# A collocational based approach to salience-sensitive lexical selection

**Leo Wanner**
GMD/Institut für Integrierte Publikations- und Informationssysteme
Dolivostraße 15
D-6100 Darmstadt, W. Germany
*e-mail:* wanner@ipsi.darmstadt.gmd.dbp.de

**John A. Bateman**
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey
CA 90292-6695, U.S.A.
*e-mail:* bateman@isi.edu

## Abstract

In this paper we address the organization and use of the lexicon giving special consideration to how the salience of certain aspects of abstract semantic structure may be expressed. We propose an organization of the lexicon and its interaction with grammar and knowledge that makes extensive use of *lexical functions* from the Meaning-Text-Theory of Mel'čuk. We integrate this approach with the architecture of the PENMAN text generation system, showing some areas where that architecture is insufficient, and illustrating how the lexicon can provide functionally oriented guidance for the generation process.

## Introduction

In natural language generation, the lexicon can be viewed, generally, as containing information for the verbalization of meanings. This information ranges over both the static organization of vocabulary — including lexical knowledge, often handled separately as "lexical semantics" (see, e.g., [Pustejovsky, 1988; Nirenburg and Raskin, 1987]) or "structural semantics" — and the process of lexical choice. To make allowance for this latter dynamic aspect of the lexical organization we will henceforth use the term **lexis** common in systemic linguistics [Hasan, 1987; Matthiessen, 1988] instead of "lexicon". Lexis thus represents lexical information at various different levels of abstraction (strata) and mapping structures that provide for the conversion between those levels. In this paper we address the organization of lexis giving special consideration to the expresssion and choice of appropriate expressions as a function of the desired salience or prominence of semantic elements. The choice set of possible configurations of prominence we call the **perspectives** of the semantic structure.

These have been addressed rarely in approaches in generation so far: for example, [Jacobs, 1985] discusses the verbs *give* and *take* as two different expressions of the same event; [Nirenburg and Nirenburg, 1988] suggest an approach to open-class lexical item selection for realization of conceptual input; and [Iordanskaja *et al.*, 1988] propose an approach to linguistic paraphrasing by adapting the Meaning-Text-Theory (MTT) [Mel'čuk and Žholkovsky, 1970] and its paraphrasing rules. Here, we make more extensive use of the MTT in order to provide a richer organization of lexis and its interaction with grammar and knowledge than has been proposed previously. Moreover, we develop this approach in the context of a concrete generation environment, the PENMAN system [Mann and Matthiessen, 1985], showing some areas where the existing architecture is insufficient and how the richer organization of lexis we propose can help.

The following set of examples gives an impression of the variety of linguistic phenomena that we include under the term perspective.[1] All the sentences can be interpreted as verbalizations of a single abstract semantic structure with differing aspects of that structure being given emphasis in each case. For example, in (4), the reader is made salient as a participant of the proposition; in (5), the 'manner of achievement' of the 'indication' is made salient; in (7), a particular temporal aspect of the process, namely the beginning, is made prominentput; and in (8), the intended purpose of the agent is made salient as a 'making clear'. While the variation that can be seen between (1), (2), and (3) can already be treated in, for example, the current PENMAN system by exercising meaning options available in the grammar (i.e., (2) exhibits passivization and (3) nominalization of 'use'), the variation shown in the remaining examples cannot be functionally motivated as possible alternate grammatical realizations of the base semantic form.

1. *We use the adjective "electronic" to indicate that the dictionaries are deeply dedicated to computers.*
2. *The adjective "electronic" is used to indicate that the dictionaries are deeply dedicated to computers.*
3. *The use of the adjective "electronic" indicates that the dictionaries are deeply dedicated to computers.*
4. *The reader gets an indication that the dictionaries are*

[1] The basic sentence given under (1) is chosen from the introductory note of a text concerning the development of electronic dictionaries in Japan [EDR, 1988].

*deeply dedicated to computers by the adjective "electronic".*

5. *The indication that the dictionaries are deeply dedicated to computers is provided by the adjective "electronic".*
6. *By the use of the adjective "electronic" we illustrate the deep dedication of dictionaries to computers.*
7. *We create an indication that the dictionaries are deeply dedicated to computers by the adjective "electronic".*
8. *By the use of the adjective "electronic" we make clear that the dictionaries are deeply dedicated to computers.*
9. *The reader should take the use of the adjective "electronic" as an indication that the dictionaries are deeply dedicated to computers.*

Some of the phenomena running through these examples have been treated as **lexical cooccurrence** [Apresjan *et al.*, 1969] or **collocation** [Firth, 1957; Halliday, 1966; Hausmann, 1985]). Most extensively they are handled by I. Mel'čuk *et al.* in the scope of the Meaning-Text-Theory by means of **lexical functions** (LFs). Our approach to generating this range of variation takes its starting point, therefore, from the notion of lexical cooccurrence relations addressed by Mel'čuk. There is a large body of descriptive work based on the notion of LFs which has been carried out for different languages [Mel'čuk and Žholkovsky, 1984; Mel'čuk *et al.*, 1988; Žholkovsky, 1970; Reuther, 1978; Janus, 1971] and we will suggest how this body of knowledge can now provide significant input to work on text generation.

