Towards the application of text generation in an integrated publication system*

Elke Teich and **John A. Bateman**
Institute for Integrated Publication and Information Systems (GMD/IPSI)
German National Research Centre for Information Technology (GMD)
Dolivostraße 15
D-64293 Darmstadt, FRG
e-mail:{teich,bateman}@darmstadt.gmd.de

Abstract

We describe the application of multilingual text generation in a system for assisting the process of publication. This system is an editor's workbench for preparation of the publication of an art history encyclopedia (the Macmillan *Dictionary of Art*), which is itself part of an integrated publication environment being developed at GMD-IPSI. We show how an editor's tasks can be facilitated by the use of NLP (natural language processing) systems and suggest the important role of text generation in future electronic publications as products. In both cases, we focus on text generation as providing an essential new mode of information presentation. Text generation provides a quality gain in which the *flexibility* of the electronic product is augmented; in particular, *views on knowledge* expressed as text, possibly in different languages are incorporated. The major prerequisite for making this possible is an explicit and systematic representation of *genres* or *text types* combined with a general interfacing method for specific domain knowledge.

1 Overview of the paper

The function of this paper is two fold. First, we present an example of a new application area for text generation: support tools for electronic publishing. Second, we provide a brief overview of a new generation architecture based closely on the 'stratified context' systemic-functional linguistic model of [Martin, 1992]. The aim of the latter introduction is primarily to promote discussion of this kind of architecture, since the details that can be presented in the space of this paper are necessarily limited. The architecture is also highly experimental at this time. The simultaneous presentation of an application emphasizes the systemic-functional linguistic commitment to the inseparability of theory and practise.

The background context of the work reported here is an electronic publishing support tool consisting of a workbench for assisting the tasks of an editor involved in the construction of a large encyclopedia of art history, the Macmillan *Dictionary of Art*. This *editor's workbench* [Möhr and Rostek, 1993, Rostek *et al.*, 1994] is part of an integrated publication environment, which has been used as a framework for the design of an electronic newspaper [Haake *et al.*, 1993]. We are currently involved in adding natural language processing capabilities to this system—including most relevantly for this paper, multilingual text generation. This is being used both to further the theoretical specification and practical implementation of the text generation system and to enable us to concretely propose the application of text generation both in the electronically-assisted construction of encyclopedic works and in the electronic version of the electronic publication as product.

In this paper, we concentrate particularly on the use of the text generation system in order to bring a new functionality to the editor's workbench. The text generation component supports a range of flexible presentation styles tailored to the editor's requirements. The applicability of the text generation system is supported by two crucial capabilities:

- the knowledge of genres or text types, which are taken to be the ultimate constraining factor for the realization of *views on domain knowledge* as text,
- a methodology for building a generic interface to domain knowledge.

We suggest how both these aspects have been strongly supported by a version of the systemic-functional linguistic model employing a stratified view of context.

The paper is organized as follows. Section 2 presents our view of the role of natural language processing in electronic publication and introduces our application scenario. We outline here the editor's workbench and describe the necessary application of NLP components, focussing on text generation. We further show how the specific linguistic paradigm we are working with, systemic-functional linguistics, supports the integrated publication scenario. Section 3.1 presents the text generation component in more detail. We argue that the

*The research described in this paper was supported in part by the European ESPRIT Basic Research Action DANDELION (EP6665) and by the US National Science Foundation Grant IRI-9003087. John Bateman is also on indefinite leave from the Penman Project, USC/ISI, Marina del Rey, California

newly gained flexibility (views on domain knowledge expressed as text) can only be made possible in practice if the genre of the text to be produced is explicitly represented. Section 4 gives an example of generation. We conclude the paper with a summary of the new opportunities offered for electronic publishing and electronic and print publications when automatic text generation is made available in the way proposed (Section 5).

2 Electronic publication and NLP

Electronic publishing is an application domain in which the acquisition, representation and presentation of knowledge are major tasks which are to be fulfilled by an editor. State-of-the-art electronic publishing technology offers a number of possibilities both to facilitate these tasks and to improve on the quality of print and of electronic publications. In publishing in general, the *document* or *text* plays a central role in that it is both source and product of the publication process. Many of the tasks of an editor in this process are of a highly recursive character and are subject to a number of publisher-specific standards an editor must use. For example, the header of a biography text usually adheres to some specified standard of what information must be included and at which place. Typically, the header first gives the name or pseudonym of the artist, then his/her place and date of birth (and death), and possibly his/her main activity (painter, sculptor, architect etc). Such highly standardized information is currently automatically classified into a net of domain knowledge by applying pattern matching methods (see [Rostek *et al.*, 1994] for examples and the methods employed).

