Ulrich Heid
Universität Stuttgart

Relating Lexicon and Corpus: Computational Support for Corpus–Based Lexicon Building in DELIS

Abstract

Lexicographers agree to favour a corpus based approach to lexicon building over one which would be based on intropspection. However, there is not much tool support for corpus based lexicography, especially when it comes to relating the observations made in corpus text with the classifications or the descriptions proposed in the dictionary. This problem is particularly relevant when it comes to relating the lexical semantic distinctions a lexicographer wants to make ("readings", "senses") with facts and data observed in the corpus.

DELIS aims at bridging this gap, at least in part, by designing, implementing and integrating tools for corpus exploration and lexicon building into a toolbox.

We give an overview of the DELIS approach and tools for corpus based lexicon building, which aims at supporting a description of lexical items at the levels of lexical semantics, syntax and morphosyntax, paying particular attention to the inter-relationship between these levels. The tools allow to create, update and modify lexical specifications and to check these against corpus material. We illustrate our work with examples from the domain of perception verbs.

1. From corpus analysis to lexical modeling

Many dictionaries[1], for both natural language processing (NLP) and human use, are based more on lexicographers' introspection than on "real text" as it occurs in newspapers, books, spoken discourse, etc. Only recent work in British lexicography (cf. work by Sinclair 1991, Atkins/Fillmore 1991), and a few dictionary projects for other languages (e.g. *Den Danske Ordbog)* are based on text corpora and accompanied by methodological work on corpus use in lexicography. In NLP, corpus based lexicon construction is also not very common; among the few examples are some of the recent ARPA projects in the USA, and, the German VERBMOBIL project.

1.1. An approach to corpus lexicography

The new, corpus–driven approach to dictionary making has not yet received much support in terms of dedicated computational tools except at COBUILD and within the HECTOR project. The main elements of the chain of corpus based lexicon building are *acquisition* (e.g. from corpus text), *formal modeling* and representation, as well as the *use* of lexical information in different applications (both for human and NLP purposes). Currently, there is a major lack of tools in two areas first: very few tools allow to explicitly relate lexicon and corpus; second: not much support is available for the spiral-wise process of dictionary enhancement, i.e. for going back and forth between the (possibly half-way finished) lexical model and the corpus texts; one way of doing this is to compare the fragment covered in the lexicon with the material found in corpora.

The starting point, in the DELIS work on lexicon building is corpus evidence, ideally tagged for part of speech and (morpho-)syntactic information. Such material serves as input and as resource of data; its intellectual (possibly tool-supported) analysis must be guided by some descriptive and theoretical expectation horizon. DELIS makes use of FILLMORE's *frame semantics* approach, for the lexical semantic description, and of a syntactic description in terms of grammatical functions (in terms of e.g. subject, object, indirect object, etc.) and of types of phrasal constructs (noun phrases, adjective phrases, verb phrases, etc.). Examples are given below, in section 3.

The *frame semantics* approach is one example of a framework for lexical semantic description by which a lexicographer might be led in his exploration of corpus text and in his model building; we do not want to justify the choice of frame semantics, here. What is more important than the choice of the framework, is the fact that we try to explicitly relate the lexical semantic description (e.g. of verbs) with a functional syntactic and a phrase structural one: the goal is to relate lexical semantic classifications with phenomena observable in corpora, e.g. in terms of the syntactic behaviour of the items described.

The construction of lexical specifications is however not seen just as a process of filling a predefined model with lexical instances. On the contrary, starting form the corpus material and from a set of initial descriptive hypotheses, specification building is conceived as spiral-wise process of modeling, checking of the current model against corpus evidence, modification and refinement of the current model, further checking, etc. This is felt to be a more realistic approach to model building, although seemingly more experimental at first sight.

Model building and checking with small but relevant quantities of lexical items is a first step; then only "mass coding", i.e. "population" of the dictionary with larger amounts of material, is possible. In the model building process, new descriptions must be added or existing ones must be merged, split, refined (e.g. by subclassification), etc.

