IJCLA VOL. 4, NO. 2, JUL-DEC 2013, PP. 121–133
RECEIVED 05/12/12 ACCEPTED 04/03/13 FINAL 09/03/13

# Medical Event Extraction using Frame Semantics — Challenges and Opportunities

DIMITRIOS KOKKINAKIS

*University of Gothenburg, Sweden*

ABSTRACT

*The aim of this paper is to present some findings from a study into how a large scale semantic resource, FrameNet, can be applied for event extraction in the (Swedish) biomedical domain. Combining lexical resources with domain specific knowledge provide a powerful modeling mechanism that can be utilized for event extraction and other advanced text mining-related activities. The results, from developing a rule-based approach, showed that only small discrepancies and omissions were found between the semantic descriptions, the corpus data examined and the domain-specific semantics provided by SNOMED CT (medical terminology), NPL (medicinal products) and various semi-automatically developed clue lists (e. g., domain-related abbreviations). Although the described experiment is only based on four different domain-specific frames, the methodology is extendable to the rest ones and there is much room for improvements, for instance by combining rule-based with machine learning techniques, and using more advanced syntactic representations.*

KEYWORDS: *Event extraction; frame semantics; semantic arguments; FrameNet.*

## 1 Introduction

Natural language understanding (NLU), is a subtopic and a long-term goal for Natural Language Processing (NLP), which aims to enable computers to derive meaning from natural language input. NLU systems require a semantic theory to guide the comprehension of any text and at the same time a suitable framework for representing lexical knowledge, preferably linked to domain ontologies and terminologies. In such a context, a semantic-oriented framework could play a vital role for alleviating the extraction of complex semantic relations and, often pre-specified, simple or composite events. Event-based, or event-template information extraction have been initiated by and explored in the MUC-4 extraction task [1]. Since then, extraction and labeling of events has also attracted attention in various activities (e.g. in the SEMEVAL framework [2] and the BioNLP shared tasks [3]). In recent years, algorithms are also developed that try to learn instead template structures automatically from raw text; *cf.* [4]. Here, we are interested in biomedical event extraction, which refers to the task of extracting descriptions of actions and relations among one or more entities from the biomedical literature.

Mining such complex relations and events has gained a growing attention in this domain; [3, 5, 6] and for several reasons. Mainly due to the existence of a publication volume that increases at an exponential level, the availability of mature NLP tools for biomedical text analysis, large lexical/terminological/ontological resources, and various manually annotated samples with semantic information. All these factors have resulted in an explosion of event-related research in the domain (*cf.* <http://nactem.ac.uk/genia/>, <https://www.i2b2.org/>). Semantically driven literature analysis and literature-based knowledge discovery provide a lot of challenging research topics and a paradigm shift is taking place in the biomedical domain, from relation models in information extraction research to more expressive event models, *cf.* [7].

Our approach is closely related to information extraction (IE), a technology that has a direct correlation with frame-like structures as described in the FrameNet. Templates in the context of IE are frame-like structures with slots representing event information. Most event-based IE approaches are designed to identify role fillers that appear as arguments to event verbs or nouns, either explicitly via syntactic relations or implicitly via proximity. In this paper we argue that frame semantics is such a framework that can facilitate the development of text understanding and as such can be used as a backbone to NLU systems. We present results from experiments using domain-specific FrameNet extensions for the automated analysis of meaning in Swedish medical texts. With this approach we aim to develop and apply automatic event extraction in the Swedish medical domain in a large scale and in the long run, we are particularly interested in developing a set of tools to support health care professionals and researchers to rapidly identify, aggregate and semantically exploit relevant information in large textual repositories.

