The role of cognitive modelling in general and that of frames in particular in terminology theory and practice

WILLY MARTIN
Lexicology/Terminology Research Group
Vrije Universiteit Amsterdam

1. INTRODUCTION AND SURVEY

In this paper *cognitive modelling* will be understood as the systematic representation of knowledge within a certain subject or knowledge field. The term *model* will be taken here as a background against which, or a system by means of which, knowledge can be organised. A TD (Terminological Database) will, among others, be regarded as the description of a knowledge field by means of a terminology. A terminological dictionary, as opposed to a terminological database, will be considered as secondary to the latter, in other words, dictionaries will be regarded as derivations (front-ends) of underlying databases.

The paper itself will consist of three parts. First of all, different levels of cognitive modelling will be distinguished and illustrated. Secondly, a frame-based approach to one of the levels, viz. that of terms/concepts, will be discussed. Thirdly and finally, the approach advocated will be confronted with a set of important terminological issues in order to situate and evaluate it.

2. LEVELS OF COGNITIVE MODELLING

In order to function well (so that knowledge can be [easily] acquired and [properly] used), I will argue that a TD needs to be organised at, at least, three levels, viz.

- that of the domain,
- that of the database and
- that of the terms/concepts.

In what follows therefore I will deal with

- domain modelling (modelling at the higher level, the so-called macrostructural level);

- data modelling (modelling at the intermediate level, that of the entities and relations in a database, the so-called mediostructural level);
- term/concept modelling (modelling at the lowest level, that of the terms and the concepts, the so-called microstructural level).

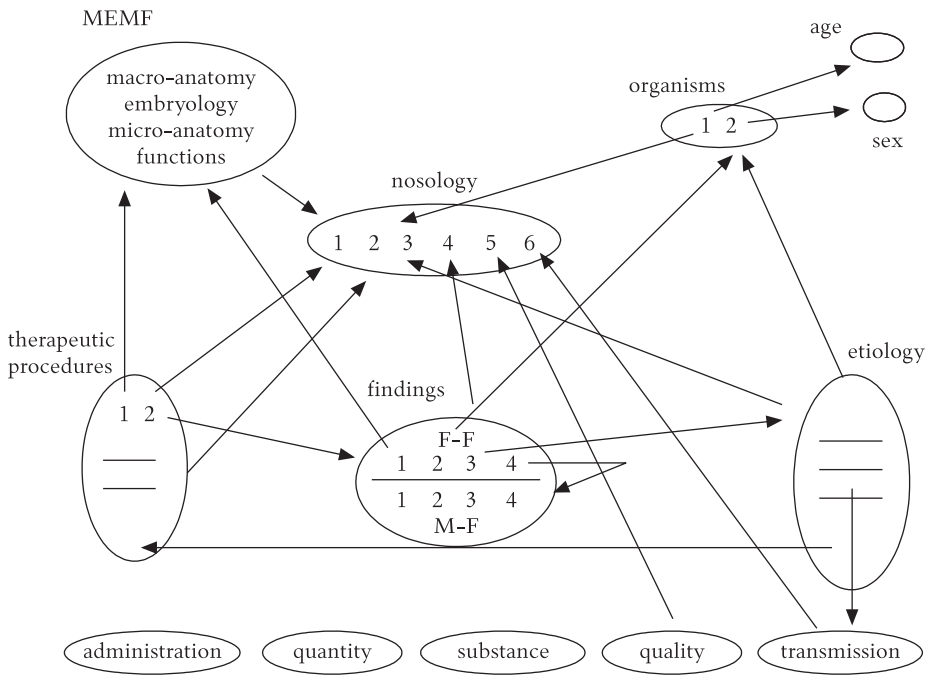
In order to make clear what is meant I will illustrate the respective levels one by one in the next sections.

2.1. Domain modelling

If a terminological database is meant to deal with knowledge and with its management, then it has to provide for a model representing the way how the '(sub-)world', the domain, is organised/structured. So, for instance, in *medicine* one basic concept, viz. that of *disease (nosology* in figure 1 below) structures the whole field. It is the central organising principle within this 'world'. If one talks about body-parts here, it is because of the fact that they are/can be affected; if one talks about organisms the same applies; therapeutic procedures only make sense when they refer to diseases and so are symptoms (findings), causes (etiology) etc. Everything in the medicinal world is linked directly or indirectly (via other concepts) to the central concept *disease*. The amount and granularity of the information given about other concepts is defined by this direct or indirect relationship. Domain modelling, therefore, is *a conditio sine qua non* without which it is impossible to construct a terminological database. In figure 1 a simplified schematic representation is given of such domain modelling for *medicine* [for more details see Martin e.a. 1991].

