Book reviews

PÖRKSEN, Uwe: **Deutsche Naturwissensehaftssprachen, historische und kritische Studien** (German special science languages; historical and critical studies). Tübingen: Gunter Narr Verlag 1986. 251p. = Forum für Fachsprachenforschung, 2.

La publication de ce libre est l'aboutissement d'un projet à long terme et, sans aucun doute, la réalisation d'un r' ève de l'auteur. Au fait, Uwe Pörksen (Univ.Fribourg i.B.) nous apprend que ce volume réunit des articles publiés cà et là entre 1972 et 1984 et destinés dès le début à ' être les chapitres d'un libre consacré à l'histoire de la langue des sciences naturelles. Les huit exposés de ce volume sont groupés en trois parties. La première partie (un chap.) relate le passage de la langue scientifique latine au langage scientifique allemand et la création de nouvelles langues spécifiques pour la biologie, la physique et la chimie. Ce chapitre parle également de l'accro' issement constant du vocabulaire scientifique et de la relation entre la langue scientifique et le langage de chaque jour. Les six études qui constituent la deuxième partie sont consacrée à des sujets très divers, entre autres: le passage du latin, langue savante, à l'allemand scientifique (notons la comparaison du nombre des livres édités soit en latin, soit en allemand, à différentes époques). Cette 2me partie traite également des rapports entre le langage courant et la langue scientifique et analyse le langage métaphorique de Linné, Darwin, Goethe et Freud, ainsi que la terminologie de la psychanalyse. Le chapitre qui expose la "naissance" d'un livre scientifique populaire est remarquablement informatif. La troisième partie développe des idées particulièrement intéressantes. Parlant de l'afflux des termes scientifiques dans la langue allemande, l'auteur se pose des questions à propos de la pureté de celle-ci. Uwe Pörksen aborde également ici le problème de la formation des enseignants du secondaire et discourt sur le nombre croissant des institutions d'enseignement supérieur.

Lorsque l'auteur passe en revue les différents moyens d'exprimer un nouveau concept, il mentionne e.a. les emprunts au matériel lexical d'autres langues. En notant des centaines de mots, il les classe soit chronologiquement, soit par sujet; on peut regretter l'absence de classification selon l'origine. A plusieurs reprises, Pörksen insiste sur l'origine latine ou grecque de beaucoup de néologismes; il nous semble qu'il ait tort de ne pas assez accentuer l'importance des mots hellénogènes (Griechisch n'est m' ème pas mentionné dans le registre...).

Puisque ce volume réunit des textes écrits au cours d'une période d'une douzaine d'années, il n'est pas étonnant que certains sujets sont abordés plus d'une fois. On remarque d'ailleurs un triple fil rouge à travers les différents exposés: que le latin des savants a été remplacée par des langues nationales en laissant beaucoup de traces dans celles-ci, que le vocabulaire scientifique s'accro' it de facon vertigineuse et qu'une partie de ces nouveaux mots est bien acceptée par la langue de chaque jour.

On est toutefois quelque peu surpris qu'un seul volume réunisse des sujets si divers. Mais l'intér' èt et la compétence de l'auteur s'étendent à des domaines aussi nombreux que différents, non seulement à ceux qui répondent au titre du livre, tels la chimie, la botanique, la physique, la biologie, les mathématiques, mais aussi à d'autres matières telles que la linguistique, la littérature, la philosophie, la psychanalyse, la pédagogie et la didactique.

C'est pourquoi ce libre pourrait donner à certains lecteurs une impression d'éparpillement; d'autres cependant lui accorderont un caractère d'interdisciplinarité. Nous nous rangons parmi ces derniers.

Henri Leclercq

Catholic University of Leuven
B-3000 Leuven, Belgium

HILDRETH, Charles R.: **Intelligent Interfaces and Retrieval Methods for Subject Searching in Bibliographie Retrieval Systems.** Washington, DC: Library of Congress/Cataloging Distribution Service 1989. III,120p. ISBN 0-8444-0626-0. = Advances in Library Information Technology, 2.