## 2 Theoretical Background

The FrameNet approach is based on the linguistic theory of frame semantics [8] supported by corpus evidence. A semantic frame is a script-like structure of concepts which are linked to the meanings of linguistic units and associated with a specific event, situation or state. Each frame identifies a set of frame elements, which are frame specific semantic participants and roles/arguments (both core and non-core ones). Furthermore, roles may be expressed overtly, left unexpressed or not explicitly linked to the frame via linguistic conventions (null instantiations). In this work, we only deal with the first type of such roles. FrameNet documents the range of semantic and syntactic combinatory possibilities of frame evoking lexical units (LU), phrases and clauses by abstracting away from syntactic differences. A LU can evoke a frame, and its syntactic dependents can fill the frame element slots, in turn, the various semantic types constrain the types of frame element fillers. Since a LU is the pairing of a word with a meaning, each sense of a polysemous word belongs to a different semantic frame, Moreover, since a single frame element can have different grammatical realizations it can enhance the investigation of combinatorial possibilities more precisely than other standard lexical resources such as WordNet.

### 2.1 *The Swedish FrameNet*

The Swedish FrameNet (SweFN++) is a lexical resource under active development, based on the English version of FrameNet constructed by the Berkeley research group. The SweFN++ is available as a free resource at <http://spraakbanken.gu.se/swefn/>. Most of the SweFN frames and frame names correspond to the English ones, with some exceptions, as to the selection of frame elements including definitions and internal relations. Compared to the Berkeley FrameNet, SweFN++ is expanded with information about the domain of the frames, at present the medical and the art domain. Since frame classification is based on general-domain frame semantics, several efforts have been described to domain adaptations even for English [9, 10].

As of November 2012, the SweFN++ covered 754 frames with around 24,000 lexical units, while 30 frames are marked as medically-oriented; [11]. The lexical units are gathered from SALDO, a free Swedish electronic association lexicon [12]. FN facilitates modeling the mapping of form and meaning within these structures in the medical discourse through manual annotation of example sentences and automatic summarization of the resulting annotations. Some of the medical frames in SweFN include: *Addiction*; *Cure*; *Recovery*; *Experience_bodily_harm*; *Falling_Ill*; *Administration_of_medication* etc. For instance, the Cure frame describes a situation involving a number of core roles such as: *Affliction*, *Healer*, *Medication*, *Patient* etc., and a number of non-core roles such as *Degree*, *Manner* and *Time*, and it is evoked by lexical units such as to *cure*, to *heal*, *surgery*, and to *treat*. The word in bold face below evokes the *Cure* frame: "`[Steloperation av fotleden]-TREATMENT {`**`lindrar}`**`-CURE [smärta]-AFFLICTION [väl]-MANNER men medför en del komplikationer`" (litt. 'Lumbar fusion operation of the ankle reduces pain well, but entails some complications').

## 3 Experimental Setting

Our approach uses the annotation results produced from the application of adapted entity and terminology taggers; as a semantic theory the use of specifically designed medical frames, with associated manually annotated textual samples, and, finally, various manually developed frame related regular expression patterns. The domain-specific medical frames we have been using are: *Administration_of_medication*, with core frame elements such as *Drug*, *Patient* and *Medic* (112), *Medical_Treatment*, with core frame elements such as *Treatment*, *Affliction* and *Patient* (102), *Cure*, with core frame elements such as *Healer*, *Affliction* and *Body_Part* (115) and *Falling_Ill*, with core frame elements such as *Patient*, *Symptom* and *Ailment* (116); the figure in parenthesis refers to the number of manually annotated sentences, randomly extracted from a large available Swedish biomedical corpora [13]. All annotated samples are available from the following addresses: http://demo.spraakdata.gu.se/brat/#/[sweFNCure_dk; sweFNMed-Treatment_dk; sweFNFallingIll_dk; sweFNAdminOfMed_dk].

### 3.1 *Relevant Resources*

We have been using a number of relevant resources (textual, terminological, etc.) for modeling pattern matching rules, i.e. complex regular expressions. Some of the most important resources have been used for both extracting relevant text samples and also aiding the recognition of relevant frame elements in the samples. The main source for medical terminology has been the Swedish Nomenclature of Medicine, Clinical Terms (SNOMED CT), since it is the largest available source of medical terminology in Swedish, approx. 300,000 terms. Medication names are provided by the National Repository for Medicinal Products (NPL, <http://www.lakemedelsverket.se>) which is the official Swedish product registry for drugs, approx. 12,000 terms.