The construction of a taxonomic inheritance hierarchy that this allows already offers new opportunities for accessing the information included and can be used to assist some of the above mentioned tasks of an editor. If, however, one would like to be able to apply these methods improving information access and assisting the editor's tasks to other kinds of information that are less standardized, the question arises of how to acquire these and integrate them in an object net. For example, in a biography article, the typical information included is the artist's education, influences on his/her work by other artists and art styles, his/her major works etc. This is information that can appear in any place in the text—or even in several texts—and can be realized linguistically in various ways. Extracting such information from source texts requires more sophisticated methods and representations that reflect the facts contained in the source text. The most adequate way to acquire this knowledge is automatic NL text analysis.

For an example of how this enriches the information represented in the object net see Figure 1, taking the example of the main activities and creations of the architect Peter Behrens.[1] This is mostly a network obtained by the KONTEXT analysis system for German [Haenelt, 1994], although some nodes and arcs have been left out or duplicated here for presentation purposes.

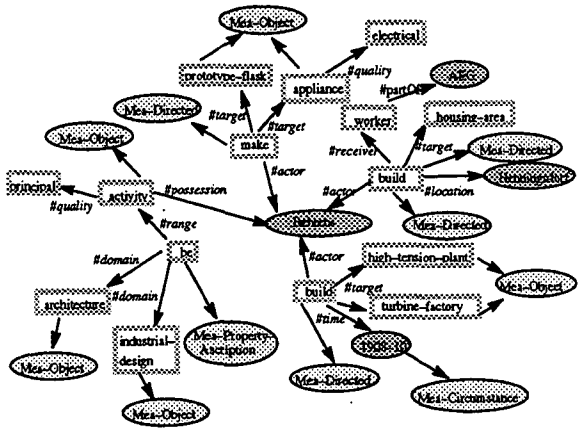

Figure 1: Knowledge acquired by NL text analysis

We assume that information of this kind and degree of complexity will become increasingly common in knowledge-based systems. We also assume that systems that previously worked with full-texts will increasingly become knowledge-based systems. However, having made available such a rich variety of potentially relevant information in this way, the complementary problem of *presenting* that information again to the user becomes the limiting factor. In our present case, the editor's workbench already offers a number of presentation styles for the information contained, such as the net or parts of the net (as shown in Figure 1, internal definitions associated with a node, article text associated with a node and cross-references to other objects, etc. Representation of complex information in the net style, however, quickly becomes too complex to be generally appropriate (as the rather simple example of Figure 1 probably already demonstrates). With a net augmented both in types of information and in quantity, other presentation styles as *views* on the information contained are essential for enabling an editor to deal with such complex information. One such presentation style that can flexibly express retrieved information is *automatically generated text*. Information including that displayed in Figure 1 (and other information not shown) is better expressed as the following text.

Sample Text 1

```
Behrens' principal activities were
industrial design and architecture. He
designed prototype flasks and electrical
appliances. As an architect, he built the
turbine factory and the high tension plant
for AEG (1908-10). For the workers of AEG
he built a housing area in Henningsdorf.
Behrens created a number of monumental
buildings, such as the Mannesmann
administration building in Düsseldorf and
the German embassy in St. Petersburg.
```

For further examples and motivations for using text generation in this kind of electronic publishing scenario,

[1] Domain-independent 'meaning model' (see below) upper structure types are given in light grey ellipses; named, unique objects and domain-specific objects are given in dark grey ellipses; instances are given in boxes.

see [Teich and Bateman, 1994].

3 The text generation architecture

3.1 Origins

For the kind of functionality set out in the previous section to be both achievable in theory and worthwhile in practise, we consider two features of a text generation component to be decisive. First, the text generation system must achieve a high degree of domain-independence in order to maximally re-use its resources. Such resources are too expensive to develop to allow redesign for different applications. And second, it must make provision for the systematic representation of genres and text types that it will be called to generate. The representation of genre not only offers a major constraint on the text generation process, including its access to domain knowledge, but also allows highly restricted fine-tuning of a system's generation capabilities.

The generation system being used in our application scenario is one under development in two continuing directions. First, it is *multilingual* in the extreme sense defined by [Bateman *et al.*, 1991a, Bateman *et al.*, 1991b]. The system thus supports extensive *linguistic resource sharing* on a functional basis. Second, it is intended to incorporate sensitivity to *register* as targetted by [Bateman and Paris, 1989, Bateman and Paris, 1991]. The system processes are based on the PENMAN generation system [Mann and Matthiessen, 1985] as extended for multilinguality by [Zeng, 1992], and includes work on multilinguality and text planning undertaken within the KOMET project [Bateman *et al.*, 1993]. The resulting system therefore has the working title KPML (KOMET-Penman multilingual).

