

Dutch Sublanguage Semantic Tagging combined with Mark-Up Technology

P. Spyns ^[1], NT. Nhàn ^[2], E. Baert ^[1], N. Sager ^[2], G. De Moor ^[1]

(1) Division of Medical Informatics, University Hospital Gent
De Pintelaan 185 (5K3), B-9000 Gent, Belgium

(2) Courant Institute of Mathematical Sciences, New York University
251 Mercer Street, NY 10012, New York, USA

{Peter.Spyns,Erik.Baert,Georges.DeMoor}@rug.ac.be, {nhan,sager}@cs.nyu.edu

Abstract

In this paper, we want to show how the morphological component of an existing NLP-system for Dutch (Dutch Medical Language Processor - DMLP) has been extended in order to produce output that is compatible with the language independent modules of the LSP-MLP system (Linguistic String Project - Medical Language Processor) of the New York University. The former can take advantage of the language independent developments of the latter, while focusing on idiosyncrasies for Dutch. This general strategy will be illustrated by a practical application, namely the highlighting of relevant information in a patient discharge summary (PDS) by means of modern HyperText Mark-Up Language (HTML) technology. Such an application can be of use for medical administrative purposes in a hospital environment.

1 Introduction

Medical patient reports consist mainly of free text. While numerical data can be stored and processed (relatively) easily, free text is rather difficult to process by a computer, although in many cases it contains the most relevant information.

The use of natural language does not facilitate the automation of these tasks and hinders access to the wealth of medical information. However, natural language still is the most frequently used and easiest way to transmit complex messages (Scherrer et al., 1989). Hence, some authors consider the study and application of Natural Language Processing (NLP) in Medicine (Scherrer et al., 1989), (McCray et al., 1995), (Chute, 1997) as one of the most challenging issues in the field of medical information retrieval (Baud et al., 1992a), (Friedman and Johnson, 1992).

Up till now, not many NLP-driven systems have actually been implemented (Spyns, 1996b). A concise overview of NLP-based information retrieval techniques for clinical narrative can be found in (Hersh, 1996, chapter 11, pp. 211-323).

A possible environment for (medical) information retrieval is the Medical Registration Department of a hospital, and more in particular the medical encoding service. Clinical data in free text format are replaced by a set of numerical codes that summarise the content of the entire document. In general, the patient discharge summary (PDS), being a synthesis of the patient stay, is used for the encoding and abstracting task instead of the entire medical record (Duisterhout, 1996). An important aspect of medical encoding consists of a thorough review of the PDS in order to discover the relevant words (diagnoses, surgical deeds, interventional equipment etc.) (Bowman, 1996, p.216). The aim of the NLP-based HTML application presented below is to speed up the reviewing process by displaying a PDS and highlighting the keywords.

The following sections provide details about some aspects of NLP systems for medical English (section 2.1: LSP-MLP) and Dutch (section 2.2: DMLP), and how results can be exchanged between them (section 2.3). Only some parts of the DMLP and LSP-MLP systems will be presented, namely those that are of importance for the experiment described below. Next to the NLP back-end, the user interface is described as well (section 2.4). The limitations of the current test are described in section 3 and some future directions for research are provided in the fourth and final section.

2 Material and Methods

2.1 The Linguistic String Project - Medical Language Processor

The Linguistic String Project - Medical Language Processor (LSP-MLP) of the New York University

is the first (and up till now the longest lasting) large scale project about NLP in Medicine (Sager et al., 1987), (Sager et al., 1995a). The LSP-MLP has also been ported to French and German, which illustrates the general applicability of its methodology and approach (Nhàn et al., 1989), (Oliver, 1992). The reason of its generality lies in the use of a well defined underlying linguistic theory (distributionalism) (Harris, 1962), (Sager et al., 1981) and a scientifically based sublanguage approach (Grishman and Kittredge, 1986).

