

## Extending the NegEx Lexicon for Multiple Languages

Wendy W. Chapman<sup>a</sup>, Dieter Hillert<sup>a</sup>, Sumithra Velupillai<sup>b</sup>, Maria Kvist<sup>c</sup>, Maria Skeppstedt<sup>b</sup>, Brian E. Chapman<sup>a</sup>, Mike Conway<sup>a</sup>, Melissa Tharp<sup>a</sup>, Danielle L. Mowery<sup>a</sup>, Louise Deleger<sup>d</sup>

<sup>a</sup>Division of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA

<sup>b</sup>Dept. of Computer and Systems Sciences (DSV), Stockholm University, Stockholm, Sweden

<sup>c</sup>Dept. of clinical immunology and transfusion medicine, Karolinska University Hospital, Sweden

<sup>d</sup>The Computer Sciences Laboratory for Mechanics and Engineering Sciences, CNRS, F-91043 Orsay, France

### Abstract

We translated an existing English negation lexicon (NegEx) to Swedish, French, and German and compared the lexicon on corpora from each language. We observed Zipf's law for all languages, i.e., a few phrases occur a large number of times, and a large number of phrases occur fewer times. Negation triggers "no" and "not" were common for all languages; however, other triggers varied considerably. The lexicon is available in OWL and RDF format and can be extended to other languages. We discuss the challenges in translating negation triggers to other languages and issues in representing multilingual lexical knowledge.

### Keywords:

Natural language processing, knowledge representation.

### Introduction

One study estimated that approximately half of all clinical conditions described in narrative reports are negated [1], indicating the importance of differentiating conditions that are present from those that are absent in natural language processing (NLP) applications. A variety of negation algorithms have been developed and evaluated on English texts [2-5]. Fewer algorithms have been developed for other languages. However, the NegEx algorithm has been ported, applied, and evaluated on a few European languages [6, 7].

Our objective was to compile a shareable lexicon for NegEx that enables lexical representations in other languages. In this paper, we describe the process we applied and the challenges we encountered in translating NegEx lexical cues into three European languages (Swedish, French, and German), as well as the knowledge representation format we chose to use for the NegEx Multilingual Lexicon. We also present an initial comparison of the lexicon on clinical corpora of the four languages.

### Background

#### Negation Detection

Recently, negation identification has been the focus of shared tasks including the 2010 i2B2/VA Challenge for clinical text [8], CoNLL 2010 for biomedical texts [9], and BioNLP 2009 for biological texts [10]. Many negation algorithms have shown high performance at identifying negation of clinical conditions in English text [2-5]. Regardless of the algorithm for assigning negation status to a concept, lexical cues for negation like "no" and "without" are a critical ingredient in determining whether a patient suffers from a clinical condition. One negation algorithm, NegEx [2], provides a rich list of

lexical cues that indicate negation in clinical text, and that lexicon has been leveraged in a variety of both rule-based and supervised machine learning applications for English clinical text [11]. NegEx has also been integrated into open-source information extraction systems such as *MetaMap* [12], *cTAKES* [13], and *HITex* [14].

#### The NegEx Algorithm

The NegEx algorithm relies on three types of lexical cues: negation triggers that indicate a negation (e.g., "denies"), pseudo-negation triggers that contain negation triggers but do not negate the clinical condition (e.g., "no increase"), and termination terms that stop the scope of the negation trigger (e.g., "but"). Any clinical condition within the scope of a negation trigger is negated. All NegEx lexical items have an action—negation and pseudo-negation triggers can modify information to the right of the term (i.e., forward in the sentence) or to the left of the term (i.e., backward in the sentence). Termination terms have an action to terminate scope, which otherwise terminates at the end of the sentence. Pseudo-negation triggers attempt to compensate for NegEx's lack of syntax by listing exceptions to the occurrence of negation triggers such as "no" in phrases like "no previous". Because the lexicon is the keystone of the algorithm, adaptation of NegEx to other languages relies mainly on translation of the lexical cues. Our goal was to develop a representation of the lexicon that facilitates translation to other languages.

