

MULTITALE: linking medical concepts by means of frames

Isa Maks and Willy Martin

Vrije Universiteit Amsterdam

Dept. of Lexicology

De Boelelaan 1105, 1081 HV Amsterdam

makse@jet.let.vu.nl

martin_w@jet.let.vu.nl

ABSTRACT

In this paper MULTITALE, a system for the semantic tagging of medical neurosurgical texts and for the semi-automatic expansion of the medical lexicon, will be presented. Given the textual information explosion (in particular in, though not restricted to, specialized domains) there is an urgent need for tools enabling to exploit the information available in natural language texts. MULTITALE has been devised therefore primarily with the aim to make explicit semantic information in medical texts, which should lead to more refined information retrieval results. By making "educated guesses" the system moreover has a possibility to expand its own lexicon of medical terms so to be able to cope with new texts.

I. INTRODUCTION

MULTITALE has been developed as part of an EU project (MLAP 93-04) which has been started in 1994 and has been completed recently. The English part of it was carried out by the Belgian partner (Office Line Engineering NV, Zonnegem; RAMIT, Gent), the Dutch part by the Dutch partner (Lexicology Research Group, Free University Amsterdam¹). Although both groups share the same starting point and objectives, the methods followed show some idiosyncracies, therefore if in what follows MULTITALE is mentioned, actually the MULTITALE Dutch module is meant.

II. SEMANTIC MODEL

The semantic tagging is based on the CEN/TC251-model for Surgical Procedures (CEN,1994). This model is a classification and coding system of medical procedures. It distinguishes the following concept types: CC_Surgical_Deed (indicating the surgical intervention), CC_Anatomy (indicating anatomical concepts), CC_Pathology (indicating pathological concepts), CC_Interventional_Equipment (indicating the instrument), CC_Combi (a term which has a medical meaning only in combination with another -medical- term),

CC_modifier_bodyside,CC_modifier_extent, CC_modifier_number (terms which modify other medical terms.

The Surgical deed concept is classified into 12 subtypes, among others: CS_remove,CS_close, CS_create, CS_close, CS_install, CS_make_appear. The Surgical deed concept is considered as the nucleus of the surgical procedure and may have different types of relationships with the other medical concepts:

the R_Direct_object indicates the object on which the surgical deed is carried out;

the R_Indirect_Object indicates the object to or from or in which the surgical deed is carried out;

the R_Location indicates the place where the Surgical Deed is carried out;

the R_means indicates that with which the Surgical Deed is carried out;

the R_manner indicates how the Surgical Deed is carried out.

The next example illustrates the CEN/TC251-model. The input is taken from a report of a neurosurgical intervention; the output is generated by the MULTITALE system.

(ex. 1)

INPUT:

Enkele fragmenten discus worden nog verwijderd, dan worden met een beiteltje de osteofyttaire randen van de dekplaat weggenomen.

(Some fragments of the discus are removed, thereafter the osteophytic edges of the cover plate are taken away with a chisel)

OUTPUT:

```
Enkele fragmenten discus
      R_dir_object  CC_anatomy      S_NP
Enkele
fragmenten CC_combi
discus     CC_anatomy

worden                                           s_verb
nog                                               s_adverb
verwijderd  CC_surg_deed CS_remove  s_verb
            +R_dir_object Enkele fragmenten discus
-----
dan
worden                                           s_verb
met een beiteltje
      R_means      CC_interv equip      S_PP
met
```

¹ T. Frizzanin (syntax), A. Kramer (lexicon), I. Maks (syntax and semantics), W. Martin (overall supervision)

(ex. 3)

Het aneurysma wordt afgeclippt met een rechte clip.
(*The aneurysm is cut off with a straight clip*)

Considering a sentence like example 3, the system has to decide which noun phrase can be related to the verb and what is the nature of that relationship. The possible Semantic Links correspond to the CEN Semantic Links. So, in relation to the verb *afclippen* (*to cut off*) we find - among others- Direct Object and Means. Like for Case Grammar, there are markers, like prepositions or the lack of prepositions, which point to a certain Semantic Link. In the above example, the preposition *met* (*with*) is an indicator for Means and it is the absence of a preposition which points to the Direct Object. For identifying the Semantic Links, three particular kinds of information are needed:

1. the surgical deed concept and its possible Semantic Links
2. the NP and its concept type
3. the Prepositions and their values (I_values).