To support these procedures, DELIS builds tools for the management of lexical specifications: not only for feeding lexical material into the existing model (we call this *"population"),* but also for modifying existing (and maybe populated) parts of the model (we call this *"reclassification").* The task of verifying coverage and adequacy of the model by testing it against corpus data is supported by another DELIS tool which produces corpus queries out of the functional syntactic and phrase type descriptions contained in the lexicon.

1.2 The lexical model: representation and dictionary architecture

1.2.1 The representation language

Lexical information in DELIS is represented in typed feature structures. We use the TFS language (see for details Emele 1993 and Zajac 1992, Emele/Zajac 1990) which combines concepts from unification grammar, from constraint logic programming and from knowledge representation.

DELIS has two types of knowledge sources and related tools: corpora and lexical specifications.

Lexical specifications are represented as lattices of typed feature structures: lexical classes from different levels of description are represented as types, whereas individual lexical items are instances of the types. This allows to organize dictionaries as class hierarchies. In a similar way, the analyzed corpus evidence can be stored in hierarchies modeled along the same lines as the dictionary.

This is possible because the lexicon and the analysed corpus sentences are interrelated, conceptually and technically. The conceptual link is provided by the use of a common descriptive framework: we have devised a *Corpus Evidence Encoding Schema* (CEES, see below, section 3.2) which accounts for properties of the sentences analyzed (e.g. polarity, tense, embedding, etc.), but also for properties of the "keyword", i.e. the lexical item illustrated by the sentence, including the main aspects of its lexical semantic, syntactic and morphosyntactic description. The descriptive vocabulary used to describe the keyword in CEES is shared with that of the lexical specifications. Technically, cross–reference links between the two resource types can be installed.

The architecture of the DELIS dictionaries in schematically represented in Figure 1.

1.2.2 Lexical specifications: the architecture of Delis dictionaries

DELIS lexical specifications are characterized by the following properties.

Modularity: The DELIS lexical specifications follow HPSG's sign–based approach (cf. Pollard/Sag 1987, Pollard/Sag 1993): for each targeted level of description, separate (classificatory) models and well–formedness conditions exist. An individual lexical item is an object with interrelated partial descriptions from all levels. The levels considered are lexical semantic roles and role configurations, grammatical functions, and phrase types. In addition, particular lexical properties of the described items are recorded, such as collocates, typical adverbs, etc.

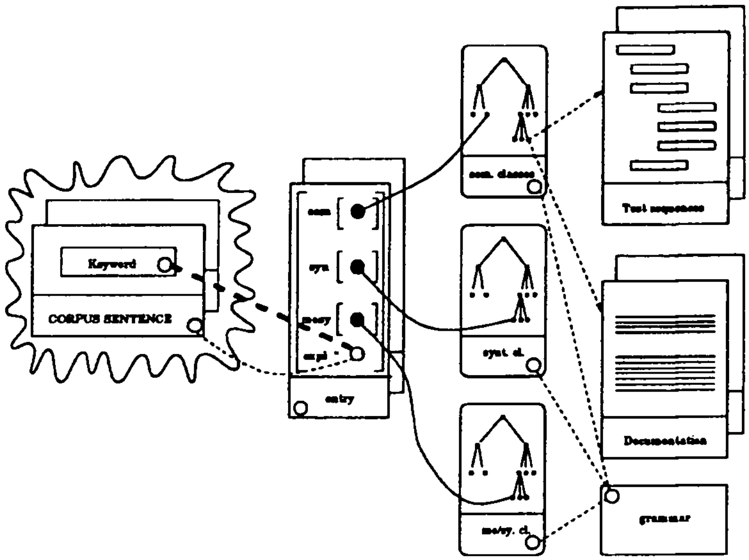

Figure 1: Architecture of the lexical specifications: corpus, dictionary entry lists, class–wise specifications for each level of description, documentation material

A classificatory approach: In order to express generalizable information only once, each level–specific model is organized as a monotonic specialization hierarchy of classes: examples are a hierarchy of subcategorization classes (in terms of grammatical functions (cf. work in the ACQUILEX project by Sanfilippo 1993)), or hierarchies of partial meaning descriptions based on predicate–argument structures and on role configurations of frame semantics.