#### NegEx performance on other languages

NegEx has been recently ported and evaluated on clinical texts in Swedish [6] and French [7] and has shown good performance (recall 82%; precision 75%) for Swedish assessment sections of the Stockholm EPR corpus and better performance (recall 85%; precision 89%) for French cardiology notes. In both studies, the adapted NegEx systems achieved comparable recall to the English NegEx with observable differences in precision (differences of -9.3% and 4.4%, respectively). Error analyses from these studies suggest that increasing lexicon coverage, improving scope detection, and including uncertainty assertion could boost the algorithm's porting performance. Although not evaluated yet, the NegEx lexicon has also been translated to German and is included in our lexical analysis.

#### Creating multilingual lexical knowledge resources

Formal ontologies can be distinguished from lexical or terminological ontologies [15,16]. Unlike formal ontologies, which focus on axioms and logically defined relations between concepts, lexical ontologies are concerned with the various ways concepts can be instantiated in language.

### Existing models for representing lexical ontologies

The Simple Knowledge Organization System, SKOS, provides a mechanism for representing lexical information using OWL and RDF vocabularies that can be used to represent multilingual lexical ontologies [17]. SKOS supports the construction of a taxonomic hierarchy using *broader* and *narrower* relations and a built-in mechanism for representing *prefLabels*, *altLabels*, and *hiddenLabels* (preferred terms, alternative terms, and misspellings, respectively). A key feature of SKOS is that it supports the encoding of multilingual terms using the “@” syntax (e.g. @en for English terms). Like SKOS, LexInfo is an RDF model that facilitates modeling of multilingual lexicons. However, unlike SKOS, LexInfo provides the means to model linguistic information associated with lexical items (e.g., part-of-speech and morphological information) [18].

### Multilingual lexical ontologies

EuroWordNet [19] is a multilingual lexical ontology that was developed under the auspices of the European Union to support machine translation between eight European languages (Dutch, Spanish, Italian, English, French, German, Czech & Estonian). Like the original Princeton WordNet [20], the basic organizing unit of EuroWordNet is the synset (a group of interchangeable synonyms terms that represent a concept in a given context). A key feature of EuroWordNet is the use of an Inter-Lingual Index as a means of relating synsets in different languages, a technique also used by the Biocaster ontology, a multilingual infectious disease ontology designed to support public health surveillance [21]. In the medical domain, Markó et al. [22] developed a comprehensive multilingual lexicon as part of a multi-site EU funded project.

## Materials and Methods

In this paper, we describe and discuss (1) the processes used to translate the NegEx lexicon to the three additional languages, (2) our approach for developing a representation format that facilitates cross-linguistic application of NegEx while retaining the simplicity that makes NegEx appealing and usable, and (3) an initial analysis and comparison of the NegEx lexicon on clinical corpora in English, Swedish, French, and German.

**(1) Translation** - Translations were performed by research groups that included clinicians and informaticists/linguists. French and Swedish translations have been validated to some extent through evaluation on clinical corpora.

**(2) Knowledge Representation** - Because the NegEx algorithm requires not only a multilingual lexicon but also encoded regular expressions, information about an action (e.g., forward), and other information not available in existing models like SKOS and LexInfo, we decided to develop a bespoke OWL representation incorporating elements of existing RDF models. See Figure 1 for a diagram representing some of the relations for the concept “cannot be excluded”.

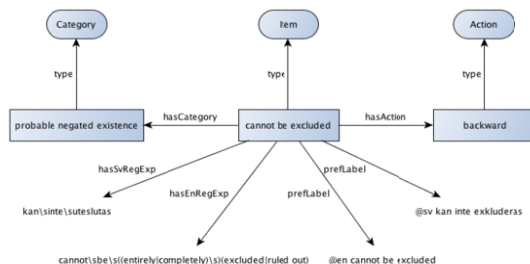


Figure 1 – Partial concept for “cannot be excluded” in the Multilingual NegEx Lexicon, with Swedish translation

**(3) Lexical Analysis** - We counted the frequency of the three types of negation triggers in clinical corpora of four languages (see Table 1) and compared them across languages. Only the longest phrase was kept in cases of overlapping triggers (e.g., “no” and “no increase”). For the German corpus, stemming was applied before string matching. We ranked triggers by frequency in the corpus and defined frequently occurring triggers as those that comprise 80% of all trigger instances.

