



ASp
la revue du GERAS

4 | 1994
Actes de l'atelier New Technologies du congrès
ESSE/2

Organising information for multimedia learning

David Bickerton



Electronic version

URL: <http://journals.openedition.org/asp/4146>

DOI: 10.4000/asp.4146

ISSN: 2108-6354

Publisher

Groupe d'étude et de recherche en anglais de spécialité

Printed version

Date of publication: 1 July 1994

Number of pages: 75-92

ISSN: 1246-8185

Electronic reference

David Bickerton, « Organising information for multimedia learning », *ASp [Online]*, 4 | 1994, Online since 20 January 2014, connection on 01 May 2019. URL : <http://journals.openedition.org/asp/4146> ; DOI : 10.4000/asp.4146

This text was automatically generated on 1 May 2019.

Tous droits réservés

Organising information for multimedia learning

David Bickerton

- 1 I shall begin with a kind of naive ontological query. What, we might ask, is the information I am about to deliver? Is it the notes of some key ideas written some time ago on two pieces of paper; or maybe the word-processed markings on sheets produced before leaving Plymouth; or is it not such physical data at all but rather the ideas in my mind provoking, if we are fortunate, similar ideas in your minds? And, supposing “my paper” exists before I have said it today, and that it lives on even after it has been said, is it my information that I am delivering or rather the thoughts of others that I have acquired for myself and transformed?¹
- 2 This must seem a very philosophical opening for a paper in the New Technologies section of the ESSE Conference, but in fact my paper serves as a paradigm for the information we now use in the world of education and, par excellence, in language learning. Whilst we recognise that languages are best learned through contact with human agents, in reality learners of second languages make far greater use in our own society of “documents”, or what in France I have heard being called “*documents intégrés*”. We might perhaps think of these entities as simply language objects if we wish to avoid the term “document” and get away from that term’s paper-oriented connotations. However, the etymological roots of the word are in instruction, which is very suitable for language learning applications, and here I shall use the term “document” to refer to any structured physical data chosen to promote learning and mediated by technology, irrespective of their acoustic or visual properties.
- 3 Many of these learning documents have very doubtful existential authority if judged from a conventional standpoint, and little real stability. It is a belief widely shared, indeed, that this instability will grow and that within fifteen to twenty years the libraries we know today will have developed a dual role: as historical repositories for pre-21st-century documents, and as management centres for the new technologies required to handle the learning that concerns us here. This learning will be based almost entirely upon the transient and dynamic data of “documents” like this paper in its different stages of

gestation. We are all aware of presently living through the most significant change in information management since paper-based printing began. As a result, we need to adjust our institutions (our resource collections, laboratories, libraries, “*espaces langues*”, mediatecs) so that they can handle radically new forms of information, information in multiple media, retrieved according to unpredictable criteria, data that is dynamic in their nature - that is to say assembled during the process of interrogation itself.

- 4 This paper is, therefore, an attempt to describe a number of ways in which we currently organise information for language learning and to clarify the issue of how best to handle the information we do and shall require in multimedia-based work. What is that information; how do we process it, locate it and retrieve it? And how will the new technologies challenge our present procedures?

1. The data of language learning

- 5 In the present context “information” is a catch-all noun describing several parts of a complex communicational process.² The communication of meanings depends upon shared knowledge and conventions; “information” is conveyed using a medium like paper and within a context like social relations. The medium and context affect both the encoding and the decoding of the message. They also change in time: media change as technological systems develop, and linguistic contexts change with the world around us. Thus, the process of matching information to a learner’s needs is hedged around with difficulties and one can see the benefits of rendering data interactive and dynamic — i.e., responsive in time and place to a declared need.
- 6 For present purposes it seems sufficient to draw a simple distinction between physical and non-physical data. Physical data have a material existence in the form of patterns (e.g. on disks or on paper stored on shelves); they are, perhaps, to be found also in brains, although we cannot yet access the brain’s storage system directly. There is no need to list here the combinations of physical media or formats like tape, paper and hard disk segments with which we all work and which we use to transmit and store data.
- 7 On perception by a user, the physical data is interpreted and can trigger meaning. We must assume that to all intents and purposes “meaning” is non-physical. Non-physical data are created in people’s minds - they are the “content”, “subject”, and “value” (e.g. suitability for purpose) attached to or represented by the physical data. They depend, in a situation of communication, upon an interpretation of physical data by a user.
- 8 Viewed like this it will be clear that we have multiple issues to clarify when considering information in a language learning context. In the first place, there is manifest incompatibility between the subtle ways we process ideas in our minds and our crude methods for organising and retrieving data in a physical medium. The whole learning process is tied to the physical attributes of data as constrained by a physical or technological medium. The physical format imposes storage, access and retrieval methods unsuited to many of the learning uses that can be made of the data. One may think, for example, of the problems of linear access to documents held on microfilm or videotape, where non-physical data can only be accessed by serially reading all the physical data. This seems a very unnatural process compared with how learners’ minds actually operate.