The primary difference between the KOMET-PENMAN system and PENMAN is that KPML attempts to apply systemic-functional linguistic results at all levels of operation, following through some of the proposals for interaction between systemic linguistics and computational models made in [Bateman, 1988]. It therefore includes linguistic resources at various strata such as:

- the systemic functional grammar NIGEL for English [Matthiessen, 1992] and the KOMET grammars for German [Teich, 1992] and for Dutch [Degand, 1993],
- a merged upper model [Henschel, 1993] appropriate for generation with all three grammars,
- systemic-functional networks for register, based on [Martin, 1992],
- system-functional networks for genre, also based on [Martin, 1992].

Work on discourse semantics, in the sense defined in [Martin, 1992], is also in progress[2] Realization is captured in terms of relational constraints holding across features and structures defined by the networks of adjacent strata. These are still partially implemented in terms of Penman-style 'preselection', although nondirectional relations such as those suggested in [Bateman *et al.*, 1992] are both necessary and under investigation.

3.2 Genre and Register: the global-level textual resource

According to systemic-functional theory (SFL), the ultimate constraints on all linguistic expression are the extra-linguistic *contexts of culture* and *situation*. Especially in a multilingual application this must be taken into account since it is only in these extra-linguistic contexts that commonalities and differences between languages can ultimately be rooted. In early work on text in context such as [Hasan, 1978], the cultural context is reflected in language in terms of *text* as the linguistic category of *genre*. Genre is a culture-specific category textually encoding a situation as it can typically occur in a culture or linguistic community. Thus, it is at the interface between what is linguistic and what is non-linguistic knowledge. The particular view on genre and context that we instantiate in the KOMET-PENMAN architecture[3] is that proposed in [Martin, 1992]. Here context is divided into two strata, *register* and *genre*. Both are fully represented in our model.

Genre is represented in SFL by *generic structure potential* (GSP), specifying the potential typically occurring stages according to which a text belonging to a particular genre develops [Hasan, 1978]. As a brief example of a GSP and its application, a typical *field* (one of the parameters of register; cf. below) in the domain of arts is information about an artist's life. A typical situation in terms of field involves all kinds of activities and events a particular artist is or was involved in in his/her career and the circumstances under which these events take or took place. It is not possible from this information alone to construct any particular text. Only when a particular text type has been selected is it possible to determine what information from the field is relevant and how it is to be presented. One genre that usefully expresses such a typical situation is the *biography*. A typical GSP for a biography text then has the following stages, of which not all are obligatory:

GSP stages	
1	Names, birth and death
2	Education, development of career
3	Major activities, major works
4	Influences, analogies to other artists
5	Impact

Uncovering such generic structures is a large empirical task that needs to be addressed in text generation: a system such as KPML can then be seen as one candidate way of noting the results of such studies in an immediately usable fashion.

Following [Martin, 1992], we provide a systemic account of genre using the same representational means as for grammatical descriptions (system networks). This permits generic structures to be constructed according to the selection of genre features in exactly the same

[2] Particularly in the context of the DANDELION basic research project.

[3] For further detailed motivations for this selection, see [Bateman and Paris, in preparation].

```
biography
   (instantiate
      Particular (v / Mea-Person))
   (stage Name)
   (instantiate
      Name (v / Mea-Named
              actee Particular))
   (stage Birth)
   (instantiate
      Birth (v / Mea-Birth-Event
               actor Particular))
   (inter-preselect-feature
      Birth reconstruction)
   (inter-preselect-feature
      Birth activity)
   (stage Education)
```

Figure 2: Realization statements for the genre [biography]

way that grammatical structures are constructed in our grammars—i.e., by means of the selection of grammatical features in system networks. The GSPs that result then condition the further realization of register selections in the linguistic system proper. Thus situation types are related to the linguistic system in particular ways by the selection of particular genres. The systemicization of genre specifications also ensures that there can be maximal re-use of existing genre specifications when new genres, slightly differing from already treated ones, are considered.

Furthermore, as noted by [Matthiessen, 1988], generic structures are in many ways similar to [McKeown, 1985]'s rhetorical schemas. They can also be used, therefore, in order to pre-structure the information to be expressed in a text. Particular kinds of information go together with particular stages of the GSP. The specification of this information—in our case as constraints in the genre network concerning appropriate selections of field features and structures—provides the content for information retrieval inquiries that are passed on to the editor's workbench for instantiation. An example of such a specification is shown in the *instantiate* realizations for the genre-feature [biography] given in Figure 2. The realization constraints *stage* introduce generic stages in a GSP; the constraints *inter-preselect-feature* constrain a given generic stage to be realized by a particular register feature selection.