Table 1 – Corpora used for lexical analysis

Language	Tokens	Report Types
English	1 137 035	Radiology, history & physical exam, and emergency department reports
French	6 781 411	Discharge summaries, consultation reports, and surgical reports from a cardiology unit
German	100 150	Freiburg Annotated MEDical corpus (FRAMED): discharge summaries, pathology, histology, and surgery reports [23]
Swedish	4 644 850	The Stockholm EPR Corpus: Assessment entries from all clinic types [24]

## Results

The OWL and RDF versions of the Multilingual NegEx Lexicon can be downloaded from the NegEx Google Code site (<http://code.google.com/p/negex/downloads/list>). In addition to negation triggers, the lexicon contains triggers for identifying whether a clinical condition is historical, experienced by a family member, mentioned as the indication for an exam, or mentioned in a general or conditional context based on the extensions of NegEx in the ConText algorithm [25].

Figure 2 shows pie charts for the unique triggers in each corpus. Black slices indicate the most frequently occurring triggers (i.e., terms comprising 80% of trigger counts in the corpus); dark grey slices indicate the least frequently occurring triggers (i.e., terms comprising the remaining 20% of trigger counts), and light grey slices indicate terms that did not occur in the corpus. The German corpus showed the lowest prevalence of negation triggers overall.

Tables 1, 2, and 3 show the counts of negation triggers for each corpus. The count of unique trigger terms for each lan-

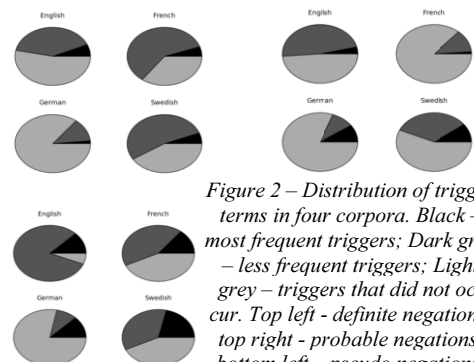


Figure 2 – Distribution of trigger terms in four corpora. Black – most frequent triggers; Dark grey – less frequent triggers; Light grey – triggers that did not occur. Top left – definite negations, top right – probable negations, bottom left – pseudo negations.

guage is shown next to the language name in the header. Counts of unique triggers for definite negation vary quite a bit among languages based on the amount of inflection that was generated. Also, triggers were translated at different time periods over the last year and were thus mostly overlapping but not identical. Row two of each table shows the number of frequently occurring unique terms in the corpus (indicated in the table by >), infrequently occurring unique terms in the corpus (indicated by <) and terms that did not occur in the corpus

(indicated by -). These numbers are reflected in the three slices of the pie charts in Figure 2. Subsequent rows list each of the frequently occurring triggers and the corresponding proportion of trigger counts in the corpus. Empty cells indicate that the concept was not one of the most frequent triggers in the corpus for that language. Two phrases in a single cell indicate that there is more than one translation for the concept. The number of unique triggers found in the corpora was much larger than the number of triggers summing to 80% coverage that are shown in the table's cells. Similar to previous findings [1], Zipf's law holds true for negation triggers such that a few phrases occur a large number of times, and a large number of phrases occur with fewer counts.

Table 2- Definite negated term counts and their occurrences in corpora<sup>^</sup>

English (60)			French (111)			German (125)			Swedish (83)		
>	<	-	>	<	-	>	<	-	>	<	-
4	24	32	7	66	39	2	16	107	5	44	34
no (0.40)			aucune (0.04) aucun (0.04)			kein (0.14)			ingen (0.15) inga (0.09)		
not (0.08)			pas de (0.17) pas d' (0.12) n'a pas (0.05)			nicht (0.70)			inte (0.31) ej (0.18)		
without (0.05)			sans (0.31)						utan (0.10)		
denies (0.28)											
<i>neither</i> F			ni (0.08)								

<sup>^</sup> Italicized English words did not occur in the English corpus but in corpus indicated by F (French), G (German), or S (Swedish). Counts in header indicate the number of unique terms in the lexicon for that language; Shaded row shows the number of frequently occurring unique terms in the corpus (>), infrequently occurring unique terms in the corpus (<) and terms that did not occur in the corpus (-).

