Jörg Schütz and Bärbel Ripplinger
IAI Saarbrücken

Controlling nlp through Terminological Information

Abstract

In the context of *niche applications* the discipline of computational terminology has received little attention in computational linguistics; an unfortunate situation given that natural language processing (NLP) systems seem to be most successful when applied to specialised domains. In this paper we present an approach that integrates an instance of computational terminology into a constraint–based NLP environment as one kind of *performance control.* Parts of this research have been carried out in the context of the ET–10/66 project supported by the Commission of the European Communities (CEC). Here, we describe the formal representation of terminological knowledge and one solution for linking this kind of information to the grammatical and lexical strata of a NLP system, with special emphasis on a concept–based translation methodology.

1. Introduction and motivation

1.1 Outline of research

The aim of our project is to provide for comprehensive terminological knowledge which is organised so that the extra–linguistic information is contained in a concise efficiently machine–tractable form, and which is formalised so as to ensure consistency across organisations of related grammatical and lexical strata. On the basis of a significant set of terms taken from the corpus of satellite communications, i.e. mainly from the *Handbook on Satellite Communications* of the *International Radio Consultative Committee* (CCIR), we have investigated existing definitions of these terms and drawn up guidelines for the formulation of new definitions which also fulfil the requirements necessary for the integration into an NLP system. Each of the chosen terms was redefined together with an expert of the domain using these guidelines. The extra–linguistic information gained from these new definitions is used to build a terminological repository which can be used within a constraint–based NLP framework. This approach has been implemented and tested in the ALEP environment (ALEP 1993), a general purpose NLP development system promoted by the CEC for the LRE action line and the yet to be defined Fourth Framework Programme.

In the first phase of the project a number of topics related to different kinds of lexical resources, including specialised dictionaries (machine–readable dictionaries (MRD) and machine–tractable dictionaries (MTD)) and terminological resources have been investigated and assessed. Another research direction of the project has concentrated on the development of the

conceptual organisation (ontology) of a part of the telecommunications domain. Its purpose is to provide conceptual information that ensures the support of the analysis and translation process of terminological expressions. The ontology is intended to represent the concepts of the selected domain and their generic relationships which correspond to the meaning of the terminological expressions contained in the selected corpus. In addition to the corpus, the information relevant for constructing the ontology was also collected through so–called *term definition forms* developed within the ET–10/66 project filled in by a human expert of the domain (cf. Section 2.3.1). The development of the ontology as one of the subgoals of the project has thus proceeded from two global research strands: on the one hand, the acquisition, organisation and representation of knowledge in lexical and terminological resources, and on the other hand, their conceptual modelling.

Since the overall purpose of this conceptual structure is to be maximally supportive for the computational processing of terminological expressions in an NLP environment, there is a third direction from which research on the topic proceeded, that is, the investigation of the linguistic realisation of terminological expressions in their sentential and textual context of the corpus and the investigation of the question what the sentential and textual context may contribute to the (human and computational) interpretation of terms. Here, the main focus is on a further conceptual and linguistic analysis of the corpus which especially takes into account the role of domain dependent and general language verbs within the specific subject field, and on how this analysis may support the entire conceptual analysis of the domain. We think that this perspective, although restricted to a special subject field, may also contribute to the general treatment of semantics within a computational environment.

The theoretical modelling of the domain and the integration into a NLP environment is inspired by two main sources: (1) Frame inheritance based knowledge representation systems of artificial intelligence (sc ai), and (2) Conceptual graphs (Sowa 1984).

The first source is essential for the description and representation of the *generic* and *partitive* knowledge of a domain; whereas the second is useful for formalising a kind of *predicate–argument* structure of the concepts of the domain to ensure the link between the conceptual organisation and the linguistic organisations (grammar and lexis). We have chosen these sources because of their adaptability to constraint–based grammar formalisms.