- 9 Secondly, there is the question of data complexity. Non-physical data in language learning applications is potentially very complex – perhaps more so than for any other information retrieval application. The reason is that a wide range of complex documents (from purpose-designed published courses to improvised simulations of language and almost any kind of linguistic realia using sound, image and/or text) are used by learners not only for the subject information they contain but also for their linguistic or other encoded features as well as for their translation and reformulation potential. The linguistic interpretation of physical data is also multi-layered: the L2 features of any learning document are interpreted by a user in terms not only of L2 but also of L1 equivalences or comparisons.
- 10 If we consider the well-established ways that learners search for learning data we can begin to analyse the potential complexity of our data. Experience dictates that the major search parameters used by language learners are as follows (see Table 1).

Table 1: Language learner search parameters

Parameter	Learner Search Objectives
linguistic	a document's language is searched in respect of structural, formal or systematic features: e.g. its grammar, vocabulary and orthography or phonetics
pragmatic	a document is evaluated in terms of its use-related features: e.g. formality, register and style, and intercultural features
psycholinguistic	the document is judged for its appropriacy for the learner's aptitude, motivation, learning style, L1/L2 experience and cognicity
didactic	the document is selected for its methodology (for courses or exercise types [e.g. encouraging active or receptive competence, drilling with morpho-syntactical patterns]) and the learning level of the material
functional	the document is analysed for its appropriacy to the learner's communicational purpose: e.g. the achievement of self-enrichment, to do a job, to read for research, to acquire a specific skill such as reading or writing
thematic	The document is selected in terms of its "subject" orientation

- 11 Finally, the multiplicity of physical media and the complexity of the non-physical data are compounded by the fact, referred to above, that many non-physical data are open to subjective interpretation and therefore vary from user to user. This is less the case for linguistic, pragmatic and thematic features, where broad consensus can exist, but the psycholinguistic, didactic and functional features of documents are largely user-determined. Any categorisation of them must therefore be derived as much from data on the user as from that on the document.
- 12 To re-capitulate the issue facing us as organisers of information for multimedia language learning, we might say that the physical data to be used by a learner are extremely varied, ranging from published and structured courses containing multiple media to

virtually any kind of realia containing language. As for the non-physical data needed by the user, these are of great inherent complexity and also vary according to user perceptions. The “meaning” and “purpose” of documents are constrained, also, by their physical medium (which can change) and by categories we invent and constantly re-invent. Ideally, any learning system must try to accommodate change and different learner requirements, but in the case of multimedia data with its complex messaging potential and easy modification and re-purposing, a flexible strategy for information management becomes an urgent priority for all language teachers.