Access–neutrality: DELIS lexical specifications can be queried with arbitrarily underspecified feature structures. Other than with lexical databases, with printed dictionaries and with most computational dictionaries, none of the descriptive levels is taken a priori as the only relevant criterion for the macrostructural organization of the DELIS lexical specifications. If queried by the orthographic form of a lemma, a semasiological (and e.g. alphabetically ordered) dictionary entry can be extracted from the DELIS lexical specifications; if queried according to the meaning classification, we get an onomasiological fragment (e.g. all elements of a lexical semantic class); if queried by subcategorization classes, we get a construction dictionary.

Avoiding an a priori commitment to one access structure is a precondition for the practical side of multifunctionality: this property is used in the export tool: different "views" on the data serve as input to different types of fragments extracted (see below, section 2.).

Documentation: Reusability of linguistic resources and, more concretely, sharability of lexical specifications between different applications, depends crucially on reinterpretability and thus on documentation. DELIS aims at supporting reinterpretation of its lexical specifications: both, classes and instances can be documented: lexical instances have links to corpus sentences. For classes, the lexicographer constructing the model can furnish a textual documentation of the criteria underlying his classification, and thus make his "guidelines" or "coding manual" available online.

Moreover, the system computes differences in terms of attributes and values between classes (e.g. the "differentia" between a class and its superclass(es), between sibling classes, etc.). The results of this computation can be presented, in the user interface for interactive acquisition, as decision trees for interactive classification of new objects; similar techniques have been used in the *Core Language Engine* and in the METAL MT system. Additional types of documentation include corpus sentences annotated according to CEES, cross-reference links between instances and acceptability tests derived from the above decision trees.

2. Tool support

The following tool components for corpus based lexicon construction, have been designed and are currently being implemented in DELIS:

- Tools for corpus exploration:
  - support for the interactive annotation and classification of corpus sentences according to the CEES specifications (frame assignment tool);
  - production of corpus queries ("search conditions") from the lexical specifications, with a possibility of processing the automatically generated search conditions with a proprietary corpus query system which is integrated into the DELIS toolbox.

- Tools for the management of lexical specifications:
  - "population": interactive data entry (adding lexical material to the model);
  - "reclassification": interactive modification of partly populated models: introduce new subclasses; split, merge or delete classes; move classes elsewhere in the hierarchy;
  - "exportation": select and extract classes and subsets of their attributes from the model, represent the selection result as feature

structures, then linearize and reorder the feature structures and finally retag the attribute and value names according to the conventions of the target application. This tool can produce SGML output which, since linearized and reordered, can be re–represented in the format of ordinary human–use dictionaries; experiments with exportation towards GESTORLEX, an SGML–based tool for the design and control of the article syntax of print dictionaries, are under way.

These tool functions are embodied in the DELIS toolbox. Its first prototype (demonstrated at the EURALEX–94 congress) integrates a type editor allowing to add and modify definitions of lexical classes and instances in the TFS language; a graph editor for navigating through and editing the (multiple) hierarchies of feature structures; a class and instance browser for TFS knowledge bases (similar in its functionality and look to the class browsers known from e.g. SMALLTALK or other object–oriented programming languages).

3. Some examples

3.1. An outline description of perception verbs

The linguistic work in DELIS has so far dealt with the description of perception verbs and nouns in English, French, Italian, Danish and Dutch. DELIS dictionaries contain information from the descriptive levels of lexical semantics, syntax and morphosyntax. This section exemplifies the lexical semantic and syntactic description, the next one the morphosyntactic description.

The descriptive vocabulary of the lexical semantic zone of DELIS dictionaries is inspired by FILLMORE's *frame semantics* approach, making use of ROLES and subtypes of roles (according, among others to the semantic properties of typical role realizations).