Every product in this registry contains information on its substances, names, dosages, producers and classifications like prescription and Anatomical Therapeutic Chemical codes (ATC). Lists of semi-automatic acquired drug / substance / disease lexicon extensions (e.g. generic expressions of drugs and diseases, misspellings etc.); lists of key words (e.g. drug forms [pill, tablet, capsule], drug administration paths [intravenous, intravesical, subcutaneous], volume units [mg, mcg, IE, mmol] and various abbreviations and variants [iv, i.v., im, i.m. sc, s.c., po, p.o., vb, T]). Finally, important pieces of information are also obtained by the application of named entity recognition, which identifies and annotates very important frame elements, particularly time expressions, various types of numerical information (such as dosage and frequency) and some terminology (such as lists of non-official drug names).

### 3.2 *Method*

As a method we apply a rather simple, rule-based approach (which can be used as a baseline for future work using other techniques) by performing three major steps. (i) *pre-processing*, that is selecting a relevant sample of sentences for each frame using trigger words (i.e. relevant LUs) for both manual annotation and pattern development and evaluation, (ii) *main processing*, which includes terminology, named entity and key word/text segment identification, (iii) *post-processing*, e.g., modeling observed frame element patterns as rules (regular expressions). All steps are applied at the sentence level, i.e. no coherent, larger text fragments are used. First, we manually annotated the sentence samples with all possible frame elements. Through the manual analysis of the annotated examples we could obtain a rather good understanding of how the examined medical events can be expressed in the data. This way we can model various rules for the task and also have annotated data for future planned supervised learning extensions. During processing, we first start by identifying and annotating the terminology (e.g. SNOMED CT terms and NPL drug names) or drug name classes (e.g., antibiotics). For the main processing step we apply named entity recognition which identifies and annotates relevant frame elements such as time expressions, various important numerical entity information types, named entities such as person and location and also non-official terminology.

These annotations are important since they are both required by the frames and appear regularly in the context of the medical frames. A number of lexical rules, as previously described, based on e.g. lists of administration paths for drug admission etc., implemented as regular expressions are applied for the recognition and annotation of relevant frame elements. Using as a guidance the order of the extracted element patterns from the annotated sample, we model those as rules. For instance, the most frequent frame element pattern in the *Administration_of_Medication* frame (10 occurrences; 20 combined with other elements) is "<Drug_name> <Drug_strength> <Frequency>", and in the *Falling_Ill* frame (22 occurrences; 46 combined with other elements) is "<Patient> <Ailment>".

An annotated example sentence with named entities, from the *Administration_of_Medication* frame, is shown below, the XML-like labels should be self-explanatory. Here, the entity tagger annotates occurrences of time ("TIMEX/TME"); frequency ("NUMEX/FRQ") and dosage ("NUMEX/DSG"): `Åtta patienter erhöll Recormon före operationen, i dosering 2 000 IE subkutant tre gånger per vecka under tre veckor` (litt. 'Eight patients received Recormon before surgery, dosage 2000 IU subcutaneously three times per week for three weeks') is annotated as `Åtta patienter erhöll Recormon före operationen, i dosering <NUMEX TYPE="MSR" SBT="DSG">2 000 IE</NUMEX> subkutant <NUMEX TYPE="MSR" SBT="FRQ">tre gånger per vecka </NUMEX> <TIMEX TYPE="TME" SBT="DAT"> under tre veckor</TIMEX>`. All labels were normalized to their frame element names at a later stage. For instance, the following example from the *Administration_of_Medication* frame, illustrates an example with normalized frame element labels: `Lugnande besked, rec <Drug_name>Tradil</Drug_name> <Drug_strength> 400 mg</Drug_strength> <Frequency>1 x 1-2 </Frequency>` (litt. 'Reassurance, rec Seractil 400 mg 1 x 1-2').