As one can observe, all main categories are centripetally related to disease (nosology): anatomy (MEMF) and organisms by the 'affect' relation, etiology by the 'caused by' relation, findings by the 'symptoms' relation, therapeutic procedures by the 'treat' relation etc. Typically then the medicinal domain is a domain which is well delineated and centralised: a domain with one central/core category to which all other categories are related. It goes without saying that taking the same objects and putting them into a different domain ('drugs' in *medicine* versus the same category in *pharmacy)* alters the structure of the field and the amount and character of the knowledge that should be expressed. That is, among others, one of the reasons why domain modelling is crucial when starting with the construction of a TD and dealing with knowledge representation.

Figure 1: Example of a 'delimited' centralised domain *(medicine)*



Not all domains show the same kind of structure though. So, for instance, in figure 2 an example is given of the domain of *educational systems,* which has an embedded or onion-like structure. There one can argue that all the tunics together form the whole (onion) and that (therefore) one simply cannot restrict oneself to, for instance, the innermost tunic, but will have to select from all tunics (circles) if one wants to come to grips with the domain of educational systems as a whole.

As the preceding examples show, subject fields/domains are not always ordered hierarchically (according to is-a or part-of relations) as one may expect at first sight, because of biological models with their strict taxonomic order. Moreover, sometimes domains are rather fuzzy. Whereas, for instance, the treated domain of *medicine* is rather well delineated, that of *business* is much less so, such as figure 3 illustrates.

Although one can observe that the domain *business* implies the interaction between a company and both its external partners and its internal parts (P = production section, F = financial section, S = selling section, A = administrative section) and although in both interactions selling

Figure 2: Example of an onion-like organised domain *(educational systems)*

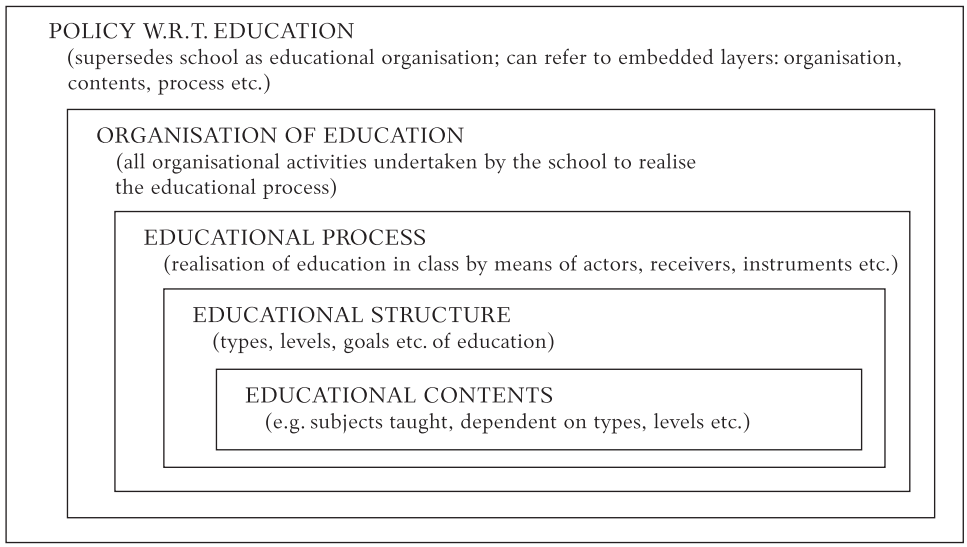
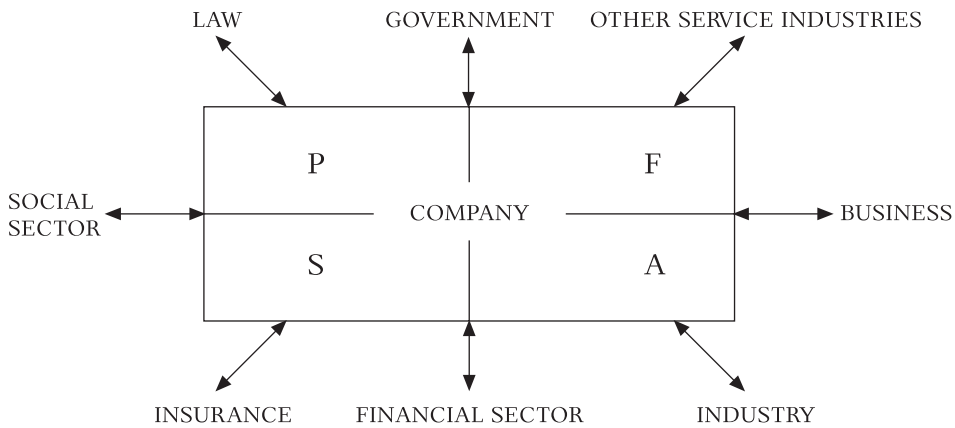