If we allow ourselves some simplification, we can distinguish on the one hand the five modalities of perception (visual (vis), auditive (aud), olfactory (olf), gustatory (gus), and tactile (tac)), and, on the other hand a small hierarchy of types of events. Following Fillmore (1993a) and Fillmore (1993b), we distinguish between *perception* and *attention* events, as well as events of the evocation vs. evaluation of a perceivable quality. In terms of roles, *perception* and *attention* events have an EXPERIENCER and a PERCEPT, and events of the evocation or evaluation of a quality have an additional JUDGEMENT role. The PERCEPT and JUDGEMENT roles have further subtypes; we distinguish, among others, between PCT–PHENOMENON *(I hear the **noise**)* and PCT–SOURCE *(I hear the **car**)*; JUDGEMENT–subtypes are indications of a perceived QUALITY *(This tastes **bitter**)* Vs. an EVALUATION *(This smells **awful**)* vs. a comparison (role: SIMILE, *this smells **like fish**)*.

The difference between perception and attention events is due to "active" or "voluntary" involvement of the experiencer in the attention–case, whereas the experiencer is not acting "voluntarily" in the perception case. The difference is best exemplified with pairs of perception vs. attention verbs such as EN *hear* (perc) ↔ *listen to* (att), *see* (perc) ↔ *look at/watch* (att), or FR *entendre* (perc) ↔ *écouter* (att), *voir* (perc) ↔ *regarder* (att). The two types are not lexically distinct in all cases: EN *feel*, *smell*, *taste* have both perception and attention meanings (see below, in Figure 2 and (Figure 3).

Type of event	vis	aud	olf	gus	tac
perception	see	hear	smell	taste	feel
attention	look watch	listen	smell	taste	feel
evoke-quality	–	–	smell	taste	feel
eval-quality	look	sound	smell	taste	feel

Figure 2: Overview of the perception field for English

The illustration in Figure 2 shows the distribution of the most common English verbs of the perception field over the lexical semantic classes stated above, and over the modalities. Simple examples of English gustatory readings are summarized in Figure 3..

Type	example
perc.	*I tasted garlic in the soup.*
att.	*Please taste this soup and tell me whether it is ok.*
evoc.	*This juice tastes bitter.*
eval	*This juice tastes good.*

Figure 3: Examples of the use of EN *taste* in the four classes

The syntactic description, at the levels of grammatical functions and types of phrasal constructs, is inspired by Lexical Functional Grammar (LFG) and Head Driven Phrase Structure Grammar (HPSG); instead of a commitment to one of these formalisms, a welldocumented descriptive vocabulary is used which allows for easy compilation into specific NLP–oriented formalisms

Type	Roles	Gramm. func.	phr. types
perc	[EXP-P PERCT]	<SUBJ OBJ>	np np
att	[EXP-A PERCT]	<SUBJ OBJ>	np np
	[EXP-A PERCT] (MNR)	<SUBJ OBJ> ADJ	np np adv
evoc	[PERCT QUAL]	<SUBJ XComp>	np ap
eval	[PERCT EVAL]	<SUBJ XComp>	np ap
	[PERCT SIMILE]	<SUBJ Pcomp>	np pp (*like*)

Figure 4: Selected readings of the English verb *taste:* roles, grammatical functions and phrase types

The table in Figure 4 contains the readings of the English verb *taste,* according to the above classifications: for each reading (perception, attention, evocation and evaluation), the grammatical functions and phrase types are indicated. Not all combinations are distinctive: the content of the table in Figure 4 is not sufficient to generate specific enough search conditions for all subtypes. However, a rough set of search conditions can be derived; for example, we could extract a set of sentences illustrating the evaluation senses of *taste* automatically.

To improve the granularity of the lexical description, additional information will be added to the dictionaries, e.g. about preferred adjuncts, about noun and adjectives classes acting as role realizations, about collocations, etc. [2]

3.2. The Corpus Evidence Encoding Schema

Above, we have given examples of the descriptive framework for lexical semantics and syntax. Tables like that in Figure 4 are created on the basis of the analyzed corpus material.

Individual sentences of the corpus have been analysed manually and grouped according to broad lexical semantics classes. Along with this work, a number of morphosyntactic properties of the keyword (i.e. in this case the perception verb) and the sentence in which it appeared, are noted. Among these are active/passive voice, mood, embedding (e.g. under modals: *"**can** you hear me?"),* polarity, etc. This way it is easy, for example, to find out that German *riechen,* although used with an OBJ(ect), very rarely at all occurs in the passive voice, in text material. This recording, along with checks across larger amounts of corpus material than can be manually analyzed, allows to come up with descriptions of the distribution of certain properties over texts; it contributes thus to the "real–text" based nature of the descriptions.

