

トーマス・マンのテキスト・データ・ベースについて

樋口, 忠治
九州大学教養部 : 教授

<https://doi.org/10.15017/6796141>

出版情報 : 言語科学. 18, pp.11-14, 1983-03-31. The Group of Linguistic Studies College of General Education, Kyushu University

バージョン :

権利関係 :



トーマス・マンの テキスト・データ・ベースについて

樋 口 忠 治

1. Th. Mann 全集のうち文学作品のデータ化がほぼ出来上りつつある状況の下で、言語研究のため、あるいはこの作家の作品そのものの研究のために、検索用の「テキスト・データ・ベース」を公開する準備が進んでいる。このデータ・ベースは九州大学大型計算機センターの「公用データ・ベース」として運用され、センターの利用資格者には利用の便宜がはかられる。そこで、このデータ・ベースの概要を簡単に紹介しておきたい。
2. 本年4月からスタートするこのデータ・ベースは次にあげる作品からなっている。
 1. Buddenbrooks
 2. Der Zauberberg
 3. Doktor Faustus
 4. Lotte in Weimar
 5. Felix Krull
 6. Tonio Kröger
 7. Der Tod in Venedig
 8. Der Herr und Hund
 9. Die Betrogene
 10. Tristan
 11. Mario und der Zauberer

* * *

テキスト・データの作成にあたっては、S. Fischer の1974年版によった。公用データ・ベースとして運用するに当たって、Fischer 書店との間に協定を結んでいるが、利用者は特に次の点に注意しなければならない。すなわち、このデータを加工していわゆる講義等の教材としたり、あるいは研究以外の目的に利用するなどの行為は禁じられている。

3. データ・ベースの利用をするにはまず、九州大学大型計算機センターの利用資格を得なければならないことはいうまでもない。更に、他のセンター地区から利用する場合には最寄りのセンターの利用資格も合わせて得る必要がある。以上、データ・ベース利用の一般的な条件を満たしてさえいれば、本データ・ベースの利用は可能である。但し、スタートしたばかりの段階では若干の制限を設ける予定である。すなわち、予め筆者宛に利用希望の届けをしていただきたいと考えている。その段階で利用に必要な「パスワード」やファイル名などを通知すること

にしたい。

4. 本データ・ベースは、トーマス・マンのテキスト・データを情報システム「SIGMA」によって利用するものである。このSIGMAについては九大センター広報 Vol. 14 No. 4 (1981) 記載の「研究者向き情報システム SIGMA について」を参照せられたい。SIGMA は元来テキスト・データ用に開発されたシステムではないが、汎用性があることから、テキスト・データの検索等にも有効である。但し一時に大規模のファイルを検索することはできないので、適当な大きさのファイルに分割しているが、長編作品の場合、順次ファイル名のみを入力すれば検索ができるようになっている。

5. 本データ・ベースの特徴は、個々の語(単語)の単位で検索できることはもちろんであるが、その他に語の組み合わせも検索できる点にある。すなわち、テキスト全体が単一のレコードとして扱われるので、複数の語の組み合わせが検索可能である。これに加えて最大の利点は、前後の一定の語にはさまれている表現では、その中間に何があっても構わないという形式で検索できる。すなわち、X...Y で表わすことができるような形の検索が可能である。予め、調査したい項目を用意しておき、同時に数項目について検索を行なうのが効果的である。どの作品について検索するのか、すべての作品について行なうのか、等は充分検討しておく必要がある。調査する項目が決りさえすれば、あとは次々にファイル名のみを投入するようになっているので、難しい点は少しもない。例えば Doktor Faustus だけについて幾つかの接続詞を用いる副文を探したい場合(—als ob 文, damit 文, obgleich 文—)を考えてみると、まず, als ob 文といっても als+定形の形式は als の用例をすべて探して取捨選択をする必要があるから、簡単にはいかない。damit 文も同様で、すべての damit を探したうえで、接続詞の用例のみを選び出す必要がある。これに対して obgleich 文はほとんど問題はないであろう。一方 um...zu Infinitiv の用例を検索によって探す場合はほぼ一義的に決まる。しかし、特別な場合もないではない。例えば um 8Uhr zu Abend essen という形の中には um...zu のパターンが含まれているので検索の網にかかるわけである。個々の単語の検索には問題はない。