Table 3- Probable negated term counts and their occurrences in corpora

English (78)			French* (81)			German (124)			Swedish (97)		
>	<	-	>	<	-	>	<	-	>	<	-
3	37	38	1	10	70	12	12	100	10	32	55
not appreciate (0.59)											
to exclude (0.10)			éliminer (0.83)								
nothing (0.20)											
<i>unlikely</i> G						unwahrscheinlich (0.06)			osannolikt (0.07)		
<i>not provable</i> G <i>cannot see</i> S <i>not obvious</i> S <i>no ground</i> S						nicht nachweisbar (0.09) nicht nachgewies (0.02)			kan inte se (0.04) ingen uppenbar (0.03) inga hållpunkter (0.30)		
<i>not known</i> G <i>no sign of</i> S						nicht bekannt (0.03)			utan tecken på (0.04)		
<i>rather</i> G						eher (0.14)					
<i>doubtful</i> G						fraglich (0.18)					
<i>absence of</i> G						fehl von (0.03)					
<i>no pain</i> G						keine beschwerd (0.03)					
<i>not think</i> S									tycker inte (0.12)		
<i>not to be</i> S									inte vara (0.09)		

<i>not a sure</i> S					ingen säker(0.05)
<i>no sure</i> S					inga säkra (0.05)
<i>cannot rule out</i> S					kan inte utslutas (0.03)
<i>no indication of/for</i> G				kein hinweis auf (0.14) kein hinweis für (0.05)	
<i>no definite</i> G				kein definitiv (0.02)	
<i>no clear</i> G				kein eindeutig (0.02)	

\* French translations were generated from an earlier version of the lexicon that did not contain as many probable negation triggers.

Table 4- Pseudo negated term counts and their occurrences in corpora

English (16)			French (14)			German (32)			Swedish (14)		
>	<	-	>	<	-	>	<	-	>	<	-
2	13	1	2	6	6	4	3	25	3	5	6
without difficulty (0.79)			sans difficulté (0.55)			problemlos (0.35)					
no change (0.20)						keine Veränder (0.18)			ingen förändring (0.54)		
<i>no increase</i> F,G			pas d'augmentation (0.30)								
<i>not only</i> S						nicht nur (0.12)			inte bara (0.24)		
<i>no irregularities</i> G											
<i>not necessarily</i> G						nicht unbedingt (0.18)					
<i>gram negative</i> S									gramnegativ (0.07)		

French showed the most diversity in its triggers: seven triggers made up 80% of all negation triggers in the corpus, and two-thirds of the triggers occurred in the corpus. German showed the least diversity: two triggers comprised 85% of negation triggers in the corpus, and only 18 of 135 triggers (13%) occurred in the corpus<sup>1</sup>. The trigger “denies” only occurred in English, and the trigger “neither” was only frequent in French.

Low frequencies of unique probable negated triggers for French (11/80) are probably due to the fact that the translations occurred before the lexicon was enhanced with additional probable negation triggers (July 2012). English showed very little diversity in its frequently occurring triggers (three triggers made up 99% of all probable negated triggers in the corpus). There was very little overlap in probable negated triggers among the languages.

## Discussion

We measured the prevalence of English NegEx trigger terms translated into French, German, and Swedish across clinical corpora. The most common negation triggers, “no” and “not”, were frequent in all four languages, but beyond those two terms, each language showed a lot of variation in the types of triggers that occurred. Probable negation triggers showed the most variation with a large number of unique triggers in English and Swedish and fewer in German, and with little overlap among the languages. Because NegEx does not use syntax to determine scope, pseudo-negation triggers are necessary to avoid false positives. Pseudo-negation triggers reflected the

<sup>1</sup> The smaller corpus size for German could be partly responsible.

prevalence of “no”, “without”, and “not” in phrases where the negation scope extends only to the following noun or adverb (e.g., “without difficulty”).