Important for the present discussion is the semantic selection level of the LSP-MLP. All the words in the LSP dictionary are characterised by labels that indicate to which sublanguage word class(es) the words belong (e.g., H-TTCHIR: "contains general and specific surgical treatment or procedure words which imply or denote surgical intervention by the physician" (Sager et al., 1987, p.268); H-TXPROC: "contains medical test words designating procedures performed on the patient and not on a patient specimen. The patient must be present to undergo the test" (Sager et al., 1987, p.264)). An overview of the actual set of labels and word classes can be found in (Sager et al., 1995a). The semantic selection module uses distributionally established co-occurrence patterns of medical word classes to improve the parse tree by resolving cases of structural ambiguity (Hirschman, 1986). Consider the sentence 63 "operatieve procedure: vijfvoudige coronaire bypass." ¹ displayed in figure 4. The word "procedure" is semantically ambiguous because it has two semantic labels: H-TTCHIR & H-TXPROC. Thanks to the co-occurrence patterns for the medical sublanguage, only the label that is valid in this context (H-TTCHIR) is ultimately selected. In another context (e.g.: test procedure: ...), another co-occurrence pattern will apply and select the H-TXPROC reading. Other examples of resolution of word sense ambiguities by means of co-occurrence patterns can be found in (Sager et al., 1987, pp.83, 95).

The very latest work includes the use of Standard Generalized Mark-up Language (SGML) and World Wide Web (WWW) Graphical User Interface (GUI) technology to access and visualise better the requested information in the text (Sager et al., 1996). It focused on the use of static SGML or HTML-code ² for displaying the results of NLP-based checklist screening of clinical documents.

¹English: surgical procedure: quintuple coronary bypass.

²"Static" HTML code eliminates the need for an on the fly conversion of the HTML file ("dynamic" HTML code) as presented in section 2.4.

2.2 The Dutch Medical Language Processor

For the Dutch medical language, an NLP system of a medium sized coverage has been designed and implemented: the Dutch Medical Language Processor (DMLP) (Spyns, 1996c). With respect to the morphological level, there is a full form dictionary stored in the relational database format (currently some 100.000 full forms that are mostly non-compound wordforms) (Dehaspe, 1993). If necessary, a recogniser characterises the unknown word forms morphologically (Spyns, 1994). Subsequently, a contextual disambiguation component tries to reduce the number of morphological readings (Spyns, 1995).

As the syntactic level uses a "logic variant" of the LSP grammar formalism (Hirschman and Dowding, 1990), the Dutch morpho-syntactic module (Spyns and Adriaens, 1992) can replace the LSP parser. Many of the LSP-MLP medical co-occurrence patterns are practically identical for English, French and German, so that the application of these patterns to Dutch parse trees can lead to interesting results, namely the feasibility of reusing the non language specific parts of the LSP-MLP for Dutch medical NLP (Spyns, 1996a).

2.3 The DMLP/LSP-MLP connection

The linguistic data are passed on from the DMLP to the LSP-MLP system via syntactic parse trees. This is due to the fact that the selection module takes syntactic relationships into account during the semantic disambiguating phase.

The linguistic information of the DMLP and the LSP-MLP systems correspond in a high degree. Semantic word class labels, which were originally not foreseen in the Dutch lexicon, had to be added. A parse tree transducer delivers nearly genuine Dutch LSP-MLP trees (Spyns, 1996a). Although on the side of the LSP-MLP some new sublanguage semantic co-occurrence patterns had to be defined, the co-occurrence patterns are highly language independent. This was in line with results earlier achieved. An example (see figure 1) shows the output of the parse tree transducer that reshapes the DMLP tree into the required LSP-MLP format. The current state of the transducer allows to transform nearly all the parse trees.

