Negation Detection in Swedish Clinical Text

Maria Skeppstedt
DSV/Stockholm University
Forum 100
SE-164 40 Kista, Sweden
mariask@dsv.su.se

Abstract

NegEx, a rule-based algorithm that detects negations in English clinical text, was translated into Swedish and evaluated on clinical text written in Swedish. The NegEx algorithm detects negations through the use of trigger phrases, which indicate that a preceding or following concept is negated. A list of English trigger phrases was translated into Swedish, taking grammatical differences between the two languages into account. This translation was evaluated on a set of 436 manually classified sentences from Swedish health records. The results showed a precision of 70% and a recall of 81% for sentences containing the trigger phrases and a negative predictive value of 96% for sentences not containing any trigger phrases. The precision was significantly lower for the Swedish adaptation than published results on the English version, but since many negated propositions were identified through a limited set of trigger phrases, it could nevertheless be concluded that the same trigger phrase approach is possible in a Swedish context, even though it needs to be further developed.

1 Introduction

Medical documentation, such as patient records, is today often stored in a digital, searchable format. This opens the possibility of extracting information, which for example could be used for disease surveillance or to find new, unknown connections between patient background, symptoms and diseases. When extracting information from a text, it is not only the words that occur in the text that are important, but also whether these words are negated or not. This is especially true when it comes to patient records, since when describing the status of a patient, the physician often reasons by excluding various possible diagnoses and symptoms.

Most work on detecting negations in medical language has been carried out for English, and very little has been carried out for other languages, as for example Swedish. This article will therefore focus on the task of finding whether a concept in a clinical text written in Swedish is negated or not.[1]

2 Related research

There are many different methods for detecting whether a concept is negated. Rokach et al. (2008) give a good overview of some approaches for detecting negations. The methods can be divided into two main groups; knowledge engineering methods and machine learning methods. Knowledge engineering methods have the advantage that a large annotated corpus is not needed, but the disadvantage that rules have to be written manually, which is often time-consuming. Negation detection based on machine learning methods, on the other hand, is faster to implement and often works better when a text is not completely grammatical, which is often the case with clinical texts. (Rokach et al., 2008)

Since little previous work has been done on negation detection in Swedish medical text, the first step

[1]This research has been carried out after approval from the Regional Ethical Review Board in Stockholm (Etikprövningsnämnden i Stockholm), permission number 2009/1742-31/5.

Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents, pages 15–21,
Los Angeles, California, June 2010.

for Swedish negation detection is to adapt a simple knowledge engineering method that is used for detecting negations in English, an algorithm called NegEx. (Chapman et al., 2001b)

2.1 The NegEx algorithm

NegEx detects *pertinent negatives* in English patient records, that is "findings and diseases explicitly or implicitly described as absent in a patient". Given a sentence and a chosen *proposition* in this sentence, NegEx determines if that proposition is negated or not. An example would be *"Extremities showed no cyanoses."*, in which the proposition is *cyanoses*. (Chapman et al., 2001b)

The NegEx algorithm uses regular expressions and three lists of phrases. The first list, the pre-negation list, consists of trigger phrases which indicate that a proposition that follows them is negated in the sentence, for example *no signs of*. The second list, the post-negation list, consists of trigger phrases that indicate that a proposition preceding them is negated, as the phrase *unlikely*. Finally, the third list consists of pseudo-negation phrases, phrases that are similar to negation triggers, but that do not trigger negation, for example *not certain if*. The algorithm judges the proposition to be negated if it is in the range of one to six words from a post- or pre-negation trigger. (Chapman et al., 2001b)

NegEx has later been further developed into *NegEx version 2*[2], for example through the addition of more triggers and by limiting the scope of the negation through a list of conjunctions.

