

Harakka T. & M. Koskela (toim.) 1996. Kieli ja tietokone. AFinLAN vuosikirja 1996. Suomen soveltavan kielitieteen yhdistyksen (AFinLA) julkaisuja no. 54. Jyväskylä. s. 33 – 48.

VALTER2 AND FORESTER: FROM A RELATIONAL TERM BANK TO HYPERTEXT ON THE WEB

Lauri Carlson

Department of Translation Studies

University of Helsinki

This paper examines the change in our conception of a dictionary caused by the electronic media, as seen through the design of a relational database for terminology. The relational approach to data storage emancipates our notion of a lexical or terminological entry, leading to a clearer separation of the actual data and its manner of presentation. This in turn can form the basis of a dynamic conception of a terminological hypertext, where different views of the same data are produced on demand from the same underlying set of base relations. A plan to implement this conception as a terminology service on the World Wide Web is reported.

Keywords: relational databases, hypertext, term banks, www term banks

ValTer2: A RELATIONAL APPROACH TO TERM BANK DESIGN

General and special language lexical work, lexicography and terminology, happens by different sets of people for different purposes. The dictionaries and vocabularies resulting from the work differ accordingly. The following list of oppositions has been adapted from Sue Ellen Wright ¹:

Lexicography

Entry identified by keyword or phrase
Different senses of a word
in the same entry
Indexed alphabetically by keyword
All parts of speech
Full grammatical detail
General vocabulary
Meaning described by paraphrase and
usage examples
Descriptive of current usage

Terminology

Entry identified by concept
Different terms for the same concept
in the same entry
Indexed by concept hierarchy
Nouns, verbs, adjectives (open classes)
Restricted grammatical detail
Domain specific technical vocabulary
Definition based on a concept hierarchy
Codifies and creates standards or norms

This list of oppositions is not arbitrary, it has an internal logic to it. The crux of the distinctions is that lexicography describes general language and terminology codifies a special language.

The semantics of general language is ambiguous and context dependent. This is not (just) a weakness of general language, but essential for its *efficiency*: in other words, for the fact that the same limited (though large) vocabulary can be used to convey not only familiar ideas but introduce, develop and convey previously unknown ones.² General language words are not tied to unique well-defined concepts; rather, they obtain meaning in context. Their meaning is negotiated in daily use through association to surrounding objects and situations, in particular through association and contrast with neighboring words. Polysemy and metaphor are therefore not aberrations of general language, but its operating principle.

If this is right, general language dictionary work cannot even in the long run fall together with terminology work: general language meaning won't reduce into language independent concepts, or free itself from contextual and verbal (what terminologists call nominal) definitions. Such definitions are not deficient or half-finished, but rather reflect what general language meanings are like. In lexicography, words are primary, and meaning is only identified through more words.

For the same reason, meaning-based dictionaries for general language (like Roget's thesaurus) do not seriously attempt to reduce their domain of discourse into well-defined concept hierarchies but content themselves to identify central themes in human experience and group words under these topics by strength of association to one another. Such associations are also being empirically extracted from general language dictionaries or corpora using statistical techniques.

In contrast, the distinctive feature of a special language is that it operates with, and strives for, a well-defined system of concepts identified by an unambiguous set of terms. Such domain specific concepts identify or classify objects, events, situations, properties and relations particular to a discipline; such types of entities are (also in general language) named by words and phrases taken from the major open word classes. The primacy of concepts before terms and the ideal of *monosemy* (one concept per one term) are among the basic tenets of terminology.³

THE HYPERTEXT CONCEPT

The hypertext concept admits of wider or narrower interpretations. On a narrow interpretation, a hypertext is a collection of passages of text and images stored in an electronic medium, with links from locations to other locations both within and between such passages that allow direct access between locations.

In this narrow sense, a hypertext is above all a faster way of browsing a dictionary of the traditional sort. Instead of taking out a volume and thumbing to a given cross-reference, one just needs to click on the link to get there. Second, hypertext allows data sharing through linking; a passage relevant in several places only needs to be stored once. Third, hypertext frees a reader from following a fixed order in perusing a set of documents. Compared to a printed dictionary, hypertext is an enhancement rather than enabling technology: it does not essentially change the way how a dictionary works, but it does make using one a lot more convenient. (The change is greater in relation to books of more traditional sort: hypertext makes all books work like dictionaries.)