2. Information organisation to date

- 13 Many classification systems have been devised in order to link the non-physical data sought by users to the physical medium upon which they are encoded and stored. Most large library collections and national union catalogues record a varied range of formal data for different document types such as books, films, pictures and manuscripts. Unfortunately, the main features of documents which librarians identify for the purpose of cataloguing items in their collections are not those of interest to most learners. A formal, physical description is generally found, accompanied by details of title, publication and author(s) together with conventionalised representations of subject. This data contains or helps generate a classification for items – generally a subject-related classification – which places the item in a logical position within the conceptual universe represented by the whole collection. The classification is in turn used to generate a “notation” (often a mixture of letters and numbers) which is used in turn to physically locate the item. Although cataloguing can be lengthy, the data it generates is only used in small part for the notation. The notation is therefore reductionist, is generally rather opaque, and, in the case of closed collections, in particular, needs above all to be comprehensible to the guardian of the collection.
- 14 The serious problem for learners is that, by definition, they are unable to know exactly what information they require, however well motivated they may be. The language learner does not use a learning library with a list of references in hand, probably has little concrete search parameters worked out, and specifically requires information that will certainly not be represented in the notation and probably not be found in catalogue entries either. Libraries have consequently devised ways of responding as well as possible to learners’ needs by complementing their classification systems with catalogues of varying sophistication and by using different indexing techniques, all of which involve expressing the “subject” of the document with a few simple nouns (and sometimes adjectives and verbs) and deriving lists of keywords or phrases with or without context.
- 15 A brief description of a small number of collections quickly reveals how well or badly the features identified in Table 1 as being of major importance to the language learner are represented in (a) catalogue data, (b) classification schemes, (c) notations used to produce class and shelf marks, and (d) subject catalogues and indexes.

2.1. LC

- 16 A conventional MARC record structure adapted to a language learning collection and adhering to AACR-2 rules would probably contain the following main fields of information introduced by numbered tags:

- 17 (1) record control number (e.g. ISBN or Library of Congress card number)
 (2) language code (including information on mixed languages and translations)
 (3) title/edition/publication data
 (4) physical description (number of parts, play time)
 (5) series (title, part number)
 (6) notes (general or relating to contents, title, audience)
 (7) subjects (e.g. Library of Congress subject headings, period and place)
 (8) authors (persons or bodies associated with the document)
 (9) shelf number/location/invoice for document.
- 18 The Library of Congress catalogue establishes data for each of the above fields, as do the catalogues of most large reference libraries. The LC notation, however, as expressed in the card number, is totally opaque and data on any didactic, functional or linguistic features can only be gleaned from the record if they are implicit in a title or happen to have been noted by the cataloguer. Whilst this is what one would expect of a reference library too large to support open access, it establishes a paradigm which is unfortunately replicated in smaller collections.

2.2. CILT

- 19 One might anticipate that the reference library of the London-based Centre for Information on Language Teaching and Research (CILT), where users are able to browse, might come closer to producing a more transparent notation or at least begin to classify teaching documents in a way that reflects their real use by learners. The scheme, which grew, historically, from a collection concerned with the English language and its teaching, was extended later to accommodate teaching/learning materials as a separate main class (arbitrarily called the “U” class). Items falling into this class of “textbooks” are given up to three amongst eight main facets, each with its own schedule.³ In addition, there are several facets for language and place (e.g. French in India) which use Dewey 400 and 900 classes for greater precision. The scheme consequently produces the following examples of notation:

UZ 430 UBR computer software for German
 UZ 491.43 UDV an Urdu course for adults

- 20 This notation is so opaque that most users are heavily reliant upon the subject index or the librarian for finding what they seek. To complement its classification CILT has produced a brief (30-page) subject index containing approximately 750 headwords. This has been regularly revised. The proportion of entries of relevance to a learner is naturally quite small (since CILT is not intended to be a “learner” library) and holdings in the teaching materials section could not be located using the search parameters identified earlier.

2.3. UGLC

- 21 One of the most recently-founded large collections of language-learning materials in the UK is at the University of Glasgow.⁴ It was decided in the mid eighties that the University’s Language Centre (UGLC) should support easy and open access to its language learning collection by members of faculty and the public alike. An automated public lending system has been built, entering into full operation in 1992. The catalogue and

circulation control system offers self-drive check-out/check-in (relying upon supervision by security staff when library staff are unavailable) and uses a catalogue record containing much of the data identified earlier as being necessary for language learners.

