Descriptive analysis of negation cues in biomedical texts

Roser Morante

CLiPS-CL University of Antwerp
Prinsstraat 13, B–2000 Antwerpen
Roser.Morante@ua.ac.be

Abstract

In this paper we present a description of negation cues and their scope in biomedical texts, based on the cues that occur in the BioScope corpus. We provide information relative to the ambiguity of the negation cue and to the type of scope, as well as examples. We show that the scope depends mostly on the part-of-speech of the cue and on the syntactic features of the clause. Although several studies have focused on processing negation in biomedical texts, we are not aware of publicly available resources that describe the scope of negation cues in detail. This paper aims at providing information for producing guidelines to annotate corpora with a negation layer, and for building resources that find the scope of negation cues automatically.

1. Introduction

Negation is a linguistic phenomenon that has been widely described from a theoretical perspective (Tottie, 1991; Quirk et al., 2000; Horn, 2001). Generally speaking, negation turns an affirmative statement into negative. *Negation cues* are words or combinations of words that express negation.

In this paper we present a list of negation cues that occur in biomedical texts and a description of their scope based on their syntactic context. Such a description aims at providing information for producing guidelines to annotate corpora with a negation layer, and for building natural language processing resources that find the scope of negation cues automatically. Thus, a theoretical study of the negation cues is out of the scope of this paper.

The description that we provide is based on the negation cues that occur in the BioScope corpus (Vincze et al., 2008), a freely available resource[1], that consists of medical and biological texts. The BioScope corpus consists of three parts: clinical free-texts (radiology reports), biological full papers and biological paper abstracts from the GENIA corpus (Collier et al., 1999). Sentences are annotated with information about the scope of negation and hedge cues. Building on previous research about negation in biomedical texts (Chapman et al., 2001b; Boytcheva et al., 2005; Goldin and Chapman, 2003; Mutalik et al., 2001; Sanchez-Graillet and Poesio, 2007; Rokach et al., 2008; Vincze et al., 2008; Harkema et al., 2009; Morante and Daelemans, 2009), we have performed a per-cue analysis of the corpus in order to determine whether the cue is ambiguous, that is to say, whether it always acts as a negation cue or not, and to dermine how the scope can be described.

Negation has been a neglected area in natural language processing. Most research on this topic has been performed in the biomedical domain (see Section 2.), due to the need to automatically process huge amounts of literature in order to extract factual knowledge. Processing negation is relevant for tasks like information extraction because extracted information that falls in the scope of negation cues can-not be presented as factual information, paraphrasing, and summarization. Although several studies have focused on processing negation in biomedical texts, we are not aware of publicly available resources that describe the scope of negation cues in detail. The aim of this paper is to make a contribution in that direction, that can be further extended with a description of negation cues based on general domain corpora.

The paper is organised as follows. In Section 2. we present related work on processing negation in biomedical texts. Section 3. contains the description of negation cues and their scope, and Section 4., the conclusions.

2. Processing negation in biomedical texts

Research on processing negation has been carried out on the biomedical domain, mostly on clinical reports, and has focused on detecting whether a medical term is negated or not.

Mutalik et al. (2001) developed Negfinder, a rule-based system that recognises negated patterns in medical documents. It consists of two tools: a lexical scanner that uses regular expressions to generate a finite state machine, and a parser. Chapman et al. (2001a) developed NegEx, a regular expression based algorithm for determining whether a finding or disease mentioned within narrative medical reports is present or absent. NegEx is publicly available at the web page of the project[2], where a list of negation phrases can be found classified as pre- or post-UMLS, depending on if they occur before or after an UMLS term. The list does not contain further details about the scope of the phrases or examples. Harkema et al. (2009) present ConText, an extension of NegEx. This system uses also regular expressions and contextual information in order to determine whether clinical conditions mentioned in clinical reports are negated, hypothetical, historical, or experienced by someone other than the patient. As for negation, a term is negated if it follows under the scope of a negation *trigger term*. In this approach the scope of a trigger term extends to the right of the trigger and ends in a termination term or at the end of

---

the sentence. Additionally for negation terms, there are 14 left-looking trigger terms that scope to the left till a termination term or the beginning of the sentence.

Sanchez-Graillet and Poesio (2007) present an analysis of negated interactions in 50 biomedical articles and a heuristics-based system that extracts negated protein-protein interactions. Elkin et al. (2005) describe a rule-based system that assigns a level of certainty to concepts in electronic health records. Negation assignment is performed by the automated negation assignment grammar as part of the rule based system that decides whether a concept has been positively, negatively, or uncertainly asserted.