### Knowledge Representation

The English NegEx lexicon comprises a text file with one string per line that has been used in rule-based and machine learning negation applications. Extending the lexicon to multiple languages necessitated devising a more complex representation. In representing the lexicon, we wanted to maintain the simplicity that encourages wide use of the lexicon, understanding there would be a cost in representational power. The simplest option would be to create a text file for strings in each language without cross-mapping among languages.

Another guiding principle was to follow best practices for multilingual lexical representation. Best practices dictate representing each trigger as a concept that can have a variety of different labels, including synonyms, misspellings, alternate phrasing, and manifestation in other languages, rather than anchoring the term to a specific language. This approach also allows us to model the relationship between a type of trigger (e.g., definite negated trigger) and the class of terms that terminate the scope of that trigger, which is necessary when applying the lexicon to other tasks, such as assigning the experiencer (patient or family member). For instance, termination terms such as “but” and “however” terminate the scope of definite and probable negation triggers, and we have assigned these terms to a class called *conjunctions*. The *conjunctions* class does not terminate the scope of experiencer triggers like “father” and “family history”. Instead, terms in the *patient* or *presenting* class terminate experiencer triggers. The OWL representation can be output in RDF, which is not as simple to process as a simple text file, but we hope it meets our target balance of representational power and simplicity. We plan to host a version of the lexicon in Web Protégé<sup>2</sup>—a cloud-based OWL ontology editor—for collaborative updates and enhancements of this community resource.

### Translating NegEx Triggers

We encountered a number of interesting issues when translating the English triggers to other languages.

- **Agglutination** – Fusional languages like German or languages with highly agglutinative features in their lexicon can add a bounded morpheme representing negation to a word, such as “diabetesfri” (diabetes+free) in Swedish or “problemlos” (problem+without) in German. Neither the NegEx algorithm nor the lexicon address this.
- **Inflections** – English is a weakly inflected language. Translating negation triggers to languages with more morpho-syntactic variation requires a number of linguistic variants for a single concept. For example, “no” can be translated in French with four different variants (“aucun” / “aucune” / “aucuns” / “aucunes”) due to gender and number agreement.
- **Ambiguity** – Some triggers in English exhibit more ambiguous usage in other languages. For instance, the Swedish translation for “without” is “utan”, which is also used as a conjunction (but). Disambiguation requires context from the sentence and is not modeled in NegEx’s lexicon or algorithm, causing false positive negations [26].
- **Language-specific terminology and usage** – Some translations from the English, although linguistically correct, are rarely or never used in other languages, which explains the large number of negation triggers that never occur in a corpus. Some triggers, like “rule the patient out”, occurred in

the English corpus but not in corpora for the other three languages. However, variations could occur. For example, whereas in English a report may say that the clinician will rule the patient out for a disease, in Swedish, the clinician would rule out the disorder for the patient.

- **Use of prepositions** – Scope in the initial version of NegEx extended for only five terms, whereas scope in the current version is extended to the end of the sentence unless encountering a termination term. One remnant of the previous scope rule is the inclusion of a variety of prepositions at the end of negation triggers to avoid consuming the window for potential content terms, as in “rule out against” and “rule out for”. Such overuse of prepositions unnecessarily complicates translation to other languages.
- **Word order** – NegEx’s approach is extremely dependent on word order in both its lexical triggers and its direction of scope (forward, backward, or bidirectional). Word order in Swedish, for example, is more flexible than it is in English. For instance, the trigger “does not have” could be translated to “har inte” or “inte har”. We therefore included variant word order when needed in translations. In addition, we sometimes had to modify the direction of scope after creating multilingual translations for a concept. For example, the concept for “declines” had a forward scope in English but required a bidirectional scope in German. The scope direction (action) is a property of the trigger concept in the OWL representation and can be different for each language (e.g., a term can have a German action and an English action).
- **Abbreviations** – A few of the English negation triggers are not easily translated, e.g., “-ve” and “neg”.