Online Public Access Catalogs (OPACs) have become part and parcel of libraries of almost any size and mission[1]. They are replacing conventional library catalogs on card or microfiche, thus passing on the benefits of a very costly enterprise called library automation to the libraries' patrons. The holdings of libraries are no longer accessible only via authors' names or titles proper; instead, the complete bibliographic descriptions as well as "enrichments" of various kinds (from subject headings to abstracts) may provide access points for queries, with hitherto unknown possibilities of linking names, keywords, or subjects by employing Boolean operators. The history of OPAC use proves that the vast majority of queries are subject oriented, rather than "known item searches" for, say, a particular author or title. The feasibility of OPACs, both in terms of library management and user satisfaction, depends on the quality of subject access. This involves two aspects, processing and storage of relevant data (subject headings, classification etc.) as well as the design of what has come to be termed the "front end", including both surface matters (menu design) and internal ones (for instance, system guidance, query correction, relevance feedback etc.).

Speaking of online public access to library files is referring to the numerous and comprehensive publications of Charles R.Hildreth. To Hildreth, OPACs never were a merely technical matter. Already in 1982, he pointed out unmistakeably that OPACs were to be considered and designed as the "human interface"[2]. Quite literally, OPACs are expected to line (or "interface") the internal, library administrative computing routines with the information needs and searching behaviour of the non-professional and, quite possibly, computer-illiterate library patron. As Hildreth's survey clearly shows: there are OPACs which are simply more intelligent than others. His monograph on OPAC design already referred to, is still semi-

nal, as is his state-of-the-art report[3]. Although both publications are "old" in terms of computing developments, librarians in this country will not find them outdated; with respect to the timelag of library automation in W.-Germany (in contrast to the United States), Hildreth's publications will certainly help avoiding the repetition of errors and pitfalls in OPAC design and implementation.

The present book was prepared for the Library of Congress with two objectves in mind. First, a state-of-the-art survey and investigation of intelligent "front end" design approaches and software for improving subject access and subject searching in today's large online bibliographic retrieval systems, including OPACs; second, a statement on the applicability of intelligent retrieval methods to a future information retrieval system implemented by the Library of Congress. The focus of this study clearly is on OPACs. 23 advanced OPACs and intelligent retrieval systems and software, including three CD-ROM systems, in the United States and the United Kingdom (plus one from France) were investigated in the course of this survey, ranging from experimental inhouse systems to commercially marketed systems which include the OPAC as a module of an integrated library system. Many OPAC features are demonstrated by helpful reproductions of screen displays.

Any investigation into the intelligence of subject access must be rooted in the data available for librarians, computing experts and end-users. As for subject access, it should be kept in mind that Anglo-American OPACs are usually based on MARC records, supplied either by the Library of Congress or the British Library; UK and US MARC records contain classification data (Library of Congress and Dewey Decimal Classification) and verbal indexing (Library of Congress Subject Headings and/or verbal features of PRECIS). In spite of several limitations, it seems that American libraries in particular will not change from their traditional instruments of subject indexing and classification, for instance from LCSH to PRECIS, but apply these instruments to an online environment[4]. In other countries, the very basis of data pertaining to subject analysis provided by the national bibliographic agency may not be quite as comprehensive to support the design of any intelligent "front end" for subject searching.

As Hildreth explains, the review of intelligent "front ends" facilitating subject access will lead to an analysis of four major issues: I) ease of use, orientation, and presentation factors, 2) vocabulary control and correlation factors, 3) more effective system-guided or automatic query formulation and retrieval techniques, and 4) meaningful engagement of the searcher in relevance assessments, query modification/expansion, and the provision of smart navigational, exploration facilities (p.5). The keynote is, of course, "smart" which is certainly synonymous with "intelligent" as far as Hildreth's sample OPACs are concerned. Other OPACs, however, are likely to render a somewhat different sense of "smart", more in the line of trial and error, getting lost, despair and pain. It may suffice to refer to the number of OPACs already implemented or to be bought off the shelf which do not permit several or indeed any of Hildreth's four major issues to be seriously reviewed at all. The checklist of standard OPAC features appended to Hildreth's study will be very helpful if disillusioning.

