Holger Nohr
Technische Universität Hamburg-Harburg

The Training of Librarians in Content Analysis

Some Thoughts on Future Necessities*

Nohr, H.: **The training of librarians in content analysis. Some thoughts on future necessities.**
Int.Classif.18(1991)No.3, p. 153-157, 12 refs.
The training of librarians in content analysis undergoes influences resulting both from the realities existing in the various application fields and from technological innovations. The present contribution attempts to identify components of such training that are necessary for a future-oriented instruction, and it stresses the importance of furnishing a sound theoretical basis, especially in the light of technological developments. Purpose of the training is to provide the foundation for 'action competence' on the part of the students. (Author)

1. Tension field between needed theoretical foundations and existing practical requirements: An introduction

Probably no subject in the training of librarians[1] has been discussed and written about as much as that of content analysis. While some training institutes use the term 'subject analysis', we will abide in the following by the expression 'content analysis', thereby already telling something about the components of this teaching subject. The fact that - in comparison with other subjects - training in content analysis is being and has been extensively discussed finds its explanation in two major, if in part oppositively-directed influences exerted on this subject field.

1) The influence of theoretical orientation

The theoretical foundations of such matters as conceptology (2), classification theory (3), the ordering of knowledge (4), linguistics[2], etc. are of major importance for a proper understanding of content analysis and therefore need to be taught. But to what extent?

2) The influence of practical orientation

The practice imposes requirements on training. Beginning librarians should be able to work in some measure with the systems and regulatory works in use. But in what measure?

The influence of practical orientation gained in weight in (West) Germany in the course of the 80s. This was a result of two developments that have found much attention in practice and will play a major part in practical content analysis in libraries in the 90s, namely: one, the elaboration of the 'Rules for Subject Cataloguing' (Regeln für den Schlagwortkatalog, or RSWK) as an instrument of central and cooperative content analysis in scientific and public libraries, and two, the increasing adoption by libraries of online applications (OPACs) as well as of other technological innovations (5).

We therefore find: Training in content analysis is exposed to the tension existing between the need for sound theoretical foundations[3] and the requirements of the practice (i.e. of the practical applications). Now as a result of the aforedescribed development in library practice - to which training of necessity had to adjust - syllabi threaten to become unbalanced. An unbalance existing even despite the impossibility to conduct adequate training in some fields, e.g. in content analysis in OPAC catalogues. The reasons will be discussed later.

Please also note: If in this contribution we speak of the training of librarians, we refer in principle to an integrated training of librarians both for scientific and for public libraries. Such integrated training will in particular be asssumed for the field of content analysis. Details would carry us too far, please consult the literature[4].

In the following sections we will try to deal more intensively with the relevant courses and to bring both influences into balance (if attainable at all) without treating them in every detail.

2. The practical orientation of training

Besides the two developments already mentioned in the field of content analysis in the library, there are undoubtedly others as well. But on the German library scene it is the elaboration of the RSWK rules (for subject cataloguing) and the conversion of library operations to OPACs which have acquired outstanding influence. An influence which should be regarded as 'quantitative' rather than 'qualitative', meaning that both instruments have become fairly widespread in German libraries without necessarily raising the level of content analysis. The reasons may become clear from Section 3.

Innovations in practical applications as well as technological innovations and their dissemination exert an influence on training, on syllabi and - as deserves to be particularly stressed here - on teaching methods and on forms of courses.

In the RSWK rules we have, for the first time on the German library scene (7), a regulatory handbook for content analysis - in this case for the elaboration of

subject catalogues - which already now has come into wide use among the libraries and shapes the planning of many libraries and library associations (cf.for instance the qualification requirements specified in many Help Wanted ads). Now what consequences does this have for training? Should we teach the RSWK rules in the same manner as this is done nationwide in (West) Germany with the nationally uniform RAK rules (Rules for alphabetric cataloguing) in formal analysis, hence teach exclusively one single regulatory handbook and its application? To give the answer right away: No, neither RSWK teaching nor teaching in content analysis should look like this! And there are sound professional reasons for this answer, reasons that go beyond the time limitations existing for the teaching of this subject in overall librarian training.