In a wider sense, a hypertext (or hypermedia) document allows access to arbitrary further electronic services from within a document, from interactive video or audio connections to running any program and showing its results as another document. In the wide sense, there is nothing a hypertext document cannot do or be, because it can borrow the capabilities of any other service. In this sense, a hypertext dictionary can engage the services of any type of data management system, including a relational database. This in turn means that there need not be any fixed document at all which is the dictionary; rather, the text of the dictionary is a protean entity recreated on demand as the result of the user's queries from the database.⁴

THE RELATIONAL VIEW: ValTer 2

The separation of the storage of data from its access formats is taken to a logical extreme in relational database technology.⁵ The main advantages of a relational database are the efficiency of storage of information, data integrity, and a universal query language.

In a relational database, each data item needs to be stored only once. The database system makes sure that the base will not contain duplicates or conflicts. The data is stored in relation tables (two-dimensional tables representing n-place relations i.e. sets of n-tuples), each type of data element in its own table (for instance, terms, concepts, definitions, administrative data) Relations between different data elements are tabulated in the same way (for instance, term-concept, concept-to-concept, element-administrative data).

The query language of a relational database implements a system of relational algebra which allows joining, selecting and projecting the tables in all imaginable ways. A relational implementation of a terminological database need not have a notion of a terminological entry or record, or not at least one unique one. The base tables allow composing a variety of different views on the same data: for instance, a term oriented lexicographical view or a concept oriented terminological view. The lexicography-terminology opposition disappears to the extent it only pertains to the layout of information. From the point of view of a relational database system a lexicographical entry and a terminological record are just two alternative report formats based on two alternative views on the same data.

Among the disadvantages of relational database systems have traditionally been constraints on basic data types, inflexible database definitions, slow access times, and primitive interface technologies.

In a relational database design, a distinction is made between the database definition and data in the database. The database definition tells what types of data can be stored in the tables and what the interdependencies are between the tables. A database user cannot change the database definition. Because terminological information varies a lot, it is best to define the database abstractly enough to allow freedom to the users to customize the system.

In ValTer2 this problem has been solved by construing a large share of the database definition technically as data stored in the database, which the user (the term bank administrator) can change. The initial database definition only defines data element types which are structurally different from one another.

Database queries tend to be slow because desired combinations of data are formed at query time by joining data from different tables. Queries can be speeded up by creating indexes for frequently looked-up tables.

The remaining two disadvantages, restrictive basic data types and primitive interface technologies, are being addressed in newer releases of commercial database systems. Database systems have traditionally had restricted support for character and textual data. In Oracle 7 release, character fields can be up to one page (2000 characters) long, which is ample for all terms and most if not all definitions. Multilingual character support has been weak but is improving.

Currently, ValTer2 user interfaces are implemented using standard (i.e. rather primitive) 4th generation database interface tools. About the only advantage is that they work over simple character-based terminal connections. However, the current version of Oracle 7 is based on client-server

architecture, which means that platform-specific client programs can now be written using graphical interface technologies. Only database queries and results need to travel over network connections.

WHAT IS A DICTIONARY MADE OF? OBJECT LANGUAGE AND METALANGUAGE

Looking at the above list, we can see that some of the oppositions refer to the language being described in a dictionary, (**object language**), others to the *way* it is described (**metalanguage**). The distinction is very easy to become confused about in lexical work, because we are using the natural language(s) as both object and metalanguage(s), and because language about language is massively type-token ambiguous. But let us give it a try.