- 22 Since easy open-access for browsing was deemed a priority, a simple and highly transparent dual facet classification was adopted consisting of a language code (GER = German; ITA = Italian)⁵ and a materiological code (A=audio recordings; S=short-life video recordings)⁶. A shelfmark based upon an accession number with a series of digits describing the number of “parts” in an item (e.g. book + 3 audio cassettes), completes the notation.
- 23 The whole focus of the UGLC library, unlike that of CILT, is upon active language learning by users. Since core facets should reflect a content analysis of the whole collection and suit its prime purpose, language and learning material type have appropriately been selected as the UGLC core facets. By 1990 some deficiencies had become apparent in the system (revealed by the rapid growth in size of the collection and new usage patterns)⁷ and this prompted a review of classification principles. It became clear that an improved learning material typology would be (1) information or reference, (2) language learning courses, (3) extension materials, and (4) realia (i.e. “raw” materials from the media and elsewhere, including all unprocessed off-air recordings).⁸ In addition, several new differential facets appeared desirable: topic, learning level, author/director.
- 24 The above modifications would entail considerable re-organisation of the UGLC collection and have not been introduced. In principle, however, such a classification system could, potentially, be used to produce facet strings as follows (see Table 2).

Table 2: Alternative UGLC classifications

Language	Material	Topic	Learning Level	Author/Director	Accession + Part Number
French	course	law	1	Beddoes	021 part 2
Basque		grammar	-	Worksop	002-

- 25 The notation for the above examples could be expressed as follows and be used to determine the shelving order:
- FRE.C/LAW1.BED/021.02
BAQ.I/GRA0.WOR/002⁹
- 26 The UGLC catalogue is fully computerised and has an extended MARC record structure with extremely fast searching. Users can retrieve entries using quite complex search criteria including language, learning level, materiological description, subject, skill aim and mode of discourse. However, the pre-requisite for finding data is that they have been entered into the records, and it has apparently proved impractical to generate complete records for many documents. For all intents and purposes, therefore, the collection cannot be searched using true learner-centred criteria.

2.4. SMILE

- 27 The Satellite Materials in Language Education Project has been running in Scotland since 1989 and is presently co-funded by most Scottish Regional Education Authorities and the

University of Glasgow.¹⁰ The Project selects small “chunks” of off-air satellite television recordings for their didactic potential at primary, secondary and tertiary levels. All chunks are held in a classified archive, the data consisting of extensive formal and sociolinguistic fields¹¹ and the recordings being held on compilation tapes. The database is accessible by subject keyword searches and other formal criteria, whilst the compilations are held according to 20 subject categories chosen to reflect the topics often used in language teaching materials today (see Table 3).

Table 3: SMILE compilation tape subject classification

1. Advertising	11. Medicine & Vet Science
2. Culture & Entertainment	12. Media
3. Economic Issues	13. People & Places
4. Education	14. Rest of the World
5. Environment	15. Science & Technology
6. Europe	16. Social Issues
7. Home Life	17. Sport & Activities
8. Ideas, Movements & Political Organisations	18. Time & Weather
9. Language, Numbers & Communication	19. Travel & Transport
10. Law	20. Other

- 28 Viewed as a subject hierarchy, the whole archive may be said to have a structure which is illustrated by the following examples (see Table 4).

Table 4: Subject hierarchy of the SMILE archive

Main Subject Category	Topic Categories	Keywords
Culture & Entertainment	arts, films, cartoons, theatre, literature, music	astrology, cinema, entertainment, horoscopes, music, quiz, songs
Economic Issues	unemployment, world of work	business, commerce, consumer, currency, international aid, jobs, interviews, prices, shopping
Home Life	the home, family, towns where you live	ages, bedroom, birthday, gift, housework, household, kitchen, parent, relatives, where people live

- 29 To date, keywords have been allocated freely by the teachers responsible for selecting the chunks, there being no suitable thesaurus or authority list. This has produced some duplication but has worked better than one might expect. Given that the archive consists of raw realia and has not been designed for learner access it is interesting to note that it is potentially better able to provide learner-centred data than most library catalogues. Its records contain data on language (+ dialect), date of production, learning level, pragmatic features, some linguistic features and theme.