In the next section we give an introduction to LFs as the means by which lexical cooccurrence dependencies are expressed within the MTT; we then describe some problems with the existing organization of LFs as defined by Mel'čuk *et al.* and show how that organization can be developed further to be used for the modeling of various perspectives. In the section following, we then discuss the influence of lexis, when enriched by perspectives, on the generation process and, subsequently, we present a detailed example of the guidance this provides.

## The Nature and Organization of Lexical Functions

Lexical cooccurrence in the scope of MTT is provided in terms of lexical functions which Mel'čuk defines as follows [Mel'čuk and Polguère, 1987]:

> *A lexical function* f *is a dependency that associates with a lexeme L, called the argument of* f*, another lexeme (or a set of (quasi-)synonymous lexemes) L′ which expresses, with respect to L, a very abstract meaning (...) and plays a specific syntactic role. For instance, for a noun N denoting an action, the LF* Oper$_1$ *specifies a verb (...) which takes as its grammatical subject the name of the agent of the said action and as its direct object, the lexeme N itself.*

The values for any particular application of a LF to a lexeme are provided by an **Explanatory Combinatorial Dictionary** (ECD); dictionaries of this type for a number of languages have already been compiled by MTT researchers. Thus, for example, the ECD for English provides for the following applications of the LF **Oper$_1$**:

*Oper$_1$ (influence) = exert*
*Oper$_1$ (punishment) = administer*

These give lexical verbs appropriate for use when the argument is to appear as a direct object to form a combination where an agent (optionally) acts upon some patient; e.g.: *He exerted influence on P...*, *He administered a punishment...*, but not, * *he exerted a punishment...*, * *he administered an influence....*

Cooccurrence relations of this kind are pervasive in natural language and need to be captured in the representation of a language's lexical resources. Such cooccurrence relations can be rather arbitrary and so are unlikely to be supportable by, for example, distinctions maintained in the knowledge base. Their *meaning* is not, however, arbitrary. An important claim of MTT is that each LF represents a particular abstract meaning which remains invariant across its various applications. Thus, for example, further LFs include: **Func$_0$** with the meaning 'something takes place' (Func$_0$ (*accident*) = {*occur, happen*}, as in the sentence: *The accident occured two hours ago.*); **Result** standing for a state following the process addressed (Result (*study*) = *master*, as in the sentence: *John mastered his subject.*); and **Liqu** expressing an active process termination. This latter LF is often used in so called *composed* LFs where a number of LFs are combined in a predefined order (LiquFunc$_0$ (*Fire*) = {*extinguish, put out*}, as in the sentence: *The fire brigade could put out the fire quickly*).

MTT classifies LFs into two general types: **lexical parameters**, which provide syntagmatic relationships (e.g., Oper$_i$ with i = 1,2,...), and **lexical substitutes**, which provide paradigmatic relationships (e.g., **Gener** — a specific generalization relation) [Mel'čuk and Žholkovsky, 1970]. However, since LFs typically correspond to knowledge at varying levels of abstraction in addition to lexical information, these classes are still very heterogenous. Previous approaches that have made use of LFs in generation (e.g., [Kittredge and Mel'čuk, 1983; Iordanskaja *et al.*, 1988; Bourbeau *et al.*, 1989]) have been hindered by this. In order to exploit the notion of LFs in generation more extensively, we need further organizational principles that emphasize the meanings distinct LFs possess.

Work in progress at IPSI suggests that the large number of heterogeneous LFs used within MTT can be organized coherently in terms of the functions and semantic distinctions that they represent. Based on this, we have defined part of a general model of lexis with a taxonomic organization underlying it, within which the most general structures provide the representations of lexical semantics and the most delicate ones lexicalization. For the purposes of this paper, we will restrict attention to the organization of LFs that are particularly relevant for modeling situation perspectives as illustrated in our examples above. In Figure 1 we set out in network form general distinctive features that classify the meaning that the LFs we discuss here cover.[2] The network explicates LFs by classifying each of them according to a particular set of semantic features. The general function of the network is thus to relate particular LFs to the functional conditions for their application. This defines the meaning that any LF expresses and so provides a functionally organized key into the LF-oriented dictionaries being developed within MTT. The network also shows the hierarchical arrangement lying behind the meaning of LFs and so reflects the relation of perspectives to one another.