2.4 The WWW interface

The basic idea was that when treating a patient, it is considered to be helpful to reread the admission history, the discharge summary, or other important parts of the medical record.

```

((SENTENCE
 (TEXTLET
  (OLDSENT
   (INTRODUCER
    (LN (TPOS (NULL))
     (QPOS (NULL))
     (APOS (ADJADJ (LAR (LA (NULL))
      (AVAR (ADJ='OPERATIEVE': ('OPERATIEF') ^ ('OPERATIEF'))
      (RA (NULL))))))
     (NPOS (NULL) ))
    (N='PROCEDURE':(F SINGULAR) ^ ('PROCEDURE'))
    (':=':': (':') ^ ( ':' )))
   (CENTER
    (FRAGMENT
     (SA (NULL))
     (NSTGF
      (NSTG
       (LNR
        (LN (TPOS (NULL))
         (QPOS (NULL))
         (APOS (ADJADJ (LAR (LA (NULL))
          (AVAR (ADJ='VIJFVOUDIGE': ('VIJFVOUDIG') ^ ('VIJFVOUDIG'))
          (RA (NULL))))
          (ADJADJ (LAR (LA (NULL))
           (AVAR (ADJ='CORONAIRE': ('CORONAIR') ^ ('CORONAIR'))
           (RA (NULL))))))
          (NPOS (NULL)))
          (NVAR (N='BYPASS':(SINGULAR) ^ ('BYPASS'))
           (RN (NULL))))))
          (SA (NULL))))
          (ENDMARK (':=':': (':') ^ ('.')))
          (MORESENT (NULL) )))
   [ ((16 [ SELECT-ATT ] H-TTCHIR) (21 [ SELECT-ATT ] H-TTCHIR H-TXPROC)
    (41 [ SELECT-ATT ] H-TMREP) (49 [ SELECT-ATT ] H-PTPART) (55 [ SELECT-ATT ] H-TTCHIR)) 1]

```

Figure 1: LSP like parse tree generated by the DMLP transducer for “operatieve procedure: vijfvoudige coronaire bypass.” [surgical procedure: quintuple coronary bypass]

The highlighting of medical concepts of interest makes it possible to scan a document quickly, focusing on a particular type of information, such as Symptoms and Diagnoses, or Treatments resolved (? , p.26).

Also for the medico-administrative activities, such a tool can also be helpful. Medical secretaries have to summarise patient discharge summaries by “translating” them into a fixed set of numerical codes of a classification (ICD-9-CM (Commission of Professional and Hospital Activities, 1978)). These codes (indirectly) serve for statistical and financial purposes. If the most important relevant terms for the encoding task (essentially the H-DIAG (diagnosis) and the H-TTCHIR (surgical deed) words) are already highlighted, the human encoder is able to detect them more rapidly so that the encoding speed can be improved.

The documents are morphologically and syntactically analysed by the DMLP first, the resulting parse trees being made conform to the LSP-format, and

subsequently passed³ on to the LSP-MLP.

The LSP subselection module generates a pseudo-HTML file consisting of semantic labels and the terminal elements of the parse trees. The file with the pseudo-HTML codes (see figure 3) could easily have been generated by the morphological component of the DMLP as well. In some occasions, it would be better to do so as the DMLP-LSP tree converter sometimes changes the word order. On the other hand, no advantage can then be taken from the sublanguage co-occurrence patterns for semantic disambiguation. Semantically ambiguous words will thus be highlighted more than once, which is bad for the precision score (more non relevant words are flagged). Without full fledged linguistic analysis, some ambiguities will not be resolved (? , p.27). As can be seen in figure 2 (and thus also in figure 3), the ambiguity for the word “procedure” in sentence 63 is resolved. The node number 2 only has the label H-TTCHIR.

³Currently, the files are transmitted by e-mail.