Figure 3: Example of a diffuse centralised domain *(business)*



(business) and (industrial) partners/parts come first, yet the domain as such remains diffuse, although it is centrally organised.

Whatever the domain, having a good insight into the super- or macro-structure of the field is a necessary condition both to better delineate the subject-field itself and to organise/represent knowledge within that field by means of a TD.

2.2. Data modelling

Under data modelling I here understand the modelling of data as in a database, implying

- the definition of the entities in the model and their relationships (e.g. terms, concepts, collocations and the relations/links that exist between them) and
- the definition of the data categories for the different entities (both the attributes and the [domains of their] values).

Indeed, in a data model one does not only have to make clear *what* one wants to represent from a contents point-of-view (see 2.1 above: the general framework or macrostructure of the world/domain to be represented), but also *how* one will do so: by means of which formal objects, entities and relations.

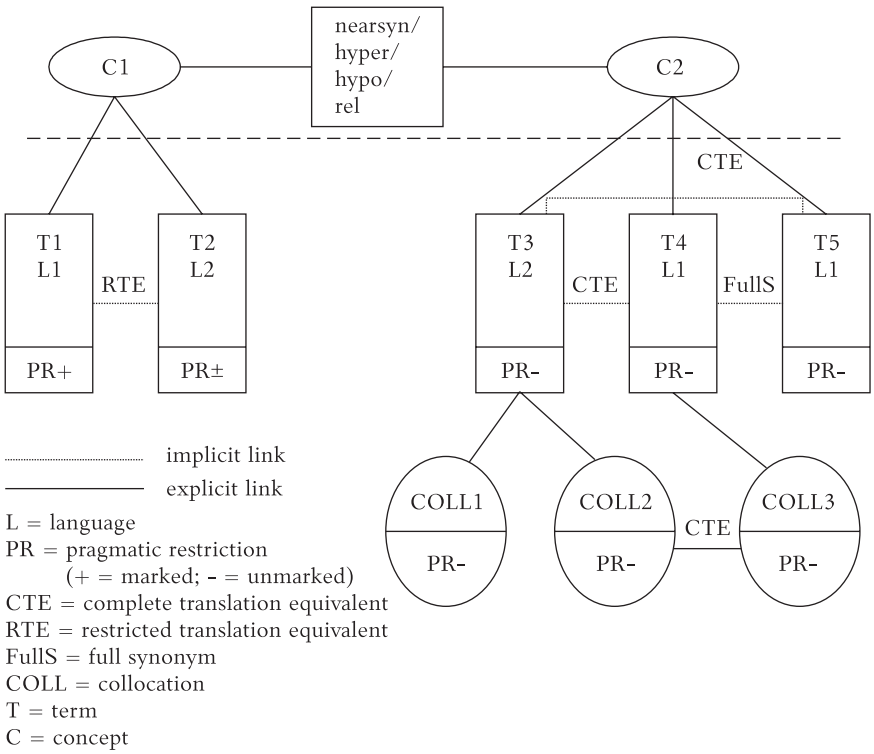
In a project called DOT (acronym for Dutch Databank OverheidsTerminologie: Database Government Terminology; see Maks e.a. 2000 and Maks e.a. 2001) the system needed as entities: concepts, terms, collocations and links in order to represent terms, the use of terms as in collocations, the relationship between terms such as (near) synonymy, (near) equivalence and the like. Figure 4 can give an idea of what is meant.

As one will observe, different graphic firms are used to distinguish between:

- concepts (C),
- terms (T) and
- collocations (COLL).