The words and phrases described in a dictionary or vocabulary⁶, their paraphrases, context examples, even definitions⁷ belong to the **object language**. There is only one of each of these: an object language word or definition either is or isn't listed in a dictionary, although it can *occur* there several times. The **first level metalanguage** describes semantic relations between object language terms and concepts. Linguistic terms like *word, term, noun, singular* and semantic terms such as *concept, synonym, equivalent, subordinate concept* characterize lexicographical and terminological metalanguage. Statements about object language terms and concepts couched in these terms: that term *T* exists, term *T* stands for concept *C*, or that *T* is synonymous with *S*, concept *C* is broader than concept *B*, concept *C* belongs to domain *D*, are first level metalanguage. Each such statement again either is or isn't found a dictionary, however many occurrences it may have there.

Thirdly, words pertaining to the organization and layout of a dictionary such as *entry, record, keyword, index, location, sort sequence* as well as administrative properties such as *source, authority, subset owner, status* belong to **second level metalanguage**, for they describe properties of particular *statements* about terms rather than properties of terms. No one can own a term, but one can claim rights on a particular description of one.

Considered from this point of view, what a dictionary consists of is first level metalanguage statements about object language terms and concepts and second level metalanguage statements about first level metalanguage statements.⁸

HOW MANY TERMS ARE THERE IN A DICTIONARY?

Defining the size of a dictionary in general is a matter of great economical interest but nowhere near settled. Although one might think that computers would make counting easier, the field only seems to have become more complex with the appearance of electronic dictionaries. The electronic medium emancipates dictionaries from the constraints of the two-dimensional textual representation, and this also allows many more ways to count words and measure dictionaries than the printed page.

Questions of identity become subtle here. An example is the definition of the term *term* itself. Surprisingly, no definition at all for it is given in many if not most handbooks in the field. In the older ISO standards, *term* is defined as an **expression** denoting a special language concept. In DIN 2334, a *term* is the **pair** of an expression and a special language concept it refers to.

The difference seems small, but in fact there are two different counting principles involved here. According to the first definition, there are as many terms as there are expressions; for instance, *term* in the sense of a special language expression, *term* in the sense of a contract clause, and *term* in the sense of a school semester count as one (polysemous) term, according to the second, they count as different terms. The distinction has practical consequences, for the term count of a given term bank may be much higher by the latter rule than by the former. ValTer2 implements both interpretations of the notion.

THE CONCEPT OF AN ENTRY REVISITED

One of the more media bound units in dictionary work is that of an *entry*, a dictionary article or a terminological record. An entry is a collection of lexical information that in some sense hangs together, pertains to a given keyword or concept. Traditionally, lexicographical entries are identified by keyword (this still leaves a lot of freedom as to when two keywords count the same or different, such as the treatment of homonyms), the several senses of a word listed in one entry. The opposite recommendation in terminology is that one terminological record contain terminological information pertaining to one concept. A terminological record may also be the locus of administrative and classificatory information (sources, domain classifications).

One problem with the notion of entry is that much of the information contained in an entry is relational, and thus pertains to more than just one keyword or concept at a time. How many synonyms does a dictionary contain? If there is a set of n synonyms in a dictionary, the n entries may contain references to the other $n-1$. Are there n synonyms in the dictionary or n squared?

In any case, it is not efficient to store the identical synonymy information in n different places, however often it is accessed.

The electronic format thus emphasizes a distinction between the information contained in a dictionary and its textual layout. An electronic dictionary may simultaneously support several different layouts, or entry formats if you like. How many entries are there then in an electronic dictionary? This question may not make any unique sense.

Once this insight has been made, the grouping and indexing of lexical information for purposes of human access and consumption can be conceptually separated from the question of its disposition in the interests of data storage. One is led to rethink the whole domain of lexical and terminological information in order to separate object level information from metalevel considerations, i.e. questions of collection, maintenance and distribution of lexical data.

ADMINISTRATIVE ASPECTS

The notion of entry is essentially a second level metalanguage concept that has to do with the way information about terms is kept together. 'Keeping together' here may mean the manner of presentation of information, but it can mean something more. A lexical entry or terminological record can be the unit of administrative work in a term bank. Items grouped together in an entry may be managed (recorded, saved, timestamped, updated), and owned (copyrighted, sold) as a unit.

