

of documents in a predetermined location according to a *filing plan/system*". The definition for FILING PLAN/SYSTEM reads, "A predetermined classification plan for the physical *arrangement*, storage and retrieval of *files* (1)..." Strangely, there is no definition for CLASSIFICATION PLAN, but the term has a see reference to FILING PLAN/SYSTEM. This produces a circular definition since FILING PLAN/SYSTEM, by inference, is defined as "A predetermined filing plan/ system for..."

There is, however, an entry for CLASSIFICATION SCHEME, which is defined as "A pattern of arrangement of *archives* (1) by *groups*, *series* and *items* (not US)". This definition calls attention to discrepancies between British and American usage - the US term for this concept is not specified. Interestingly, also, the entailed terms are somewhat confusing. GROUPS has no entry - but we do find definitions for ARCHIVE GROUP, RECORD GROUP, and five other kinds of "groups". ARCHIVE GROUP is explained as "The primary division in the *arrangement* of *archives* (1) at the level of the independent originating unit or agency".

ITEM names "The basic unit of *arrangement* and *description*..." and SERIES, "Item or documents arranged in accordance with a *filing plan/system* or maintained as a unit...". SERIES has a see also reference to CLASS, which is defined as "An identifiable and self-contained subdivision of an *archive group* consisting of a number of *items* with one or more common characteristics (UK). Generally equivalent to *series*". Thus "series" and "class" appear to be used as virtual synonyms but they are defined by reference to the origin ("provenance") of documents, rather than their subject matter. If classificationists find this confusing, it is because archivists use some of their vocabulary, but with significant shifts in meaning.

Enough has now been written to give readers a feeling for the book. On the whole it is very readable and clearly laid out, a good reference tool for archivists. It has minor mistakes and discrepancies scattered about and it overlaps classification theory, using some of its terms for different notions. Above all, its quasi-lexicographic format is a real handicap: a well conceived terminological design would make its contents more intelligible and helpful to its users.

Fred W. Riggs

Prof. Dr. F. W. Riggs, Department of Political Science,
University of Hawaii, Honolulu, HI, 96822, USA.

BUTLER, Christopher: *Computers in Linguistics*.
Oxford, GB: Blackwell 1985. IX, 266p.,
ISBN 0-631-14266-5

"Computers in Linguistics" ist eine Einführung in die

philologische Datenverarbeitung unter ganz besonderer Berücksichtigung von SNOBOL 4.

Der Autor erklärt, wie Computer im Prinzip arbeiten. Ein sehr nützliches Kapitel seines Buches gibt Auskunft über den Funktionsumfang von Softwarepaketen für die philologische Textverarbeitung: EYEBALL, OXEYE, OCP, CLOC. Auch auf Statistikpakete, wie SPSS wird hingewiesen.

Butlers Anliegen ist es, Philologen und Linguisten an die computerunterstützte Sprachforschung heranzuführen. Für Aufgaben, die mit Softwarepaketen nicht zu bewältigen sind, schlägt er als Programmiersprache SNOBOL vor, eine seit 1962 entwickelte Sprache zur Text- und Symbolverarbeitung. Sie unterscheidet sich von anderen höherem Programmiersprachen, wie ALGOL, COBOL und FORTRAN vor allem dadurch, daß sie das Pattern-Matching, die Definition von symbolischen Mustern und den Test auf deren Vorkommen in den Daten, besonders unterstützt. Auch die benutzerdefinierten Datentypen von SNOBOL sind interessant.

Den Schwerpunkt des Buches bildet eine gründliche und gut verständliche Einführung in SNOBOL. Viele Beispiele, Originalprogramme und Aufgaben mit Musterlösung machen das Buch als Hilfe auch zum Selbststudium geeignet. Erklärt werden Grundkonzepte von SNOBOL, das Pattern-Matching, die Kontrollstruktur von SNOBOL-Programmen, eingebaute Funktionen, Vergleichsoperatoren, benutzerdefinierte Funktionen und Datentypen, die Fehlersuche in SNOBOL-Programmen.

Zwei Beispiele für SNOBOL-Anwendungen schließen den Band ab: eine sprachstatistisch-stilistische Analyse der Dichtung von Sylvia Plath und eine Berechnung der lexikalischen Dichte (type-token ratio) von Texten aus dem London-Lund-Korpus des gesprochenen Englisch.

Die Beispiele kennzeichnen den Schwerpunkt des wissenschaftlichen Interesses von Ch. Butler und den Lesern, die er anspricht: Im Blickpunkt stehen die computerunterstützte quantitative Stilistik, Autoren-schaftsuntersuchungen, Wörterbucharstellung, Text-edition und computerunterstützter Sprachunterricht. Computer dienen vor allem zur Produktion von Indices und Konkordanzen, zur Auszählung von Frequenzen, zur Errechnung statistischer Kennwerte, zur Lemmatisierung. Innerhalb dieses Interessenbereiches ist die Darstellung von Ch. Butler technisch und konzeptionell modern. Der Ausblick auf andere computerlinguistische Forschungsrichtungen (automatische Übersetzung, Textproduktion, Simulation menschlichen Sprachverhaltens) bleibt begrenzt. Eine Computerlinguistik, die sich um die Gestaltung von Mensch-Maschine-Dialogen, um die Repräsentation von Texten, um die Verarbeitung nicht transkribierter, sondern akustisch aufgenommener Sprache und immer wieder um das zentrale Problem des syntaktischen und semantischen Parsings bemüht, kommt bei Ch. Butler so gut wie nicht vor.

Brigitte Endres-Niggemeyer

Prof. Dr. B. Endres-Niggemeyer, Fachhochschule Hannover,
FB BID, Hanomagstr. 8, D-3000 Hannover 91.