2.5. CML

- 30 The newly founded Centre for Modern Languages at the University of Plymouth has recently introduced a new classification system for use in its library of language learning resources.¹² The scheme's design took into account the approaches to information management described above, and sought to balance the following (partially incompatible) desiderata:

- transparency and expressivity of notation, in order to offer support for open access shelf browsing
- simplicity and hospitability of notation, in order to facilitate collection management
- capacity to describe and discriminate between documents according to the "language learner search criteria" identified earlier (see Table 1)
- ability to accommodate a wide variety of non-book materials, including unpublished or "home-grown" extension materials for multimedia resources
- potential to allow the collection to develop and change

- 31 The resulting classification scheme uses two core facets, language and subject, with subject being defined by the following schedule:

- A Reference
- B Language courses and learning materials
- C Country: Economic, Political and Social Background
- D Culture: principally the Arts and Entertainment
- E Current Affairs

- 32 The language facet uses the UK MARC notation. Any item for which no single language code is appropriate is placed in a general (GEN) section, for which a quite separate schedule has been established.

- 33 Schedules for each of the main subject sub-categories effectively incorporate differential facets responding, where most appropriate, to learner search criteria:

- learner level (a typology is used in category B)
- topic or theme (a typology is used in category B)
- title (3 letter code used in categories A, B, C, D (except for literature) and E)
- author (3 letter code used in literature, category D)

- 34 In the cases of learner level and topic, the facet is embedded numerically in the classmark; title and author codes, however, are effectively appended to the classmark to form a shelfmark. The resulting classification scheme produces notations such as the following:

- FRE.A.11 Pet (French / Reference / Dictionary / Monolingual / Petit Robert)
- GER.B.32 Ein (German / Language Course / Upper Secondary / General / Einsicht)
- SPA.D.22 Cer (Spanish / Culture / Literature / Prose fiction / Don Quijote de la Mancha by Cervantes)

2.6. Visual resources

- 35 What none of the preceding classification systems can do, however, is offer pointers to the possible meaning of images. A method of subject analysis for images, elaborated by K. Markey, uses cluster analysis on text to define images and thereby generate a subject index. The process relies upon detecting linguistic “similarities” between simple verbal and noun phrases used to describe objects, expressions and actions (events) in images. Clusters of phrases are formed computationally, and subject specialists are then required to link them to symbolic meanings. (Thus “the man nailed on a cross” may be identified as representing “The Crucifixion”.) Thus a subject retrieval system is generated which can allow searchers ignorant of the symbolism in images to access collections via concordances.
- 36 An interesting feature of the approach is the required use of non-specialists to generate the descriptive phrases. If artists or photographers are seeking to convey visual symbolic meaning to receivers of their images, they presumably assume their audience will understand the symbolism and not simply see semantically primitive entities in the pictures. It may not appear very convincing, therefore, to build a whole model of analysis of visual meaning upon ignorance (or, rather, upon the different cultural preconceptions of those who are ignorant of a given symbolism: the symbolical world of, say, mediaeval religious iconography).¹³

3. The holistic and dynamic alternative

- 37 It seems unlikely, on the basis of the systems described above, that we can successfully use small, hierarchical sets of descriptors (words) compressed into a code to retrieve the very complex non-physical data we require as language learners. The match between learner needs and library classification is simply too imprecise. There is equally little likelihood that we can use plain and simple descriptive phrases to encapsulate the non-physical data in moving images. It had become clear by 1990 that the advent of multimedia data in digital form exchangeable electronically between users in different sites meant that an alternative approach could be considered. In October 1991 the practical problem of language learning materials exchange between European Language Centres in Higher Education was discussed at an international conference in Bologna and a group of Universities established the MELODI Project (Multimedia Language Object Definition and Implementation Project, Glasgow 1992) in order to achieve the following objectives:
1. portability of electronic materials between sites and computer platforms by agreeing formats for the linguistic, pragmatic and pedagogic tagging of multimedia documents
 2. escape from the classificational reductionism currently used in describing learning resources (as described in section 2)
 3. achievement of a means of representing and storing learning data which can be viewed from a user-centred perspective
 4. establishment of a means of enabling non-published, informal learning materials to become widely accessible to teachers and learners.
 5. development of a series of software tools for presentation and analysis of documents.¹⁴