Among administrative items relating to a piece of terminological information there may be the identification of the author/creator and successive updaters of the information, update dates, or a update history; the sources and authorities from which the information was obtained, information on the status or quality of the data, and administrative notes related to the datum. Information about sources and authorities on the one hand, and creators and updaters on the other hand, both belong to the metalanguage (they are information about terminological information).

In fact, there is a full metalanguage hierarchy lurking here: the author and updaters of the dictionary, and the dictionary itself, are in turn sources and authorities for the next generation of dictionaries. Consider what happens when information from some vocabulary is stored in a term bank. The vocabulary is of course the primary source for the information. But the vocabulary itself has its own sources, and so on. How should this succession be recorded (if at all)?

One solution is only to register the primary source. Another solution is to register the secondary sources on the same footing: if A refers to B, make the term bank to refer to A and B both. This means trusting A for its reference to B. The general but complex solution is to establish a chain of references: the term bank stores a third level metalanguage statement that according to A, term T occurs in B. The usefulness of all this fidelity is doubtful. A dictionary has to take its own stand on matters, although it may quote its sources for support; it is not just the sum of its predecessors.

The main purpose of administrative data is that they should document different items in the base. For this it suffices that administrative information describes term elements, it does not have to distinguish them. Having two sources for a term element in the term bank does not require having *two entries (copies)* for the same term element in the base, it suffices to have two sources associated to the same element. A citation source or other administrative element allows picking out, from among all the various term elements stored in the base, just those that are associated to that administrative element. In this way, it is already possible to reconstitute a term entry from the base by rounding up all the term elements that share the relevant source. To the extent ValTer2 can have said to implement a notion of a terminological entry at all, it is embodied in the notion of administrative data.

In ValTer2, the administrative viewpoint is just another viewpoint on the same data comparable to the term oriented and a concept oriented viewpoints earlier. All three views can be as it were pulled out from the exact same base relations; the difference is only a choice of reporting strategy. One can choose between surveying all the information associated to a term or concept, merging data from different sources together, or viewing all the information that share some administrative property, for instance the property of coming from a particular terminological entry in a given source.⁹

DEALING WITH INFORMATION FROM DIFFERENT SOURCES

A relational term bank is designed to monitor the integrity and non-redundancy of the data entered into it. The key concept here is the concept of a key. Each database table has one or more columns (fields) singled out which suffice to identify the row and distinguish it from other rows. The rule is that no table contains two rows with identical keys.

If a term element already exists in the base, the system will not allow re-entering it. What happens if a new source is imported which contain some of the same terms, but the entries differ in detail? What happens in ValTer2 is that the net difference of the two entries (the genuinely new information) is imported, and the shared part is supplied with the administrative infor-

mation from the new source.¹⁰In this way, the new entry is in effect added to the base without duplicating any of the information already there.

CONCEPTS

Concepts are abstractions of terms. One concept can correspond to a number of terms in one or more languages. The soberest way to think of a concept may be as an equivalence class of synonymous or equivalent terms, as a reification or abstract representative of the relation of equivalence. From a point of view of information technology, the main purpose of a concept is just to act as a peg for properties and relations shared by a set of synonymous terms, so as to reduce the number of links needed to represent all those properties and relations.

In order to talk about concepts, they need to be named somehow. A common solution is to give concepts arbitrary identifiers, say a number. The disadvantage is that arbitrary identifiers are hard to remember. To motivate the ValTer2 solution, a short discussion of concepts is in order.

The bulk of traditional terminology theory deals with terms that correspond in logical terms to one-place predicates, using ideas borrowed from traditional logic of classes. Concepts in a terminological concept hierarchy form a tree that stepwise refines a domain into mutually exclusive and jointly exhaustive subsets. A superordinate concept corresponds logically to the exclusive disjunction of its immediate subordinates. Two concepts in such a hierarchy are either disjoint or nested.