Hildreth classifies today's common OPACs as "second generation", that is to say, they permit key word and phrase searching, browsing of indexes as well as hit lists, and the use of Boolean operators. They clearly support the type of exact match searching: a precisely stated query, possibly incorporating Booleans, is supposed either to prompt a match or nothing at all. This is certainly true for "known item" searches: the library either holds a particular title or it simply does not. Now, the situation gets different when it comes to subject queries. The vocabulary may be ambiguous; descriptors or subject headings may have to be looked up in a list or thesaurus for a correct search statement; or the OPAC user may not be able to state precisely what s/he is actually searching. Titles matching blurred search statements are either too few (or, indeed, none) or too many. There may be recall, yet little precision; or, what is more likely, there will be no recall at all. Booleans pretend efficience of OPAC use and document retrieval. In this context they turn out to be a blessing in disguise, though, for the formulation of good Boolean queries is an art most untrained users cannot cope with. And even if they can: Booleans appear to be the wrong approach to information retrieval. The problem of ranking creates an awareness of Boolean limitations. Whithin Boolean retrieval, any match to an OR-query which contains just one query term is as "good" as another which contains all; and any match to an AND-query which contains all but one query terms is as "bad" as another which contains none (and will, therefore, not be considered a match). As Hildreth pointedly states, second generation OPACs "represent a marriage of the library catalog and conventional online information (IR) systems familiar to librarians who search online abstracting and indexing databases via DIALOG, BRS, DATASTAR, MEDLINE, etc." (p.7). In a library environment, online search specialists are a minority, and most library patrons will find their OPAC stuck in a somewhat unfavourite marriage[5].

Is there any "intelligent" solution to that problem? In contrast to "known item" searches, the term "subject search" may be an euphemism for "uncertainty", either about the topic itself or the search terms. What matters is the capability of the OPAC to react to the varying kinds and degrees of that uncertainty. The "intelligent" answer to the Boolean dominated method is probabilistic retrieval[6]. It is based on a ranking algorithm which "orders the set of retrieved documents according to their decreasing similarity to the query" (p.48). Moreover, the idea is to provide "near matches" in those cases when the Boolean method would have produced none. Not all the terms of a search statement are necessarily of equal importance to the user, and retrieval is facilitated enormously if terms can actually be weighted according to their importance. Further improvement can be made if queries stated in natural language are automatically translated, as it were, into relevant descriptors. The same applies to automatic spelling corrections, phonetic searches ("Soundex algo-

rithms) etc.[7]. Other major issues which can only be mentioned here are navigation and browsing facilities as well as help screens. OPACs will be the more efficient and usable for non-specialist users if a kind of guidance is provided, leading the users through the net of descriptors and paving the way to subject areas which may be related to the initial search statement. "Intelligent" OPACS overcome the static situation of user's query and system's answer. Interactive systems will give "opportunity for search term and document appraisal and relevance feedback during the search process" (p.105). Many OPACs at least suggest different search modes to improve retrieval results; and some, in fact, automatically implement alternative search strategies if there is no recall to a search statement.

Hildreth's survey is not exactly a market analysis; and he has no "best buy" OPAC to recommend[8]. Almost by definition, however, by being included in the survey, all 23 systems reviewed provide more "intelligent" subject searching facilities than most common second generation OPACs. Yet it should not be overlooked that a couple of systems are repeatedly and favourably mentioned for their intelligent approaches. These systems include CITR (at the National Library of Medicine), I3R (at the University of Massachusetts), KIM (at the University of Aberdeen), and OKAPI (designed at the Polytechnic of Central London, and partially included in the integrated turnkey library system LIBERTAS marketed by SWALCAP Library Services Ltd.).

Concluding his survey, Hildreth concisely states the most important requirements of a future information retrieval system of the Library of Congress. It will have to include MARC and non-MARC files, employ multiple thesauri, rely on a traditional database structure utilizing inverted indexes, provide Boolean search mode as one variety accompanying other, probablliste, interactive retrieval modes. The most important requirement probably will be the system's efficiency for the library's clientele expected to be unfamiliar with the mysteries of Booleans and information retrieval. There will never be a totally "automatic" OPAC; yet "automatic" techniques will certainly help to make OPACs and other information retrieval systems more interactive, or, to put it in Hildreth's own words, cooperative and engaging.