Furthermore, a clear distinction is made (see horizontal broken line) between terms and concepts, implying that concept entities correspond to semantic units expressed by one or more terms in one or more languages. Term entities represent one term together with its full linguistic description including its usage. Collocation entities do the same for collocations. There are several kinds of links also: both explicit (the full lines in the scheme) and implicit links (the broken lines). An example of an explicit link (one the terminologist has to explicitly fill out) is that between a concept and a term, or that between a concept and a concept (with values such as NEARSYN, HYPER, HYPO, REL(ATED)). Implicit links are links that the system can derive automatically: because of the fact that terms are linked to concepts and that pragmatic values are specified per

Figure 4: Entities, links and relations in DOT (based on Maks e.a. 2000, see also Maks e.a. 2001 for more information)

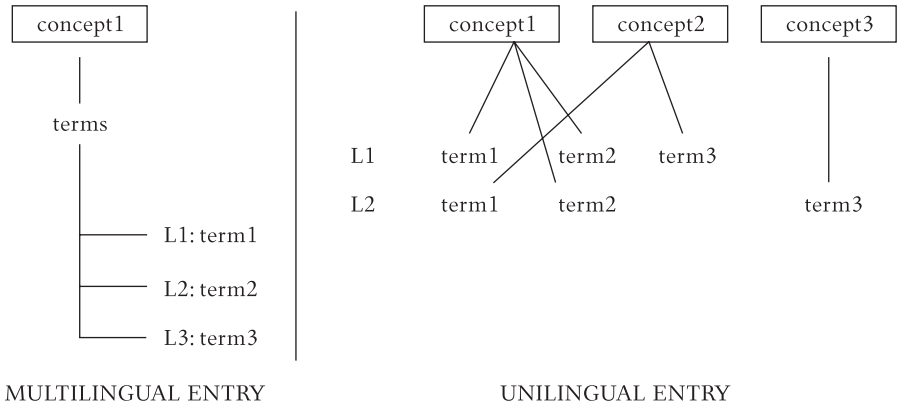


term, the relations between terms (both intra- and interlingual ones) need not be mentioned explicitly, but can be 'calculated', leading to full synonymy, complete translation equivalence, restricted translation equivalence, and near translation equivalence.

The advantage of keeping the conceptual and the linguistic (terminological) level apart is, among others, that the description of a term in one language does not influence the description of its so-called translation equivalent in another language. In other words, one can work now with unilingual entries, meaning that the terms of one language can be described independently from that of another one and yet can be linked with each other via the conceptual level.

In figure 5 the difference between unilingual and multilingual entries in a multilingual database is schematically represented.

Figure 5: Multilingual versus Unilingual Entries



Unilingual entries (entries within one language) can be linked with other unilingual entries (entries from one or more languages) without one language biasing the description of the other.

In multilingual entries one entry contains all information for all languages. The problem then is that differences at the conceptual level are blurred if terms from different languages are treated as translation equivalents without being fully equivalent.

The above not only makes clear that one cannot construct a data model without having any notion about the (terms occurring in the) domain one is dealing with, but also that one should not abstract away from the tasks one wants to carry out with the databank (as in the case of DOT: comparing law systems and translating 'governmental' texts).

Data modelling not only comprises the definition of entities, but that of the data/information categories that 'decorate' these entities as well. In the next section I will deal with one of these entities, viz. concepts.

2.3 Concept modelling

In the preceding section I have already pointed at some of the advantages of a conceptual approach. One of the problems encountered here is how to represent concepts (taken as mental building blocks to organise knowledge with). If one accepts that the (conceptual) meaning of a term is, as a rule, represented by its definition, then one could represent the meaning/definition/concept expressed by the term using a semantic network

as a model (see, for instance, Fraas 1998: 433 ss.). In Martin 1998 semantic networks are represented in the form of frames and, among others, used as definition models. In the next part I will further elaborate upon the role of frames and on that of a frame-based approach to terminology.

3. A FRAME-BASED APPROACH TO CONCEPT MODELLING

Frames are taken here in the AI sense of the word, following the Minskyan tradition (see, for instance, Minsky 1975). In this sense they are structures representing background, implicit, stereotyped knowledge which is necessary in order to understand concepts and meaning. AI frames à la Minsky have a slot-filler format. From this point-of-view a frame is a set of general conceptual categories or relations (slots) followed by specifications (fillers). In order to make clear what is meant, I will turn to a concrete example.