We will now briefly describe in semantic terms a representative set of the LFs covered by the network, showing how the network relates perspectival presentation decisions to choices of LFs.[3] Then, with the organizational network of perspectives in place and motivated, we show how it can be used to guide the generation process to produce the kinds of variation illustrated in (1)-(9).

### Situation introduction

One of the basic terms used in the scope of MTT is that of ***abstract situations***, circumscribed, roughly, as something that takes place. In the sense of Mel'čuk, abstract situations are defined by **key terms** and the participants of the situations; the key term is designated by the LF $S_0$, the participants as $S_i$ (i-th participant of the situation). We view this construct here as an abstract semantic partial specification of what is to be expressed linguistically. Thus, looking at the situation of *teaching*, the ECD for English offers us:

*$S_0$ (teaching) = teaching,*
*$S_1$ (teaching) = teacher,*
*$S_2$ (teaching) = pupil.*

The notion of situational key terms is closely related to term of *processes* in the tradition of *Systemic Functional Grammar* [Halliday, 1985; Steiner *et al.*, 1987].

When a situation is introduced, this may be done respecting a number of varying attributions of salience — e.g., the salience of particular participants of a situation or the situation itself.

The selection of particular combinations of process and participants for the realization of abstract situations according to differing attributions of salience is then provided in the scope of the ECD by LFs of various groups. For example: **Func**, which has the effect of attributing salience to the term labeling the situation; **Oper**, which has the effect of attributing salience to one of the participants; and **Labor** which has the effect of attributing salience to a combination of the participants. The selection of these broad groups is made in the network by the choices available under SITUATIONAL ORIENTATION, by the features 'situation oriented' (Func) and 'participant oriented' (Oper, Labor).

These are further differentiated according to which participants are affected; e.g.: $\text{Oper}_1$ makes the 'first' participant of the situation salient (i.e., the participant for which the LF $S_1$ provides a lexeme) and $\text{Oper}_2$ the 'second' (i.e., the participant for which the LF $S_2$ provides a lexeme): $\text{Oper}_2$ ***(influence) = be under.*** Similarly, $\text{Func}_0$ makes the key term of the situation itself salient, while $\text{Func}_1$ introduces the situation with particular respect to the first participant: $\text{Func}_0$ ***(problem) = exist***, $\text{Func}_1$ ***(problem) = come [from]***. $\text{Labor}_{12}$ makes the first and the second participant salient, the first more then the second, $\text{Labor}_{21}$, on the contrary, makes the second participant more salient, e.g.: $\text{Labor}_{12}$ ***(pressure) = bring to bear [on]***, $\text{Labor}_{21}$ ***(pressure) = get [from]***. These options are controlled by the further selections of participants to be accorded salience in the choice points SITUATION ORIENTATION and PARTICIPANT ORIENTATION.

Finally, the third option in the SITUATIONAL ORIENTATION system, 'process orientation' is responsible for the neutral LF $V_0$, which provides the most direct lexical verb for realizing the key term of a situation; e.g., ***$V_0$ (influence) = [to] influence.***

### Temporal dependency

LFs also address the global arrangement of a process on the temporal axes by the definition of its preceding and succeeding processes. These considerations are reached in the network by a feature selection of {global temporal oriented, ..., } from the alternatives of the TEMPORAL ORIENTATION choice point. These alternatives call for the application of the LFs **Prox** and **Perf**; examples of which from the ECD for English are: $\text{ProxFunc}_0$ ***(storm) = brew***, Perf ***(storm) = subside.*** In addition, the internal temporal aspects of a process, represented by its stages, are reflected by the corresponding triple of "phasal" LFs: **Incep** for the beginning, **Cont** for continuing, and **Fin** for the termination stage. These meanings are reached via the features under the 'stage oriented' option in the choice point PROCESS STAGES ORIENTATION in the network.

### Results and consequences

Situations can also be expressed so as to give salience to their *results*. The treatment of this requires a consideration of the *intended* result of the situation —

[2]The notation of Figure 1 follows that used within the NIGEL grammar of the PENMAN system for the specification of grammar possibilities. Names in capitals represent the names of choice points, and names in lower case features which may be selected: one from each choice point; also square brackets represent disjunction of features and braces conjunction. Such networks can be readily expressed in a number of distinct formalisms, e.g., FUG (cf. [Kasper, 1988]), LOOM (cf. [Kasper, 1989b]).