Frames are chosen for the implementation of the model. They have a top level which is fixed and represents things that are true of a certain situation. They have a lower level with slots, conditions on the slots and fillers, which represent things that are expected for a certain situation (Minsky, 1975). As regards the top level, three different kinds of frames are defined for Multitale.

a. standard surgical deed frame:

LEXICAL ELEMENT	surgical deed concept
PART_OF_SPEECH	verb
CONCEPT TYPE	CC_surgical_deed
CONCEPT SUBTYPE	subtype of CC_surgical_deed: CS_..

b. variant surgical deed frame:

LEXICAL ELEMENT	surgical deed concept
PART_OF_SPEECH	noun
CONCEPT TYPE	CC-surgical-deed
CONCEPT SUBTYPE	subtype of CC_surgical_deed : CS_..

c. non-surgical-deed frame:

LEXICAL ELEMENT	non-surgical-deed concept
PART_OF_SPEECH	noun
CONCEPT TYPE	non_surgical-deed concept type

The lower level of the non-surgical-deed frame will be discussed below being different from the lower level of both the surgical-deed-frames. The lower level of the *surgical_deed_frames* contains sets of slots for the identification of Semantic Links. The set consists of a slot -called ROLE- for the type of Semantic Link, a slot -called ARG- for the pointer to the element in the sentence which is linked to the surgical deed concept, a slot -called CC - for the concept type of the element which is linked and finally a slot -called IND- for the indication (I-value) of the function of the linked element. Each slot can specify conditions its filler must meet.

slots:	conditions on the fillers are marked with *:
ROLE	*one of the CEN-defined Semantic Links
ARG	
CC	*CEN-defined concept types
IND	*I-values
ROLE	

ARG
CC
IND
ROLE

The conditions on the fillers are found in the surgical deed lexicon and the type lexicon. The type lexicon gives general information for the surgical deed subtype; the surgical deed lexicon gives information for the individual token, the individual surgical deed concept.

IV.2 THE LINKING MODULE STEP BY STEP

For filling in the frame and analyzing the sentence, the system performs the following steps for each Surgical-deed-clause.

(ex. 4a)

..de catheter wordt in de wond geplaatst
(*the catheter has been installed in the wound*)

a. consultation of the surgical deed lexicon for finding the subtype entry of *plaatsen* in the surgical deed lexicon:

<lemma>	plaatsen
<cat>	verb
<concept type>	CC_surgical_deed
<concept subtype>	CS_install

b. consultation of the type lexicon for defining the frame
The following frame will be built (conditions on the fillers of the slots are marked with *):

LEXICAL ELEMENT	geplaatst
PART_OF_SPEECH	verb
CONCEPT TYPE	CC_surgical_deed
CONCEPT SUBTYPE	CS_install
ROLE	R_DIRECT_OBJECT
ARG	
CC	*CC_inter_equipment/ *CC_combi/ CC_WAY/ *CC_anatomy
IND	*I_NONE
ROLE	*R_INDIRECT_OBJECT
ARG	
CC	*CC_pathology/CC_anatomy
IND	*I_SITE
ROLE	*R_MEANS
ARG	
CC	*CC_inter_equipment/ *CC_anatomy
IND	*I_MEANS
ROLE	*R_LOC
ARG	
CC	*CC_anatomy/ CC_pathology/ *CC_combi
IND	*I_LOC
ROLE	*R_MANNER
ARG	
CC	*CC_surgical_deed
IND	*I_MANNER

c. if necessary, adaptation of the frame with information from the surgical deed lexicon will take place.

d. if necessary, adaptation of the frame will be carried out if the surgical deed concept is a noun or a non-finite form of the verb.