Semantic frames are type-bound, meaning that they are bound to certain concept types. Concept types need to have been established in the domain modelling phase (see section 2.1. above). For instance, in the domain of government terminology a type such as *allowance* will occur.

The frame-like representation for *allowance* looks as follows (see also Martin and Heid 2001: 58):

Table 1: Frame for the type *allowance*

	allowance
SLOT	PARAPHRASE OF SLOT
goal	what the *allowance* is meant for
source	who pays the *a.*
beneficiary	who receives the *a.*
reason	why the *a.* is paid
size	what the amount of the *a.* is
time	when the *a.* is paid
periodicity	how many times the *a.* is paid
way	in which form the *a.* is given
condition	under which conditions the *a.* is given

In the world of *social services* then, the concept *pension* will be regarded as a token of the type *allowance*.

The underlying frame for this type will consist of the following slots/elements:

- beneficiary — *who gets the allowance?*
- source — *who gives the a.?*
- goal — *what is the a. given for?*
- reason/ground — *on which basis is the a. given?*
- size — *what is the size of the a.?*
- periodicity — *how many times is the a. given?*
- time — *when is the a. given?*
- way — *in which form (e.g. money or other values) is the a. given?*
- condition — *which are the conditions under which the a. is given?*

A definition derived from this frame could read:

> A *pension* is an amount of money, fixed by law or (insurance) agreement, paid to someone (a *pensionable* or his *widow* or *orphans)* by someone else (a *[former] employer,* an *executive organisation),* periodically (e.g. *every month)* to provide for the cost of living, after one has retired either because of having reached the fixed age of retirement or because of invalidity, if a contribution has been paid for during the term of office.

Of course both the concrete form and contents of the definition itself strongly depend on the user they are meant for. However, if one takes for granted that the conceptual meaning of a term is, as a rule, represented by its definition, then cognitive models such as frames can certainly be of great help in systematising definitions. In the next section I will try to make clear that cognitive modelling in general and a frame-based approach in particular, go beyond that, and have an impact not only on the realm of definitions and concepts, but in that of terminological theory and practice in general as well.

4. DISCUSSION: IMPACT OF COGNITIVE MODELLING ON THEORY AND PRACTICE OF TERMINOLOGY

The basic claim put forward in this paper has been the following:
If terminology has to do with the

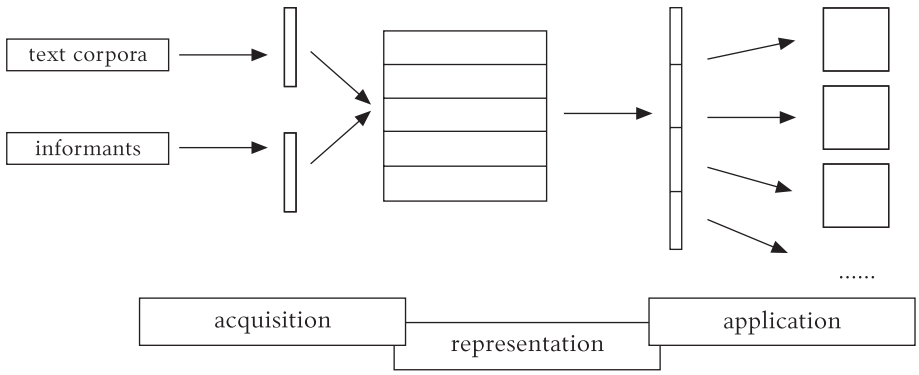
- acquisition,
- representation and
- application

of knowledge in a specific knowledge domain, preferably, but not solely, by experts in this domain, then
in order to come to good/better results in theory (design) and practice (production and use of TD's), more attention should be devoted to *cognitive modelling,* implying

- domain modelling,
- data modelling and
- concept modelling.

In other words, cognitive modelling in general and a frame-based approach in particular is claimed to have a positive impact on the three important aspects of terminology work. Figure 6 underneath schematises these three aspects, viz. acquisition, representation and application.

Figure 6: The three facets of terminology work



Elsewhere (Martin 1998) I have tried to make clear that a cognitive and, more in particular, a frame-based approach shows certain advantages over other approaches with regard to the three aspects mentioned. So, for instance, will a frame-based approach not only lead to a better access to knowledge, but also to a production/prediction of knowledge: on the basis of definitions, combinations can (in abstractor) be predicted. In the case of *pension,* for instance, one can, on the basis of the frame, expect combinations with *get* (beneficiary), *give/pay* (source), *high/low* (size) etc. This also could give rise to socio-linguistic investigations: depending on the context/communicative situations/social roles different fillers/values could be used for the same slots/attributes.