To be able to collect this information relevant for the the DELIS *Corpus Evidence Encoding Schemata* (CEES) has been defined.

CEES provides a specification for the description of keywords appearing in corpus evidence, but at the same time, this specification also forms the most general part of the lexical description used in the DELIS lexicon fragments. CEES is sufficiently detailed for the description of corpus evidence, and it is assumed that the classification appearing in CEES will be further refined in the lexical specification. The fact that such shared use of a lexical specification in corpus description and lexicon definition is at all possible, is due to a number of particular choices made in the design of CEES. These can be summarized as follows:

- CEES is a specification, not just an inventory of "labels" or "tags"; this implies that it is possible with CEES to define well-formed linguistic objects; this would not be feasible with a "traditional" tagset, as used in corpus work usually.

- CEES is hierarchically organized, the same way as the lexicon: CEES types define "shapes", or "models" of well-formed linguistic objects. CEES has been modeled in TFS (see Heid/Krüger 1993).

- CEES is close to a descriptive tagset, such as, for example, the EAGLES proposal for a morphosyntactic specification for lexicons (cf. work by CALZOLARI and MONACHINI). A descriptive tagset, other than a "physical tagset" (the LOB or BNC tagsets are examples of "physical tagsets"), abstracts away from specific "proprietary" conventions of encoding of lexical material (e.g. word forms) in a given corpus. CEES goes beyond this (towards what could be called an "abstract" tagset) insofar as it generalizes over a number of morphosyntactic phenomena and thus captures some generalizations across languages. This is possible because the linguistic objects described with CEES are sentences (and the "keywords" appearing therein) and not just individual word forms.

CEES instances include the actual text form of the sentence analyzed (in a keyword in context format), with explicit mention of the relevant word forms of the keyword and of the elements described by the semantic roles (these appear under the EXPRESSION attribute; redundancy is wanted, for reasons of easy reformatting of CEES–instances, in the preparation of comparative surveys. CEES not only exists as a TFS–encoded specification, but also as a definition of fields of a database. Thus, for instance, one can restructure CEES instances to retrieve all sentences with the keyword *taste* and list them in alphabetical order of the strings which are EXPRESSIONS of the EXPERIENCER role).

Details of the use of CEES in descriptive work can be found in Braasch (1994). CEES is described in detail in Heid/Krüger (1993). In Figure 5, a full

CEES–instance, for the sentence *The soup may have been tasted by Mary* is shown. To avoid redundancy, we have not expanded some of the attributes which usually would be expanded (e.g. ADMIN, COMMENT); in a few cases we have left the topmost type, and have chosen not to expand it.

```
[KEYWORD taste
 TEXT [IDENTIFIER ref-number-of-text
       EXPRESSION [LEFT-CTX The soup may have been
                   KW-TEXTFORM tasted
                   RIGHT-CTX by Mary]text-expr
      ]text
 GOV-CTX [GOV-PRED may
          EXPRESSION may have been]g-context
 MORPH [CATEGORY v
        PERS 3
        NUMB sg
        TENSE pres-perf
        MOOD indicative
        VOICE passive
        PASSAUX be]vfin-morph-pass
 SENT-PROPS [ASPECT asp
             INVERSION -
             POLARITY affirm
             CLAUSE-TYPE decl]s-props
 FRAME < [ROLE percept
          FUNC subject
          PHR-T np
          EXPR the soup]fr-comp ,
         [ROLE experiencer
          FUNC obl-by
          PHR-T pp (by)
          EXPR by Mary]fr-comp >
 ADMIN admin-info
 COMMENT string
]sentence
```

Figure 5: An example of an instantiated CEES structure for an English passive sentence

4. State of the art – implementation – future developments

4.1. DELIS vs. other projects

The DELIS approach to corpus lexicography is inspired by work in the HECTOR project (cf. Atkins 1992 etc.) and by Atkins/Fillmore (1991). The HECTOR tools, however, do not computationally relate lexicon and corpus (an integrated lexical knowledge base has however been built, once the descriptive work was carried out): it makes tool functions for dictionary editing and corpus exploration available, and results can be manually transferred; the control of such processes is however exclusively the lexicographer's task. HECTOR templates are similar to lexical classes of DELIS, but in DELIS the classes are organized in a class hierarchy, and the typed feature system is used for consistency control, both in population and reclassification in HECTOR, only class membership of instances is automatically verified at editing time.