The subtype frame is specified for the finite form of the verb and the past participle. If the surgical deed concept is a noun or an infinitive form of the verb, the R_direct_object is marked by the preposition *van* (I_value = I_van). Therefore, the condition on the <IND>-slot of the R_Direct_object, will be changed to I_van.

e. filling in the slots:

(ex. 4b)

.. de catheter wordt in de wond geplaatst

LEXICAL ELEMENT	geplaatst
PART_OF_SPEECH	verb
CONCEPT TYPE	CC_surgical_deed
CONCEPT SUBTYPE	CS_install
ROLE	R_DIRECT_OBJECT
ARG	-> de catheter
CC	CC_INTERVENT_EQUIPMENT
IND	I_NONE (-)
ROLE	R_INDIRECT_OBJECT
ARG	-> in de wond
CC	CC_pathology
IND	I_SITE (in)

The linking module tries to match the specifications of the elements of a Surgical-Deed-clause with the conditions on the fillers of a slot.

IV.3 THE LINKING MODULE AND PREPOSITIONAL PHRASE ATTACHMENT

For all non-surgical-deed concepts, namely CC_anatomy, CC_pathology, CC_combi, and CC_intervent_equipment, the following frame has been defined:

top-level:

LEXICAL ELEMENT	non-surgical-deed concept
PART_OF_SPEECH	NP or PP
CONCEPT TYPE	CC_anatomy/ CC_combi CC_pathology/ CC_intervent_equipment

lower level:

ROLE	R_POST_MOD
ARG	
CC	*non-surgical-deed concept type
IND	*I_VAN
POS	*+1

There is only one set of slots, expressing the link between two non-surgical-deed concepts in a sentence:

(ex. 5)

*..het intracellair gedeelte van de tumor wordt uitgereu-
teerd.*

(the intracellular part of the tumour is cleaned)

The prepositional phrase *van de tumor* modifies the noun phrase *het intra- cellair gedeelte*. The link between these two phrases is called post-modification link. The slot POS(ition) in combination with the constraint +1 requires

that the postmodifying phrase directly follows the NP for which this frame is defined.

(ex. 6-a)

non-surgical-deed frame with slots filled in:

LEXICAL ELEMENT	het intracellair gedeelte
PART_OF_SPEECH	NP
CONCEPT TYPE	CC_combi
ROLE	R_POST_MOD
ARG	-> van de tumor
CC	CC_pathology
IND	I-VAN
POS	+1

(ex. 6-b)

surgical deed frame with slots filled in:

LEXICAL ELEMENT	uitgecuureteerd
PART_OF_SPEECH	verb
CONCEPT TYPE	CC_surgical_deed
CONCEPT SUBTYPE	CS_clean
ROLE	R_DIRECT_OBJECT
ARG	-> het intracellair gedeelte van de tumor
CC	CC_pathology
IND	I_NONE (-)

We have established an order for the matching of the Semantic Links, giving priority to these Links which connect a surgical deed concept with another concept.

(ex. 7)

*..waarna de frontale lob van zijn adherenties wordt vrijge-
maakt*

(.. after that the frontal lobe has been freed ...)

The specifications of *van zijn adherenties*, CC_pathology and I_van meet both the conditions on the filler for the R_indirect_object of the surgical deed concept and the conditions on the filler of the R_POST_MOD of the non-surgical-deed concept. Since the R_Indirect_Object precedes the R_Post_Mod *van zijn adherenties* will be linked - correctly - with *vrijgemaakt*.

V. THE GUESSING MODULE

V.1 INTRODUCTION

The guessing module of the Multitale system deals with the semi-automatic augmentation of the concept lexicons (=lexicons of surgical deeds and non- surgical deeds). The performance of the tagger depends for a great deal on the completeness of the lexicon. If the lexicon does not contain a medical term, the tagger cannot assign a semantic link to this unknown term and another one in the sentence. The guessing module is an important help for the augmentation of the concept lexicon, and consequently an important part of the Multitale system when tagging unknown texts. The function of the module is twofold:

1. - generation of a list of words which are likely to be

medical terms and CEN concepts. The list does not present just a list of words unknown to the system but a selection of words relevant to CEN.

2. - suggestions regarding the concept type for each word of the generated list. The suggested concept types are

CC_surgical_deed (without subtype) ,CC_anatomy, CC_pathology, CC_intervent_equipment and CC_way.

The module works semi-automatically: the list of unknown words is generated in an automatic way, but the user of the system has to decide whether the suggestion is correct or not before adding it to the lexicon.