Content analysis is more than just the RSWK rules, more than a single regulatory handbook, even if this handbook should become nearly as important as the RAK rules. The RSWK rules offer only one possibility of verbal content analysis, only one possibility of solving a content analysis problem; besides them there is also classificatory content analysis, which, with the advent of online catalogues, will grow in importance. At this point we will not repeat existing criticism of the RSWK rules, as the positions are sufficiently well known. What matters, rather, is that such criticism exists, that - evidently for methodical reasons - the RSWK rules are not unchallenged. Also important is the fact that there are alternatives, alternatives also from the field of verbal content analysis. And it is important, finally, that the RSWK rules - like any other system of regulatory handbook - are not immutable once and for all. All this means: there is, in content analysis, rarely a 'True' or 'False', but mostly only a 'Better' or 'Worse'. This must be taken into account in training. Syllabi must include alternatives, must show where existing regulatory handbooks have their weaknesses, must try to furnish the basis for 'action competence'.

Now what is 'action competence'? The graduates of a training institute for librarians should acquire the competence to apply in their field of practical activity - be it a public or specialized library, a private documentation center or the preparation of a selective bibliography (the list of possible fields may easily be expanded) the knowlege they have acquired, and to apply it to specific problems. This does not and cannot mean, however, that during training a maximum number of application fields should be gone through and of practically existing solutions should be presented and 'learned'. This would be a foolish undertaking. We must abstract, i.e. limit ourselves to imparting the essential foundations of verbal documentation languages (keyword languages, thesauri, index compilation, etc.), of classificatory analysis systems (universal, special and other classifications) and of mixed methods. And these we must impart in a way which is independent of special application fields but can be applied to them nevertheless. In this process, actually existing applications (besides the RSWK rules: e.g. the General Classification for Public Libraries (ASB), the UDC, existing thesauri) are not so much taught as introduced as concrete examples of fundamental phenomena having been discussed in this connection.

What we have said so far - in summary: a fairly strict limitation, after all, of practical influences on training in content analysis - will now be supported by a further argument: Any course of study should also pursue the principle of carrying the field of concern ahead, of pointing to new directions, of laying the foundations for new practical applications, at least of discussing them. If we limit ourselves in training to the teaching of existing concrete applications we will find out that no new innovations will be forthcoming from the new librarians thus trained. However, we cannot be satisfied in our libraries with things as they are. If e.g. we take the criticism of the RSWK rules seriously, it seems logical to suspect that there were shortcomings in the theoretical foundation of their elaboration.

Let us look now at the second type of development mentioned in the beginning: the technological innovations. The conversion of library catalogues from conventional card or COM catalogues to OPACs (let CD-ROM technology be included here) confronts training with problems of a wholly different nature. The new catalogue technique, expanding as it does the possibilities of access, potentially constitutes a major improvement of the services rendered[5]. PC-assisted databank systems enable small libraries, too, to perform fairly complex content analysis (9). The technology of knowledge-based systems will make its entry into libraries and documentation centers, and the expansion of this knowledge base is a genuine problem field of content analysis.

Here, now, we are faced with major problems. The methods of content analysis in the OPACs realized so far are traditional in nature, i.e. were developed for conventional, linearly-ordered catalogue media. The simple adoption of online systems does not yet produce online-adequate, content-oriented access possibilities (10); methods of content analysis need be viewed under a completely new aspect when seen against the background of the online environment. A pertinent example of current interest is the discussion around the application of the RSWK rules in online operations, in post-coordinated retrieval according to precombined keyword strings. In a future-oriented training this aspect must be given particular attention. thus, e.g., in traditional systematic catalogues it is immaterial whether we use enumerative or structure-mapping notation systems, whereas in online retrieval this is wholly different. Today we cannot conduct training on and for catalogue forms which in the future will be encountered more and more rarely in library practice. In other words: in the development of OPACs in our libraries the graduates of the coming years will play an essential part, especially in smaller units. The foundations for this must be laid in

today's training.

This development in the practice must of necessity lead us to other forms of teaching. We must conduct training on online systems, teach the fundamentals and particularities on online retrieval. Of importance here is a familiarization with the 'behavior' of online systems being accessed, especially with a view to the use of particular aids (Boolean operators, truncations, etc.) and the effects of these aids on retrieval results in their application to natural-language designations. Now the question presents itself whether this should form part of the course in content analysis? The answer should be affirmative. System 'behavior' in online retrieval - with the help of certain aids - according to content analysis data, frequently controlled vocabularies, should not be separated from the analysis process itself. Content analysis and the access to the data obtained through it form a unity which in training should be treated as such.