- 38 To this end several proposals have since been written attempting to implement different aspects of the MELODI programme: SMART-TALK (Pisa, 1992)¹⁵ focussing upon the exploitation of linguistic corpora, HARMONI (Glasgow, 1993)¹⁶ addressing modes of interchange between Language Centre Libraries, and STACCATO (Antwerp, 1993)¹⁷ dealing directly with language-teaching/learning needs. STACCATO has now received significant financial support from the LINGUA programme and work on devising and implementing the coding scheme for multimedia language documents which had commenced at the MELODI Conference in Glasgow in April 1993 is now set to move much further forward.
- 39 Core partners in MELODI are the Universities of Antwerp, Glasgow, Pisa (including the Centre for Computational Linguistics) and Plymouth, and there are beta testers in France, UK, Germany, Switzerland and Spain ready to evaluate the initial coding schemes and software products that will flow from the Project.¹⁸
- 40 The kind of documents being created today for multimedia education are central to the alternative solution (to that of using traditional classification systems) at the heart of the MELODI programme. It has long been the case in language education that most input materials are either self-made or modifications of existing realia. Linguistic realia offer an authentic, plentiful and renewable resource which is vital in a sphere where the medium and context of communication change so fast. And even when language teachers select materials from structured and published courses it is customary for exercise extensions to be developed and for modifications to be made. In fact, one can begin to identify a developmental pattern behind the generation of language learning documents: a piece of realia, or an item of manufactured language created to look like a piece of realia, is selected and forms the core to which exercises, explanatory material and other stimuli are added, often over a period of time. The core is often television material, or film (on video), and the support materials use text and sound. This process is also common to the generation of much general multimedia-based educational material today, where the availability of hypertextual programming tools encourages a constant process of change and growth in which learning materials are never “complete” but rather evolve. The situation is, of course, helped by the relative ease with which electronic data can be copied and re-purposed, and it is this which encouraged me to question in my introduction the very status of information as a complete and well-defined entity.
- 41 It is around this paradigm of the language learner’s document that the MELODI Project has developed its approach. It adopts a holistic view in which entities are seen as part of a group of linked items forming large and extensible documents. This is done by embedding tags within entities themselves so that a potential user can extract information about the data from the data (rather than from a file of metadata separate from the data) and link them dynamically with others. Thus, a film sequence may be tagged so that the shot sequences and edit decisions are stored with the images. The script, or a transcript of the dialogue, or a series of script director’s notes could be equally be attached to each element in the film sequence. There is no necessary pre-supposition, however, that the shots list or the script are held in the same file of data or location as the images, so long as the image file is organised according to an agreed format and can be referenced unambiguously. In this way an image data file (first created in Pisa, for example) can “know” where to find related script data (in Plymouth) and also “knows” the nature of the relationship. A collaborating institution in Paris devising an exercise based on the image file must subsequently be able to append references to its new data in the master file on the host machine in Pisa.

- 42 In the case of a more elaborate production – e.g. a fully developed course comprising stimulus materials in text, sound and image, exercises, explanations, glossaries and grammar - the process is identical. The course will necessarily contain a structure. If the course is accessible on a digital medium, the structure will be explicit in the text mark-up and indexing used to define the whole document. It may be that some important features of entities within the course will require further analysis in order to establish, say, data on theme, grammatical structure or lexical density. It is a major objective of the MELODI Project to develop a suite of software tools to facilitate the automatic analysis of “untagged” data.
- 43 In this way it is hoped that we shall be able to move away progressively from our present reliance upon rigid and reductionist representations or notations for obtaining non-physical data. Instead, the user will be given access to the data within the limitation of the access speed of his or her computer link. This is not to say that conventional subject access via an index becomes redundant. Where large numbers of documents are available it is most efficient to scan a simple subject analysis code in order to identify a suitable match for one’s search criteria. However, once having identified potential choices one must then be able to conduct further analyses of the data as a whole. This requires the whole data to have been tagged in a way that supports different forms of enquiry.
- 44 It also seems important that the process of accessing information should involve providing the user with “expert” tools to enable him or her to participate dynamically in the retrieval process. Meeting a requirement can become an exploration so long as the prime data is organised in a way to support the analysis. In this respect we should try and replicate the way in which expert judgements are formed. If a learner is trying to assess the “difficulty” of a piece of stimulus material which seems to meet other search criteria he should not be directed solely to the learning level field in the document record for an authoritative answer. Rather he must be given the tools to view and assess difficulty for himself.