[3]In the full version of this network, the consequences of each possible selection of features for LF selection is specified; space precludes a detailed discussion at this point, although examples are given below.

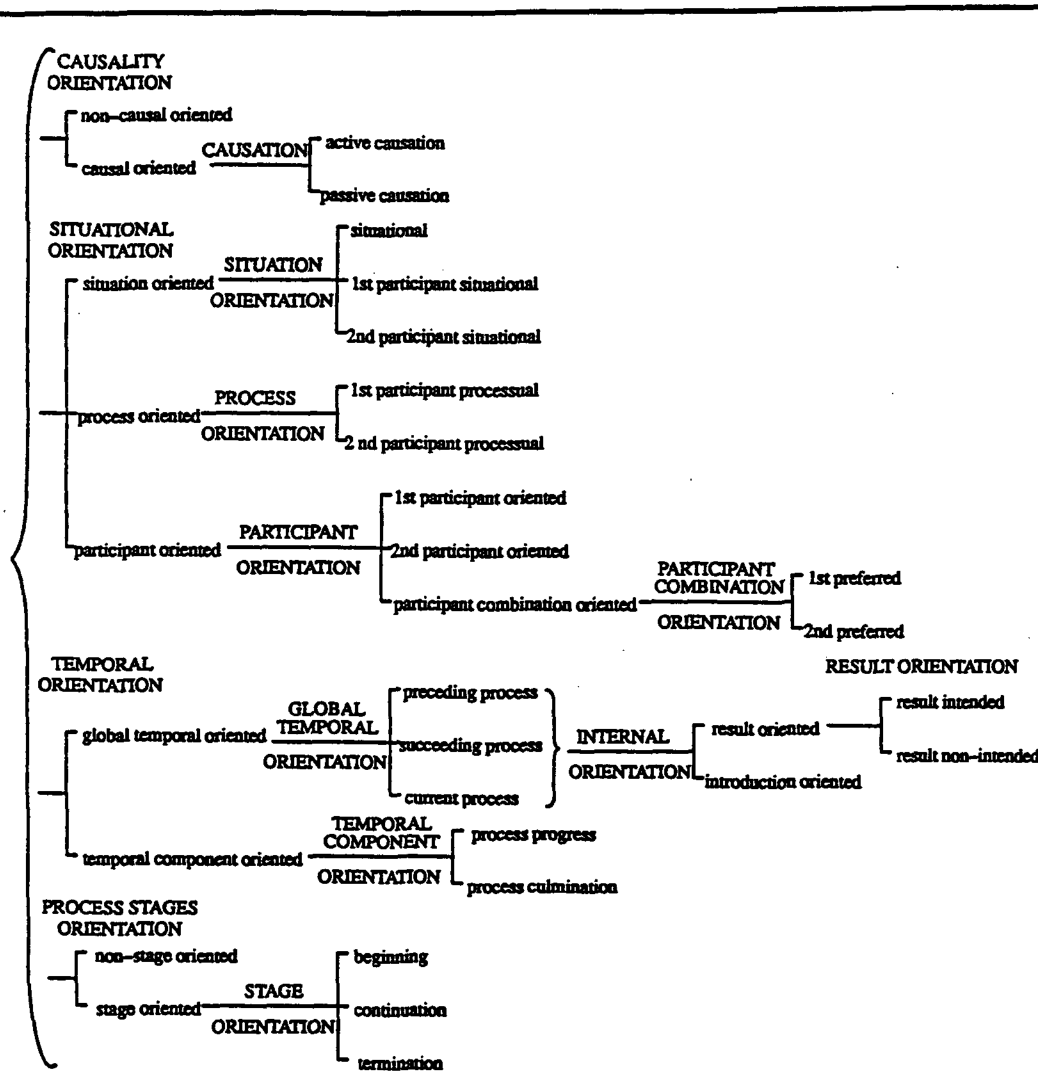


Figure 1: A hierarchical organization of the meaning underlying lexical functions in network form

the actual LF chosen depends on whether that result was achieved or not. These options are found under the choice point INTERNAL ORIENTATION and RESULT ORIENTATION. If the result of the process was the intended result (i.e., the 'purpose' of the carrying out the process), then the **Real**$_i$, **Labreal**$_{ij}$, and **Fact**$_i$ groups of LFs are applicable; in the opposite case, the **AntiReal**$_i$, **AntiLabreal**$_{ij}$, and **AntiFact**$_i$ groups apply. Each of these groups provide further the salience either of the key term of the situation itself or of the various participants of the situation as determined by the simultaneous selections of features made under (in this sense Real$_i$ and AntiReal$_i$ correspond to Oper$_i$, Labreal$_{ij}$ and AntiLabreal$_{ij}$ to Labor$_{ij}$, and Fact$_i$ and AntiFact$_i$ to Func$_i$). For example: Real$_2$ (*order*) = *obey*, AntiReal$_2$ (*order*) = *defy*.