V.2 GUESSING NON-SURGICAL-DEED CONCEPTS

The guessing module uses the frames of the linking module. For the guessing of the non-surgical-deed concepts, it uses the constraints given for the fillers of the slots of the surgical deed frame. The general rule is the following: if a phrase (noun phrase or prepositional phrase) has a Semantic Link with a surgical deed concept, at least one of the words of the phrase is a CEN- concept. Suppose a sentence contains a surgical deed concept, but the system is not able to make a semantic link between the surgical deed concept and another concept in the surgical-deed-clause. In most cases, this is due to the fact that the concept type of the terms is not known, for example:

(ex. 8)
.. de tumor wordt verwijderd ..
 -de tumor CC-?
 -verwijderd CC_surgical_deed CS_remove
 (*.. the tumour will be removed..*)

Suppose that *tumor* is not present in the lexicon, then the system is not able to meet the conditions of the slots of *verwijderen* and cannot indicate the R_direct_object. The I_value of *de tumor* meets the required value for the slot <IND>, but none of the concept classes required for the <CC> slot can be matched.

(ex. 8a)

LEXICAL ELEMENT	verwijderd
PART_OF_SPEECH	verb
CONCEPT TYPE	CC_surgical_deed
CONCEPT SUBTYPE	CS_remove
ROLE	*R_DIRECT_OBJECT
ARG	
CC	*CC_pathology/ CC_combi/ *CC_anatomy/ *CC_intervent_equipment
IND	*I_NONE

The guessing module then 'relaxes' the conditions set and now considers the syntactic function of a noun phrase or a prepositional phrase (expressed by its I_value) as a sufficient indication for a semantic link. In other words: if the I_value of an element in the surgical-deed-clause satisfies one of the I_values required for an <IND> slot, then the element will be linked.

(ex. 8b)

LEXICAL ELEMENT	verwijderd
PART_OF_SPEECH	verb
CONCEPT TYPE	CC_surgical_deed
CONCEPT SUBTYPE	CS_remove
ROLER_DIRECT_OBJECT	
ARG	-->de tumor

CC	?
IND	I_NONE

The next step is to make a guess about the concept type of the filled-in element. The constraints - CC_pathology, CC_combi and CC_anatomy (see frame for *verwijderen*) - of the <CC> slot, are considered as good candidates. To be able to make a choice for one of them, the constraints are connected with priority numbers, obtained by corpus observation:

(ex. 8c)
 part of the entry CS_remove of the type lexicon:
 priority number:

<conceptClass>	CC_surgical_deed	
<conceptSubclass>	CS_remove	
<rol>	R_DIRECT_OBJECT	
<cc1>	CC_pathology	1
<cc2>	CC_anatomy	2
<cc3>	CC_intervent_equipment	3
<cc4>	CC_combi	2
<prep1>	I_NONE	
<rol>..		

The numbers are based on the occurrences of combinations of concept types in the corpus. These occurrences are translated into priority numbers for the constraints on the <CC> slots which are registered in the type and surgical deed lexicon. The concept type with the highest occurrence (in combination with the given surgical deed concept and the given Semantic Link) was marked with the highest priority number (namely 1). The concept type with the highest number is considered the most likely candidate for the filled-in element. So, the system will suggest the concept type CC_pathology for *tumor* in (ex. 8). In most cases the element, for which the concept type is guessed, consists of more than one word:

(ex. 9)
.om de laterale tumorale expansie te kunnen verwijderen..
 -de laterale tumorale expansie
 CC-? R_DIRECT_OBJECT
 -verwijderen CC_surgical_deed CS_remove
 (*.. to be able to remove the tumoral expansion*)

The guessed concept type is suggested for all the nouns and adjectives, being the meaningful words in the phrase. Words which have a meaning in general language are marked <GEN>, to indicate that the guess is more questionable than in other cases.

results of the guessing module for (ex.) 9:

laterale	CC_pathology?
tumorale	CC_pathology?
expansie	CC_pathology?

The suggestion is only correct for *tumorale*, whereas *laterale* is of the type MC_bodyside and *expansie* of the CC_combi. In a later phase we intend to correct these cases of overgeneration.