The overall leading idea for the integration (linking) process is to control a *competence grammar* for general language with conceptual (terminological) information, which we call *performance control*. As one possibility this is done entirely on a lexical basis by so–called *terminological anchors* which link the different dimensions of the conceptual information to general semantics. The advantage we gain from this approach is that we have the *full grammar* as a *fall–back* in cases were the conceptual information cannot contribute to the disambiguation process, due to

ambiguities inherited from the domain itself. In this approach terminological information is used as one instance of a *sublanguage constraint.*

1.2 Major results

In our research we have established a methodology for computational terminology that is not based on intuitive grounds about what is and is not *true* about the considered subject field, but instead on empirical and practical concerns, namely what data, information and knowledge the processes of natural language analysis and translation require in order to allow for a performance control that enables the resolution of domain specific ambiguities which a competence grammar for general language is unable to perform.

The conceptual knowledge of the specific domain obtained in a knowledge acquisition phase, i.e. the corpus analysis phase, is effected in three steps:

1. Building a taxonomy of conceptual realisations required to describe the domain (ontology construction).
2. Organising the domain entities in terms of the conceptual structure and their linguistic realisations.
3. Integrating the result into a constraint–based NL description in order to enable the control of domain specific organisations.

After the knowledge acquisition process and the formalisation of that knowledge (organisation and representation), the resulting knowledge structures contain, declaratively and explicitly represented, those distinctions required to control the analysis process in an effective way and to provide for a concept–based transfer. In addition, the implementation demonstrates the feasibility of the chosen approach which can be characterised by:

- Built–in 'fall–back', i.e. the competence grammar for general language will cover unconstrained situations.
- Generic knowledge of one sublanguage is included.
- Integration of further sublanguage information can be performed easily.

The results will also have implications for the emerging *language technology*, in so far as our methodology provides a solid foundation for: advances for linguistic engineering, NLP efficiency, sublanguage handling, improving robustness.

Finally, our framework also allows for the integration of other performance models, for example, it could be combined with statistical and/or connectionist approaches.

2. Conceptual analysis of domain

2.1 Characteristics of the domain

The selected domain, telecommunications or satellite communications, of the project is concerned with the advanced techniques and facilities allowing for very fast communications around the world.

This includes various cable, e.g. fiber optics, and satellite techniques which allow us, for example, to make private phone calls across continents, to receive TV programmes and to use effective and reliable electronic mailing utilities via computer systems.

In this domain processes, such as transmitting, receiving, amplifying, modulating and demodulating of different kinds of telecommunication products (signal, wave and data), perform a stand out function, in so far as they trigger the active part of the domain. They operate by using a certain instrument that could be conceptualised as a specific equipment of the domain, for example, converters, modems, amplifiers and multiplexers. The entire process can be facilitated by a certain method, such as digital speech interpolation, time division multiple access and frequency division multiple access, and each process usually has a specific type of product as input and output, such as digital data, intermediate frequency signal and radio frequency signal, which can be conceptualised as specific products, as well as can be characterised by specific properties, such as frequency rate, noise temperature, bandwidth and number of channels. Each of these domain specific interdependencies has to be expressed by a formal description, the conceptual structure, in order to allow for on the one hand, a classification of domain concepts, and on the other hand, a particular linguistically motivated NL realisation, for example, verb phrases, noun phrases and prepositional phrases. In the particular domain instruments are often introduced by the preposition *by*, locations by the preposition *at* and purposes by the preposition *for*. Besides temporal, causal and spatial relations between conceptual elements there are also relationships between concepts which share one common superconcept, for example, digital transmission and analogue transmission as concepts may have the concept transmission as superconcept. This relationship is the taxonomic (generic) IS–A relation known in term subsumption systems, and realised in the project's ontology (cf. Section 2.3.1).

2.2 Representation formalism

For the formalisation of an expert's knowledge *conceptual analysis* provides general techniques for analysing knowledge of any subject field. Thus, it is a method of analysing informal knowledge expressed in natural language as a preliminary stage to encoding it in a knowledge representation

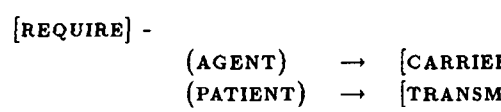
language. In our research we have used conceptual graphs (Sowa 1984; Sowa 1991) as the primary knowledge representation language, but the techniques could be applied to any other AI language.

Conceptual graphs are a version of semantic networks designed as a complete system of logic, including modal and higher–order forms: they have a direct mapping to and from natural language; they can be translated to and from other AI formalisms; and they can support automated knowledge acquisition tools. As an example of the mapping from language to conceptual graphs, consider the following sentence, which appears in the telecommunications corpus: *A transmission chain is required for each carrier.*

The analysis starts with the *content words* in the sentence: *transmission, chain, require*, and *carrier*. In any database or knowledge base, these words would map to some significant feature. In conceptual graphs they map to *type labels* for the associated concepts which are organised in a *type hierarchy.*

The other words in the sentence – *a, is, for*, and *each* – are usually called *function words.* Unlike the content words that map to concepts, function words map to conceptual relations or to quantifiers inside a concept node. The words *a* and *each* represent quantifiers; the auxiliary verb *is* and the ending *–ed* mark the passive voice which indicates that the concept REQUIRE is linked to TRANSMISSION_CHAIN by the patient relation; and *for* indicates that REQUIRE is linked to CARRIER by the agent relation.

This situation can be depicted graphical or in a linear notation:



This notation can be mapped to the predicate calculus by the *formula operator OE* which assigns a constant or a quantified variable to each concept node. Type labels map to one–place predicates and conceptual relations map to predicates with as many arguments as there are slots attached to relations:

$$(\forall x)(\exists y)(\exists z)(carrier(z) \supset (transmission\_chain(y) \wedge require(z) \wedge agent(z, x) \wedge patient(z, y)).$$

The conceptual graph representation also allows for the translation into the representation formalisms used in constraint–based NLP system (cf. section 4).

2.3 Domain modelling

2.3.1 Domain ontology

The major concepts of the domain, i.e. those concepts which are mainly realised by nouns and nominalised verbs, are represented in the domain's ontology; they are characterised by *descriptors* which list the properties of the real world thing the concept denotes, e.g. *input, output, location,* etc.

The ontology constructed within the project shall serve as representation device for extra–linguistic knowledge to be integrated in an NLP system and as a classical knowledge base (in the AI sense). The information source the ontology is based on are a set of term descriptions, the so–called term definition forms. For each term occuring in the sample corpus, such a form exists describing the inherent properties of that term and its relationships to others. The whole set of these forms represents a kind of concept system known from traditional terminology. Such systems have the advantage compared with knowledge bases known from AI that they are more or less application independant. To transform such a terminological concept system in an ontology allows to meet the two requirements described above: it will represent common domain knowledge together with the linguistic information.

As the most appropriate formal device to realise such a kind of knowledge base the concept of an *interface ontology* (cf. Bateman, 1992's classification of ontologies) was chosen. It's basic model for the representation is a subsumption lattice over types with a mechanism corresponding to structured inheritance of attribute information associated with the concepts.

The domain considered in the project, telecommunications, gives clear hints about the possible structure of the knowledge base to be used. It is mainly *process–based* around the core concepts TRANSMISSION, RECEPTION, MODULATION, AMPLIFICATION and MULTIPLEXING. Additional concepts to describe them are added: Each of these processes deals with an object (here mostly a SIGNAL) which belongs to the class of PRODUCT, uses an instrument (member of the EQUIP class) and a certain method (belonging to the METHOD class).

The overall structure must be monotonic, acyclic and therefore a tree because the type system of the ALEP framework requires a strictly hierarchical order. The whole ontology is spanned over a root node TCOMM and consists of the supertypes PROCESS, EQUIP, PRODUCT and METHOD. All other concepts are classified and arranged in the lattice according to their classification and *isa*–feature of the corresponding term definition form. The subtree for PROCESS, for example, looks as follows (in ALEP notation):

```
tcomm > {process > {transmission > {analogue_transmission,
                                    digital_transmission},
                    multiplexing > {fdm,
                                    tdm},
                    demultiplexing,
                    modulation   > {frequency_modulation,
                                    amplitude_modulation,
                                    phase_modulation      > {phase_shift_keying}},
                    demodulation,
                    amplification,
                    conversion   > {down_conversion,
                                    up_conversion},
                    propagation,
                    formatting,
                    reception}
        }
```

A concept in the lattice is considered to be determined by its extend and its intent: the extend consists of all objects belonging to a concept while the intent is the collection of all attributes shared by the objects. The term definition forms deliver also the characterisation of the concepts: The features and their values describing a concept's inherent properties are also directly derived from the definition forms (however, not all slots have a corresponding feature in the type declaration due to the fact that the definition forms are not only for machine use but also for human translators).

Due to the fact that ALEP's type system facility provides no special data structure to represent relations, these have also realised by means of feature–values pairs. Therefore the manifold interrelations existing between the concepts of one class and the interrelations between the classes can not be described in greater detail. This and the fact that *multiple inheritance* is not possible the current ontology shows only one description dimension although almost all concepts can be classified in more than one way.

2.3.2 Conceptual templates

To model something it must have a well–defined term, i.e. a word or a phrase, that denotes it. Based on the domain's ontology and a further conceptual analysis of the verbs used in the corpus, we defined *conceptual templates* as the knowledge structure for the information repository to be used in the NLP system to trigger the performance control. Such a template consists of a number of properties that characterises either a type, i.e. a thing that can have instances, or a class which governs types that specialise the class. The common classes are *entities, situations* and *properties. Entities* are those types that can have real world instances and which are realised linguistically as nouns and nominalised verbs.

Situations are facilitated by types that express time and place relations, and that identify participants, agents and result roles. Properties are types that denote verbs involving a thing as subject (actions and states), modifiers (adjectives) that describe details of a thing (e.g. measures), relationships that identify relational properties to other things, types that denote attributes that (partially) describe a thing, and types that denote constraints which are logical assertions that impose some restrictions on one or more properties of a thing.

For the actual specification of the information repository we have analysed the text part selected for the demonstrator implementation in terms of conceptual graphs. From this representation we derived the general semantics and the specific terminological descriptions to be used for the templates. In order to provide for a complete terminological description we have also used the classification schema of the term definition forms, i.e. the class typology, and the explicite term/concept definitions which were checked against their definition according to ISO and DIN norms, as well as existing de facto standards in the subject field. For the semantic descriptions we have used the semantic relations approach developed in EUROTRA–D; the domain specific information was associated to these relations in a separate information slot to permit the testing and evaluation of different terminological templates in an appropriate way. Thus, the general organisation for the semantic template is:

$$sem\_fs \begin{bmatrix} gov & \Rightarrow & gov\_fs\ [] \\ args & \Rightarrow & args\_fs\ [] \\ mods & \Rightarrow & mods\_fs\ [] \\ term & \Rightarrow & term\_fs\ [] \end{bmatrix}$$

The terminological template is:

$$term\_fs \begin{bmatrix} term\_info & \Rightarrow & term\_info\_fs\ [] \\ concept & \Rightarrow & tcomm\ [] \\ concept\_roles & \Rightarrow & concept\_roles\_fs\ [] \\ concept\_modify & \Rightarrow & concept\_modify\_fs\ [] \end{bmatrix}$$

The *term_info* structure contains general terminological information, i.e. the classification schema and the concept definition. The concept feature identifies the *concept* according to the ontology of the domain; the *concept_roles* structure specifies the role slots of the concept, derived from the situation class and parts of the property class, and the conceptual modifiers are listed in the *concept_modify* structure, also derived from the property class.

In addition, in a further information structure the overall syntactic information is specified. On the one hand, this organisation establishes the global structure of a knowledge base entity, and, on the other hand, the

complete template representing terminological information.

Within the domain, i.e. in the reference chapter of the telecommunications corpus, we have identified the following conceptual relations: *agent, experiencer, patient, state, result, point_in_time* and *location.* Verbs that denote actions are linked to their subject by the *agent* relation; verbs that denote states are linked by the *experiencer* relation if they are mental states, and by the *state* relation if they do not depend on any mental experience. If an object is created as a result of an action then it must be indicated by the *result* relation instead of the patient relation. With this information organisation the natural language processing, i.e. the NL analysis, can be controlled by:

- *Selectional restrictions* based on co–occurence patterns of general semantic patterns and specific domain patterns.
- *Subcategorisation frames* based on general semantic and domain specific information.
- *Conceptual classification* information (ontology).

The testing for selectional restrictions and the subcategorisation frames can be done by feature unification operations as used in constraint–based grammar formalisms; the use of conceptual classification information must be based on inferences over the ontology of the domain, i.e. type deduction during runtime.

3. Terminological information and translation

In terminography the focus is on concepts and their linguistic form expressed in terms which are extracted from texts (term identification). In translation the focus is on *production*, i.e. a dynamic process, concerned with the movement from the textual substance in one language to the textual substance in another language. Inside this process there is a procedure in which units of meaning of one culture are matched with those of another before finding their textually and situationally appropriate linguistic expression. In view of terminology these units are not of interest because they are temporary and casual collocations of concepts brought into a particular relationship by an author.

Translation has to work with concepts and terms in context, whereas terminology isolates terms from context (decontextualisation) and then associates them to concepts, i.e. matching between term and concept vs. matching between textual units.

Concept correspondence is discovered when comparing the terminologies of different languages, subject fields, school of thoughts etc. Based on this assumption there are thus four possibilities for the process of translation

based on the intension of a conceptual representation:

1. *Complete co–incidence* of intensions.
2. *Inclusion* of one intension in the other.
3. *Overlapping* of intensions.
4. *No co–incidence* of intensions.

By *intension* we mean the set of characteristics, i.e. the formal (or mental) representation of the properties of an object serving to form and delimit its concept, which constitutes a concept. The latter case of the concept based translations is called *conceptual mismatch* or *intensional mismatch.* No coincidence of intensions might be caused on social and cultural backgrounds, although the conceptual structures are not bound to particular languages.

Case 1 needs no specific translation rule; cases 2 and 3 need inferencing capabilities over the concept systems of source and target language and rules that can trigger the inference procedure; case 4, however, needs explicit translation rules.

4. Demonstrator implementation

In the previous section we have established the theoretical framework for the integration of terminological knowledge into the analysis and translation process of an NLP system; in this section we describe the actual implementation in the ALEP framework.

4.1 Implementation overview

The general architecture of our analysis module is based on a staged processing that was also suggested in the ET–6.1 study (Pulman (ed.) 1991) for efficiency reasons. In our approach analysis is composed of two steps:

1. *Shallow syntactic analysis* for efficient parsing with a competence grammar for German.
2. *Semantico–terminological refinement* of the parsing result as performance control.

With the second step we achieve a semantic filtering and a domain–specific filtering of the parsing results. For parsing we have used a grammar and a lexicon for general language which includes the terms of the domain, but without any particular domain information; for the refinement process (filtering) we have used a lexicon with general semantics and domain–specific information where necessary; in this step the grammar remains the same. As source language we have chosen German for the implementation of the demonstrator.

For the transfer module which has been designed for mapping German analysis output (so–called *linguistic structures*) to English synthesis input, we have adopted the option foreseen in ET–6.1 and, thus, in ALEP, that translation may be called on a specific type contained in the top–most feature structure of the input linguistic structure, i.e. the semantic and terminological (sub–) feature structures. Compared to the German analysis module, the transfer module as well as the English synthesis modules have a very limited coverage. This is mainly due to the fact that the focus of the project is on the conceptual organisation of the domain and the use of terminological information within the analysis process.

4.2 Type and sort system

The formal specifications for the conceptual and sortal organisation as described in Section 2.2 can be directly expressed in terms of the type system facility of the ALEP formalism. Since we could not use directly the properties specified for the entities of the domain, due to internal system restrictions on the overall size of a type system, we specified the domain's ontology only by its partitive structure in the type system's subsystem for terminological information.

4.3 Grammar and lexicon for parsing

In the parsing grammar we have specified the information distribution of the semantic feature structure which includes as a substructure the conceptual knowledge organisation about the domain. During parsing these information slots are opened and during refinement they are filled in by the appropriate information by unification. Unification failure then triggers the disambiguation process in the analysis phase and thus the performance control in analysis.

4.4 Lexicon for refinement

In the refinement lexicon we have stated the selectional restrictions for different semantic and conceptual reading distinctions, as well as the appropriate subcategorisation frames. This information is used during the refinement process to identify valid parsing results by unification. The result of the refinement process is a fully specified analysis representation according to the selected semantic and conceptual information.

Figure 1 shows the lexicon entry for *adaptieren* (adapt).[1] In the entry the semantic subject *agent* is linked to the conceptual role *agent* which is of type EQUIP which is a type of the domain's ontology, and the semantic object *affected* is linked to the conceptual role *patient (ptnt)* which is of type SIGNAL.

Selectional restrictions based on specific domain information for nouns are linked to the noun's subcategorisation frame and which can be associated

with the prepositions, like *von* (of) and *mit* (with) which have a specific interpretation in the domain. Similar to these, domain dependent restrictions can be formulated for the prepositions used in the sample corpus.

4.5 Results

All sentences of the sample corpus were analysed successful. Since we could not use constraints on the conceptual organisation, there are still ambiguities left which are due to the phenomenon of subject/object topicalisation. For example, in *Fernübertragungsausrüstungen umfassen auch Modulationsgeräte*, we got two readings. A constraint stating that the concept associated to the subject must be less specific than the concept associated to the object would resolve this ambiguity. The same holds for sentences were conceptual relations are explicitly expressed, for example, the *part_of* relation. However, the demonstrator implementation has shown the advantage we gain from introducing domain specific information in addition to general semantics into the analysis process.

4.6 Transfer relations

In the following, we restrict the description of translation to cases of conceptual mismatches (case 4 above). Within the transfer module there is one rule for initialising the translation process. Once translation is called on the semantic and thus the terminological (sub–) feature structures specified as the value of the linguistic structure's top–most *sem*–attribute, translation is called recursively on type SEM_FS and all subordinate types respectively. When translation is called on type SEM_FS the predicate string specified by the *pred*–attribute within the governor feature structure is translated from one language into the other. For the translation of the appropriate conceptual information rules for the different conceptually dependent arities are used. This approach allows for a straightforward account of instances of complex transfer where changes have to be performed according to the argument structure of the predicate that has to be translated (this applies to the general semantics of SEM_FS only).

```
adaptieren '
mLEIde_SIGN_refine[
      mLEIde_SYN_MAJOR[
            -,
            mLEIde_HEAD_V[],
            m_SUBJ[
                  sign:{m_COMPL_N[nom,ARG1]}],
            m_SUBCAT_1[
                  sign:{m_COMPL_N[acc,ARG2]}]],
      m_SEM_term_yes[
            m_GOV_V[adaptieren,action],
            m_ARGS_BI[
                  m_ARGselec[agent,ARG1,sem_fs:{term=>term_yes:{concept=>equip:{}}}],
                  m_ARGselec[affected,ARG2,sem_fs:{term=>term_yes:{concept=>product:{}}}]],
            term_yes:{
              term_info=> term_info_fs:{
                          class       => class_fs:{
                                          c1_type => tcomm,
                                          c2_type => tranm,
                                          c3_type => process},
                          definition  => _,
                          form        => no_mvt},
              concept     => modulation:{},
              concept_roles => bi_c_role:{
                               concept_role1 => conceptual_fs:{
                                                concept_role  => agnt:{},
                                                concept_type  => equip:{},
                                                concept_descr => term_fs:{}
                                                                 },
                               concept_role2 => conceptual_fs:{
                                                concept_role  => ptnt:{},
                                                concept_type  => product:{},
                                                concept_descr => term_fs:{}
                                                                 }
                                             },
              concept_modify => _ }]].
```

Figure 1: Refinement entry for '*adaptieren*'.

A domain–specific role structure of a concept, identified by the terminological attribute *concept_roles*, is translated by a rule dedicated to the relevant subtype of type CONCEPT_ROLES_FS. For instance, the role structure assigned to the predicate *senden* is translated by a rule operating on the subtype TRI_C_ROLE and calling recursively for translation on type CONCEPTUAL_FS which is the type assigned to the roles of a concept.

Type CONCEPTUAL_FS will, then, be translated by a rule which, in turn, calls for translation on type TERM_FS again, since the value of *concept_descr* is a terminological feature structure. Accordingly the semantic argument structure is translated.

The translation of the modifier–list of a concept in SEM_FS and in TERM_FS, finally, is performed by distinct rules with each of them accounting for a

specific number of elements specified in the modifier list (including the empty modifier list).

4.7 Synthesis

Ideally, the basic *sign* feature structure and, more specifically, the semantic and terminological feature structures should be the same for all languages. With this assumption, it should be only the syntactic feature structure which has to be revised in designing the type and feature specification for an English synthesis grammar.

Since no refinement can be applied in synthesis, the English synthesis grammar must operate on a single lexicon which contains fully specified lexical entries including terminological information too (i.e. specific syntactic realisation information).

5 Limitations and conclusions

In this paper, our focus was essentially on how conceptual information can act as a performance control of a competence grammar and how terminological information can contribute to this task as well as to better translation output. We have described our approach which is entirely lexicon based. There are also limitations in the actual implementation which are due to the formalism used in this research. At present the ALEP formalism does not allow for multiple inheritance which in knowledge representation is an asset. We have circumvented this problem through the introduction of additional relations which are realised in the attribute slots of a feature structure type.

Another area which is not supported by the ALEP implementation is type deduction during runtime. This facility would be useful for completeness and coherence tests on the generated information structures and the handling of conceptual mismatches during translation, and would thus be an additional sort of performance control for the system.

Although the ALEP formalism lacks essential features for our purpose and in contrast to other recent formalisms and systems, such as ALE, we have decided to use this formalism because of its intended ability to serve as a general purpose NLP platform for large–scale linguistic engineering and as a kind of standard for future NLP systems. In this context it should be noted that ALEP is still under development (under the supervision of the CEC and well reputed NLP researchers from academia and industry), and that our research results will thus act as a stimulus for future ALEP development work.

Finally, our approach to a knowledge base for terminology shares some similarities with the COGNITERM project of the University of Ottawa (Meyer et al. 1992) facilitated by the knowledge management system CODE (Skuce and Lethbridge 1993). However, the main focus of this project is the design

of a terminology knowledge base without any reference to an integration into an NLP system.
This task is foreseen for a future collaboration between our institutes.

Notes

1 In order to give a real clue of the structure we have used the macro notation of the ALEP formalism instead of the fully expanded notation.

References

ALEP, 1993: *ALEP Documentation Package*, Vol. I and II. CEC and PE International, Luxemburg.
Bateman J., 1992. "The Theoretical Status of Ontologies in Natural Language Processing". In: *Proceedings of the International Workshop on Text Representation and Domain Modelling*, TU Berlin.
Meyer I., Bowker L. and Eck K., 1992. "Cogniterm: An Experiment in Building a Terminological Knowledge Base". In: *Proceedings of the 5th Euralex International Congress*, Tampere, Finland.
Pulman S.G. (ed.), 1991. *ET–6/1 Final Report. CEC, DG–XIII*, Luxembourg.
Skuce and T. C. Lethbridge. 1993. *Code4: A Multifunctional Knowledge Management System.* Department of Computer Science, University of Ottawa, Canada.
Sowa J.F., 1984. *Conceptual Structures: Information Processing in Mind and Machine.* Addision Wesley, Reading, MA.
Sowa J.F., 1991. Towards the Expressive Power of Natural Language. In: J. F. Sowa (ed.). *Principles of Semantic Networks: Explorations in the Representation of Knowledge.* Morgan Kaufmann Publishers, San Mateo, CA.