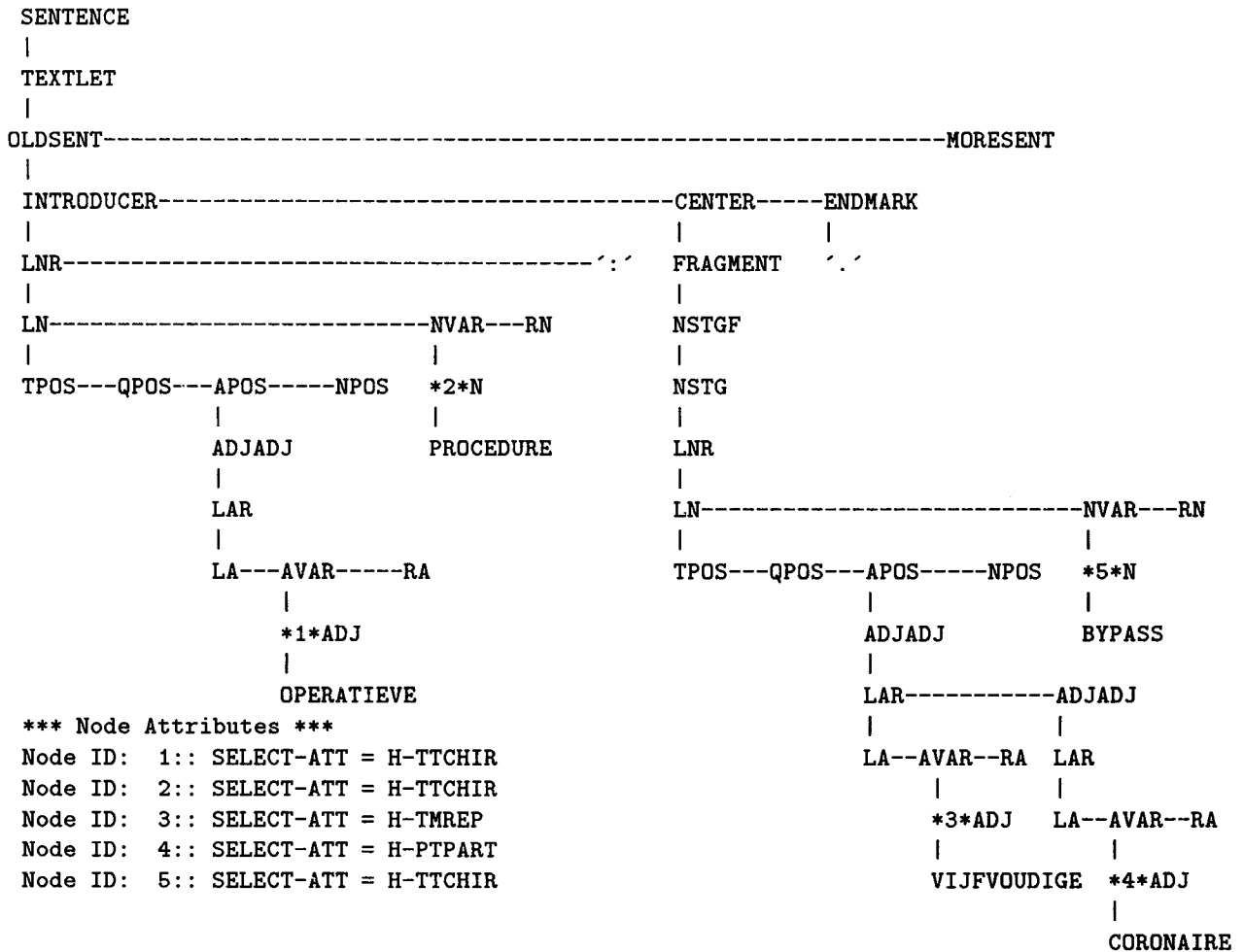


Figure 2: LSP-MLP parse tree generated after sublanguage processing for sentence of figure 1

No actual HTML-codes were furnished but the semantic labels are noted according to the HTML-style (see figure 3). The NLP processing of a load of PDSs can be done in batch during the night so that the throughput of the encoder is not affected in the negative sense.

```

63
< H - TTCHIR> operatieve < /H - TTCHIR>
< H - TTCHIR> procedure < /H - TTCHIR>
:
< H - TMREP> vijfvoudige < /H - TMREP>
< H - PTPART> coronaire < /H - PTPART>
< H - TTCHIR> bypass < /H - TTCHIR>

```

Figure3: pseudo-HTML code generated after joint DMLP/LSP-MLP processing for the sentence in figure 1

The GUI consists of two WWW-pages. The first page is conceived as a menu window. Two selection boxes allow the medical encoder to choose a text and the semantic labels. Currently, the set of PDSs is limited to nine texts. In the future, HTML-files for an unrestricted and varying number of PDSs will have to be produced. Before the encoder can start to view the NLP-processed PDSs, the HTML-code of the menu-page needs to be updated to include all the (path)names of the files concerned. This can easily be achieved by activating before each encoding session a C-shell script that scans a subdirectory and creates an actualised HTML-file for the menu page. Only the “< OPTION >< /OPTION >” lines of the first choice box need to be adapted.