Much work on corpus query has concentrated on wordform based keyword in context production; this is contained in HECTOR, or, for example, in the German COSMAS System (cf. Justen/al–Wadi 1992); both include a morphology component allowing to expand a lemma to all its relevant forms, before corpus search. However, it is rarely possible to query the corpus text with search expressions containing both word forms and annotations. The DELIS search condition production tool and the XKwic tool used to retrieve corpus material according to the search conditions produced (cf. Christ 1993) are innovative in this respect.

Tool support for interactive data entry is standard in lexical and terminological databases. Such functions exist also for typed feature structure based systems. Examples are the feature structure viewers and editors by Kiefer/Fettig (1993) and Groenendijk (1993), as well as, for lexicographic purposes, the LEMMING system, which was developed in the LEXIC project (cf. Fokker 1992). These systems mostly support the syntax of typed feature structures, but only LEMMING also checks to some extent the well-formedness of incoming descriptions with respect to the existing model. None of these tools includes any facilities for model evolution and full consistency checking of the model at any given point in time. LEMMING, HECTOR and system QUIRK (a lexicographer's workstation produced in part in MULTILEX; cf. Holmes–Higgin et al. 1993) do not support model evolution at all; the CODE4/COGNITERM system, a tool for knowledge engineering in terminology and knowledge representation (cf. Skuce 1993) supports certain non–destructive changes. This tool however, does not straightforwardly handle complex feature structures, as they are needed for lexical representation. Work in other lexicon projects, such as ACQUILEX and MULTILEX has not dealt with the process of lexical modeling so far. In corpus projects, the use of exploration tools for actual dictionary building has not been in the focus.

4.2. Implementation – next steps

The DELIS toolbox integrates the relevant tools under a common graphical user interface (GUI); it is implemented in C/C++ on UNIX workstations using the X Windows System and OSF/Motif. The toolbox provides a client/server architecture. The server module is the kernel of the lexical knowledge base (representation in TFS) and provides access at the level of classes and instances. The TFS language is implemented in Common Lisp and available for different platforms. An existing search tool XKWIC (cf. Christ 1993) has been integrated under the common GUI.

A first toolbox prototype is available as of May 1994. A version which will integrate early feedback from professional users will be demonstrated at the EURALEX–94 congress. A second version is foreseen for the end of 1994.

Notes

1 The author would like to thank the project members for their discussion of previous versions of this paper. Special thanks go to Katja KRÜGER (IMS) who has contributed to the part on CEES (cf. section 3.2). All remaining errors, imprecisions and misconceptions are of the author's responsibility.

2 The abbreviations in the illustration in Figure 4 should be read as follows: EXPER–A and EXPER–P denote the "active" and "passive" experiencer; PERCT, QUAL and EVAL are abbreviations of the roles PERCEPT, QUALITY and EVALUATION. With (MNR) we indicate that manner adjuncts are frequent with the **att** reading of *taste*. For grammatical functions, we use LFG's terminology: SUBJ(ect), OBJ(ect), POBJ for prepositional objects, XCOMP and PCOMP for a predicative complement. ADJ denotes an adjunct.

References

Atkins, B.T.S. 1992. "Tools for computer-aided corpus–lexicography: the Hector Project", in F. Kiefer et al. (Eds.): *Papers in computational lexicography–* COMPLEX '92, Budapest Linguistic Institute, Hungarian Academy, pp. 1–59.

Atkins, B.T.S., C. Fillmore 1991. "Starting where the dictionaries stop – The challenge of corpus lexicography", in B.T.S. Atkins, A. Zampolli (Eds.): *Computational Approaches to the Lexicon,* Oxford, to appear.

