = Die Welt)

Delmaterial

				Allt i ng ANSI, NO LABEL		
	Band	fil		Band nr	Art nr	Post Block längd längd
1	C210	3	Lemma-index Z	B418,	1600	49,7840
2	C212	1	Lemma	B397,	1600	120,6000
3		2	Lemma-Index	B407,	1600	108,7560
4		3	Index Z	C002,	1600	54,7992
5	C214	1	Lemma-Index, W, rulle 1	C001,	6250	49,7840
		2	" 2			
6	C215	1	Index W, rulle 1	C012,	6250	54,7992
		2	" 2			
7-11	A427	1-5	Löpande Z, art kat 1-5	B307,	1600	80,8000
12		6	" 0			
13		7	" hela			
14-18		8-12	" W, art kat 1-5 1-2			
19		13	" 0 3-5,0	B351,	1600	80,8000
20		14	" hela	B367,	6250	80,8000
	C216	1-8	div programfiler			

På massminne

ART-KAT-W	Tabell artikelkategorier W	C 020, 1600 fil 2	80,8000
" Z	" Z	C 020, 1600 fil 1	80,8000
WSZ	Programfil		
MINPROG-W	"		
" Z	"		

Inger Rosengren

**Ein
Frequenzwörterbuch
der deutschen
Zeitungssprache**

**Die Welt
Süddeutsche
Zeitung**

1

CWK Gleerup Lund Schweden

Lunder germanistische Forschungen 41

**begründet von
Erik Rooth**

**herausgegeben von
Inger Rosengren**

Gedruckt mit Unterstützung von
Statens Humanistiska Forskningsråd

© 1972 Inger Rosengren
Villadsen & Christensen, Köpenhamn, 1972

Vorwort

Das Frequenzwörterbuch besteht aus zwei Bänden. Der hier vorliegende erste Band bietet das Material mit Ausgangspunkt von den Wortformen, der zweite mit Ausgangspunkt vom Wort, vom Lemma (s. S. XVII ff.).

Das Material, das aus einer repräsentativen Stichprobe aus den beiden Zeitungen DIE WELT und SÜDDEUTSCHE ZEITUNG vom 1. Nov. 1966 bis zum 30. Okt. 1967 besteht, wurde wegen des dort vorhandenen Streifenlesers an der Rechenanlage *Medicindata* (Saab D 21) in Göteborg bearbeitet. Als Eingabe dienten nämlich die von den Zeitungen in der automatischen Setzerei verwendeten 6-Kanal-Streifen (TTS-Streifen). Die Lochstreifen wurden von den beiden Zeitungen freundlichst zur Verfügung gestellt.

Ein Teil der Programme, die in DAC und Algol-Genius geschrieben sind, wurden für das Projekt *Datamaskinell undersökning av tidningsprosa, Språkdata*, unter der Leitung von Dozent Sture Allén, geschrieben, mußten aber mehr oder weniger modifiziert werden, ehe sie von uns verwendet werden konnten¹. Professor S. Abrahamsson, Ingenieur G. Häggstrand und Dipl.-Ing. L. Andersson, *Medicindata*, Göteborg, schrieben und verbesserten die Programme. Ingenieur J. Haarde führte die maschinelle Bearbeitung durch.

Das Grundmaterial wird aufbewahrt und kann für andere Untersuchungen verwendet werden.

Fil.mag. M. Brandt war bis 30. Juni 1971 Assistentin an der Arbeitsstelle und hat an der Ausformung des Wörterbuchs aktiv teilgenommen. Frau K. Franzén, Frau H. Helmers, Frau A. Kling, Fräulein Ruth Kneip, Frau M. Persson, Frau G. Pettersson, Frau J. Skowronski, Frau G. Deppe, Herr B. I. Andersson, Herr H. Bartel, Herr H. Kessler, Herr H. J. Lüssenhop, Herr S. Pettersson und Herr M. Schonebohm, der seit 1. Juli 1971 Assistent an der Arbeitsstelle ist, haben u.a. das Material identifiziert, katalogisiert und korrigiert. Universitätslektor J. Lanke, Institut für mathematische Statistik, Lund, war statistischer Berater.

Mit Dozent Sture Allén, fil.lic. Jan Thavenius und fil.lic. Lars Åhlander habe ich viele Probleme diskutiert. Dozent Sture Allén hat außerdem die für sein Projekt geschriebenen Programme freundlichst zur Verfügung gestellt. Allen diesen möchte ich an dieser Stelle danken.

Mein Dank richtet sich auch an den damaligen Leiter des Instituts für deutsche Sprache, Lund, Professor Tage Ahldén, der der Arbeits-

¹ S. hierzu Rosengren, I. Ein Frequenzwörterbuch der modernen Zeitungssprache – wie und wozu? in: Beiträge zur Linguistik und Informationsverarbeitung 14, 1967, 9 f und dies., Projektet Modernt tyskt tidningsspråk in: Språklig databehandling, hrsg. von S. Allén und J. Thavenius, Lund 1970, 64.

stelle Platz bereitete, an die Universität Lund, die die Rechenzeit und die Mittel zur Programmierung und zum Stanzen bewilligte, und an *Humanistiska forskningsrådet*, *Statens råd för samhällsforskning*, die Pädagogische Hochschule in Malmö, *Nathhorstska stiftelsen*, *Långmanska kulturfonden* und *Anslaget för biträdeshjälp m.m. åt docenter*, die durch ihre finanzielle Hilfe die Durchführung der Arbeit ermöglicht haben. Es ist unmöglich, eine Arbeit wie die hier vorliegende fehlerfrei durchzuführen. Wir planen für den zweiten Band eine Errata-Liste und sind deshalb dankbar für alle diesbezüglichen Hinweise.