## Causality

Situations can also be expressed so as to make the *causality* relationships that the situation enters into explicit or not; these options are considered by the choice point CAUSALITY ORIENTATION, which is responsible for application of either the LF **Caus** or **Perm**.

- The Caus function provides an active causation of the situation, as in the case of *problem*; e.g., CausFunc$_0$ (*problem*) = *pose*.
- Perm presupposes a 'permission', or allowance or acceptance, of the occurrence of the situation; e.g., PermFunc$_0$ (*problem*) = *tolerate*.

## Guiding the Generation Process by Lexis

The concrete generation system in which we are realizing organizational structures of lexis in order to support the emphasis of various salience aspects of semantic structures is the PENMAN system [Mann and Matthiessen, 1985]. The linguistic core of PENMAN is a large systemic-functional grammar of English, the NIGEL grammar [Matthiessen, 1983]. The semantic interface of NIGEL is defined by a set of *inquiries* which mediate the flow of information between the grammar and external sources of information. PENMAN provides structure for some of these external sources of information, including a conceptual hierarchy of relations and entities, called the *Upper Model* (UM) [Bateman *et al.*, 1990; Bateman, 1990]; the UM hierarchy classifies concepts according to their possibilities for realization in natural language and may be used as an interface between the organizational structures of Domain Knowledge (DK) and the grammar's inquiries. PENMAN accepts demands for text to be generated in the notation of the PENMAN *Sentence Plan Language* (SPL) [Kasper, 1989a]. SPL specifications are partial semantic representations of what is to be realized through the grammatical resources of NIGEL. More formally, SPL expressions are lists of terms describing the types of entities and the particular features of those entities that are to be expressed in English. The features of SPL terms are either semantic relations to be expressed, which are drawn from the upper model or from domain concepts subordinated to the upper model, or direct specifications of responses to NIGEL's inquiries.[4]

To generate any of the sentences (1)-(9) above using PENMAN, therefore, we must define appropriate SPL input. However, as mentioned in the introduction, these input specifications do not, at present, capture the generalization that these sentences share significant aspects of their meaning. To capture this, while still maintaining complete functional control of the generator, we introduce a more abstract input specification, from which particular SPL specifications are constructed depending on additional salience-oriented semantic distinctions. These semantic distinctions are specified in terms of the hierarchical organization of the meanings of LFs shown in the network of Figure 1. This organization provides a statement of semantic feature interdependencies that represents the perspectives available and the functional motivations for choosing one perspective over another. Each of the decision points in this network may place constraints on the mapping between the abstract input level and SPL. These decisions themselves need to be made by a text planning component — the network represents the *capability* of generating variation under control rather than the control process itself. In this sense, lexis as the stratum containing perspective information provides a controlling mechanism for the generation process entirely analogously to the grammatical network defined by NIGEL.

## Example of perspective-guided generation

We now illustrate the realization of some chosen perspectives in detail. Consider the clauses (1), (4), (6), and (7) given above. The SPL input specifications necessary to generate each of these clauses are set out in Figure 2.[5] As we can see, there is no connection between these since the generalization that they refer to the same situation is captured neither within the grammar, nor the upper model. Our new level of abstract input to the generation process, which corresponds more with Mel'čuk's conception of 'abstract situation' introduced above, provides this connection as follows. Abstract situations are represented in terms of a general type and a set of participants drawn from the lexemes defined with respect to the Domain Knowledge; for example, the abstract input for the situation underlying sentences (1), (4), (6), and (7) may be set out thus:[6]

$$\begin{bmatrix} S_0 & \textit{use} \\ \uparrow S_1 & \textit{we} \\ \uparrow S_2 & \textit{adjective 'electronic'} \\ \uparrow S_3 & \begin{bmatrix} S_0 & \textit{indication} \\ \uparrow S_2 & \textit{reader} \\ \uparrow S_3 & \begin{bmatrix} S_0 & \textit{(deep) dedication} \\ \uparrow S_1 & \textit{dictionaries} \\ \uparrow S_2 & \textit{computers} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

In order to generate sentences from this specification, we need to construct appropriate SPL expressions. This we achieve by following the semantic alternatives made in the LF network of Figure 1, applying the constraints that it specifies to compose a mapping between the abstract input and SPL.