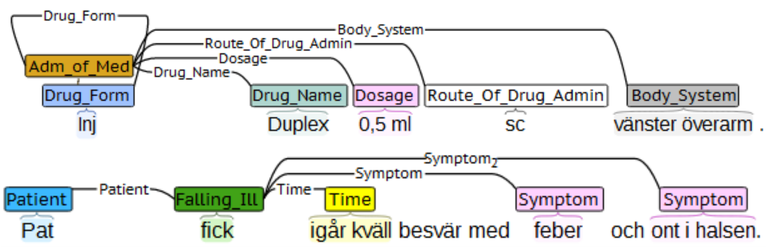


**Fig. 1.** Examples of manually annotated data with the frames *Administration_of_Medication* (top) and the *Falling_Ill* (bottom) using the *brat* annotation tool [14].

## 4 Results and Discussion

Table 1 shows the evaluation results (complete match) for the top 4 frame elements (most occurrences in a test set of 30x4 sentences) for the four examined domain frames. Some of the no-core frame elements could not be found in the sample, while some had very few occurrences and this is the reason we chose not to formally evaluate all of those at this stage. This vertical level evaluation assess the extraction of each frame element individually. A number of problematic issues still remain. For instance, certain elements are difficult to capture using regular expressions, such as <Purpose>, <Outcome> and <Circumstance>. These seem the most problematic since these element shows great variability and expressed by common language patterns. Perhaps syntactic parsing needs to be exploited in such cases because these elements are often described by lengthy, complex noun or prepositional phrases and clauses.

For instance, the following example shows a prepositional phrase complex with four prepositions (in bold face): <Circumstance> **Vid** klart skyldig blindtarmsinflammation **av** varierande grad upp till kraftigare inflammation **med** tecken **på** vävnadsdöd i blindtarmen </Circumstance> administreras antibiotika Tienam 0,5 g x 3 (litt. 'In clear-cut case appendicitis of varying degree up to stronger inflammation with signs of necrosis in the cecum antibiotic Tienam 0.5 g x 3 is administered'). Another problematic aspect is observed for many cases where there is an ellipsis, that is, clauses where an overt trigger word is missing (often a predicate belonging to the frame). For instance, the following example shows such an ellipsis, lack of an overt trigger, a verb, in the last clause marked in italic: Av journalblad framgår att han behandlats med digitalis , såväl i injektion som per os, *samt med kinidin tabletter*. (litt. Of the record sheet it is shown that he has been treated with digitalis, both injection and per os, and with quinidine tablets.)

**Table 1.** Evaluation of the most frequent frame elements in the test sample.

| Frame | Frame elements | | | |
|---|---|---|---|---|
| Admin. of_Medic. | Drug_Name<br>92,6%(Pr)<br>81,2%(R) | Dosage<br>96%(Pr)<br>90,1%(R) | Frequency<br>98,7%(Pr)<br>91,9%(R) | Route_Of_Drug_Admin<br>100%(Pr)<br>97,1%(R) |
| Cure | Affliction<br>94%(Pr)<br>92,9%(R) | Treatment<br>83,1%(Pr)<br>79,2%(R) | Patient<br>100%(Pr)<br>100%(R) | Medication<br>94%(Pr)<br>89,2%(R) |
| Falling_Ill | Patient<br>100%(Pr)<br>95%(R) | Ailment<br>88,9%(Pr)<br>91,1%(R) | Symptom<br>78,9%(Pr)<br>83.4%(R) | Time<br>100%(Pr)<br>100%(R) |
| Medical _Treatment | Patient<br>100%(Pr)<br>100%(R) | Affliction<br>93,2%(Pr)<br>91%(R) | Medication<br>97,9%(Pr)<br>95%(R) | Time<br>100%(Pr)<br>100%(R) |