In the evaluation of NegEx, the propositions consisted of UMLS[3] phrases that belonged to any of the UMLS categories *finding*, *disease or syndrome* or *mental or behavioral dysfunction* and that could also be found in the describing text of an ICD-10 code[4]. Sentences containing these UMLS phrases were extracted from discharge summaries. Thereafter, 500 of the extracted sentences that contained at least one negation trigger and 500 sentences that did not contain a negation trigger were randomly selected. A few sentences that contained phrases that were suspected to sometimes indicate a negation, but that were not in the three lists, were included in the first group. The sentences were then categorised by physicians into containing an *affirmed proposition*, a *negated proposition* or an *ambiguous proposition*. The inter-rater agreement was almost 100%. For the NegEx evaluation, the categories *affirmed* and *ambiguous* were grouped into the category *not negated*. The results showed a precision of 84% and a recall of 82% for sentences in the group with negation triggers and a negative predictive value of 97% for sentences in the group without triggers. Of the correctly found negations, 82% were triggered by only three negation triggers; *no*, *without* and *no evidence of*. Moreover, only 15 of the 35 negation triggers were found in the test set. The trigger *not* had a precision of 58%, which was much lower than the precision for the other common triggers. (Chapman et al., 2001b)

An evaluation of the NegEx algorithm on ten other kinds of reports has also been carried out. The average precision of NegEx was 97%, and 90% of the detected negations were triggered by only seven negation phrases, with the four most frequent being *no*, *denies*, *without* and *no evidence*. (Chapman et al., 2001a)

In a later study by Goldin and Chapman (2003), a Naive Bayes classifier and a decision tree were used to classify which occurrences of the trigger *not* that indicated a negation, based on features such as surrounding words and their part of speech. Both these methods resulted in an increased precision.

3 Research Question

An evaluation was carried out on how the NegEx algorithm performs on health records written in Swedish, compared to health records written in English. The hypothesis was that the results for Swedish would be similar to the results for English, since the two languages are grammatically close. This comparison could give an indication of whether it is possible to adapt more advanced methods for negation detection into Swedish, and the results could also be used as a baseline for comparing the results of other methods.

4 Translation and adaption method

In order to use NegEx on a Swedish text, there must be a list of Swedish phrases that trigger negation.

[2] http://www.dbmi.pitt.edu/chapman/negex.html

[3] See Bodenreider (2004) for a description of UMLS

[4] http://www.who.int/classifications/icd/en/

4.1 Translating trigger phrases

The triggers for Swedish were obtained by translating the phrases for *NegEx version 2*. The translations were made with the help of a web-based English-Swedish dictionary[5] and with the help of Google translate[6]. In the cases where there was a good translation neither in the dictionary nor in the Google translation, the negation was translated by the author of this article. When it was not possible to find a good Swedish translation, the phrase was omitted. A total of 148 phrases were translated. Almost all negation phrases were general English terms. However, in a few cases they consisted of specific medical terms, and in these cases the translation was made by a physician. In many instances the dictionary offered many translations, and in other cases the same translation was offered for different English phrases. In the cases where several translations were offered, all of them were added to the list of Swedish negations.

4.2 Expanding the translated trigger phrases

English and Swedish are both Germanic languages (Crystal, 1997) and they have a similar grammar. Nevertheless, there are some grammatical differences that have to be taken into account through an expansion of the list of translated trigger phrases.

Swedish has two grammatical genders (common gender and neuter gender), whereas the English language lacks grammatical gender. Adjectives and some quantifiers in Swedish have a gender concord, as well as a number concord (Dahl, 1982). To compensate for this, the English negative quantifier *no* was translated into three different forms of the corresponding Swedish negative quantifier, namely *inga*, *ingen* and *inget*. Inflections of all adjectives in the trigger phrases were also generated. This was accomplished by invoking the Granska inflector[7].

The English combinations of aspect and tense do not always correspond directly to a Swedish verb form (Dahl, 1982). Therefore, a direct translation of the different forms of a verb in the trigger phrase list was not performed. The lemma form of the verb was instead added to the list of negation triggers in Swedish, and from this all inflections of the verb were generated, again using the Granska inflector.

The difference connected with the do-construction did not need to be taken into account. When negating a non-auxiliary verb in English, the do-construction is used. This type of construction does not exist in Swedish. The phrase *han vet* (*he knows*) would for example be negated as *han vet inte* (*he knows not*) (Svartvik and Sager, 1996). However, the NegEx algorithm only checks if the proposition is less than six words to the right of the word *inte* (*not*), and when it is, it will consider the proposition to be negated. The lack of a do-construction should therefore not affect the results.[8]

Swedish has a word order inversion in subordinate clauses. The position of the negating adverb is changed, and it is instead positioned immediately before the verb (Holmes and Hinchliffe, 2008). When stressing the negation, there is also the possibility of using this word order in the main clause (Sells, 2000). A version with reversed word order was therefore generated for trigger phrases containing some of the most common adverbs. From the translation of the trigger phrase *has not*, a version with the word order *not has* was for example generated.