If we accept these contents - contents based on online systems in the library - into our training, this must also find expression in specific forms of courses[6].- Training must not be merely theoretical, it must take place on the OPAC or other device, at least in order to demonstrate examples. This, however, presents difficulties: only few library OPACs are accessible via the DATEX-P network; the availability of devices is deficient here and there; information from the practice is not flowing in sufficient measure to afford the teacher an adequate overview. Practical exercises on the devices are necessary for the study of the 'retrieval behavior'; theoretically this can be done only to a limited extent. These practical exercises must be offered in addition to the traditional classes, as they require smaller groups and more intensive attention to individual students, e.g. in thesaurus work with PC-assisted software.

3. Theoretical orientation of training

In the preceding chapter we already saw that there must be limits to the practical orientation of training. Training must be oriented to the practice without teaching the situation 'as is' exclusively.

Now how about the theoretical orientation? In our content-describing catalogues we retrieve potentially the contents of all literature (but also of other information carriers) and thus potentially all knowledge. This knowledge may also be found in works of fiction (belles lettres). We can make no predictions as to in what framework or direction our knowledge will enlarge. Within the framework of training in content analysis we are therefore obliged to pay attention to the organization of knowledge (rather than of books of or simply existing contents). This means, among other things, that the classification systems employed in libraries must be 'knowledge classifications'. To the extent that we speak of 'library classifications' this can merely mean to express the application field, not a particular type of classification[7].

In the pre-language stage, knowledge is 'represented' in concepts; communication on knowledge and concepts takes place by means of natural language. Concepts are thus represented by designations in natural language or, as in classification systems, by a unequivocal sign system (notational system). Knowledge classification systems may therefore also be termed concept classification systems (11).

For documentation languages of a verbal nature the same is true in principle: the verbal designations of these languages likewise represent concepts. The verbal designations of these artificial languages derive, however, from the natural language; they are, therefore, not unequivocal. Linguistic problems in content analysis are the result. The methodological bases, however, regardless of whether we treat verbal or classificatory systems, is - with certain exceptions - identical.

For the training in content analysis we obtain from the above, in summing up, questions from the following three fields:

1) From the *field of philosophy*: fundamental questions of knowledge organization (world order, reality), e.g. investigation of the system concept and of the classification concept (12).

2) *Conceptual foundations* of the ordering and the ordered presentation (concepts and their relationships) of knowledge. Importance and identification of categorial (semantic) concepts for content analysis purposes[8].

3) *Linguistic foundations*, e.g. synonymy, near-synonymy, homonymy, polysemy, grammar as regulatory system of a language (e.g. syntax, fields of morphology), etc., the relationship between concept and designation as well as the treatment of artificial languages.

In the introduction we spoke also of the fundamentals of classification theory; meanwhile we have seen, however (in Sect.2), that a practically oriented training means that matters of principle should be abstracted from and be demonstrated by means of concrete examples. This, in combination with the above points 1) and 2), comprise these necessary fundamentals.

If a study of library science is divided into basic and advanced courses[9], some of the theoretical fundamentals outlined in the above will have to be taught in the basic stage to form the basis for a proper understanding of subjects taught later (e.g. large parts of the conceptual fundamentals), while others can serve as an introduction to the field of content analysis as such (questions of the ordering of knowledge).

It presents difficulties to determine (and specify) in detail what in effect the theoretical part of the syllabus should acquire. In the training stage, practically-oriented and theoretical contents often cannot be clearly separated, nor is it necessary here to do this. The conceptual foundations of content-analyzing activities (to use these as an example here) will figure prominently throughout the entire study, but will have to be centered in the basic stage. They form the foundation for the

entire subject field, e.g. for class formation or the identification of relationships between classes or descriptors. It is recommended to treat this field in detail, covering also DIN Standards 2330 and 2331 as well as further relevant technical literature and sufficient technical exercises. At this stage linguistic foundations also play their part, e.g. the relationship between concept and designation and the problems connected with the natural-language representation of concepts.

4. Subdivision of the subject 'Content Analysis' in the training of librarians

The subdivision of the subject 'Content Analysis' within the framework of the training of librarians can only be summarily outlined here. For more detailed information a look at the *Materials for Content Analysis* (cf. Annex) is recommended which were compiled for use in training at the Hamburg Polytechnic for Library Science. The table of contents of the various volumes represents at the same time a subdivision of the subject matter to be taught.