The use of GSPs for knowledge selection is further enhanced by our requirement that domain information is already classified according to a general conceptual hierarchy for organizing domain information that is specified by KOMET-PENMAN internally. This guarantees that the retrieved information is in a form which can be readily interpreted by the other resources of the linguistic system maintained internally to KOMET-PENMAN. This provides the definitions for the semantic types used in the SPL-like second parameter to the 'instantiate' constraints shown in Figure 2. At present, information is retrieved from the editor's workbench which matches the these second parameters. Work in progress is exploring more sophisticated information retrieval operations.

The representation of genre therefore offers a major constraint on the text generation process, including its access to domain knowledge. As mentioned before, not all of the stages of a GSP are obligatory. This gives rise to a number of subtypes of the biography genre. For example, a short biographical entry, as it typically occurs in an encyclopedia realizes all stages and tends to be activity-focussed and timeline organized. Our sample text 1, on the other hand, did not realize all stages; it presents more the roles Behrens took on as an artist (designer, architect) and his major works in one of these roles (the text therefore focuses on activities of *creation* and the results of these creative processes and is not explicitly timeline organized). Both kinds of texts are, however, useful for an editor (and for an information seeker) in particular contexts and so both should be available presentation styles. Unless the text generation component allows the ready definition of such genres and their linguistic realizations with maximal re-usablity of existing resources, such flexibility will be compromised.

The next strata in the system, register, is further described in terms of the three parameters of *field, tenor* and *mode*. Field describes the states and events and participants occurring in a particular situation. Tenor refers to the roles and statuses of the participants in a particular situation. Mode refers to the symbolic organization of the situation as text, including the channel of communication and the rhetoric goals [Halliday and Hasan, 1989, p12]. Specific sets of values that realize field, tenor and mode bring about *situation types* or registers. Particular situation types accordingly constrain the kind of language that may occur in those situations.

Certain aspects of register are constrained by the situation that the text generator is to understand itself as being in—thus, whether the text is to be written or spoken, whether the relationship between the hearer and the speaker/writer is distant or close, etc. must be selected externally. The provision of such features in the register specification automatically provides a high degree of parameterization. Moreover, the field information provides a natural home for many aspects that would traditionally be included under 'domain knowledge'. This information is quite removed from its expression in natural language. In order to interface with such information, however, we provide a general subsumption hierarchy analogous to previous usages of the *upper model* within PENMAN (cf. [Bateman, 1990]). Further details of the motivation and development of this hierarchy are given in [Bateman *et al.*, 1994]. This (domain-independent) 'upper structure' of the domain knowledge is motivated by the semantics defined by our NLP components (e.g. [Kunze and Firzlaff, 1993, Bateman *et al.*, 1990, Henschel, 1993]). We see examples of the use of the hierarchy in Figure 1 in the net objects prefixed by *Mea-*.

The hierarchy can in general be considered to be a system network with realization constraints (further conditionalized by the contributions of feature selections from the mode and tenor hierarchies) effecting

the possible semantic and grammatical realizations that may be adopted. However, the interfacing with domain knowledge is further supported by allowing the mixing of KOMET-PENMAN internally specified system networks and *externally* specified resource organizations, such as, e.g., that for domain knowledge. Most of the upper structure for the domain knowledge is therefore represented externally in SFK (the 'Smalltalk FrameKit' [Fischer and Rostek, 1993]), which is the implementation language for the editor's workbench. This is linked back to a KOMET-PENMAN internal portion of the hierarchy which directly attaches to the top-most field features and positions the combined field hierarchy with respect to tenor and mode. When an external field feature has consequences for linguistic realization, it is 'imported' into the KOMET-PENMAN internal systemic network by means of *external-system* definitions. These function entirely analogously to internal systems as far as the connectivity of the network is concerned, but the selection of features in such systems is linked directly to the state of the corresponding external representation. In practice this permits modules of external 'system networks' to be incorporated in the text generation process.

The constraints of the genre specification as GSP and register on local-level discourse organization of the text are manifold. For example, with a GSP for a biography text as sketched above, the thematic development is typically taking the person the text is about as *macro-theme*. Also, very often biography texts are organized along a time line — circumstantials of time such as dates, years or time adverbials are therefore often thematic or part of theme. We are currently extending the set of such constraints based on further analysis of text types that are useful in our application context.

Finally, as already noted, there can also be subtypes of the biography genre that do not have all of the stages listed above, but focus on one or two of them, leaving out others (see again sample text 2). They therefore realize only parts of the information available in the field of context (i.e. the domain knowledge). This contributes to the formation of distinct *views* on the information maintained in the editor's workbench and electronic publications.[4] The organization of the possible genres in terms of a systemic classification hierarchy then guarantees that we are able to capture similarities between related genres and to represent the constraints they share in an efficient fashion.