6. 言語研究および文学研究のための TSS によるテキスト・データ・ベースは我が国はもちろん、外国にも未だ例を見ないものである。これはテキスト・データの作成が困難であることその他に、TSS のネットワークが存在する国が少ないことによる。幸い我が国にはこうしたシステムが設置されており、研究者に広く開放されている。しかし、言語の研究者、あるいはドイツ語の研究を行う人々はこうしたシステムの利用にはなじみがないので、当初はいささかのとまどいがあるかも知れない。

いわゆるデータ・ベースの利用自体は簡単なコマンド類の使用法をマスターすれば、誰にでもできることであるから、さほど難しく考える必要はない。一人の作家の作品全体について、任意の語や句、あるいは文などを具体的に検索する方法として初めての試みである本データ・ベースにはまだ不十分な点があるかも知れない。そうした点は今後改善していくつもりである。

7. 検索の具体的な方法については、運用開始後、九大センターの広報に紹介する予定であるから、これを参照していただくとして、利用者のために予め運用の開始をここに紹介し、できるだけ多くの方々に利用していただくようお願いしておきたい。

○本データ・ベースに関する問い合わせ：

810 福岡市中央区六本松 4 丁目 2 番 1 号

九州大学教養部 独語科

樋口忠治

電話：092-771-4161(代) 内線323または333

○九州大学大型計算機センター利用の問い合わせ：

812 福岡市東区箱崎 6 丁目10番 1 号

九州大学大型計算機センター共同利用掛

電話：092-641-1101(代) 内線2505

また九大以外の各センターへの利用の問い合わせは下記へされたい。

北大センター共同利用掛（以下同じ）	011-711-2111（内2951）
東北大センター	0222-27-6200（内3214, 2745）
東大センター	03-812-2111（内2717）
名大センター	052-781-5111（内6503, 6505）
京大センター	075-751-2111（内7406, 7407）
阪大センター	06-877-5111（内2817）

（利用許可がおりた時点で、端末操作等について当方へ照会があれば、具体的な案内をする）

なお、本年7月下旬にはこのデータ・ベースの利用の仕方とコンピュータ・ネットワークについての講習会を東京大学大型計算機センターにおいて開催する予定である。参加希望者は5月末日までに日本独文学会あてに申し込まれたい。

8. 以下の例は全集8巻中 **Herr und Hund** について **daß** 文中の **würde** を文単位でリストアップしたものである。検索の結果は6例あり、各文の先頭の番号（最初の2桁は巻数、次の3桁はページ数、最後の2桁は行数を示す）によって、実際にテキストに当たって確かめることができるし、このままでも文章単位として出力されるので用法を確認するには十分であろう。なお、文番号#0859026と#0861628は該当しないことがわかる。

LOGON TSS ~~*****~~

KDS700011 ~~*****~~ LAST ACCESS AT 14:49:17 ON 83.039

KEQ564551 ~~*****~~ LOGON IN PROGRESS AT 15:21:30 ON FEBRUARY 8, 1983

JOB NO = TSU5007 CN(01)

KEQ569511 NO BROADCAST MESSAGES

<USER MODE> OK

READY

SIGMA

SIGMA> SEARCH

VERSION (D/E)?

RECORD DELIMITERS

D1:=#

D2:=

KEYWORDS

A1:= DA\$... W=URDE

A2:=

LOGICAL FORMULAE

Z1:=A1

Z2:=

FILE:=S. MANN. HERR

RETRIEVED TEXTS

RETRIEVED

TOTAL = 6

QUESTION 1 (Z1) = 6

FILE:=

LIST OF RESULTS (N/Y)?

DO: REF

QUESTIONS:=1

TOTAL RECORDS = 6

RECORD DELIMITER:=#

NUMBERING (N/Y)?