Thus, for example, consider the context of use where a text planner has determined, in addition to expressing the situation shown in the abstract input, that that situation is to be presented textually as one in which the process is introduced neutrally, without respect for what preceded or succeeded, and with the process and the first participant (*we*: $S_1$) made relatively more salient. This corresponds to the set of LF network features {non-causal oriented, non-stage oriented, global temporal oriented, current process, introduction oriented, process-oriented, 1st participant processual}.

[4] For full details of the PENMAN system and its components, see the PENMAN documentation [The Penman Project, 1989].

[5] Note that in this figure, in order to save space, we share the varables `we`, `N1`, `A1`, `AS1` across the distinct SPL specifications; this would not normally be done.

[6] The notation $\uparrow S_i$ is used to indicate that the value given is *not* the value of the LF $S_i$ itself, it is rather the value of the role that the LF delivers; i.e., $S_1$ (*use*) is *user*.

```
(C1 / use
    :actor (we / person)
    :actee (N1 / adjective
                :name electronic)
    :purpose (A1 /indicate
                 :actor we
                 :subject-matter
                     (AS1 / dedicate
                          :domain (dictionaries / thing)
                          :range (computers / thing)
                          :manner (deep / sense-and-measure-quality))))
```

SPL specification for sentence (1)

```
(C2 / get :actor (r / reader) :means N1 :actee A1)
```

SPL specification for sentence (4)

```
(C3 / illustrate :actor we :actee AS1 :means N1)
```

SPL specification for sentence (6)

```
(C4 / create :actor we :actee A1 :means N1)
```

SPL specification for sentence (7)

Figure 2: SPL specifications for differing perspectives on a situation

This set of features governs the selection of the LF $V_0$, which is applied to the key-term of the situation, i.e., to $S_0$ of the input form: the lexeme associated with the DK concept *use*. The ECD for the language then supplies a candidate lexical item — in this case, the process *use*.

We integrate the information from the ECD by requiring lexical items to be linked to concepts which are subordinated to the PENMAN upper model. It is then possible to determine, by inheritance, the particular set of upper model/semantic role relations that are appropriate for a process of any type. The concept for *use* is classified as a *nondirected-action* in the upper model and so the role-set `:actor, :actee` is inherited. The fillers of these roles are then selected from the ordered set of participants specified in the abstract input under $S_1, S_2$. The process then recurses for the complex filler of $S_3$ — filling, in this case, the `:purpose` upper model relation — and the SPL given in Figure 2 for sentence (1) is constructed.[7]

If the text planning component had determined that a different set of presentational LF features were necessary, then a different LF would be selected for application to the key-term of the abstract input. Thus, with the selection of the features {non-causal oriented, non-stage oriented, global temporal oriented, current process, result oriented, result intended, participant oriented, 1st participant oriented}, which expresses the intended effect of the process *use* with salience on its first participant, the LF $\text{Real}_1$ is selected and, here, the ECD gives the process *illustrate*. This term is then, again, selected as the main term in the corresponding SPL specification and, as before, since it is also linked into the upper model, we know that the relevant role set is `:actor, :actee, :means`. The further mapping of situational roles $S_i$ to available UM-roles then provides the necessary fillers for the slots in the SPL. This gives the SPL for sentence (6).

In sentences (4) and (7), the interaction between the lexical network and the situation subordinated under $S_3$ in the abstract input is shown.[8] For the situation of 'indication', then, when the LF features: {non-causal oriented, non-stage oriented, global temporal oriented, current process, introduction oriented, participant oriented, 2nd participant oriented} are required, expressing that the situation is introduced with emphasis on its internal composition and participants and that the second of those participants is the more salient, then the LF $\text{Oper}_2$ is selected for application to the filler of $\uparrow S_3$ (i.e., *indication*). The ECD in this case supplies *get*. Note that here, the LF $\text{Oper}_2$ also has consequences for the latter mapping between situational roles and upper model roles; the key-term itself, $S_0$, is now associated with the role `:actee`. This provides the SPL specification for sentence (4).

[7] The association of the abstract situational roles $S_i$ and the roles drawn from the upper model in fact offers another significant source of presentation variability which may also be addressed in terms of LFs. We do not discuss this further within the confines of the present paper however.

[8] Work elsewhere (e.g., [Bateman and Paris, 1989]) has shown that *propositionally* embedded components of an input specification can be linguistically realized under certain textual conditions as unembedded, or as dominating, constituents. This is the case here, although space precludes a more thorough discussion.