In order to make the impact of cognitive modelling more tangible I will now have a look at the answers which our approach gives to questions/issues such as those raised in the talk given by R. Kocourek at the Workshop on Terminology during the 17th International Congress of Linguists (Prague, July 2003, R. Kocourek 'Theory of Terminology and Specialized Language: Criteria and Choice'). I have selected a dozen of questions which can be regarded to be 'hot' issues in terminology theory and practice and I have provided for short answers to them, tackling the question from a cognitive modelling point-of-view. What follows can then be read as a short question/answering dialogue.

- Question: Should we also study non-terms occurring in specialised texts?
  Answer: A terminologist should study what is relevant from a cognitive point-of-view, that what is interesting for the knowledge domain he is describing; in this respect also non-terms should be object of study.
- Q: Should we try to examine terminology in one language, in two languages, in groups of more than two languages, or in all languages?
  A: As we have made clear in section 2.2., unilingual entries are to be preferred to multilingual entries. That does not mean that terminology should not be multilingual, it only implies that terminologies should not be linked so that they become biased to one language. In other words, a terminology from one language should not be biased but linkable to that of other languages.
- Q: Should terminology include proper names [...] in texts?
  A: Proper names can be important data in knowledge fields. They should not be excluded therefore. Also see the renewed interest in the NLP-field for named entities and NER (named entity recognition).
- Q: If denotative meaning (things) and pragmatic aspects of terminology (users) are to be considered, what methods should be used?
  A: It is indeed the case that terminology is governed both by conceptual and by pragmatic considerations. In this respect it is of primary importance that a TD not only takes into account domain/data or concept modelling. It is of equal importance to provide for so-called user profiles which delineate search routes and information

packages adapted to specific user groups. On figure 6 in this section the right hand rectangles standing between the database and the appliances suggest such user profiles.

- Q: Even though we cannot think of constructing a pseudo-symbolic formal language, should we introduce a system of acronyms, [...] signs and symbols or similar devices? How could we construct such a system?
  A: Frames are typically a representation language which is half way formalisation. With some effort, the slots or relations for a (sub)world can be fixed and standardised. As to the fillers, if one could restrict them (both formally and semantically), formalisation would come closer by. A first step could be to restrict fillers to terms. If this were the case, a real *termnet* could become established.

- Q: How do we obtain the primary data for description and analysis of terms: from corpora of specialised texts, from conceptual systems of the subject disciplines, or from both?
  A: There should be an interaction between the two such as figure 6 in this section suggests. The acquisition side (corpora, informants) influences the representation system and vice versa. In other words, terms in texts (partially) define the design of the representation model (acquisition → representation), whereas the representation model steers the selection of terms and the information about them (representation → acquisition).

- Q: How do we distinguish from non-terms in specialised texts?
  A: As a term functions on three levels – the conceptual, the linguistic and the sociolinguistic – it should obey the criteria set for each of them. In other words, terms should express concepts which function/are established within a certain (sub)world and are accepted by the linguistic system and the sociolinguistic community.

- Q: How do we distinguish multiword terms from combinations of terms and complex noun phrases?
  A: See the preceding question/answer. Actually, cognitive modelling shows what is conceptually possible/computable/combinable within the (sub)world under description; linguistic and sociolinguistic factors define whether what is possible conceptually, is also used and accepted.

- Q: How do we study terminological collocations and phraseology?
  A: In a frame-based approach the slots do not only function in defining the concept, but, as they are relations, they also show which kind of combinations (with other concepts) could occur. See for instance the *pension* case which can (in Dutch) get such fillers as:
  - hoog; laag, klein: connected to the slot *size,*
  - betalen, uitkeren, toekennen, uitbetalen: connected to the slot *source,*
  - ontvangen, genieten, krijgen: connected to the slot *beneficiary*.