Through the HTML SUBMIT command, the options selected by the medical encoder are passed (via a FORM and CGI-SCRIPT) to an external C-program. The C-program takes the filename and the requested sublanguage label(s) as parameters and generates

a new HTML-file by replacing the occurrences of the concerned label(s) by a genuine HTML-code (`< STRONG > & < /STRONG >`) around the relevant words). This temporary file is directly fed into the browser and displayed as a second WWW-page ("PDS-page"). The words marked (= belonging to the selected semantic sublanguage word class) are displayed in boldface. As the pseudo-HTML codes are ignored by the browser, the rest of the PDS is displayed in a "neutral" way.

Figure 4 shows the menu-page and PDS-page in which words concerning the diagnosis (H-DIAG), the surgical procedure (H-TTCHIR) and the bodypart (H-PTPART) are marked. The PDS-page is the bottom right part of the figure and partly overlaps the menu-page, which shows the selected PDS and labels ⁴.

3 Evaluation & Results

Before a large scale validation involving "a gold standard" and various statistical metrics (e.g. see (Hripcsak et al., 1995)) is set up and conducted, a modest formative evaluation (Hirschman and Thompson, 1995) allowed to rapidly assess the functionality of the application from the point of view of the actual user. A limited validation test has been set up. A sample of 100 Dutch sentences of varying length and syntactic complexity was selected. All the words in the dictionary covering the 100 sentences were manually tagged with LSP semantic word class labels. The medical doctor supervising the medical registration activities was asked to provide some combinations of semantic labels relevant from the

⁴The translation of the document PDS6 is as follows:

- 61 On 21/1/87 your patient has been operated in our cardiovascular surgery unit.
- 62 Pre-operative diagnosis: coronary sclerosis.
- 63 Operative procedure: quintuple coronary bypass.
- 64 Reconstruction of the left arteria mammaria on the LAD.
- 65 Venal jump graft from the aorta to the diagonalis, further to the LAD.
- 66 Venal jump graft from the aorta to the first branch of the circumflexus, further to the second branch of the circumflexus, till the RDP .
- 67 Single venal bypass from the aorta to the AV-Sulcusbranch.
- 68 After the procedure, the patient has been admitted to the Intensive Care unit.
- 69 Enclosed you can find the operation report.

viewpoint of a medical encoder (using ICD-9-CM), and to evaluate the system's responses.

For all the 100 sentences, pseudo-HTML code was generated. The recall was 100 % (all the labels concerned were flagged). The precision ranged from 66% to 100 % depending on the label combination. Nevertheless, these figures are temporary as examination of the sentences showed that very few words had more than one semantic label so that the medical subselection stage did not have a big impact. A larger test set needs to be processed in order to provide more conclusive results. Probably, recall will drop while precision could raise. Nevertheless, the experience did prove to be valuable as the collaborating doctor, who had never heard of NLP before, said he was "positively surprised and impressed" by the capabilities of the system. He also judged the tool to be an interesting utility and consented in setting up a larger experiment to measure exactly the impact of the tool on the daily routine of the medical encoders. The evaluation procedure of this large test will be organised to comply as much as possible with the evaluation criteria recently proposed by Friedman and Hripcsak (Friedman and Hripcsak, 1997).

4 Future Research

In order to demonstrate the full power of the LSP-MLP, the same sentences could be processed by the joint DMLP/LSP-MLP systems and stored in a RDB table - as is done in other experiments involving the LSP system (Hirschman et al., 1981). Specific SQL-queries can then return the ID-number of the sentence with the relevant information instead of the information itself. If the ID-number is added to the original document as a pseudo-HTML code, the same mechanism as mentioned above can be used to highlight the sentences containing the relevant information. Several variants on this base scheme can be thought of.

Following the line of research of Sager (Sager et al., 1995a) and Wingert (Wingert et al., 1989), classification codes could already be generated automatically (see also (Lovis et al., 1995)) and presented on the screen next to the original text. But the human encoder would remain responsible for the ultimate selection of the exact codes.