Braasch, A. 1994. "There's no accounting for taste – except in dictionaries ...". Internal paper: Copenhagen: CST, 1993, this volume.

Christ, O. 1993. "The XKwic User Manual", internal report, Stuttgart.

Emele, M.C. 1993. "TFS – The Typed Feature Structure Representation Formalism". In H. Uszkoreit, ed., *Proceedings of the EAGLES workshop on implemented formalisms,* DFKI report 1993.

Emele, M., U. Heid 1993. "Formal specification of a typed feature logic based lexical representation language" – Deliverable D–V–2 of DELIS (LRE 61.034).

Emele, M.C. and R. Zajac 1990. "Typed unification grammars". In *Proceedings of the 13th International Conference in Computational Linguistics (CoLing90),* ed. H. Karlgren, Helsinki, August.

Fillmore, Charles 1993a. "A Cognitive–Frames Approach to the Vocabulary of Sensation and Perception in English", Draft, ms., 8pp, University of California in Berkeley, 1993.

Fillmore, Charles 1993b. "Frame semantics and perception verbs", unpublished contribution to Dagstuhl Seminar 9313 *Universals in the Lexicon,* 1993.

Fokker, J. 1992. "LEMMING User Manual", INF/DOC–92–04, Rijksuniversiteit Utrecht, Vakgroep Informatica.

Groenendijk, M. 1993. "Xmfed User Guide", P–E International.

Heid, U., M. Emele, L. Humphreys 1993. "Towards a linguistic architecture and tool system architecture for DELIS" – Descriptive choices, dictionary organization and tool system outline, Deliverable D–I–2 of DELIS (LRE 61.034).
Heid, U. and K. Krüger 1993. "On the DELIS corpus evidence encoding schema (CEES)", DELIS internal report of November 3rd, 1993.
Holmes–Higgin, P. 1993. "User Guide: System Quirk", Workpackage 5, Task 5 of Esprit project 5304 Multilex, University of Surrey.
Justen, K., D. al–Wadi 1992. "COSMAS – COrpus Storage, Maintenance and Access System", internes Papier des Instituts für deutsche Sprache, Mannheim, 1992.
Kiefer, B. and T. Fettig 1993. "FEGRAMED – An Interactive Graphics Editor for Feature Structures", DFKI report, Saarbrücken, April 1993.
Olivier J. 1993. "DELIS specifications", deliverable D–V–1 of DELIS (LRE 61.034).
Pollard, Carl and Ivan Sag 1987. "*An Information–based Syntax and Semantics*, vol. I, CSLI lecture notes no. 13, 1987.
Carl Pollard, Carl and Ivan Sag 1993. In press. *Head–Driven Phrase Structure Grammar.* Chicago: University of Chicago Press and Stanford: CSLI Publications, in press.
Sanfilippo Antonio 1993. "LKB Encoding of Lexical Knowledge", in: Briscoe, T., Copestake, A. and de Paiva, V. (eds.): *Inheritance, Defaults and the Lexicon*, Cambridge, 1993.
Sinclair, J. 1991. "Corpus, Concordance, Collocation." Oxford University Press, Oxford.
Skuce, D. 1993. "A Wide Spectrum Knowledge Management System", to appear in *Knowledge Acquisition.*
Sperberg–McQueen, C.M. & Lou Burnard (Eds.): ACH/ACL/ALLC. *Guidelines for Electronic Text Encoding and Interchange.* TEI P2. Part I–VIII. Chicago, Oxford. In preparation; partly released 1992–93.
Voutilainen, Atro, Juha Heikkila, Arto Anttila 1992. "Constraint Grammar of English: A Performance–Oriented Evaluation", Publication No. 21, University of Helsinki, Department of General Linguistics, 1992.
Wijers, M., 1993. "Specification of the DELIS TFS–to–SGML export tool for dictionary production", appendix to deliverable D–V–1 of DELIS (LRE 61.034).
Zajac, R., 1992. "Inheritance and Constraint–Based Grammar Formalisms." *Computational Linguistics*, 18(2), pp. 159–180.