### Implications for NegEx Lexicon

Translating the NegEx lexicon to European languages has proven to be mostly straightforward and has demonstrated comparable negation performance in French and Swedish. Translation is a tedious task with hundreds of trigger terms, many with multiple variations. The strikingly small ratio of frequent triggers (black slices in Figure 2) to unseen triggers (light grey) suggests that the lexicon could be substantially reduced before translation without a large effect on performance. It remains to be seen whether the lexicon can be effectively translated and applied for non-European languages.

### Limitations

The main limitations in our study relate to the preliminary analysis we performed across corpora of different languages. Each corpus differed greatly in size and in character. The German corpus exhibited different patterns of prevalence for negation triggers, which could have been partly due to its substantially smaller size. This was an opportunistic, retrospective study that did not allow us to control for similarity of the corpora. Perhaps more important than size differences was the difference in the report genres in the corpora. Previous studies have shown that negation triggers differ significantly across report types [1, 27] (e.g., negation triggers in radiology reports differ from triggers in emergency department reports). Hidden variation due to report genre could explain some of the variation among frequently occurring triggers across languages. With limited accessibility of clinical reports and differences across countries and even individual hospitals, a study that matches report genres across languages may not be realistic, making this comparative study worthwhile in spite of the limitation. Finally, our analysis did not address the amount of actual negation in the corpora but only the presence of strings in the negation lexicon as a surrogate; we do not make any claims about accuracy of the lexicon across languages.

<sup>2</sup> <http://webprotege.stanford.edu/>

## Future Work

We are building a web application called pyConTextKit for experimenting with the algorithm and lexicon. The user interface allows lexicon editing and real-time evaluation on a corpus. We plan to extend the capability of the interface to enable user extensions of the lexicon to other languages, including functions for generating machine translations and for retrieval of sentences containing lexical items in a relevant corpus for quick validity checking of proposed triggers.

## Conclusion

We translated the NegEx lexicon from English to French, German, and Swedish, modeled the lexicon in OWL, and performed a preliminary analysis of multilingual negation triggers. Lexical cues varied greatly across languages as did the number of unique triggers found in the corpora. We hope others will continue to extend the lexicon and experiment with its usefulness for detecting clinical negation.

## Acknowledgments

We thank Udo Hahn for counts on the FRAMED corpus and Kevin Zogg for German translations. We acknowledge IR01LM010964, U54HL108460, and the Stockholm University Academic Initiative. Research approved by Regional Ethical Review Board in Stockholm permission 2012/834-31/5.