Finally, with the selection of LF features {non-causal oriented, stage oriented, beginning, participant oriented, 1st participant oriented, global temporal oriented, current process, introduction oriented}, the LF $\text{IncepOper}_1$ is selected. When this is applied to *indication*, the ECD gives the process *create* and the SPL for sentence (7) is set up accordingly.

## Conclusion

We have shown how lexical cooccurrence relations can be used to express the salience of particular aspects of abstract semantic structures and how their underlying meaning and communicative function can be organized in order to influence the generation process. A specification of perspectival presentatation features as defined in the network of Figure 1 makes it possible to generate rather varied surface realizational forms. We can view this network as a candidate for the *textual organization* [Matthiessen, 1988] of lexis — which complements the more traditional 'propositional' organization found in lexical discrimination nets (e.g., [Goldman, 1975]) and thesauri. The textual/communicative functional meanings of LFs we propose, although arguably inherent in the MTM, have not formerly been extracted as an explicit principle of organization. We suggest that this kind of organization may substantially enhance the information collected by MTM researchers. Finally, although we have restricted ourselves in this paper to details that are particularly relevant for modeling situation perspectives, we are working towards a general model of lexis including, e.g., a semantically motivated classification of verbs, relations, etc. as proposed by, for example, [Matthiessen, 1988; Hasan, 1987] and pursued in a computational context by [Fawcett and Tucker, 1989]. For this we also use a set of further LFs represented on various levels of abstraction.

**Acknowledgments.** We would like to thank Elisabeth Maier, Hans Müller, Erich Steiner, and Elke Teich for fruitful discussions. We are also grateful to Igor Mel'čuk and Alain Polguère for comments on an earlier draft of this paper. John Bateman acknowledges the additional financial support of IPSI during the development of the ideas reported here.

## References

[Apresjan *et al.*, 1969] Yu.D. Apresjan, A.K. Žholkovsky, and I.A. Mel'čuk. On a possible method of describing restricted lexical cooccurrence. *Russkij Jazyk v Nacionalnoj Shcole*, 6:61–72, 1969.

[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 *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, 1990] John A. Bateman. Upper modeling: organizing knowledge for natural language processing. In *5th. Natural Language Generation Workshop, June 1990.*, Pittsburgh, PA., 1990. Also available as technical report of USC/Information Sciences Institute, Marina del Rey, CA 90292.

[Bourbeau *et al.*, 1989] L. Bourbeau, D. Carcagno, R. Kittredge, and A. Polguère. Text synthesis for marine weather forecast. Technical report, Odyssey Research Associates Inc., Montréal, 1989. Final Report, FOG project.

[EDR, 1988] Electronic dictionary project. Technical report, Japan Electronic Dictionary Research Institute, Ltd., 1988.

[Fawcett and Tucker, 1989] Robin P. Fawcett and Gordon Tucker. Prototype generators 1 and 2. Technical Report COMMUNAL Report Number 10, Computational Linguistics Unit, University of Wales College of Cardiff, 1989.

[Firth, 1957] J.R. Firth. Modes of meaning (1951). In J.R. Firth, editor, *Papers in linguistics 1934-1951*, pages 190–215. Oxford University Press, 1957.

[Goldman, 1975] Neil Goldman. Conceptual generation. In R. C. Schank, editor, *Conceptual Information Processing*. North-Holland Publishing Co, Amsterdam, 1975.

[Halliday, 1966] Michael A.K. Halliday. Lexis as a linguistic level. In C. E. Bazell *et al.*, editor, *In Memory of J.R. Firth*. Longman, London, 1966.

[Halliday, 1985] Michael A.K. Halliday. *Introduction to Functional Grammar*. Edward Arnold, London, 1985.

[Hasan, 1987] Ruqaiya Hasan. The grammarian's dream: lexis as most delicate grammar. In Michael A.K. Halliday and Robin P. Fawcett, editors, *New Developments in Systemic Linguistics. Volume 1.* Frances Pinter, London, 1987.

[Hausmann, 1985] F.J. Hausmann. Kollokationen im deutschen Wörterbuch: Ein Beitrag zur Theorie des lexikographischen Beispiels. In H. Bergenholtz and J. Mugdan, editors, *Lexikographie und Grammatik, Akten des Essener Kolloquiums zur Grammatik im Wörterbuch*, pages 118–129. 1985.

[Iordanskaja *et al.*, 1988] Lidija Iordanskaja, Richard Kittredge, and Polguère Alain. Lexical selection and paraphrase in a meaning-text generation model, July 1988. Presented at the Fourth International

Workshop on Natural Language Generation. Also appears in selected papers from the workshop: Paris, Swartout and Mann (eds.)(1990)(*op. cit.*).