In Table 1, Precision measures the amount of elements correctly labeled, out of the total number of all elements labeled by the rules; while Recall measures the amount of elements correctly labeled given all of the possible elements in the sample. The evaluation results are based on sentences for each frame that were annotated separately from the annotated sample used for the creation of the pattern matching rules (these sentences were annotated and evaluated by the author). Nevertheless, it should have been advantageous if (trained) experts, e.g. physicians, could annotate the test data but that was prohibitive at the moment, but will be considered in future, larger scale evaluations and method combinations.

As previously discussed, some of the frame elements could not be found in the annotated samples, while some had very few occurrences and were not formally evaluated, for instance the element *Place* in the *Falling_Ill* frame. Moreover, the manual annotation gave us the opportunity to revise some of the frame elements and in a revised version of the frames in SweFN++, some of the domain frames will be divided in two. Thus in order to get even more accurate and precise semantics (arguments) some frames would require more *specialization*. For instance, the *Administration_of_medication* would be required to be divided between *Administration_of_medication_conveyance* (where the procedures that describe the administration of medicine will be the focus of the frame; e.g. `Normalt ska en salva eller kräm strykas på tunt`; litt. "Normally, an ointment or cream will be thinly applied") and *Administration_of_medication_specification* (where the focus should be on the specifications concerning administration of medicines; e.g. `Tegretol 20 mg/ml, 30 ml x 1`).

## 5 Conclusion and Future Work

We have presented a set of experiments using a rule-based approach on automatic semantic role labeling, and in particular event-based information extraction, using frame semantics modeled in the Swedish FrameNet. We have investigated the use and efficacy of a rule-based approach for the recognition and labeling of the semantic elements, on a specialized textual domain, namely biomedicine. So far we have been working with four different frames and experimenting with simple pattern matching approaches in order to use as a baseline for future experiments. The driving force for the experiments is the theory of frame semantics, which allows us to work with a holistic and detailed semantic event description than it has been previously reported in similar tasks or in efforts using, for instance, most traditional methods based on relation extraction. Moreover, event extraction is more complicated and challenging than relation extraction since events usually have internal structure involving several entities as participants allowing a detailed representation of more complex statements.

Due to the small amount of labeled data, we have not yet attempted to apply a machine learning approach, since such as classifier would suffer from feature sparsity. However, annotating sentences is very time-consuming and we will thus have to live with small training sets for the near future. Still, this problem can be addressed in several ways; for instance through the use of cross-frame label generalization and by adding cluster-based features. In a similar fashion, Johansson et al. [15] have shown that such methods result in clear performance improvements. This way, traditional, lexicalized approaches may lead into other research paradigms, such as semi-supervised approaches [16] and the inclusion of automatically produced training data [17]. In the near future we intend to investigate the validity of the medical frames by manually annotating authentic samples for all available medical frames and also combine the pattern-based approaches with supervised learning for automatic extraction and labeling of frame elements. Note, however that we have observed that in some cases/frames, such as *Administration_of_Medication,* simple means implemented as regular expressions are enough for accurate identification of frame elements, since such a frame contains a plethora of numerical information and domain-specific abbreviations and acronyms that require less advanced techniques in order to obtain good coverage. In other cases, such as in the *Cure* frame, other means seem more appropriate, such as syntactic parsing.

Event recognition at the moment is performed at a sentence level using a nearly homogeneous corpus of biomedical Swedish and also overuse of trigger words. One of the future challenges is of course to treat the problem of event detection as a classification one where one could strive to rely less on the presence of such trigger words. On the other side rule-based methods on domain-specific events and frames with a limited set of vocabulary (lexical units) can be as efficient or even outperform classification accuracy. Moreover, it has been shown that the most effective classification approach is dependent on the target event type [18]. Events that can be described by a large set of lexical units (many synonymous, near-synonymous etc.) are more suitable for training purposes and thus more efficient using a classification approach, while for events using a limited set of vocabulary a triggers' based classification system produces better results. Therefore, in the future, we plan to compare which technique is most appropriate for which type of frame.