4 Example of generation

In this section we present a very brief example of generation in the context of the editor's workbench. The system is configured according to the overall architecture shown in Figure 3. The requirement that text be generated is embedded in the editor's interaction with the editor's workbench. A range of possible types of presentation are offered: currently these are restricted

[4]Other contributions are the currently focussed portion of the information net that the editor has established during his/her interaction with the system and possible user-profiles of interest.

```
activityOrCreationPatternCauser: aString target: aType
  "| res |
  (res := self activityOrCreationPatternCauser: 'behrens' target: meaObject) = nil
    ifFalse: [res do: [:x | Transcript show: '*'; cr.x kometTextPrintOn: Transcript]]"
  | genconc attrib |
  genconc := DoAGeneralConcept! aString.
  attrib := genconc at:
      (SlotPath ~ #meaObject ~ #sfkRoleOf ~ #factorOfOrCauserOf ~ #target).
```

Figure 4: SFK retrieval pattern

to several subtypes of biographies. The text generation process is then started with the constraint that some set of genre features hold; in addition, some set of register features are also selected in order to characterize the communicative situation. The text generation component then constructs a full GSP specification for the text type. This sets up staged sets of constraints on knowledge to be selected from the domain. That information is retrieved and classified according to the complete register classification network. This establishes

- a set of constraints for the semantics and grammar,
- a set of pools of contextual information that are to be realized linguistically.

The latter are then organized rhetorically, thematically, ideationally, etc. according to the constrained semantics, which results in sets of rhetorically organized semantic specifications that can be passed to the constrained grammar for final realization.

More specifically, the text generation component carries out the following operations. First, the editor's workbench provides a focus for the textual summary that is to be constructed for the internal information represented and triggers generation. A specific genre must also be selected at this point: for example, one classified by the set of genre features: {`activity-focused partial-biography third-person-recount particularized factual-genre`}. These features give rise to the concrete GSP:

`(DEVELOPMENT ACTIVITIES)`

Each of these stages has particular information needs, which are then instantiated with respect to the knowledge base. For example, for the stage `ACTIVITIES`, the information needed consists of events and activities the artist dealt with is involved in as actor or causer. This is specified in an SFK retrieval pattern with the artist (*Behrens*) as starting point and providing the slot path along which the retrieval takes place; this is shown in Figure 4.

An example of the result of this information request is as follows:

```
(SFK-337 / MEACREATION
  ACTOR
  (SFK-338 / (BEHRENS MEAOBJECT
```

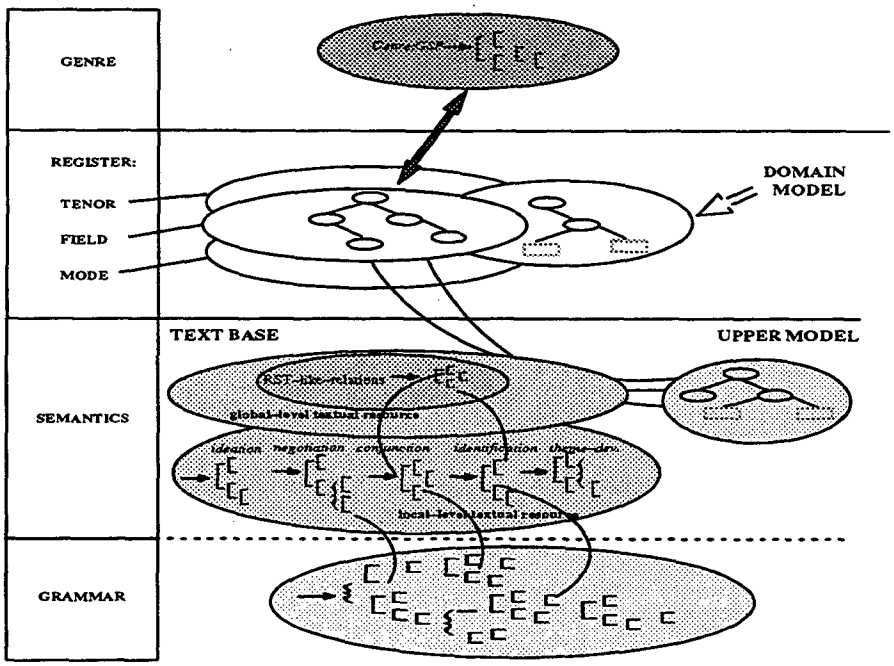

Figure 3: Generation architecture respecting genre

```
          MEANAMED MEAMALE))
TARGET
(SFK-339 / (MEAOBJECT DECOMPHOUSINGAREA))
SPATIALLOCATION
(SFK-340 / (MEA-THREEDIMENSIONALOBJECT
            HENNINGSDORF))
BENEFICIARY
(SFK-341 / (MEAPERSON PERWORKER MEAPARTOF)
         PARTOF
         (SFK-342 / (AEG MEAOBJECT
                     MEANAMED))))
```

Here a number of domain instances (SFK-337, etc.) with particular types (e.g., MEACREATION, MEAOBJECT, etc.) are returned. These types are as defined in the field conceptual hierarchy. Instances with particular types also have associated roles (e.g., ACTOR, TARGET, PARTOF). The instance specification as a whole represents a *partial* register-stratum structure. This partial structure is then filled out to a maximal description by traversing the register-stratum classification networks. This traversal is partially constrained by the GSP, by the instantial domain structure, and the specified context of generation. The result of this traversal is a full set of register features, and a set of corresponding constraints on the semantic and discourse realization of the generic stage. An extract from a register 'profile' resulting from such a traversal is shown in Figure 5.