The frequency of the Swedish trigger phrases was counted on a text other than the test set, and the most frequent trigger phrases were selected. The number of selected phrases was two more than used in the English NegEx evaluation, to compensate for Swedish gender and number concord[9].

5 Evaluation method

5.1 Construction of test data

Propositions to use for evaluating the performance of the Swedish version of NegEx were taken from the Swedish translation of the ICD-10 codes. However, the description in the ICD-10 code list often contains both the name of a symptom or disease and a clarification or specification of it, which has the

[5] http://www.norstedtsord.se

[6] http://translate.google.com

[7] http://www.csc.kth.se/tcs/humanlang/tools.html

[8] When negating the actual verb on the other hand, the position of the word *not* is different in English and Swedish. In order for the Swedish NegEx to handle verb phrase propositions, this difference has to be accounted for.

[9] The triggers that were used can be downloaded from http://people.dsv.su.se/∼mariask/resources/triggers.txt

effect that simple string matching would not find some of the most common symptoms and diseases. An automatic pre-processing of the ICD-10 code list was therefore first accomplished, where for example text within parenthesis and clarifications such as *not specified* or *other specified forms* were removed. To find more names of symptoms and diseases, additional lists were also added, including the KSH97-P[10], an adaption of the ICD-10 codes for primary care, and the MeSH terms under the sections *diseases* and *mental disorders*.

The test data was extracted from a set of sentences randomly chosen from the assessment part of Swedish health records from the Stockholm EPR Corpus (Dalianis et al., 2009). From this set, sentences that contained any of the propositions in the proposition list were extracted, also when the proposition was part of a compound word. Neither the preprocessing of the ICD-10 code list nor the detection of a proposition in a compound word was perfect and therefore some words that were not comparable with *findings*, *diseases or syndromes* or *mental or behavioral dysfunctions*, were added to the list of propositions. Sentences containing these were manually filtered out from the test data.

The chosen sentences were ordered in a list of pairs, consisting of the sentence and the proposition. If a sentence contained more than one proposition, the sentence was added to the list one time for each proposition.

In order to be able to compare the English and Swedish versions of NegEx, the same evaluation method was used, and two groups of test sentences were constucted. The first group, *Trig*, contained 202 sentences with at least one of the trigger phrases. The second group, *Non-Trig*, contained 234 sentences without any of the trigger phrases.

5.2 Classification of test data

The propositions were manually classified into the categories *affirmed*, *negated* and *ambiguous* by a rater without medical education. The categories *affirmed* and *ambiguous* were thereafter collapsed into the category *not negated*. The results are presented in Table 1.

Of the 202 sentences in group *Trig*, 70 were also classified by a physician. The inter-rater agreement between the physician and the other rater with respect to the two groups *negated* and *not negated* was 80%.

	Negated	**Not negated**	**Total**
Trig	90	112	202
Non-Trig	10	224	234

Table 1: Number of sentences manually classified as *negated* and *not negated* for each of the groups *Trig* and *Non-Trig*. Group *Trig* only contains sentences with trigger phrases and Group *Non-Trig* only contains sentences without trigger phrases.

The majority of the sentences where there was disagreement were judged as negated by the physician rater and ambiguous by the other rater, or ambiguous by the physician rater and negated by the other rater. There was no evident systematic tendency to judge the propositions as more or less ambiguous by either of the two raters.

When there was a difference in opinion of how to classify the proposition, the classification made by the physician was chosen. Also sentences that were subjectively judged by the rater as not possible to rate without deep medical knowledge, were rated by the physician.

6 Results

The Swedish version of NegEx was executed with the sentences in group *Trig* and the sentences in group *Non-Trig* as input sentences.[11] As shown in Table 2, group *Trig* had a precision of 70% and a recall of 81%. Group *Non-Trig* had a negative predictive value of 96%, as shown in Table 3.

When comparing Swedish and English results for recall using the χ^2-test, no significant difference was found between them. (p-value $>> 0.1$). When comparing the results for precision using the χ^2-test, it was significantly lower for Swedish. ($p < 0.001$).

The precision of each trigger was also counted and the results are shown in Table 4.

[10] http://www.socialstyrelsen.se/publikationer1996/1996-4-1

[11] http://code.google.com/p/negex/updates/list is the web location of NegEx (negex.python.zip, 2009). NegEx could be used in a Swedish context without any major modifications.