SORT ON:

DO: LOOK

#0853706 ABER DEN <KINDERN HATTE DIE VIERBEINIGE <TR=UBSAL UNTER DEM <TISCH E-S NAT=URLICH ANGETAN , UND WIR <ERWACHSENEN GABEN UNS VERGEBENS DIE <MIENE , - IHRE <WAHL- UND <URTEILSLOSIGKEIT ZU BEL=ACHELN : AUCH WIR F=UHLTEN DEN <STACHEL IM <HERZEN UND SAHEN WOHL , DA\$ ES UNS SCHWERFALLEN W=URDE , DAS <BILD DER ARMEN <LUX WIEDER AUS UNSERM <GED=ACHTNIS ZU TILGEN .

#0853726 ABER WIR SAHEN WOHL , DA\$ DIE <SACHE , WIE ALLES STAND UND LAG , EINEN ANDERN <AUSGANG KAUM W=URDE NEHMEN K=ONNEN .

#0854318 ICH SEHE IHN WOHL AUCH VON MEINEM <TISCHE AUS AUF DEM ERH=OHTEN <WEGE DR=UBEN , ZWISCHEN DEN ALTEN <ESPEN , IN NACHDENKLICHEM <BUMMELTRABE SICH HINBEWEGEN ; DOCH SOLCHE <PROMENADEN SIND NUR EIN MATTER <ZEITVERTREIB , OHNE - <STOLZ , <GL=UCK UND <LEBEN , UND V=OLLIG UNDENKBAR BLEIBT , DA\$ <BAUSCHAN SICH ETWA AUF EIGENE <HAND DEM HERRLICHEN <JAGDVERGN=UGEN HINGEBEN K=ONNTE , OBGLEICH NIEMAND IHN DARAN HINDERN W=URDE UND MEINE <GEGENWART , WIE SICH ZEIGEN WIRD , NICHT UNBEDINGT ERFORDERLICH DAZU W=ARE .

#0859026 EINEN <FITTICH GESPREIZT , MIT WEGGEDEHNTEM <HALSE LAG DER <FASAN IM- <GRASE UND SCHRIE , SCHRIE OHNE <PAUSE , DA\$ ES KLANG , WIE WENN IM <GEB=USCH EINE <GREISIN GEMORDET W=URDE , UND ICH HERZUEILTE , UM ETWAS <GR=ASLICHES - ZU VERH=UTEN .

#0859314 DER <PROFESSOR , EIN <MANN AUF DER <H=OHE DER <JAHRE , IN WEISEM <OPERATIONSMANTEL , MIT GOLDNER <BRILLE , EINEM LOCKIGEN <SCHEITEL UND VON SO KUNDIGER , LEBENSFREUNDLICHER <MILDE DES <WESENS , DA\$ ICH IHM UNBEDENKLICH MICH SELBST UND ALLE DIE <MEINEN IN JEDER <LEIBESNOT ANVERTRAUT HABEN W=URDE , L=ACHELTE W=AHREND MEINES <VORTRAGES V=ATERLICH AUF DEN VOR IHM SITZENDEN UND - VON SEINER <SEITE VERTRAUENSVOLL ZU IHM AUFBLICKENDEN <KLIENTEN HINAB .

#0861628 WENN DU GLAUBST , DA\$ ER DICH WISSENSCHAFTLICHER <BEOBACHTUNG =UBERGEHEN WIRD , FALLS ES DIR EINF=ALLT , OKKULT ZU BLUTEN , ODER DA\$ DU ALS SEIN - <HUND F=UR NERV=OS UND AN=AMISCH ERKL=ART WERDEN WIRST , SO GEH NUR ZU IHM , - DOCH K=ONNTE ES SEIN , DA\$ DU DICH IM <IRRTUM WIEGEST IN <HINSICHT AUF DAS <MA\$ VON <ACHTUNG , DAS DIESE <ART <HERR DIR ENTGEGENBRINGEN W=URDE !

DO: