Medical Informatics Europe '96
J. Brender et al. (Eds.)
IOS Press, 1996

Automated extraction of neurosurgical procedure expressions from full text reports: the Multi-TALE experience

Ceusters W [a, b], Deville G [a, c], De Moor G [b]

[a] *Office Line Engineering NV, Hazenakkerstraat 20, B-9520 - Zonnegem, Belgium*
[b] *RAMIT VZW, University Hospital, De Pintelaan 185, B-9000 Gent, Belgium*
[c] *FUNDP, Rue de Bruxelles 61, B-5000 Namur, Belgium*

The purpose of the Multi-TALE syntactic-semantic tagger is to identify surgical procedure expressions in full text reports, and to assign semantic tags to its composing elements according to the CEN/TC251/PT002S prestandard. Ten English reports - 5 from the original training sample and 5 from a testing sample - have been used for validation. In total 2139 syntactic units (to be mapped into 6 categories) and 857 semantic-contextual entities (to be mapped into 8 categories) were to be retrieved. Information recall (I-R) and information precision (I-P) were measured for both samples. I-R and I-P for the syntactic units were respectively 93.3% and 94.4% for the training sample, and 95.7% versus 95.7% for the testing sample. I-R and I-P for the semantic tags were respectively 91.3% and 92.0% versus 89.3% and 94.8%.

1. Introduction

For data-collection in healthcare, two major issues are generally well recognised. Firstly, coherent models for the representation of data, information and knowledge are urgently needed. Indeed, only when such models are available, data coming from different sources may be compared and used for various purposes. Secondly, despite the increasing use of healthcare information systems in which data are registered in a structured and standardised way, the clinical narrative remains an important source of information when delivering optimal care to patients. However, the format in which this information is expressed, cannot readily be used for automatic data-processing, and is inefficient in terms of computer assisted medical management, data transfer between different information systems, quality assurance and surveillance. The Multi-TALE syntactic-semantic tagger has been developed as a tool for factual information retrieval from neuro-surgical procedure reports, hence making this information suitable for further processing. In this paper, we describe the underlying semantic model and the overall technical architecture of the prototype, sketch the results of two validations, and compare the results with those of similar systems.

2. The CEN/TC251 ENV on Surgical Procedures

A project Team (PT002S) of CEN/TC251 (Technical Committee 251 on Medical Informatics of CEN (Comité Européen de Normalisation) was given the task to develop a structure for classification and coding of surgical procedures. The purpose of this standard is to *identify the concepts within the text of surgical procedure notes and to structure them to represent the concepts and their internal relations* [1]. According to this standard, a surgical procedure is conceptually composed of a *surgical deed* (deed that can be done by the operator to the patient's body during the surgical procedure) which is semantically linked to the concept fields *human anatomy*, *pathology* and *interventional equipment*. Potential semantic links are *direct object* (referring to that on which the surgical deed is carried out), the *indirect object* (referring to the site of the surgical deed) and the *means* (referring to the means with which the deed is carried out). Although the standard was developed for the description of elementary surgical procedure expressions as they can be found in classification and coding systems, one of the hypotheses of the Multi-TALE project was that the same structure could be used to represent the particular tasks described in full text reports. To test this hypothesis, a functional approach to the medical language was adopted, recognising that the language structure is not to be separated from language use. The functional dimension of natural language has been convincingly introduced by Dik in theoretical linguistics [2], and successfully applied and adapted by Deville for the modelisation of administrative language [3]. By using the same approach, we could semantically model surgical procedure expressions as structures that consist of a predicate (surgical deed) with an adequate number of terms, specified by means of a semantic function or case, and functioning as arguments of that predicate. We identified the concept field *surgical deed* as predicate primitive and the semantic links *direct object, indirect object and means* as the expression of semantic functions labelled case roles that can be ascribed to the arguments of a predicate (i.e. concept of surgical deed). We also redefined the notion of combinatorial rule as a specification at surgical deed level, constraining on semantic grounds the case roles that are prototypical of a given surgical deed. We finally defined 12 classes of surgical deeds by appealing to the notion of predicate primitive. To each class of surgical deed corresponds a cluster of arguments with specific semantic constraints [4].