These constraints are then applied during the further lexicogrammatical expression of the information retrieved from the knowledge base in order to construct an appropriate text. This is mediated by the local-discourse semantics that groups the information of the generic stage into rhetorically organized semantic specifications. These specifications consist of a statement of the propositional content, plus discourse constraints. An example is shown in Figure 6.

The final (English) text then generated is, in this case, as follows.

" -- Behrens's principal activities were architecture and industrial design. -- He made electrical appliances and prototype flasks. -- He built the high tension plant and the turbine factory for AEG in 1908 - 1910. -- He built a housing area for the workers of AEG in Henningsdorf. -- He created a number of monumental buildings, such as the administration building of Mannesmann in Duesseldorf and the German embassy in St. Petersburg."

The text could clearly be improved in a number of ways, many of which form active areas of research. Similar texts are created for German and Dutch.

5 Conclusions

In this paper we have presented the experimental interfacing of a multilingual text generation system, KOMET-PENMAN, to domain knowledge about arts and art history in an editor's workbench assisting the publication of an art encyclopedia. The major improvement that the application of text generation offers is a gain in flexibility with regard to the presentation of *views* on the domain knowledge. The text generation system itself organizes the access to the domain knowledge in a linguistically-motivated way: The representation of *genre* constrains information access, thus predefining, as it were, the *textual views* that are possible to be taken on the information contained in biographical entries in the encyclopedia. The method proposed for the editor's workbench

```
REGISTERIAL PROFILE FOR GENERIC STAGE: (ACTIVITIES)
===================================

     Following general constraints hold on realizations of this stage:

((MACRO-THEME . INST-484)
 (LEXICOGRAMMAR :EVENT-Q OBJECT))

     Semantic image created for this stage includes:
          concept:WORKER                  is a subconcept of concept:PERSON
          concept:HOUSINGAREA             is a subconcept of concept:DECOMPOSABLE-OBJECT
          concept:AEG                     is a subconcept of concept:NAMED-OBJECT
          concept:HENNINGSDORF            is a subconcept of concept:THREE-D-LOCATION
          concept:BEHRENS                 is a subconcept of concept:PERSON
          concept:BUILD                   is a subconcept of concept:CREATIVE-MATERIAL-ACTION

     Following registerial instances selected:

-----------------------------------------------------------------
Unit with:
          syntagmatic structure:
          (ACTOR TARGET FIELD-ACTIVITY)
          with register features selected:

(SPATIALLY-LOCATED TEMPORALLY-UNLOCATED HUMANITIES EXPLORATION
WRITTEN-TRANSMISSION CENTRAL-PARTICIPANT
MEACREATION-REALIZATION DOA-ACTIVITIES ACTIVITY-EXPECTANCY
TIMES-PAST ACTIVITY-SEQUENCE FIELD UNMARKED-AFFECT ONE-OFF
UNINVOLVED-CONTACT UNEQUAL-STATUS TENOR NONPSEUDO-UNPROJECTED
UNPROJECTED GENRE-STRUCTURED SOLIDIFIED VISUALLY-OBJECTIFIED INFORMING
DOCUMENTATION PUBLIC-GENERAL PUBLIC AURAL-NONE VISUAL-NONE MODE
REGISTER MEACREATION START)

With substructure:

...
```