Heiner Schnelling

1 Mitev, Nathalie N., Efthimiadis, E.N.: A classified bibliography on online public access. London: British Library 1987.
2 Hildreth, Charles R.: Online Public Access Catalogs: the human interface. Dublin, OH: OCLC 1982.
3 Hildreth, Charles R.: Online Public Access Catalogs. In: Annual Review of Information Science and Technology, 20(1985)p.233-285. Reprinted in: Helal, A.H., Weiss, J.W.(Eds.): Future of Online Catalogues. Essen: Universitätsbibliothek 1986. p.3-55
4 See Markey, Karen: Dewey Decimal Classification Online Project: Evaluation of a library schedule and index integrated into the subject searching capabilities of an online catalog. Final report to the Council of Library Resources. Dublin, OH: OCLC 1986. See also Cochrane, P.A.: Improving LCSH for use in online catalogs. Littleton, CO: Libraries Unlimited 1986. 348p.
5 Hildreth, Charles R.: Beyond Boolean: Designing the next generation of online catalogs. Libr. Trends 35(1987)p.647-667
6 See Salton, Gerard; McGill, Michael J.: Introduction to modern information retrieval. New York: McGraw-Hill 1983.
7 See Walker, Stephen; Jones, Richard: Improving subject retrieval in online catalogs. London, GB 1987.
8 See Leeves, J.: Library systems: a buyer's guide. Aldershot: Gower 1987. 2nd ed.forthcoming in 1989.

Dr.H.Schnelling. Universitätsbibliothek Giessen, D-6300 Giessen

NEET, Hanna E.: **A la recherche du mot clé. Aualyse documentaire et l'indexation alphabétique.** (In search of the keyword. Document analysis and alphabetical indexing). Genève: Les Editions IES, Institut d'Etudes Sociales 1989. 187p. ISBN 2-88224-014-7 = Les Cours de l'I.E.S., 2

Hanna E.Neet, lecturer at the library school in Geneva and author of "Assoziationsrelationen in Dokumentationslexika für die verbale Sacherschließung" (Geneva 1984) presents a concise introduction to document analysis and alphabetical indexing; she gives an overview of the subject, based, for the most part, on a lecture script put at her student's disposal. Ms.Neet's book is a quick, sometimes even hasty tour d'horizon, reviewing the major instruments of subject indexing (such as alphabetical subject catalogue, thesaurus, KWIC-index, KWOC-index, abstract, online searching) and discussing most of the central concepts and categories (e.g., terminological control, difference between pre-coordination and post-coordination, permutation). It is, in other words, a rapid overview in which "categorisations fines" (p.135) are mostly spared.

The bilingual Franco-German origin of the Swiss author and her legitimate recognition of the powerful Anglo-Saxon influences on international library research account for a welcome internationalism necessary for the multilingual information, documentation and library landscape of post-92 Europe. The technical terms, for example, referring to the semantic relations between the descriptors of a thesaurus, are given in English, French and German (p.133, 136). Unfortunately, such a polyglot view over the borders of national traditions of subject indexing is rather atypical of library research, and, all too often, the far too willing acceptance of the Anglo-American hegemony within international information and library science tends to suppress the urgently needed, careful consideration of all systems of subject indexing not germane to the librarianship of the United States and of Great Britain.

Freeing herself (and her readers) from the shackles of provincialism, the author displays a large, international set of systems of subject indexing such as the French subject authority file "Répertoire alphabétique de matières, encyclopédique, automatisé et unifié" (R.A.M.E.A.U.), the German "Regeln für den Schlagwortkatalog" (RSWK), the British PRECIS (chapters VIII-X). Unfortunately, the presentations of individual codes, authority files and alphabetical subject indexes are lined up in a rather incohesive fashion. Comparisons and cross-references are missing and the requirements of comparative librarianship are hardly met. From time to time the transnational bias of the script, however welcome it may be,