3. The English Multi-TALE architecture

The function of an automatic *tagger* is mainly to perform the first and essential pre-processing for any natural language processing application: labelling words in sentences with their grammatical (sub-)categories, only taking into account a limited context. The output of an automatic tagger is a list of the words of a text with the appropriate labels. This output can be used for further processing, for example by a parser. With *parsing* is meant here performing a complete syntactic analysis of a sentence, distinguishing syntagms (noun phrase, verb phrase, etc.) with their syntactic functions (subject, direct object, main verb, etc.). In addition to this syntactic tagging, the Multi-TALE prototype also provides semantic tags to the texts. More specifically, semantic decoration is provided on the basis of the CEN\TC251model for Surgical Procedures as explained above.

The existing DILEMMA-1 tagger-lemmatizer for English general language is used as a starting-point for the analysis [5]. Its output, a list of lemmatised words with their basic syntactic categories (adj, noun, verb, ...) is then processed by the semantic tagger. This tagger uses a rule-set to combine in a bottom-up approach words and word groups in the sentences to be analysed, to more complex elements, up to the level of a predicate, or one of its arguments. Each rule contains information regarding the syntactic and semantic conditions to be fulfilled by the word or word group in focus, and its left and right neighbours within in a specified distance. For example, the rule:

sc("16start","np".[scope("break","",1)],["np"],join(["prep"]),["np"],[],,"anat","
path","inte","mcr"].[],[],"@1").

specifies that when in a sentence the combination noun phrase - preposition - noun phrase is found, and the second noun phrase denotes an anatomy, pathology, instrument or macroconcept, and the first noun phrase is the first unit in the sentence, then the np-prep-np combination may be aggregated in one unit. The new unit gets the syntactic value of "np", and inherits the semantics of the first np. This rule has as result that in the sentence "The catheter in the third ventricle was removed", "the catheter in the third ventricle" is recognised as being one unit, while in "Complete removal of the tumour could not be realised", no aggregation will occur as "removal" is of type surgical deed, and not anatomy, instrument, pathology, or macro-concept.

Other resources used by the Multi-TALE prototype are a small syntactic lexicon to correct the systematic mistakes of DILEMMA-1 [6], and a semantic lexicon of 2000 words (base-forms) with additional information on case-assignment and usage of prepositions.

As an example, the entry:

```
lex( "cut",    [m("M-14110","path","","noun",[],[]),
                m("P1-01020","surd","","noun",[],[]),
                m("P1-01020","surd","make_looser","verb",
                      [role("do",["path","anat"]),role("io",["anat"]), role("m",["inte"])],
                      [prep("with","m"), prep("from","io")])])
```

specifies that the lemma "cut" may indicate a noun or a verb. As noun, it may denote a pathology or a surgical deed, each having a different SNOMED-code. To the verb-form one can associate a case-frame in which the direct-object candidates are of type anatomy and pathology, indirect object candidates are of type anatomy, and candidates for the mean case are of type instrument. Additionally, the entry specifies that if the preposition "with" is used, it probably refers to the mean, and that "from" refers to the indirect object.

Finally, a heuristic scoring procedure is used to select the best solution when more than one analysis for a given sentence is possible.

For example: the sentence "After insertion of the ventricular catheter, removal of cerebral fluid was performed.", is analysed by Multi-TALE as:

Table 1: typical output of Multi-TALE

Semantic link	Semantic type	Syntactic tag	Rule	Syntagm	Meaning (Snomed)
		prep		After	G-4004
action	install	sg		insertion	P1-05500
		prep		of	
do	inte	detnoun	4	the ventricular catheter	(T-A1600,A-26800)
		comm		,	
action	remove	sg	bac	removal was performed	P1-03000
		prep		of	
do	anat	adjnoun	2	cerebral fluid	(T-A0110,fluid)

4. Methods

Two types of validation have been carried out: one to test the intermediate performance of the system with fine-tuning as primary objective, and a second one to assess the results of the final modifications. Ten surgical reports (138 sentences) from the corpus, five having been used for the development of the syntactic-semantic rule-base (training sample), and five for which this is not the case (testing sample), have been manually validated. Two human experts (physicians) were used as gold standard. Recall (number of entities retrieved and relevant over number of relevant entities in the report) and precision (number of entities retrieved and relevant over number of entities retrieved (only for second validation)) were calculated separately for testing and training sample. Calculations were based on the correct recognition of syntactic entities such as sentences, segments (surface form of the predicates), clauses (surface form of predicate arguments), simple and complex noun phrases and verb phrases, as well as on semantic information (types and semantic links correctly identified). In total 2139 syntactic units (to be mapped into 6 categories) and 857 semantic-contextual entities (to be mapped into 8 categories) were to be retrieved.

5. Results

The next table shows the results for both validations (* indicates test not performed):

Table2 : Validation results

	First Evaluation		Second Evaluation			
	Known	Unknown	Known		Unknown	
	Recall	Recall	Recall	Precision	Recall	Precision
Sentences	100	99	100	100	98.7	100
Segments	88	86	98.0	88.3	98.5	92.3
Clauses	91	80	91.9	92.6	95.2	93.3
Simple NP's	93	85	92.8	100	94.3	100
Complex NP's	79	56	88.6	93.9	86.7	100
VP's	93	85	92.6	98.9	95.2	100
Tot. Syntax	**91**	**82**	**93.3**	**94.4**	**95.7**	**95.7**
Deeds	*	*	96.5	94.3	98.1	97.1
Anatomy	*	*	93.9	95.8	98.4	98.4
Pathology	*	*	100	88.6	94.7	100
Instrument	*	*	87.5	97.2	71.1	96.4
Tot. Sem. Types	**90**	**71**	**94.6**	**94.2**	**93.4**	**97.8**
Action	97	82	96.5	90.1	97.1	93.5
Direct Object	84	69	83.3	88.7	77.4	92.9
Indirect Object	78	61	75.0	80.0	60.0	70.6
Means	68	50	70.6	100	77.8	93.3
Tot. Semantics	**83**	**71**	**91.3**	**92.0**	**89.3**	**94.8**

The first validation revealed an acceptable performance for syntactic tagging (except for complex noun phrases) and semantic type recognition, with only very modest results for case assignment. In addition, recall in the training sample appeared to be much higher than in the test sample. Fine-tuning of the system turned out to be very effective, and led to syntactic recall for the test sample of 95.7% (precision being 95.7% also) and 89.3% for semantic recall with a precision of 94.8%. Semantic labelling still appeared to be more successful than case-assignment (recall 93.4% versus 77.4%, 60.0% and 77.8%, $p < 0,01$).

6. Discussion

With these figures, the English Multi-TALE prototype equals (and mostly outperforms) of comparable systems designed for factual information retrieval from clinical narrative. When the LSP system [7] was used to find 13 important details of asthma management in a total number of 31 discharge summaries (testing set), recall appeared to be 82.1% (92.5% counting only omissions instead of errors) and precision 82.5% (98.6%) [8]. Haug et al. report recall and precision rates of 87% and 95% for detecting clinical findings in 839 chest x-ray reports by using SPRUS [9], and rates of 95% and 94% respectively for the detection of diagnoses [10]. The CAPIS system was able to recall 92% of the relevant physical findings (156 in total in 20 reports on patients with gastro-intestinal bleeding), with a precision of 96% [11]. Improvement of the performance of Multi-TALE was primarily due to (i) enhancing the recognition of complex noun phrases, and (ii) expanding the semantic lexicon. The latter was realised by merely adding the words found in the neuro-surgery section of the SNOMED procedure axis. Instruments are seldom mentioned in that axis, hence recall for that semantic type remained low in the second validation (71.7% versus at least 94.4% for the other Semantic types). Multi-TALE being a tagger and not a parser, it is also obvious that case-assignment (for which a sentence needs to be analysed as a whole) performed less good.