V.3 GUESSING SURGICAL DEED CONCEPTS

The general rule for the guessing of surgical deeds is: each verb that has a Semantic Link with a CEN-concept, is a

surgical deed concept. For finding the unknown surgical deed concepts, MULTITALE makes use of the frames as well. For each verb in the text that is not in the concept lexicon, a frame is built. This frame is called CS_neutral. Its semantic constraints - the allowed concept types - and its syntactic constraints - the I_values - are less strict than the constraints which have been specified for the frames of the surgical deeds belonging to a specific subtype. Because of the 'neutral' character of the frame, no priority information can be given, so every constraint is labelled with the same degree of priority(=1).

(ex. 10)

entry of CS_neutral in the type lexicon:

```
<conceptClass>   CC_surgical_deed
<conceptSubclass> CS_neutral
<role>           R_DIRECT_OBJECT
<cc1>           CC_pathology           1
<cc2>           CC_anatomy             1
<cc3>           CC_intervent_equipment 1
<cc4>           CC_combi              1
<prep1>         I_NONE                1
<role>           R_INDIRECT_OBJECT
<cc1>           CC_pathology           1
<cc2>           CC_anatomy             1
<cc4>           CC_combi              1
<prep1>         I_SOURCE              1
<prep2>         I_SITE                1
<role>           R_MEANS
<cc1>           CC_anatomy             1
<cc2>           CC_intervent_equipment 1
<prep1>         I_MEANS               1
<role>           R_MANNER
<cc1>           CC_surgical_deed      1
<prep1>         I_MANNER              1
<endtype>
```

If the verb has at least one of the Semantic Links of the entry CS_neutral, it will be considered as a surgical deed concept:

(ex. 11)

```
..wordt de peritoneale drain intercutaan getunneld ..
-peritoneale drain [CC_intervent_equipment
                   R_DIRECT_OBJECT?]
-getunneld [CC_surgical_deed, CS_neutral?]
(the drain is .. connected)
```

(ex. 12)

```
.. wordt losgemaakt door wegboren ..
-lossgemaakt [CC_surgical_deed, CS_neutral?]
-door wegboren [CS_remove, R_MANNER?]
(.. freed by removing..)
```

VI. EVALUATION AND CONCLUSION

By way of conclusion we will mention the main results obtained until now. MULTITALE has not yet been extensively tested, yet when confronted with new texts, results look quite satisfactorily and promising. The following table is based upon 5 new medical reports (each some 200 wordtokens in length), the words not being a priori in the lexicon.

syntax	
present Nps	56
correctly assigned	49
successrate	87%
concept type assignment	
present medical concepts	121
correctly assigned	114
successrate	94%
concept linking	
links present	53
correctly assigned	45
successrate	85%

Although these results should be confirmed by further tests and although the restricted character of the domain, no doubt, has got an influence on the score, yet we hope to have shown that the approach as such to semantic/conceptual tagging of medical reports seems both to be promising and worth while of further exploration.

References

- CEN: European Committee for Standardisation\TC251\PT002s. 1994. *Terminology and coding systems of medical procedures and surgical procedures*.
- Deville, G. 1989. *Modelization of task-oriented utterances in a man-machine dialogue system*. PHD-thesis, Antwerp.
- Fillmore, C. J. 1968. The case for case. In Bach, E. and Harms, R. T., eds., *Universals in Linguistic Theory*. Holt, Rinehart and Winston, New York.
- Martin, W. 1992. On the parsing of definitions. In *Euralex-Proceedings-92*, pp. 247-256.
- Martin, W. 1992. Concept-oriented parsing of definitions. In *Coling-92 Proceedings*, pp. 988-992.
- Martin, W. 1994. Knowledge-representation schemata and dictionary definitions. In Carlon, K., ed., *Perspectives on English*. Peeters, Leuven.
- Minsky, M. 1975. A framework for representing knowledge. In Winston, P. H., ed., *The psychology of Computer Vision*. Mc_Graw-Hill, New York.
- MULTITALE. 1996. *Final Report*, VU Amsterdam, RAMIT Gent, Office Line Engineering Zonnegem. (in press).
- Pinkhof-Hilfman. 1992. *Geneeskundig woordenboek*. Bohn Stafleu van Loghum, Houten/Zaventem.
- Wegner, I. 1985. *Frame-Theorie in der lexicographie*. Niemeyer, Tuebingen.