It is a problem indeed that there does not exist any textbook for content analysis in libraries with special regard for conditions in Germany. The *Materials for Content Analysis* listed in the Annex should be regarded as a substitute for such textbooks, they comprise the entire width of the subjects to be taught. Examples for demonstration (such as sections of classification systems, thesauri, retrieval examples) are contained in them as well as texts and exercises. The volume for the basic course (a) offers an introduction to the subject, to cataloguing, to the conceptual foundations, to verbal and classificatory content analysis (next to the UDC and DDC also an introduction to systems and methods of relevance in German public and scientific libraries), and content analysis of fiction.

For the advanced courses three volumes have been prepared according to the contents of the courses: classificatory content analysis (b), verbal content analysis (c), and OPACs (d). The foundations of coordinate and syntactic indexing are treated in (c), as well as the history of the subject (headings) catalogue, the RSWK, PRECIS, and also the important work of the British Classification Research Group (CRG), and the thesaurus principle, accompanied by different examples.

Volume (b) offers an introduction into the principles of faceted classification, inspects somewhat closer the work of the British CRG, gives an introduction into Colon Classification and Bliss Classification as well as different special classification systems, and an excursion to historically interesting classification systems.

The application of verbal and classificatory methods of content analysis and access to these data are treated in volume (d) which includes also techniques and problems of free text searching. Some OPACs are presented also.

Each volume contains a selection of important texts, (e.g.standards) and includes also a comprehensive list of references for further study.

In the *basic courses,* primary attention is paid to theoretical fundamentals.

- Organization and ordering of knowledge, the system concept (can be regarded as some sort of introduction to the entire field as such),
- Subject cataloguing: forms and types of catalogues, order in catalogues, retrieval in catalogues, indexes and inverted lists, fundamentals of online retrieval,
- Conceptual and linguistic foundations of content analysis,
- Introduction to verbal and classificatory content analysis (e.g. Facet classification, UDC, RSWK, PRECIS, etc.),
- Historical problems: catalogue history, disputes around alphabetical subject vs. systematic catalogues, etc.)

Computer-relevant subjects are not so much stressed in the basic course - especially not at its beginning as in the advanced courses. The subjects 'Content analysis', 'Introduction to Electronic Data Processing', or 'Introduction to online-supported retrieval' demand primarily a teaching of basics, such as an introduction to the handling of the devices. Admittedly, already during the first semester online catalogues are taught in the course on cataloguing (however, more on their existence than on their functioning). Computer courses should be limited to their specific domain; applications such as also content analysis must be treated in the corresponding courses.

The *advanced courses* have the task of deepening and expanding the foundations of verbal and classificatory content analysis laid in the basic ones. Here, the given subject matter is to be gone into more deeply, with due and increasing consideration of the world of online catalogues and co-operative content analysis. In reaching beyond the present state of affairs, desirable directions of development and alternatives for the future should likewise be outlined.

So far we have roughly outlined what subjects should be taught in those courses and classes which we might term *required courses*, hence those that are mandatory for all students. Special subjects can be offered as optional subjects and in seminars in which specific methods and regularitory handbooks can be treated in detail (e.g. RSWK, the AMP Method for the analysis of fiction - A.Mark Pejtersen, PRECIS, Eppelsheimer method, UDC, and many others).

At the Library and Information Department of the Hamburg Polytechnical Institute the following quantitative conditions are available for training in content

analysis: in the three-term basic study, two class hours per week (altogether thus 6 hours per week) are obligatory. One semester comprises the time of 4 months, the time for the courses, however, is usually only 13 weeks or sometimes even less. In addition practical exercises supplementing the basic courses taught may be optionally chosen. This basic study s concluded with an oral examination in the subject.

Basic and advanced study are separated by a term (a semester) in which practical work is done.

In the advanced part of the study (also three terms), two courses of two hours per week must be taken obligatorily during each term, which in addition practical exercises in preparation of the oral examination concluding the training in the subject may be taken voluntarily.

In addition, seminars and exercises on various topics of the overall subject (Content Analysis) and of various scope are offered which can be attended at every stage of the study.