4. A Final note on authority

- 45 Under the MELODI scenario a new view of authorship, of authorial “authority” and of copyright seem to impose themselves. Currently, authors protect their intellectual and financial rights by trying very hard to prevent users from taking their work and re-purposing it. This understandable desire to defend one’s property normally leads to the property being immobilised or subject to carefully-controlled distribution. Such an approach is clearly challenged by a dynamic view of documents which are accessible electronically and subject to constant modification and can be replicated ad infinitum without degradation of quality.
- 46 One can foresee the day when most information will be distributed electronically and assembled at the point of use. Modern printers have developed such high performance that one can anticipate hard copy production on demand from electronic files. It will be necessary for communications products to guarantee an audit trail facility so that electronic distribution can be monitored and charged. But much educational use will depend upon the facility to change data rather than simply access it.
- 47 Here we enter an area in which traditional notions of an author’s authority seem challenged. To this there is no answer beyond requiring authors themselves to recognise

that their words and ideas may well have come from elsewhere. Which is where we began.

BIBLIOGRAPHY

A Language Teaching Classification and Subject Index. 1988. London: CILT.

Bickerton, D.M. 1992. "The European context for the production of resources for language learning". In Favretti, R. R. (ed.), *Il Computer nell' apprendimento e nell' auto-apprendimento linguistico*. Bologna: Monduzzi Editore, 25-33.

Bickerton, D.M. & S.W. Woodruff, S.W. "The challenge of hypermedia for language teaching: From pilot demonstration to generalised application". *Actes of Bordeaux CERCLÉS conference 1992* (electronic publication by DVLP, Université Bordeaux 2).

EUDisE Multilingual Thesaurus. *Alphabetical list of descriptors*. Mouton, Genthiner Strasse 13, D 1000 Berlin 30.

Knopf, P. *Vocabulary for Indexing Documents on Modern Language Teaching in connection with the EUDisE Thesaurus*. Lerchenstrasse 29, CH 4059 Basel.

Markey, Karen. 1986. *Subject Access to Visual Resources Collections: A Model for Computer Construction of Thematic Catalogs*. Connecticut: Greenwood Press.

Rowley, J. E. 1987. *Organising Knowledge*. Gower.

Universal Decimal Classification, Abridged English Edition, BS 1000A, 1961 (Schedule and Subject Index).

Weihls, J. 1989. *Non-book materials: The organisation of integrated collections*. 3rd ed. Canadian Library Association. 088802 240 9.

NOTES

1. To dispel some ambiguity, the present document was first fully word-processed some days after being spoken. The text is based upon the notes used and ideas generated during oral delivery in Bordeaux on Wednesday 8 September 1993. It appeared in electronic form in October 1993 and was subsequently revised extensively for paper publication in April 1994. Perhaps only this is the real paper?

2. To operate, the process relies on the use of symbols which have meanings conventionally ascribed to them by both a sender and a receiver. The sender has no real control over who receives the information but can influence the way it is processed by the web of implication or internal self-reference set up within the document. The process of symbolisation (and representation) holds for images as well as sounds and writing.

3. Available facets in the CILT classification scheme for "textbooks" are: materiological ("UB", e.g. slide sets: "UBP"), learning level ("UD", e.g. pre-school: "UDN"), special aspect of language ("UK", e.g. spelling: "UKT"), special uses of language ("UB", e.g. business: "UBM"), special classes

of student (“UN”, e.g. blind: “UNM”), readers (“UR”, e.g. advanced: “URG”) and texts and anthologies (“US”, e.g. song books: “USJ”).

4. For up-to-date information on the UGLC Audio-Visual Library contact the Librarian, Metta McLeod, mml@uk.ac.gla.lang

5. The language facet appears first in the notation and determines shelf location. It is based upon the MARC list, adapted from the LC notation scheme for written languages, employing a three-letter suffix. The only exception allowed is the suffix LCT used to classify items which are not language specific such as multilingual dictionaries and books on general topics.

6. A small schedule could be devised to improve the divisions of the main learning materials facet. Within “information” one could, for example, distinguish general reference books from periodicals and dictionaries, and the schedule could also subdivide realia along materiological or media lines (e.g. magazines, radio, television, CD ROM, film, slides, paper realia).