[Jacobs, 1985] P.S. Jacobs. A knowledge-based approach to language production. Technical Report UCB/CSD 86/254, Univ. of California at Berkeley, 1985.

[Janus, 1971] E. Janus. Five polish dictionary entries... *Naučno-techničeskaja informacia*, 2(11):21–24, 1971.

[Kasper, 1988] Robert T. Kasper. An Experimental Parser for Systemic Grammars. In *Proceedings of the 12th International Conference on Computational Linguistics, August 1988*, Budapest, Hungary, 1988. Association for Computational Linguistics. Also available as Information Sciences Institute Technical Report No. ISI/RS-88-212, Marina del Rey, CA.

[Kasper, 1989a] Robert T. Kasper. A flexible interface for linking applications to PENMAN's sentence generator. In *Proceedings of the Darpa Workshop on Speech and Natural Language*, 1989. Availabe from USC/Information Sciences Institute, Marina del Rey, CA.

[Kasper, 1989b] Robert T. Kasper. Unification and classification: an experiment in information-based parsing. In *Proceedings of the International Workshop on Parsing Technologies*, pages 1–7, 1989. 28-31 August, 1989, Carnegie-Mellon University, Pittsburgh, Pennsylvania.

[Kittredge and Mel'čuk, 1983] R. Kittredge and I.A. Mel'čuk. Towards a computable model of meaning-text relations within a natural sublanguage. In *Proceedings of the IJCAI 1983*, 1983.

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

[Matthiessen, 1983] Christian M.I.M. Matthiessen. Systemic grammar in computation: the nigel case. In *Proceedings of the First Annual Conference of the European Chapter of the Association for Computational Linguistics*, 1983.

[Matthiessen, 1988] Christian M.I.M. Matthiessen. Lexico(grammatical) choice in text generation, July 1988. Presented at the Fourth International Workshop on Natural Language Generation. Also appears in selected papers from the workshop: Paris, Swartout and Mann (eds.)(1990)(*op. cit.*).

[Mel'čuk and Polguère, 1987] I.A. Mel'čuk and A. Polguère. A formal lexicon in the meaning-text theory (or how to do lexica with words). *Computational Linguistics*, 13(3-4):276–289, 1987.

[Mel'čuk and Žholkovsky, 1970] I.A. Mel'čuk and A.K. Žholkovsky. Towards a functioning meaning-text model of language. *Linguistics*, 57:10–47, 1970.

[Mel'čuk and Žholkovsky, 1984] I.A. Mel'čuk and A.K. Žholkovsky. *Explanatory Combinatorial Dictionary of Modern Russian.* Wiener Slawistischer Almanach, Vienna, 1984.

[Mel'čuk *et al.*, 1988] I.A. Mel'čuk, N. Arbatchewsky-Jumarie, L. Elnitsky, and A. Lessard. *Dictionnaire explicatif et combinatoire du francais contemporain.* Presses de l'Université de Montréal, Montréal, Canada, 1988.

[Nirenburg and Nirenburg, 1988] Sergei Nirenburg and Irene Nirenburg. A framework for lexical selection in natural language generation. In *Proceedings of COLING 88*, Budapest, Hungary, August 1988. Association for Computational Linguistics.

[Nirenburg and Raskin, 1987] Sergei Nirenburg and Victor Raskin. The subworld concept lexicon and the lexicon management system. *Computational Linguistics*, 13(3-4), 1987.

[Paris *et al.*, 1990] Cécile L. Paris, William R. Swartout, and William C. Mann, editors. *Natural Language Generation in Artificial Intelligence and Computational Linguistics.* Kluwer Academic Publishers, 1990.

[Pustejovsky, 1988] James Pustejovsky. Event semantic structure. Technical report, Brandeis University, Waltham, MA., 1988.

[Reuther, 1978] T. Reuther. Plädoyer für das Wörterbuch. *Linguistische Berichte*, 57:25–48, 1978.

[Steiner *et al.*, 1987] Erich H. Steiner, Ursula Eckert, Birgit Weck, and Jutta Winter. The development of the EUROTRA-D system of semantic relations. Technical Report Eurotra-D Working Papers, No. 2, Institut der angewandten Informationsforschung, Universität des Saarlandes, Saarbrücken, West Germany, 1987.

[The Penman Project, 1989] The Penman Project. The PENMAN documentation: User guide, primer, reference manual, and nigel manual. Technical report, USC/Information Sciences Institute, Marina del Rey, CA, 1989.

[Žholkovsky, 1970] A.K. Žholkovsky. Materials for a russian-somali dictionary. *Mashinnij perevod i prikladnaja lingvistika*, 13:35–63, 1970.