Concepts are picked apart and compared using *features*.¹¹ The distinction of concepts and features is an abstraction of the grammatical distinction between nouns and adjectives. Concepts stand for objects and features for their properties. A concept is the intersection (logical conjunction) of its defining features¹². Related concepts share features and a subordinate concept inherits all the features of its superordinate.¹³

The distinction between concepts and features is a relative one. From a Boolean point of view, features are indistinguishable from concepts (both have classes of objects as their extension); in fact, notions that act as features in one system of concepts may be treated as a system of concepts of their own. In ValTer2, concepts and features are named the same way. Borrowing an idea from recent linguistic theory, ValTer2 has *typed features*, which are formally just triples of form

type:attribute=value

where *type* names a covering concept (a class for which the feature is meaningful 14), the *attribute* names a property or criterion of classification with a range of possible values, and *value* is a possible instance of that property.

Using typed features to name concepts should not in any way restrict imagination. For instance, classical taxonomic names of species in natural sciences (say *homo sapiens*) can be represented in this form (most straightforwardly as *homo:species=sapiens*).

Entirely arbitrary codes or numbering systems are equally possible (in the style of *concept system 69:concept number=4711*). Types, attributes and values can be arbitrary character strings. The main advantage of regimenting the naming with the typed feature discipline is that it is easier to maintain and retrieve several simultaneous classification systems.

PROPERTIES, RELATIONS AND OPERATIONS AMONG CONCEPTS

ValTer2 is able to represent typed binary relations among concepts. This means that the user is allowed to name new relationships and link concepts with them freely. There are no built-in concept types, which on the other hand means that there is little built in logic for specific types of relationships. The only element of relation algebra built into in ValTer2 is that it maintains converse relationships. When a new binary relation is introduced, the user also enters a name for its converse. This ensures that relations are visible from both ends, and symmetric relationships are correctly represented.

Beyond this, it is up to the user to see that the relationships she defines obey logical constraints natural to them. There is currently no special provision for transitivity, functional relationships, or equivalence relations. For instance, ValTer2 does not mind assigning a concept two different or contradictory features.

There is no special representation for Boolean operations among concepts (such as conjunction, disjunction or negation). It is possible to define a binary relation like defining feature between a concept and its defining features, but it is up to the user to see that this relationship behaves as it should. Altogether, as far as concept representation are concerned, ValTer2 is still more of a dumb database than an artificial intelligence knowledge base.¹⁵

TERMS

Terms in the linguistic sense ValTer2 are identified as pairs of a language and a keyword. Different language homographs are treated as different terms, but homonyms in a given language are lumped together. Any properties distinguishing homonyms must be represented as properties of different term senses. This is not accurate, but it reflects the relative unimportance of grammatical distinctions in terminology.¹⁶

For the DIN 2334 notion of an interpreted term ValTer2 has a separate data type called *term sense*. A term sense links a term to a concept and a domain where that sense is valid. Because ValTer was originally designed for multinational administrative terminology, the domain identifier is by default a country code (or a code for an international organization like EU or UN).

TERM SENSES

As was pointed out at the outset, one of the distinguishing features of lexicography and terminology is that the ideal of terminology is a one-to-one relationship between concepts and terms. In lexicography, an expression and its meaning are not separated, but expressions identify their own meanings. It suffices to keep different senses of a word apart. There is little use (or hope) for a separate representation of conceptual structure.

In ValTer2, this difference in approach is recognized by making a distinction between **exact** (terminological) term senses and **inexact** (lexicographical) ones. A term sense is exact when it links a term to a concept it exactly corresponds to. Terms exactly linked to a concept are exact synonyms.

An inexact (lexicographical) term sense links a term to some feature that distinguishes it from any other senses of the same term. The feature need not exhaust the meaning of the term in the intended sense. Terms linked to the same feature only share that feature.

ValTer concepts are universal, while term senses are bound to a given language, term and domain. Term senses provide an alternative, language dependent level where to define semantical relationships between terms. For instance, the concept of 'head of state' could be defined independently of language by some conditions shared by rulers of all countries. This universal concept can then be associated to its closest realizations in different countries through inexact term senses. Each such national variant of the universal notion can be given a separate definition and be related to other similarly localized concepts.

CLASSIFICATION OF TERMS AND TERM SENSES

ValTer2 features are also available for creating arbitrary classifications of terms and term senses. The user can freely define typed features like *noun:number=singular* or *pre-term: project=ForesTer* and use them to define and access subsets of terms. There is no upper limit to the number of features defined in ValTer2 (their number is likely to be of the same order as that of terms in fact, given that they are also used to represent concepts).