Figure 5: Extract from register profile

```
...

 Setting local context according to constraints:
((CONSTRAINTS ((:EVENT-Q OBJECT))) (MACRO-THEMES (BEHRENS))
                                   (DISC-INDIVIDUALS
                                     (HENNINGSDORF AEG
                                                   DARMSTADT
                                                   BEHRENS)))
Semantic input to lexicogrammar:

((V-626 / (BUILD PROCESS)
        :SPATIAL-LOCATING
        (V-620 / (HENNINGSDORF THREE-D-LOCATION NAMED-OBJECT OBJECT)
               :NAME HENNINGSDORF)
        :BENEFICIARY
        (V-623 / (WORKER OBJECT)
               :PART-OF
               (V-622 / (AEG GROUP NAMED-OBJECT OBJECT)
                      :NAME AEG)
               :SINGULARITY-Q NONSINGULAR
               :MULTIPLICITY-Q MULTIPLE)
        :ACTEE
        (V-624 / (HOUSINGAREA OBJECT)
               :SINGULARITY-Q SINGULAR
               :MULTIPLICITY-Q UNITARY)
        :ACTOR
        (V-625 / (BEHRENS NAMED-OBJECT MALE OBJECT)
               :NAME BEHRENS)
        :TENSE PAST))
 Discourse Semantics:  an individual is being referred to:
   (BEHRENS NAMED-OBJECT MALE OBJECT)
 Already mentioned locally; dynamically changing reference strategy.
 Discourse Semantics:  an individual is being referred to:
   (AEG GROUP NAMED-OBJECT OBJECT)
 Discourse Semantics:  an individual is being referred to:
   (HENNINGSDORF THREE-D-LOCATION NAMED-OBJECT OBJECT)

"He built a housing area for the workers of AEG in Henningsdorf."
```

Figure 6: Extract from discourse-sensitive sentence realization

can be taken further to apply to the overall publication scenario and to a possible electronic publication, offering the newly added functionality to the reader also.

Future work within the scenario described here will now include:

- expansion of the genres available,
- expansion of the discourse semantics to improve textuality,
- an investigation of the use of a genre representation for also structuring a (restricted) NL query interface,
- an investigation of the user of the generalized types of the domain knowledge upper structure for facilitating information retrieval,
- a coordination of the textual realization of views on domain knowledge with a graphical presentation mode—thus other presentation styles to be automatically generated should include graphics, tables, figures etc.; some early steps towards this in the editor's workbench context are presented in [Kamps, 1993].

Acknowledgements

We gratefully acknowledge the help of Lothar Rostek of the PAVE/PUBLISH group in the work preparatory for this paper.

References

[Bateman and Paris, 1989] John A. Bateman and Cécile L. Paris. Phrasing a text in terms the user can understand. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, Detroit, Michigan, 1989. IJCAI-89.

[Bateman and Paris, 1991] John A. Bateman and Cécile L. Paris. Constraining the development of lexicogrammatical resources during text generation: towards a computational instantiation of register theory. In Eija Ventola, editor, *Recent Systemic and Other Views on Language*. Mouton, Amsterdam, 1991.

[Bateman and Paris, in preparation] John A. Bateman and Cécile L. Paris. Register theory: the generic basis for contextualized natural language generation. Technical report, USC/ISI, in preparation.

[Bateman *et al.*, 1990] John A. Bateman, Robert T. Kasper, Johanna D. Moore, and Richard A. Whitney. A general organization of knowledge for natural language processing: the PENMAN upper model. Technical report, USC/Information Sciences Institute, Marina del Rey, California, 1990.

[Bateman *et al.*, 1991a] John A. Bateman, Christian M.I.M. Matthiessen, Keizo Nanri, and Licheng Zeng. The re-use of linguistic resources across languages in multilingual generation components. In *Proceedings of the 1991 International Joint Conference on Artificial Intelligence, Sydney, Australia*, volume 2, pages 966 - 971. Morgan Kaufmann Publishers, 1991.

[Bateman *et al.*, 1991b] John A. Bateman, Christian M.I.M. Matthiessen, Keizo Nanri, and Licheng Zeng. Multilingual text generation: an architecture based on functional typology. In *International Conference on Current Issues in Computational Linguistics*, Penang, Malaysia, 1991. Also available as technical report of the department of Linguistics, University of Sydney.

[Bateman *et al.*, 1992] John A. Bateman, Martin Emele, and Stefan Momma. The nondirectional representation of Systemic Functional Grammars and Semantics as Typed Feature Structures. In *Proceedings of COLING-92*, volume III, pages 916 - 920, 1992.

[Bateman *et al.*, 1993] John A. Bateman, Liesbeth Degand, and Elke Teich. Multilingual textuality: Some experiences from multilingual text generation. In *Proceedings of the Fourth European Workshop on Natural Language Generation, Pisa, Italy, 28-30 April 1993*, pages 5 - 17, 1993. Also available as technical report from GMD/Institut für Integrierte Publikations- und Informationssysteme, Darmstadt, Germany.

[Bateman *et al.*, 1994] John A. Bateman, Elke Teich, and Beate Firzlaff. A linguistically oriented methodology for domain model construction in knowledge-based systems. Technical report, GMD/IPSI, 1994.

[Bateman, 1988] John A. Bateman. From Systemic-Functional Grammar to Systemic-Functional Text Generation: escalating the exchange. In Eduard H. Hovy, David D. McDonald, Sheryl R. Young, and Douglas E. Appelt, editors, *Proceedings of the 1988 American Association for Artificial Intelligence Workshop on Text Planning and Realization*, pages 123–132, St. Paul, Minnesota, 1988. Also available as ISI Reprint Series report RR-89-220, USC/Information Sciences Institute, Marina del Rey, California, April 1990.