Group *Trig*	English	Swedish
recall (sensitivity)	82.00 %	81 %
specificity	82.50 %	71 %
precision (ppv)	84.49 %	70 %
npv	80.21 %	82 %

Table 2: Group *Trig*, 500 English sentences and 202 Swedish sentences. *Recall*: No. of correctly detected negated propositions divided by no. of manually rated negated propositions. *Specificity*: No. of propositions correctly detected as not negated divided by no. propositions that were manually rated as not negated. *Precision*: No. of correctly detected negated propositions divided by total no. of propositions that NegEx classified as negated. *Negative predictive value*: No. of propositions that NegEx correctly did not classify as negated divided by total no. of propositions that NegEx did not classify as negated. (Figures for English from Chapman et al. (2001b).)

Group *Non-Trig*	English	Swedish
npv	96.99 %	96 %

Table 3: Group *Non-Trig*, 500 English sentences and 234 Swedish sentences. (Figures for English from Chapman et al. (2001b).)

7 Discussion

The comparison between the English and Swedish evaluations is complicated by the fact that the Swedish test data had lower inter-rater agreement, which adds uncertainty to the Swedish results. This difference could perhaps be partly explained by the different types of health records; the English version was evaluated on discharge summaries, whereas the Swedish version was evaluated on the assessment part of a health record, which possibly contains more reasoning and thereby perhaps more ambiguous expressions.

Also, the fact that group *Trig* in the evaluation of the English version also included some sentences not containing trigger phrases complicates the comparison.

It could, however, be concluded that the precision is lower for Swedish. The following error types could at least account for some of this difference:

It is difficult to draw a line between what is an ambiguous expression and what is a negation, both for the raters and for the NegEx program. The above-mentioned difference in type of evaluation data could have resulted in lower precision and recall for the Swedish version.

Phrase		Precision	Occur.
inga tecken	(no signs of)	89 %	9
ingen	(no)	89 %	27
ej	(not)	75 %	8
inga	(no, plural)	67 %	15
utan	(without)	63 %	8
inte har	(not have)	60 %	5
inte	(not)	57 %	21
icke	(non-, not)	0 %	4

Table 4: The most frequent triggers, their precision and the number of times they occur in the sentences.

It is a common construction for a name of a disease, or a version of a disease, to have a name that starts with the word *icke* (*non-, not*), for example *icke allergisk astma*. The disease is present in the patient, even though the word *icke* is interpreted as a negation trigger by NegEx. In the test data, all the occurrences of the word *icke* are constructions like this, thus having a negative impact on precision.

The Swedish word for *without* (*utan*) has a double meaning. It is also a conjunction meaning *but*. This gives rise to a few instances where the program incorrectly classifies a proposition as negated, resulting in lower precision.

Other error types were also identified. These were, however, not specific for Swedish or for the type of test data, and could therefore not account for the difference in precision between the English and Swedish versions of NegEx. Examples are when the negation of the proposition occurs in a conditional clause or when the scope of the trigger should be less than the NegEx scope of six words, for example when the scope is limited by a conjunction.

7.1 Identified negation triggers

In the test set, only 16 of the 39 negation triggers were found, and among them, only 12 correctly negated a proposition. This is close to the English version where 15 of 37 triggers were found. None of the post-negation triggers were found in the Swedish test data.

In the English version of NegEx, 82% of the cor-

rectly found negations were triggered by the three negation phrases *no*, *without* and *no evidence of*. In the Swedish version, the three most common triggers were the common gender version of *no* (*ingen*), *not* (*inte*) and the plural form of *no* (*inga*). Together, they constitute 63% of the total number of correctly identified negations. If the trigger in fourth place, *no signs of*, is also counted, they make up 75% of the correctly negated propositions. In both English and Swedish there are thus a small number of negation triggers that are very common.

It can also be noted that both in Swedish and English, the precision of the trigger *not* (*inte*) is low.

No other common negation triggers were found in the test data. The only re-occurring trigger that was not included in any of the three lists were different forms of the phrase *rule out*.

8 Conclusion

The Swedish version of the NegEx algorithm had a significantly lower precision than the English version, and for the recall no significant conclusions could be drawn. Not taking the uncertainty of the low inter-rater agreement into account, the Swedish version has a precision of 70% and a recall of 81% for sentences containing the trigger phrases and a negative predictive value of 96% for sentences not containing any trigger phrases. As for the English version, a small number of trigger phrases accounted for the majority of detected negations.