Lund im August 1972
Inger Rosengren

Inger Rosengren

**Ein
Frequenzwörterbuch
der deutschen
Zeitungssprache**

**Die Welt
Süddeutsche
Zeitung**

2

CWK Gleerup Lund Schweden

Lunder germanistische Forschungen 43

begründet von

Erik Rooth

herausgegeben von

Inger Rosengren

LiberLäromedel
Box 1205
221 05 Lund

CWK Glerup ist das Impressum für wissenschaftliche
Veröffentlichungen des Verlages LiberLäromedel Lund

© 1977 Inger Rosengren och LiberLäromedel Lund
Villadsen & Christensen, Köpenhamn 1977
ISBN 91-40-04470-X

1. Einleitung

1.1. Allgemeines

Das Frequenzwörterbuch besteht aus zwei Bänden. Der erste Band bietet das Material mit Ausgangspunkt von den Wortformen. Der hier vorliegende zweite Band basiert auf einer Lemmatisierung des ganzen Materials des ersten Bandes.

Das Material der beiden Bände ist zusammengesetzt aus zwei unterschiedlich großen repräsentativen Stichproben aus der WELT bzw. der SZ vom 1. November 1966 bis zum 30. Oktober 1967. Eine nähere Beschreibung der beiden Stichproben findet sich im ersten Band. Dort wird ebenfalls die Worteinheit, d.h. die im Text gezählte Einheit, die den Ausgangspunkt auch für die Lemmatisierung bildet, näher definiert. Nur an einer Stelle wurde von der dort gegebenen Definition abgewichen, nämlich bei der Behandlung der Verbzsätze. Diese Abweichung führt eine Änderung der Totalfrequenz der laufenden Wörter mit sich (s. S. XIX).

Die schon im ersten Band durchgeführte und dort beschriebene Aufteilung des Materials auf fünf Kategorien (Meinung, Politik, Kultur, Wirtschaft und Vermischtes) wurde im zweiten Band beibehalten. Die statistischen Maße sind ebenfalls dieselben wie im ersten Band und werden dort eingehend beschrieben. Fehler im ersten Band wurden selbstverständlich beseitigt.

In diesem Band werden deshalb nur die neu hinzugekommenen Lemmatisierungsprinzipien und die einzelnen Listen des zweiten Bandes, die aufgrund der Lemmatisierung zuweilen von den Listen des ersten Bandes abweichen, näher behandelt.

1.2. Die Lemmatisierungsprinzipien

1.2.1. Die Hauptprinzipien

Eine Voraussetzung einer konsequent durchgeführten Lemmatisierung ist eine Definition des Lemmas, des Wortes, die subjektive Überlegungen so weit wie möglich ausschaltet. Aus Gründen, die in Rosengren (1969) ausführlich diskutiert werden, gehen wir von Flexionsreihen oder Paradigmata aus. Eine Flexionsreihe besteht aus einer Reihe paradigmatisch zusammengehörenden Wortformen mit normalerweise demselben Stamm (s. hierzu Rosengren 1969, 109), aber verschiedenen derselben Flexionsklasse angehörenden Flexionsmorphemen, die mit dem syntaktischen Auftreten des Stammes variieren.



LUNDS UNIVERSITET

Tyska institutionen
Germanistisches Institut
der Universität Lund

Datum 26.4.94

Jris Alfredsson
SSD

Hj!

Har nu talat med Inger Rosengren om era
frågor kring rätt tidningsmaterial (Die Welt,
Süddeutsche Zeitung).

Jag skickar dig här kopior på titelblad,
förord m.m. ur band 1 och 2 av "Eine
Frequenzwörterbuch der deutschen Zeitungs-
sprache", som undersökningen av ovanstående
material resulterade i.

När det gäller löpande text - som är en del av
ett material - är materialet från "Die Welt" i Oräthet
skick, medan texterna från "Süddeutsche Zeitung"
bör finnas i en korrigerad version i Mannheim (vid Institut
für deutsche
Konkordanser finns här hos oss i Lund. ^{Speache} - JdS)

Postadress
S-223 62 LUND

Gatuadress
Helgonabacken 14

Telefon (vx)
046-10 70 00

V. G. V.

Om Du tycker att de här uppgifterna är
för knapphändiga, får Du återkomma.

Hälsningar

Britt-Mari Ek

Tyska inst.

tel. 046/109164

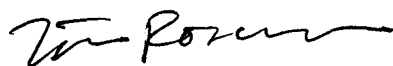
fax. 046/104211

Jag nås lättast hemma, tel 046-158373, och min adress är

Jörgen Rosengren
Dag Hammarskjölds v. 4A:310
223 64 Lund

Om du har några frågor av mer organisatorisk karaktär, kan du fortfarande tala med Wolfgang Koch på tel. 046-109162 (arb).

Med vänlig hälsning,



Jörgen Rosengren