Another possibility is the creation of "views" or "masks". HTML files can be generated with "hard coded" instructions to emphasise fixed combinations of semantic labels. Buttons in the menu-page allow to display very rapidly the selected view on the PDS. Several experiments for English have already been successfully carried out (?) on the use of "static WWW-technology". Interesting as well is the cre-

ation of Document Type Definitions (DTD) that associate a particular layout with a specific semantic label (see also (Zweigenbaum et al., 1997)). The DTDs can act as a locally defined view (GUI aspect) on common SGML data (NLP aspect).

Other potential applications in the medical domain for the DMLP/LSP-MLP combination are e.g., the determination of patient profiles (Borst et al., 1991), quality assurance (Lyman et al., 1991) and extraction of sign/syptom information for medication (Lyman et al., 1985). Overviews of the possible utilisation in the healthcare area of NLP based systems, irrespective of their theoretical background, can be found in (Baud et al., 1992b) & (Sager et al., 1987, chapter2).

But before any application of such an extent can be envisaged for Dutch, the words of the dictionary database all have to receive the appropriate semantic label(s). Luckily, this process can be automated. The LSP-team has implemented such routines (Hirschman et al., 1975) but other techniques could be applied as well (see (Habert et al., 1996)).

From a technical point of view, it would be better to group all the involved software modules (NLP, RDBMS, WWW) on the same platform to optimally exploit the potentialities offered by the combination of the components mentioned. Ultimately, a client/server architecture (separating language specific from domain specific issues and the linguistic aspects from user interface aspects) will be the best architecture for a real life application.

We can conclude that the application presented above shows the feasibility to integrate Electronic Medical Record (EMR) systems with NLP applications. This is the kernel message of the DOME project (Bouaud et al., 1996) that advocates the use of SGML - and HTML-technology for EMR systems. The above presented WWW-application could thus be integrated in such a hypertextual EMR system.


References

- R. Baud, A.-M. Rassinoux, and J.-R. Scherrer. 1992a. Natural language processing and medical records. In K.C. Lun, editor, *Proc. of MED-INFO 92*, pages 1362 - 1367. North-Holland.
- R. Baud, A.-M. Rassinoux, and J.-R. Scherrer. 1992b. Natural language processing and Semantical Representation of Medical Texts. *Methods of Information in Medicine*, (31): 117 - 125.
- F. Borst, M. Lyman, N.T. Nhàn, L. Tick, N. Sager, and J.-R. Scherrer. 1991. Textinfo: A Tool for Automatic Determination of Patient Clinical Profiles Using Text Analysis. In P. Clayton, editor, *Proc. of SCAMC 91*, pages 63 - 67. McGraw-Hill. New York.
- J. Bouaud, B. Séroussi, and P. Zweigenbaum. 1996. An experiment towards a document centered hypertextual computerised patient record. In *Proc. of MIE 96*, pages 453 - 457, Amsterdam. IOS Press.
- E. Bowman. 1996. Coding and classification systems. In M. Abdelhak, S. Grostick, M-A. Hanken, and E. Jacobs (eds.), editors, *Health Information: Management of a Strategic Resource*, pages 214 - 235. W.B. Saunders Company, Philadelphia.
- C. Chute, editor . 1997. Preprints of the IMIA WG6 Conference on Natural Language and Medical Concept Representation. Jacksonville .
- Commission of Professional and Hospital Activities. 1978. *The International Classification of Diseases, Ninth Revision, Clinical Modifications (ICD-9-CM)*. Ann Arbor, Michigan.
- L. Dehaspe. 1993. *Report on the building of the sc menelas lexical database*. Technical Report 93-002, K.U. Leuven.
- J. Duisterhout. 1996. Coding and Classifications. In J. van Bommel, editor, *Handbook of Medical Informatics*, pages 83 - 94. Bohn, Stafleu, Van Loghum, Houten/Diegem, preliminary version.
- C. Friedman and S. Johnson. 1992. Medical text processing: Past achievements, future directions. In M.J. Ball and M.F. Collen, editors, *Aspects of the Computer-based Patient Record*, pages 212 - 228. Springer - Verlag, Berlin.
- C. Friedman and G Hripcsak. 1997. Evaluating Natural Language Processors in the Clinical Domain. In (Chute, 1997), pages 41 - 52.
- R. Grishman and R. Kittredge, editors. 1986. *Analyzing Language in Restricted Domains: Sublanguage Description and Processing*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- B. Habert, E. Naulleau, and A Nazarenko. 1996. Symbolic word classification for medium-size corpora. In *Proc. of COLING 96*, pages 490 - 495.
- Z. Harris. 1962. *String Analysis of Sentence Structures*. Mouton, The Hague.
- W. Hersh. 1996. *Information Retrieval, A Health Care Perspective*. Springer Verlag, New York.
- L. Hirschman and J. Dowding. 1990. Restriction grammar: a logic grammar. In P. Saint-Dizier and S. Szpakowicz, editors, *Logic and Logic Grammars for Language Processing*, pages 141 - 167. Ellis Horwood.