## References

- [1] Chapman W, Bridewell W, Hanbury P, Cooper G, and Buchanan B. Evaluation of Negation Phrases in Narrative Clinical Reports. AMIA Annu Symp 2001: 105-109.
- [2] Chapman WW, Bridewell W, Hanbury P, Cooper GF, and Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001: 34 (5): 301-310.
- [3] Mutalik P, Deshpande A, Nadkarni P. Use of general-purpose negation detection to augment concept indexing of medical documents a quantitative study using the umls. J Am Med Inform Assoc 2001: 8 (6): 598-609.
- [4] Aronow D, Feng F, Croft WB. Ad-hoc classification of radiology reports. J Am Med Inform Assoc 1999: 393-411.
- [5] Elkin PL, Brown SH, Bauer BA, Husser CS, Carruth W, Bergstrom LR, Wahner-Roedler DL. A controlled trial of automated classification of negation from clinical notes. BMC Med Inform Decis Mak 2005: 5 (13).
- [6] Skeppstedt M. Negation detection in swedish clinical text: An adaption of negex to swedish. J Biomed Semantics 2011: (Suppl 3): S3
- [7] De'eger L. and Grouin C. Detecting negation of medical problems in french clinical notes. In: Proc. 2nd ACM SIGHIT symposium on International health informatics, ACM 2012: pp. 697-702.
- [8] Uzuner O, South B, Shen S, DuVall S. 2010 i2b2/va challenge on concepts, assertions, and relations in clinical text. J Am Med Inform Assoc 2011: 18 (5): 552-556.
- [9] Farkas R, Vincze V, Mo'ra G, Csirik J, Szarvas G. The conll-2010 shared task: Learning to detect hedges and their scope in natural language text. In: Proc. 14th CoNLL, Uppsala: ACL 2010: pp. 1-12.
- [10] Kim JD, Ohta T, Pyysalo S, Kano Y, Tsujii J. Overview of bionlp'09 shared task on event extraction. In: BioNLP '09, Stroudsburg: Association for Computational Linguistics, 2009; pp. 1-9.
- [11] Uzuner O. Recognizing obesity and comorbidities in sparse data. J Am Med Inform Assoc 2009:16 (4): 561-70.
- [12] Aronson A and Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010: 17: 229-236.
- [13] Savova GK, Masanz JJ, Origen PV, Zheng J, Sohn S, Kipper-Schuler KC, and Chute CG. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010: 17: 507-513.
- [14] Zeng Q, Goryachev S, Weiss S, Sordo M, Murphy SN, and Lazarus R. Extracting principal diagnosis, comorbidity, and smoking status for asthma research: evaluation of a natural language processing system. BMC Med Inform Decis Mak 2006: 6:30.
- [15] Lacasta J, Nogueras-Iso J, Soria FJZ. In: Terminological Ontologies: Design, Management & Practical Applications. London: Springer-Verlag, 2010.
- [16] Falquet G, Guyot J. Ontologies & Multilingualism in Ontologies in Urban Development Projects. In: Advanced Information and Knowledge Processing London: Springer-Verlag, 2011.
- [17] Maa X, Carranza EJM, Wub C, van der Meera FD, Liub G. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps. Computers & Geosciences 2011: 37(10): 1602-1615.
- [18] Cimiano P, Buitelaar P, McCrae J, Sintek M. LexInfo: A Declarative Model for the Lexicon-Ontology Interface. Web Semantics: Science, Services and Agents on the World Wide Web 2011: 9 (1): 29-51.
- [19] Vossen P. Introduction to EuroWordNet. Computers & the Humanities 1998: 32: 73-80.
- [20] Fellbaum C. WordNet and Wordnets. In: Brown, Keith et al. (eds.), Encyclopedia of Language and Linguistics, 2<sup>nd</sup> ed. Oxford: Elsevier, 2005; pp. 665-670.
- [21] Collier N, Goodwin RM, McCrae J, et al. An Ontology-driven System for Detecting Global Health Events. In Proceedings of the 23<sup>rd</sup> International Conference on Computational Linguistics (COLING), Stroudsburg: Association for Computational Linguistics, 2010; pp. 215-22.
- [22] Markó K, Baud R, Zweigenbaum P, Borin L, Merkel M, Schulz S. Towards a multilingual medical lexicon. AMIA Annu Symp Proc 2006: 534-538
- [23] Wermter J, Hahn U. An annotated German-language medical text corpus as language resource. In: Proc 4th Intl LREC Conf 2004. Lisbon, Portugal; pp. 473-476.
- [24] Dalianis H, Hassel M, and Velupillai S. The Stockholm EPR Corpus-Characteristics and Some Initial Findings. In Proceedings of ISHIMR 2009, Kalmar, Sweden, Oct 14-6.
- [25] Chapman B, Lee S, Kang H, Chapman WW. Document-level classification of CT pulmonary angiography reports based on an extension of the ConText algorithm. J Biomed Inform 2011: 44 (5): 728-737.
- [26] Velupillai S, Skeppstedt M, Kvist M, Mowery D, Chapman BE, Dalianis H, Chapman WW. Porting a rule-based assertion classifier for clinical text from English to Swedish. In: 4th Int Workshop on Health Document Text Mining and Information Analysis: Sydney (Louhi 2013).
- [27] Chapman BE, Wei W, Chapman WW. The frequency of ConText lexical items in diverse medical texts. Proc of the IEEE HISB Conference, 2012.

## Address for Correspondence

Wendy Chapman: wendy.w.chapman@gmail.com.