RELATIONS BETWEEN TERMS AND TERM SENSES

Variant term relationships are language dependent relations of synonymous terms. An example of a variant relationship is the relation between the long and short forms of a term. Here, it is in general not enough to say that both terms relate to the same concept, and add that one of the forms is long and the other short. There may be many more terms related to the same concept, and we want to know which form is short for which term. (For instance, *The United States of America* has two short forms, *the United States* and *America*. Although all three terms refer to the same concept, *the U.S.* is an abbreviation only of one of them.

Related term relationships are typed relations between term senses. They allow specifying sense dependent relations between individual terms, for instance, that a certain legal term (say *solicitor*) in Britain corresponds to another term in the USA (say *lawyer*), without having to claim that they represent the same concept, in fact without doing any conceptual analysis on them at all.

DESCRIPTIONS

Free text descriptions can be associated in ValTer2 not only to terms, concepts, and term senses, but also to the various relationships between concepts and terms. Thus it is possible to associate qualifications to a suggested conceptual relationship, give context examples or constraints to translation relationships between terms in different countries, or attach usage notes to variants. This ensures that textual descriptions remain closely associated to the elements they really modify. When a translation relationship is deleted, related examples disappear as well. Maintenance of the coherence of an entry through successive changes is made easier.

ForesTer: A WWW SERVICE BASED ON ValTer2

As a relational database system, ValTer2 supports a variety of views on the same data, but the Oracle based terminology manager's interface is not suited for casual access to data by nonprofessional users of terminology. For this purpose, the hypertext interface technology provided by current WWW clients is a good candidate. The current ValTer2 implementation already supports producing reports from ValTer2 in HTML format that can be directly presented in a WWW browser. An EU project proposal called ForesTer was submitted in January 1995 with the purpose of establishing a ValTer2 based forestry terminology service on the Web.

The ForesTer forestry term exchange is primarily designed to meet the terminology needs of forest researchers and educators and personnel in forest companies. The exchange not only offers professional users an immediate networked access to up-to-date, multilingual forest terminology but provides a forum for international harmonization and discussion.

As an Internet service, it can also offer casual users networked access to forestry and environment concepts and terms in a variety of languages.

The project plan includes building a demonstrator server shell around ValTer2 using current Internet programming techniques (an Oracle WWW interface and Java net programming tools). The demonstrator consists of a version of ValTer running on a Unix platform, plus a set of WWW interfaces to it for exchanging data through Internet. Through these interfaces, term material can be presented to users in Internet in hypertext format and reactions and discussions on terminology fed back to the term bank.

Examples of alternative formats for the same data

en term
~~concept:domain=terminology~~
~~en definition *special language expression standing for a well-defined concept*~~
~~source ISO 1087, NN~~
~~concept:domain=education~~
en definition *period in an educational institution*
 source MM
~~concept:domain=law~~
en definition *condition in a contract*
 source LL

Monolingual term oriented view on English *term*. The underlined words are links to alternative views.

ISO 1087:concept number=000
 superordinate ~~concept:domain=terminology~~
~~en definition *special language expression standing for a well-defined concept*~~
en term
 source ISO 1087, NN
sv term
 noun:gender=utrum
 noun:plural=er
de Terminus
 noun:gender=masculine

Multilingual concept oriented view on the terminological concept *term*

ISO 1087:concept number=000
~~en definition *special language expression standing for a well-defined concept*~~
~~source ISO 1087~~
en term
 source ISO 1087

Entry oriented view on *term* (the entry for *term* in ISO 1087)