- L. Hirschman and H. Thompson. 1995. Overview of evaluation in speech and natural language processing. In J. and Mariani, editor, *State of the Art in Natural Language Processing*, pages 475 – 518.
- L. Hirschman, R. Grishman, and N. Sager. 1975. Grammatically-based automatic word class formation. *Information Processing and Management*, pages 39 – 57.
- L. Hirschman, G. Story, E. Marsh, M. Lyman, and N. Sager. 1981. An experiment in automated health care evaluation from narrative medical records. *Computers and Biomedical Research*, (14):447 – 463.
- L. Hirschman. 1986. Discovering sublanguage structures. In (Grishman and Kittredge, 1986), pages 211 – 234.
- G. Hripcsak, C. Friedman, P. Alderson, W. Dumouchel, S. Johnson, and P. Clayton. 1995. Unlocking Clinical Data from Narrative Reports: A Study of Natural Language Processing. *Annals of Internal Medicine*, vol. 122 (9): 681 – 688.
- C. Lovis, P.-A. Michel, R. Baud, and J.-R. Scherrer. 1995. Use of a conceptual semi-automatic ICD-9 encoding system in a hospital environment. In *Artificial Intelligence in Medicine, Proc. of AIME 95*, pages 331 – 339. Springer-Verlag.
- M. Lyman, N. Sager, C. Friedman and E. Chi. 1985. Computer-structured Narrative in Ambulatory Care: Its Use in Longitudinal Review of Clinical Data. In *Proc. of SCAMC 85*, pages 82 – 86.
- M. Lyman, N. Sager, L. Tick, N.T. Nhan, Y. Su, F. Borst, and J.-R. Scherrer. 1991. The application of natural-language processing to healthcare quality assessment. *Medical Decision Making*, (11 Suppl): S65 – S68 .
- A. McCray, J.-R. Scherrer, C. Safran, and C. Chute (eds.). 1995. Special Issue on Concepts, Knowledge, and Language in Healthcare Information Systems. *Methods of Information in Medicine* (34) 1/2 .
- N.T. Nhan, N. Sager, M. Lyman, L. Tick, F. Borst, and Y. Su. 1989. A medical language processor for two indo-european languages. In *Proc. of SCAMC 89*, pages 554 – 558.
- N. Oliver. 1992. *A sublanguage based medical language processing system for German*. Ph.D. thesis, Dept. of Computer Science. New York University.
- N. Sager, C. Friedman, and M. Lyman. 1981. *Natural Language Information Processing: a computer grammar of English and its applications*. Addison Wesley, Reading, Massachusetts.
- N. Sager, C. Friedman, and M. Lyman. 1987. *Medical Language Processing: Computer Management of Narrative Data*. Addison Wesley, Reading, Massachusetts.
- N. Sager, M. Lyman, N. Nhan, and L. Tick. 1995a. Medical language processing: Applications to patient data representation and automatic encoding. *Methods of Information in Medicine*, (34):140 – 146.
- N. Sager, N. Nhan, M. Lyman, and L. Tick. 1996. Medical Language Processing with SGML display. In *Proc. of the 1996 AMIA Annual Fall Symposium*, pages 547 – 551 .
- J.R. Scherrer, R. Côté, and S. Mandil (eds.). 1989. *Computerized Natural Medical Language Processing for Knowledge Representation*. North Holland.
- P. Spyns and G. Adriaens. 1992. Applying and Improving the Restriction Grammar Approach for Dutch Patient Discharge Summaries. In *Proc. of COLING 92*, pages 1164 – 1168.
- P. Spyns. 1994. A robust category guesser for Dutch medical language. In *Proc. of ANLP 94*, pages 150 – 155. ACL.
- P. Spyns. 1995. A contextual disambiguator for Dutch medical language. In *Proc. of the BeNeLux Workshop on Logic Programming BENELOG 95*, pages 20 – 24, Gent.
- P. Spyns. 1996a. Medical language processing and reusability of resources: a case study applied to Dutch. In *Proc. of MIE 96*, pages 1147 – 1152, Amsterdam. IOS Press.
- P. Spyns. 1996b. Natural language processing in medicine: An overview. *Methods of Information in Medicine*, (35):285 – 302 .
- P. Spyns. 1996c. *Natural Language Processing in the bio-medical area: Design and Implementation of an Analyser for Dutch*. Ph.D. thesis, Dept. of Computer Science, K.U. Leuven.
- F. Wingert, D. Rothwell, and R. Côté. 1989. Automated indexing into SNOMED and ICD. In (Scherrer et al., 1989), pages 1 – 5.38.
- P. Zweigenbaum, J. Bouaud, B. Bachimont, J. Charlet, B. Séroussi, J.F. Boisvieux. 1997. From Text to Knowledge: a Unifying Document-Oriented View of Analyzed Medical Language. In (Chute. 1997). pages 21 – 30 .