Notes

*Revised text of a contribution in *Auskunft* (Information), Newsletter published by Hamburg Libraries 10(1990)No.3, p.272-281

1 To understand the author of such a training cf. (1)
2 cf.(2)
3 cf.(3)
4 See (6), esp.p.191-194 and (1)
5 See (8) as well as its comprehensive bibliography
6 cf.(1), p.1045-1046
7 cf.(3), p.77-78
8 cf (11)
9 cf.(1)

References

(1) Gödert, W., Nohr, H.: Bibliothekarisches Fachhochschulstudium im veränderten Umfeld. Bibliotheksdienst 23(1989)No.10, p.1027-1048
(2) Greiner, G.: Sachliche Erschließung und Ordnung als Unterrichtsinhalt bei der dokumentarischen Aus- und Fortbildung. In: Studien zur Klassifikation, Bd.16. Frankfurt: Indeks Verlag 1986. p.64-72
(3) Gödert, W.: Gedanken zu Ausbildungsinhalten im Fach bibliothekarische Sacherschließung. In: Studien zur Klassifikation; Bd.16. Frankfurt: Indeks Verlag 1986. p.73-83
(4) Sträter, H.: Structures of knowledge, information transfer and library education. In: Rajagopalan, T.S. (Ed.): Proc.Int.Conf.on Ranganathan's Philosophy, Assessment, Impact, and Relevance. New Delhi: Vikas 1986. p.615-621 (the book is available via Indeks Verlag, Frankfurt)
(5) Nohr, H.: Subject access in German-language OPACs: A survey. Int.Classif.16(1989)No.2, p.95-100
(6) Gödert, W.: Educating the public librarian, educating the special librarian - is there a common future? Educ.for Inform. 5(1987)p.187-197
(7) Neue Regelwerke zum Schlagwortkatalog: Einführung in RSWK und PRECIS. Köln: Greven 1985. p.1-25
(8) Nohr, H.: Sachliche Erschließung in Online-Publikumskatalogen. Beschreibung und Analyse ausgewählter deutschsprachiger Systeme. Hannover: Koechert 1989
(9) Nohr, H.: Thesaurusarbeit mit LIDOS. Nachr. Dok. 42(1991)No.3, p.217-220
(10) Nohr, H.: Klassifikation als Basis einer Sacherschließungs-komponente im OPAC. Anforderungen an Elemente und Struktur. Bibliotheksdienst 25(1991)No.5, p.664-670
(11) Weinberger, O.: Begriffsstruktur und Klassifikation. In: Studien zur Klassifikation, Bd.9. Frankfurt: Indeks Verlag 1980. p.316-326
(12) Diemer, A. (Ed.): System und Klassifikation in Wissenschaft und Dokumentation. Meisenheim am Glan: Hain 1968.

Annex

Materialien für den Unterricht (Materials for Courses)

(The following references to 6 volumes "Materialien zur Inhaltserschließung" (Materials for content analysis) have all been issued by the Fachhochschule, Fachbereich Bibliothekswesen, Grindelhof 30, D-2000 Hamburg 13, Germany. They are originally in German, we are citing them in English only)

(a) Introduction to content analysis. Compiled by W.Gödert in collaboration with C.Hübscher and H.Nohr. 1989. 400p.

(b) Rules for the Subject Catalogue. Compiled by W.Gödert. 1989. 152p.

(c) Classification systems: Theory and application for shelving in scientific libraries, systematic catalogues, bibliographies and online-retrieval. Compiled by W.Gödert, ed. by H.Nohr, 2nd ed. 1990. 2 vols. together 419 p.

(d) Methods and procedures of verbal content analysis and their application in libraries. Comp.by W.Gödert, ed. by H.Nohr. 3rd ed. 1990. 354 p.

(e) Subject retrieval in the OPAC. Interplay between documentation language, indexing principle, data base construction, and retrieval language. By W.Gödert. 1989. 37p.

(f) Online Public Catalogues (OPACS). Methods of content analysis in online public catalogues. Comp. by H.Nohr. 2nd ed. 1991. 158p.

Address: Holger Nohr, Dipl.Bibl.,
Technische Universität Hamburg-Harburg, Arbeitsbereich Städtebau II, Schwarzenbergstr.93c, D-2100 Hamburg 90 and Fachhochschule FB Bibliothek und Information, Grindelhof 30, D-2000 Hamburg 13.