[Bateman, 1990] John A. Bateman. Upper modeling: organizing knowledge for natural language processing. In *5th. International Workshop on Natural Language Generation, 3-6 June 1990*, Pittsburgh, PA., 1990. Organized by Kathleen R. McKeown (Columbia University), Johanna D. Moore (University of Pittsburgh) and Sergei Nirenburg (Carnegie Mellon University).

[Degand, 1993] Liesbeth Degand. Towards a systemic functional grammar of dutch for multilingual text generation. Technical report, GMD/Institut für Integrierte Publikations- und Informationssysteme, Darmstadt, Germany, 1993. (Available in abbreviated form in the Proceedings of the Fourth European Workshop on Natural Language Generation, Pisa, Italy, 28-30 April 1993, pp143–147).

[Fischer and Rostek, 1993] Dietrich Fischer and Lothar Rostek. SFK: A Smalltalk Frame Kit. Technical report, GMD/Institut für Integrierte Publikations- und Informationssysteme, 1993.

[Haake *et al.*, 1993] Anja Haake, Christoph Hüser, and Klaus Reichenberger. The Individualized Electronic Newspaper: An example of an Active Publication. Technical report, Arbeitspapiere der GMD No. 799, Institut für Integrierte Publikations- und Informationssysteme, 1993.

[Haenelt, 1994] Karin Haenelt. Das Textanalysesystem KONTEXT. *Sprache und Datenverarbeitung*, 1994.

[Halliday and Hasan, 1989] Michael A.K. Halliday and Ruqaiya Hasan. *Language, Context and Text: a social semiotic perspective.* Oxford University Press, London, 1989.

[Hasan, 1978] Ruqaiya Hasan. Text in the Systemic-Functional Model. In Wolfgang Dressler, editor, *Current Trends in Text Linguistics*, pages 228–246. de Gruyter, Berlin, 1978.

[Henschel, 1993] Renate Henschel. Merging the English and the German Upper Model. Technical report, GMD/Institut für Integrierte Publikations- und Informationssysteme, Darmstadt, Germany, 1993.

[Kamps, 1993] Thomas Kamps. Automatic visualization for Hypermedia Publications. Technical report, GMD/Institut für Integrierte Publikations- und Informationssysteme, 1993.

[Kunze and Firzlaff, 1993] Jürgen Kunze and Beate Firzlaff. *Sememstrukturen und Feldstrukturen*, volume XXXVI of *Studia Grammatica*. Akademie Verlag, Berlin, 1993.

[Mann and Matthiessen, 1985] William C. Mann and Christian M.I.M. Matthiessen. Demonstration of the Nigel text generation computer program. In James D. Benson and William S. Greaves, editors, *Systemic Perspectives on Discourse, Volume 1*. Ablex, Norwood, New Jersey, 1985.

[Martin, 1992] James R. Martin. *English text: systems and structure*. Benjamins, Amsterdam, 1992.

[Matthiessen, 1988] Christian Matthiessen. Organizing Text: Rhetorical Schemas and GSP. Technical report, USC/ Information Sciences Institute, Marina del Rey, 1988.

[Matthiessen, 1992] Christian M.I.M. Matthiessen. Lexicogrammatical cartography: English systems. Technical report, University of Sydney, Linguistics Department, 1992. Ongoing expanding draft.

[McKeown, 1985] Kathleen R McKeown. *Text Generation: Using Discourse Strategies and Focus Constraints to Generate Natural Language Text*. Cambridge University Press, Cambridge, England, 1985.

[Möhr and Rostek, 1993] Wiebke Möhr and Lothar Rostek. TEDI: An Object-Oriented Terminology Editor. In *Proceedings of the Third International Congress on Terminology and Knowledge Engineering*, 1993.

[Rostek *et al.*, 1994] Lothar Rostek, Wiebke Möhr, and Dietrich H. Fischer. Weaving a web: The structure and creation of an object network representing an electronic reference network. In *Proceedings of Electronic Publishing (EP) '94*, 1994.

[Teich and Bateman, 1994] Elke Teich and John A. Bateman. Selective information presentation in an integrated publication system: an application of genre-driven text generation. Technical report, GMD/IPSI, 1994.

[Teich, 1992] Elke Teich. Komet: grammar documentation. Technical report, GMD/Institut für Integrierte Publikations- und Informationssysteme, Darmstadt, West Germany, 1992.

[Zeng, 1992] Licheng Zeng. Ml-penman: implementation notes. Technical report, GMD/IPSI and University of Sydney, 1992.