7. The learning materials facet is divided into the following main series: A = audio recordings; B = books; C = compilations (from audio or video sources); M = multimedia collection (items with a variety of media); S = short-life off-air recordings; X = editable off-air recordings; Y = video for distribution (e.g. schools programme). This is a purely formal classification, and much of the notation is arbitrary.

8. It became clear that as the collection grew the notation based upon 2 facets + accession and part number was insufficiently expressive and discouraged browsing. Nevertheless, users appear to have continued to prefer browsing, however difficult, rather than use the computer catalogue searching made available to them.

9. Since there are very few level 1 French courses in law and, equally few Basque grammars, the level and author facets are non-discriminatory and are therefore redundant. Without them the notation could be simplified to read as follows: FRE.C/LAW/021.02 and BAQ.I/GRA/002. One consequence is that documents with the topic facet GRA and no learning level would form a general class containing both grammars and books on grammar.

10. For up-to-date information on the SMILE Project contact the Project Officer, Maggie Kirk, mk@uk.ac.gla.lang.

11. Additional information on the socio-linguistic record in SMILE data is given in Bickerton (1992).

12. For all enquiries and further details contact the CML librarian, Belinda Torlot, at the University of Plymouth, PLYMOUTH PL4 8AA, UK.

13. See Markey, 1986. One may question two major assumptions in the method proposed: (a) that clusters of words with words in common constitute examples of linguistic similarity; and (b) that there is validity in the semantic theory of E. Panofsky (*Meaning in the Visual Arts* (1955), New York, Doubleday Anchor Books) whereby “expressional qualities” are deemed to be unambiguous in their meaning.

14. For the first published report on this Project see Bickerton & Woodruff (1992).

15. SMART-TALK: System for the MARK-up of mulTi media language TeAching Linguistic Corpora; for all enquiries contact Anthony Cafazzo (tonypisa@icnucevm.earn).

16. HARMONI: High-speed Access to Resources (Multimedia) using ON-line Interrogation; for all enquiries contact Stephen Woodruff (sww@uk.ac.gla.lang).

17. STACCATO: SysTeem voor Adequate Codering en Conversie van Alle taalmaterial voor multimediaal TaalOnderwijs: System for the adequate coding and conversion of all language materials for multimedia language teaching; for all enquiries contact Pol Cuvelier (cuvelier@be.ac.uia.reks).

18. Members of institutions wishing to join the beta testing programme should contact any of the researchers managerially responsible for the Project. For general enquiries

concerning MELODI, contact David Bickerton (David.Bickerton@uk.ac.plym.school-of-computing).

ABSTRACTS

Text, sound and image resources used by language learners are now being created and distributed in digital form. This paper examines a representative sample of principally British classificatory systems currently employed to organise documents for learning purposes. It suggests that present techniques for organising highly structured “bibliographic” information must make way for systems in which the data sought by the learner can be accessed directly and flexibly using learner-specific criteria expressed at the point of use. The approach to dynamic retrieval adopted by researchers in the MELODI Project is outlined, and conclusions are drawn which have important implications for all institutions involved in the electronic storage, retrieval and communication of language learning data.

Les supports texte, sons et images pour apprenants en langues qui sont créés et distribués aujourd’hui peuvent l’être sous forme numérique. Cet article examine un échantillon de systèmes en vigueur (surtout en Grande-Bretagne) pour la classification de documents à des fins didactiques. Il suggère que nos techniques de gestion de données « bibliographiques » fortement structurées devront faire place à des systèmes permettant un accès direct et flexible selon des critères spécifiques à l’apprenant et à ses besoins immédiats. Cette conception d’un accès dynamique aux données définit l’approche de l’équipe de recherche du projet MELODI et mène à des conclusions importantes pour toute institution qui souhaite gérer et échanger des documents à destination d’apprenants de langues.

INDEX

Mots-clés: apprentissage des langues, bibliothèque, classification, multimédia, projet Melodi

Keywords: classification, language learning, library, Melodi project, multimedia

AUTHOR

DAVID BICKERTON

University of Plymouth. david.bickerton100@plymouth.ac.uk