Netscape: Medical String Tagging

File Edit View Go Bookmarks Options Directory Window Help

Location: 

Select a text to process:

- Le Pds.html
- PDS1.html
- PDS2.html
- PDS3.html
- PDS4.html
- PDS5.html
- PDS6.html
- le pds1.html
- le pds2.html

Select the semantic category: *[multiple selection is allowed]*

For a full description of the categories: see N. Sager, M. Lyman, N.T. Nhan, L.J. Tick, **Medical Language Processing: Applications to Patient Data Representation and Automatic Encoding**, in *Methods in Information in Medicine* 34 (1/2): 140-157.

Or go to the next paragraph containing a short definition of the labels (Reprinted from the above mentioned reference).

Netscape: Medical String Tagging

File Edit View Go Bookmarks Options Directory Window Help

H-PTPALP
 H-PTPART
 H-PTSPEC
 H-PTVERB
 H-TXCLIN
 H-TXPROC
 H-TXSPEC
 H-TXVAR
 H-TTCHIR
 H-TTCOMP

The document: *PDS6.html* was searched using the labels:

- *H-PTPART*
- *H-TTCHIR*
- *H-DIAG*

61 UW PATIENTE WERD OP 21 1 . 87 OPERATIEF BEHANDELD OP ONZE DIENST VAN CARDIOVASCULAIRE HEELKUNDE .

62 PREOPERATIEVE DIAGNOSE : CORONAIRE SCLEROSE .

63 OPERATIEVE PROCEDURE VIJFVOUDIGE CORONAIRE BYPASS .

64 RECONSTRUCTIE VAN DE LINKER ARTERIA MAMMARIA OP DE LAD .

65 VENEUZE JUMP GRAFT VAN DE AORTA NAAR DE DIAGONALIS , VERDER NAAR DE LAD .

66 VENEUZE JUMP GRAFT VAN DE AORTA NAAR DE EERSTE CIRCUMFLEXTAK , VERDER NAAR DE TWEEDE CIRCUMFLEXTAK , TOT DE RDP .

67 ENKELVOUDIGE VENEUZE BYPASS VAN DE AORTA NAAR DE AV-SULCUSTAK .

68 DE PATIENTE WERD IN AANSLUITING MET DE PROCEDURE OP DE INTENSIEVE THERAPIE EENHEID OPGENOMEN .

69 HIERBIJ INGESLOTEN VINDT U HET OPERATIEVERSLAG .

[back to start](#)