7. Conclusion

From this work, two conclusions can be drawn. First, tagging, as a more simple procedure compared to parsing, may be an effective strategy to extract factual information from clinical narrative. This has convincingly been shown by the English Multi-TALE system in the domain of neurosurgical procedures. Second, the CEN/TC251/PT002S standard for surgical procedure classifications can be used for syntactic-semantic tagging of neurosurgical procedures provided that the concept type *surgical deed* is thought of as predicate primitive and the semantic links *direct object, indirect object and means* as the expression of semantic functions labelled case roles that can be ascribed to the arguments of a predicate (i.e. concept of surgical deed). In this context, Multi-TALE is in line with the World Health Organisation's view in that *the challenge is now to clarify how we can pass from traditional encoding of medical data to automatic encoding of natural language, and how the universally accepted classifications with their advantages and disadvantages can be used in this context* [12].

8. Acknowledgements

The Multi-TALE project was partly funded by the Commission of the European Communities in MLAP-II contract 93-04.

9. References

[1] CEN ENV 1828:1995. Medical Informatics - Structure for classification and coding of surgical procedures.

[2] Dik S. : A Theory of Functional Grammar. Foris. Dordrecht. 1989

[3] Deville G. Modelisation of Task-Oriented Utterances in Man-Machine Dialogue System. Ph.D. Thesis, University of Antwerp. 1989

[4] Deville G, Ceusters W, Mommaerts JL, Martin W. Semantic Analysis of the CEN/TC251/PT002S-model for Surgical Procedures: Applicability and Extendibility. Multi-TALE deliverable D2-2, 25-06-94.

[5] Paulussen, Hans & Willy Martin, *DILEMMA-2: a lemmatizer-tagger for medical abstracts*, in Proceedings of the Third Conference on Applied Natural Language Processing (ACL), Trento, 141-146, (1992).

[6] Mommaerts JL, Ceusters W, Deville G. Are taggers for general language useful for medical sublanguage ? A case-study with DILEMMA (Dutch). In: Beckers WPA, ten Hoopen AJ (eds) Proceedings of MIC'94, Velthoven, The Netherlands, 25-26/11/94, 283-290, 1994.

[7] Sager N, Friedman C, Lyman MS: Medcial Language Processing: Computer Management of Narrative Data. Reading, MA: Addison - Wesley, 1987.

[8] Sager N, Lyman MS, Bucknall C, Nhan N, Tick LJ: Natural language processing and the representation of clinical data. J. Am Med Infomatics Assoc, 1, 142-160, 1994.

[9] Ranum DL. Knowledge based understanding of radiology text. Proceedings of the 12th Annual Symposium on Computer Applications in Medical Care, Washington, DC, 141 - 145, 1988.

[10] Haug PJ, Ranum DL, Frederick PR: Computerized extraction of coded findings from free-text radiologic reports. Radiology, 174, 543 - 548, 1990.

[11] Lin R, Lenert L, Middleton B, Shiffman S. A free-text processing system to capture physical findings: Canonical Phrase Identification System (CAPIS). Proc-Annu-Symp-Comput-Appl-Med-Care. 1991: 843-7

[12] Jardel JP. Opening address for the International Working Conference on Natural Language Processing of the International Medical Informatics Association. In: Scherrer JR, Côté RA, Mandil SH (eds.) Computerized Natural Language Processing for Knowledge Engineering. Elsevier Science Publishers, p 1 (1989).