- 1 The comparison is from a draft background document of the TIF terminology interchange format (MS).
- 2 The efficiency of natural language is discussed in Barwise and Perry (1983).
- 3 This commonplace needs qualification. Systematic ambiguity is not harmful. Compare for instance first names, local phone numbers or relative file names, all of which are multiply ambiguous but routinely complemented with longer identifiers as needed. The converse principle of mononymy, i.e. one term per concept, is even less essential. Synonymy is pervasive, in fact essential, in most formal languages. The reason synonymy may create confusion is rather the tendency to avoid synonymy operative in general language: synonyms tend to develop semantic distinctions just to keep terms apart.
- 4 Cf. Meyer et al. (1991) where many of the same points are made about a similar setup (a hypertext system based on Oracle).
- 5 For a somewhat outdated survey of RDB technology in terminology, cf. Sager (1990:175ff).
- 6 This formulation intentionally begs the question of individuating object language words: whether homonyms are the same word with two meanings or two different words, for instance. Such problems are orthogonal to the object language-metalanguage distinction.
- 7 In logic, a(n explicit) definition is an object language formula of the form of an identity or equivalence where the left side is the definiens and the right side the definiendum. In terminology, the term definition usually refers just to the definiens (i.e. right-hand term) of such an equation. This may be because terminology insists on defining concepts rather than terms, which means that the definition must be the same for all terms associated to a given concept. On the other hand, it is said that a definition is an object language expression which can be substituted for the term(s) referring to the concept defined. In multilingual terminology this entails that a language independent concept has different definitions for each language concerned. In both disciplines, a metalanguage definition like a term denoting ... synonymous with ...' is not considered a formally correct definition (though it may entail one).
- 8 The distinction between two metalanguage levels can be compared to the distinction between descriptive and administrative elements in the ISO terminology interchange format (TIF/ETIF/MATIF) standard proposal.
- 9 There are a few samples of such alternative views at the end of the paper.
- 10 A commercial term bank shell where the problem of merging information from different sources has been specifically addressed is the Trados Multiterm.
- 11 The term *feature* is systematically ambiguous: it can denote an attribute (a finite function with a number of mutually exclusive values) such as color, or a particular value from the attribute e.g. (the color) blue).
- 12 This wording includes the closest superordinate concept among the defining features, and thus presupposes the ValTer2 solution of treating concepts and features on a par. Note that a feature is a simpler (thus wider) concept than the concept it defines.
- 13 In Boolean terms, the logic of features is the dual of the logic of concepts: a concept can be defined either from below as the disjunction (union, join) of its subordinate concepts, or from above as the conjunction (intersection, meet) of its features. The key property of features is that the values of each feature are exclusive but values from different features cross-classify. The exclusion property is preserved in intersections and thus inherited to the concept system defined through features. Cf. e.g. Boolos (1974).
- 14 In set theoretic terms, values of a typed feature are guaranteed to have nonempty intersections only within its type.
- 15 This is susceptible to change in subsequent versions of the system.
- 16 The IBM Translation Manager/2 dictionary distinguishes a separate homonym level where grammatical distinctions among homonyms can be made.

REFERENCES

- Barwise, Jon and Perry, John. *Situations and Attitudes*. MIT Press, Cambridge, Mass 1983.
- Brace, Colin. Termbase design and TIF. *Language Industry Monitor* 14, 1993, p.7.
- Carlson, L. ValTer: a multilingual term bank system for terminology work. In *Proceedings of TAMA '94*. Infoterm, Vienna 1995.
- DIN 2342. *Begriffe der Terminologielehre*. Berlin: Beuth 1986.
- Halmos, Paul R. *Lectures on Boolean algebras*. Springer-Verlag, New York 1974.
- ISO 1087-1989. *Vocabulary of terminology*. Geneva: ISO1989.
- ISO-DIS 12200. *Computational aids in terminology - Terminology Interchange Format (TIF)- An SGML application*. ISO 1994.
- Meyer, R., Geer, A, Hanne, K-H. On how to bring hypertext to termbanks. In *TKE '90: Terminology and Knowledge Engineering: Applications*. INDEX Verlag, Frankfurt 1990, 437-446.
- Sager, J. *A practical course in terminology processing*. John Benjamins Publishing Company 1990.
- Wallmansberger, Josef. Hypertext approaches to terminological information processing. In *TKE '90: Terminology and Knowledge Engineering: Applications*. INDEX Verlag, Frankfurt 1990, 222-229.