Since a limited set of triggers can be used to identify many negations also in Swedish, this simple approach of the NegEx algorithm can be used as a base method for identifying negations in Swedish. However, even for use in a system without high demands on robustness, the method needs to be further developed.

From the relatively low inter-rater agreement, especially with respect to concepts that might be classified as either ambiguous or negated, it can be concluded that it is a difficult task also for a human rater to determine what is an ambiguity expressed as a negation or an actual negation.

9 Limitations

The most important limitation of this study is the relatively low inter-rater agreement, and the fact that most of the sentences were rated by a person who did not have a medical education. The lack of medical knowledge may have lead to mistakes when classifying the test data and could probably also partly explain the low inter-rater agreement.

Another limitation is that errors in the module for selecting sentences lead to that a few test sentences did not contain a symptom, disease or equivalent. Consequently, these sentences had to be filtered out manually.

As in the study by Chapman et al. (2001a), no analysis has been made of the occurrences of negations that stretch over sentence boundaries.

10 Future work

To automatically distinguish an ambiguity from a negation is not always trivial. However, the errors originating from the other error types mentioned could be limited through the use of more advanced natural language processing methods. The cases where the phrase *icke* does not trigger a negation, could probably be detected by a simple regular expression rule. Which meaning of the phrase *utan* that is intended could perhaps be detected by the machine learning methods used by Goldin and Chapman (2003). A list of conjunctions that limit the scope of the negations, as in *NegEx version 2*, could also be used to increase the precision, and a similar method could be used to detect when the proposition is negated in a conditional phrase.

It would also be interesting to use the complete list of negation triggers that was constructed for this study, instead of limiting the size to that of the NegEx trigger list, and to evaluate this list on a larger test set. This evaluation could also determine whether there are any common Swedish negation triggers that were not obtained by translating the English trigger list.

Acknowledgments

I would like to thank my supervisors Hercules Dalianis and Gunnar Nilsson for valuable comments on this paper, and specifically Gunnar for the help with the classification of the sentences. I would also like to thank Birgitta Melin Skeppstedt for initial help with the statistical calculations and Sumithra Velupillai for the support on the early stages of the

work. Many thanks also to the three anonymous reviewers of the paper.

References

Olivier Bodenreider. 2004. The unified medical language system (umls): integrating biomedical terminology. *Nucleic Acids Res*, 1;32(Database issue).

Wendy W Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001a. Evaluation of negation phrases in narrative clinical reports. *Proc AMIA Symp*, pages 105–109.

Wendy W. Chapman, Will Bridewell, Paul Hanbury, Gregory F. Cooper, and Bruce G. Buchanan. 2001b. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*, 34(5):301–310, Oct.

David Crystal. 1997. *The Cambridge encyclopedia of language*. Cambridge University Press, second edition.

Östen Dahl. 1982. *Grammatik*. Studentlitteratur.

Hercules Dalianis, Martin Hassel, and Sumithra Velupillai. 2009. The Stockholm EPR Corpus - Characteristics and Some Initial Findings. In *Proceedings of ISHIMR 2009, Evaluation and implementation of e-health and health information initiatives: international perspectives. 14th International Symposium for Health Information Management Research, Kalmar, Sweden*, pages 243–249.

Ilya M. Goldin and Wendy W. Chapman. 2003. Learning to detect negation with 'not' in medical texts. *ACM SIGIR '03 Workshop on Text Analysis and Search for Bioinformatics: Participant Notebook, Acknowledgements Toronto, Canada: Association for Computing Machinery;*.

Philip Holmes and Ian Hinchliffe. 2008. *Swedish: An Essential grammar*. Routledge.

Lior Rokach, Roni Romano, and Oded Maimon. 2008. Negation recognition in medical narrative reports. *Information Retrieval*, 11(6):499–538, December.

Peter Sells. 2000. Negation in Swedish: Where it's not at. In *Online Proceedings of the LFG-00 Conference. Stanford: CSLI Publications. (At http://csli-publications.stanford.edu/LFG/5/lfg00.html)*.

Jan Svartvik and Olof Sager. 1996. *Engelsk universitetsgrammatik*. Liber.