  In other words, from semantic frames and frame-based definitions abstract collocational patterns can be derived and open up interesting perspectives to study combinations (see Martin 2003 in this respect).
- Q: What methods will be employed for the semantic analysis of terms (definition, synonymy, ambiguity, hyperonymy, hyponymy, antonymy, metonymy and other relations, etc.)?
  A: In a frame-based approach it is by far the 'other relations' that play a prominent role. The predominantly structuralistic relations (synonyms etc.) are considered neither to be sufficient nor to be expressive enough.
- Q: Will the approach be semasiological or onomosiological?
  A: A frame-based approach to definition starts from concept types and is in this sense onomasiological.
- Q: What linguistic research methods do we use: structuralist, generativist, poststructuralist or other methods?
  A: See the question with regard to the semantic analysis of terms. I consider in particular the structuralist approach as too restrictive and have it therefore superseded by a cognitive one.

5. CONCLUDING REMARK

In this article I have tried to make clear that cognitive modelling is a conditio sine qua non for terminology in this sense that it creates the necessary general framework without which it is not possible neither to work consistently nor to distinguish between what is relevant and not.

BIBLIOGRAPHY

Fraas C. 1998: Lexikalisch-semantische Eigenschaften von Fachsprachen. – *Fachsprachen – Languages for Special Purposes*; L. Hoffmann, H. Kalverkämpfer & H. E. Wiegand (Eds.), Berlin/New York: de Gruyter, Vol. 1, 228–238.

Maks I., Martin W. & Groesz I. 2000: The DOT Project: Data and Datastructures for a Bank of Governmental Terminology. – *Proceedings of the Ninth Euralex International Congress – Euralex 2000*; U. Heid, S. Evert, E. Lehmann & C. Rohrer (Eds.), Stuttgart: IMS, 297–303.

Maks I., Martin W. & Groesz I. 2001: *DOT Final Report (part 1)*. Amsterdam: Vrije Universiteit (Free University).

Martin W. 1998: Frames as Definition Models for Terms. – *Proceedings of the International Conference on Professional Communication and Knowledge Transfer*; A. Munteanu (Ed.), Vol. 2, Vienna: Termnet, 189–221.

Martin W. 2003: Definitions and Collocations in Dictionaries: the GWDS Compared to the Van Dale Groot Woordenboek der Nederlandse Taal. – *Lexicographica, Series Maior*, Vol. 113, Tübingen: Niemeyer, 3–23.

Martin W., Cremers C. & Vliet H. van der 1991: *Over Atlex, Relset, Conceptor e.a.* Amsterdam: Vrije Universiteit (Free University).

Martin W. & Heid U. 2001: Frame-based Definitions and the Selection of Multiword Term Candidates in DOT. – *TIA 2001: actes des quatrièmes rencontres terminologie et intelligence artificielle*, Vandoeuvre: INIST-CNRS, 55–65.

Martin W. & Vliet H. van der 2003: Design and Production of Terminological Dictionaries. – *A Practical Guide to Lexicography;* P. van Sterkenburg (Ed.), Amsterdam/Philadelphia: Benjamins, 333–349.

Minsky M. 1975: A Framework for Representing Knowledge. – *The Psychology of Computer Vision*; P. H. Winston (Ed.), New York: McGraw-Hill, 211–277.

KOGNITYVINIO MODELIAVIMO APSKRITAI IR FREIMŲ KONKREČIAI VAIDMUO TERMINOLOGIJOS TEORIJOJE IR PRAKTIKOJE

Šiame straipsnyje išskirti, apibrėžti ir pavaizduoti skirtingi kognityvinio modeliavimo lygmenys – srities modeliavimas (makrostruktūrinis lygmuo), duomenų modeliavimas (mediostruktūrinis lygmuo) ir termino ir (arba) sąvokos modeliavimas (mikrostruktūrinis lygmuo). Kalbant apie pastarąjį lygmenį pristatomas freimais besiremiantis sąvokų modeliavimo būdas. Pagrindinis šiame straipsnyje iškeltas principas – kognityvinis modeliavimas apskritai ir freimais besiremiantis požiūris konkrečiai turi teigiamos įtakos tam tikros srities žinių gavimui, pateikimui ir taikymui. Akivaizdžiai parodoma, kad kognityvinis modeliavimas yra terminologijos būtinybė, nes jis suteikia reikalingą struktūrą, be kurios neįmanoma nei nuosekliai dirbti, nei atskirti tai, kas svarbu, ir kas ne.

Gauta 2010-08-09

Willy Martin
Faculteit der Letteren, Vrije Universiteit
De Boelelaan 1105, NL-1081 HV Amsterdam, Nederland
E-mail w.martin@let.vu.nl