ACKNOWLEDGMENTS This work was supported by the Centre for Language Technology, http://clt.gu.se/, at the University of Gothenburg and the Swedish Research Council project: *Swedish FramNet++ — a lexical infrastructure for Swedish language technology* (2011–13).

## References

1. Rau, L., Krupka, G., Jacobs, P., Sider, I., Childs, L.: MUC-4 test results and analysis. 4th Message Understanding Conf. (MUC-4), 1992.
2. Ruppenhofer, J. et al.: SemEval-2010 Task 10: Linking Events and Their Participants in Discourse. The NAACL-HLT 2009 Workshop on Semantic Evaluations: Recent Achievements and Future Directions. Colorado, USA, 2010.

3. Kim, JD., Pyysalo, S., Ohta, T., Bossy, R., Tsujii, J: Overview of BioNLP Shared Task 2011. Proceedings of the BioNLP 2011 Workshop Companion Volume for Shared Task 1–6, 2011.
4. Chambers, N., Jurafsky, D.: Template-based information extraction without the templates. Proc of the 49th Annual Meeting of ACL: HLT. 976–986. Oregon, USA, 2011.
5. Ananiadou, S., Pyysalo, S., Tsujii, J., Kell, DB.: Event extraction for systems biology by text mining the literature. Trends Biotechnol. 28(7):381–90, 2010.
6. Miwa, M., Thompson, P., McNaught, J., Kell, DB., Ananiadou, S.: Extracting semantically enriched events from biomedical literature. BMC Bioinformatics, 13:108, 2012.
7. Björne, J., Ginter, F., Pyysalo, S., Tsujii, J., Salakoski, T.: Complex event extraction at PubMed scale. Bioinformatics 15;26(12):i382-90, 2010.
8. Fillmore, CJ., Johnson, CR., Petruck, MRL.: Background to FrameNet. J of Lexicography. 16(3), 2003.
9. Dolbey, A., Ellsworth, M., Scheffczyk, J.: BioFrameNet: A Domain-Specific FrameNet Extension with Links to Biomedical Ontologies. 2nd Workshop on Formal Biomed. Knowledge Repres. (KR-MED). Baltimore, USA, 2006.
10. Tan, H., Kaliyaperumal, R., Benis, N.: Ontology-driven Construction of Corpus with Frame Semantics Annotations. Proceedings of the 13th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing). New Delhi, India, 2012.
11. Borin, L., Dannélls, D., Forsberg, M., Toporowska Gronostaj, M., Kokkinakis, D.: The past meets the present in the Swedish FrameNet++. In Proceedings of EURALEX, 2010.
12. Borin, L.: Med Zipf mot framtiden - en integrerad lexikonresurs för svensk språkteknologi. LexicoNordica, 17 (In Swedish), 2010.
13. Kokkinakis, D.: The Journal of the Swedish Medical Association - a Corpus Resource for Biomedical Text Mining in Swedish. The Third Workshop on Building and Evaluating Resources for Biomedical Text Mining (BioTxtM), an LREC Workshop. Turkey, 2012.
14. Stenetorp, P., Pyysalo, S., Topic, G., Ohta, T., Ananiadou, S., Tsujii, J.: brat: a Web-based Tool for NLP-Assisted Text Annotation. Proc of the Eur. ACL. 102–107. Avignon, France, 2012.
15. Johansson, R., Friberg Heppin, K., Kokkinakis, D.: Semantic Role Labeling with the Swedish FrameNet. Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC). 3697–3700. Istanbul, Turkey, 2012.
16. Fürstenau, H., Lapata, M.: Semisupervised semantic role labeling. In Proceedings of the 12th Conference of the European Chapter of the ACL (EACL). 220–228, Athens, Greece, 2009.
17. Johansson, R., Nugues, P.: A FrameNet-based semantic role labeler for Swedish. In Proceedings of Coling/ACL, Sydney, Australia, 2006, 436–443.

18. Naughton, M., Stokes, N., Carthy, J.: Sentence-Level Event Classification in Unstructured Texts. Tech. Report UCD-CSI-2008-07. School of Computer Science and Informatics, University College Dublin, Ireland, 2008.

## Appendix

The *Falling_Ill* frame. Domain [*domän*]: Medicine; Semantic Type [*semantisk typ*]: Change_of_State; Core Elements [*kärnelement*]; Non-Core Elements [*periferielement*]; Examples [*exempel*]; Lexical units [*saldo*]; Comments [*kommentar*].

### Falling_ill

| | |
|---|---|
| **ram** | **Falling_ill** |
| **domän** | Med |
| **semantisk typ** | Change_of_state |
| **kärnelement** | **Ailment**, **Patient**, **Symptom** |
| **periferielement** | **Cause**, **Circumstance**, **Degree**, **Depictive**, **Manner**, **Modal_polarity**, **Name**, **Outcome**, **Place**, **Time** |
| **exempel** | Kim insjuknade i influensa igår.<br>Eva drabbades av cancer.<br>Hon gick in i depression.<br>Hon fick hög feber och frossa.<br>På väg till toaletten föll hon och ådrog sig en komplicerad femurfraktur.<br>Fler och fler drabbas av Alzheimers sjukdom.<br>Han vet att Agda själv och Agdas närmaste kan bli drabbade av HIV, aids, autism, Parkinson, Alzheimer och så vidare.<br>Efter några dagar fick pojken ont i halsen, feber och svårigheter att svälja.<br>Far insjuknade i hjärtinfarkt i 42 års ålder.<br>Röntgenundersökning visade att han ådragit sig en subtrokantär dislokerad femurfraktur, som fixerades med spik vid en operation samma dag.<br>Patienten försämrades hastigt, fick kramper och avled 4 månader efter symtomdebut.<br>Har stora besvär med ryggsmärtor o har också smärtor i hö knäled där han har utvecklat en artros efter en knäfraktur på 80-talet.<br>Kanske skulle Fredrika till och med sluppit gå in i en depression.<br>Distriktsläkaren misstänkte att flickan drabbats av borrelia, vilket var sannolikt. |
| **sms-exempel** | |
| **saldo** | **vb:** insjukna..1 sjukna..1<br>**nn:** demensutveckling..1 insjuknande..1 utvecklande..1 utveckling..1 |
| **saldo (nya)** | **vb:** få..4 utveckla..2<br>**vbm:** dra_på_sig..2 drabba..2 gå_in_i..2 ådra_sig..1 åka_på..1 |
| **kommentar** | Ny ram, en pendang till Recovery ramen.; utveckla (sjukdom), gå in i depression, dra på sig en förkylning;; OBS: få..1 fungerar som ett support-verb (SUPP), snarare än en äkta semantisk LU, och bildar en kollokation med det substantiv som syftar på en sjukdom eller symtom. Övriga roller (FEs) taggas med hänsyn till det semantiskt bärande ordet. Explicit uppmärkning av LU/SUPP bör övervägas.;; SALDO drabba..2 [P Fler och fler] [LU drabbas] [A av [N Alzheimers] sjukdom].Han vet att [P Agda själv och Agdas närmaste] kan bli [LU drabbade] [A av HIV], [A aids], [A autism], [A/N Parkinson], [A/N Alzheimer] och så vidare. |

**DIMITRIOS KOKKINAKIS**
CENTRE FOR LANGUAGE TECHNOLOGY
AND THE SWEDISH LANGUAGE BANK,
UNIVERSITY OF GOTHENBURG,
SWEDEN
E-MAIL: <DIMITRIOS.KOKKINAKIS@SVENSKA.GU.SE>