The systems mentioned above are essentially based on lexical information. Huang and Lowe (2007) propose a classification scheme of negations based on syntactic categories and patterns in order to locate negated concepts, regardless of their distance from the negation cue. Additionally, Boytcheva et al. (2005) incorporate the treatment of negation in a system, MEHR, that extracts from electronic health records all the information required to generate automatically patient chronicles.

The above-mentioned research applies rule-based algorithms to negation finding. Machine learning techniques have been used in some cases. Averbuch et al. (2004) developed an algorithm that uses information gain to learn negative context patterns. Golding and Chapman (2003) experiment with Naive Bayes and Decision Trees to distinguish whether a medical observation is negated by the word *not* in a corpus of hospital reports. Goryachev et al. (2006) compare the performance of four different methods of negation detection, two regular expression-based methods and two classification-based methods trained on 1745 discharge reports. They show that the regular expression-based methods show better agreement with humans and better accuracy than the classification methods. Rokach et al. (2008) present a new pattern-based algorithm for identifying context in free-text medical narratives.The originality of the algorithm lies in that it automatically learns patterns similar to the manually written patterns for negation detection.

In contrast with the above mentioned work, which focuses on finding negated terms, Morante and Daelemans (2009) developed a machine learning system that finds negation signals and their scopes. The system is trained on the BioScope corpus (Vincze et al., 2008), which, as far as we know, is the only corpus annotated with scopes.

3. Description of negation cues and their scope in biomedical texts

A description of negation cues can be useful for annotating corpora with negation information and for building systems that process negation. Even if the number of negation cues is not large in comparison to the number of hedge cues, annotating their scope is not a straightforward task. Vincze et al. (2008) report F1 inter-annotator agreements for the BioScope corpus of 76.29 for clinical records, 92.46 for abstracts, and 70.86 for full papers. Morante and Daelemans (2009) report experimental results of a scope finding system of 70.75 for clinical reports, 66.07 for abstracts, and 41.00 for full papers, in terms of percentage of correct scopes.

As Mutalik et al. (2001) put it, negation cues are heterogeneous, since they can be single words, simple phrases or complex verb phrases, they may preceed or suceed the words they have scope over, and the scope is not always continuous. An additional characteristic of negation cues is that some cues are very frequent, whereas the rest occurs only occasionally.

This section contains a list of negation cues and the description of their scope. We provide the following information per cue:

- Morphological type: verb (*lack*), adverb (*not*), adjective (*absent*), determiner (*no*), noun (*absence*), conjunction (*neither*), preposition (*without*).

- Characteristics of the scope: the scope can be total over the clause, or partial over certain phrases. The cue can scope to the right or to the left and it can change depending on the syntactic construction. For example, passive constructions change the direction of the scope.

- Ambiguity: some negation cues express negation always, but others do not, like *not* in *not only* and *not always*. We describe in which cases certain negation cues do not express negation. Some negation cues participate in multiwords that express speculation (*no evidence of*). These cases are also described.

- Examples from the BioScope corpus where the negation cue is marked with <> and the scope with [ ]. Each example starts with a number that identifies the text and the sentence in the corpus.

In order to decide whether a cue expresses negation or not, we do not take as criterium the factuality of the statement, but the function of the cue. For example, in cases of double negation a cue is under the scope of another cue, with the result that the facts that fall under the embedded cue are not negated. In these cases we mark the two cues as negation cues. In the examples below, *exclude* is marked as a negation cue, despite the fact that it falls under the scope of *not* (1) and *cannot* (2).

(1) 99587991–02: This does [<not> [<exclude> the diagnosis of pertussis]].

(2) 99590672–02: [<Cannot> [<exclude> a pleural effusion]].

Certain characteristics of the context cause that the facts that fall under the scope of a cue are not negated. In (3), the future tense causes that the information under the cue *exclude* is not negated, and in (4) the use of the verb *recommend* has the same effect. However, we mark the negation cue in both cases because we differentiate between the task of finding the scope and the task of finding negated information.

(3) 99714033–02: Voiding cystogram will be performed to [<exclude> reflux].

(4) 99715153–04: Bladder wall thickening, clinical correlation recommended to [<exclude> infection or neurogenic bladder].

When a cue, like *fail* in (5) falls under the scope of hedge cues, like *it is likely that* , the facts under its scope are not completely negated. In these cases we also mark the negation cue.

(5) 1471-2105-8-249–224: In such cases, *it is likely that* a voting–based ensemble will follow the crowd and [<fail> to find the true binding site].

Table 1 below contains a summary of the description of the scope per negation cue that is presented in the next subsections.

3.1. Absence

The noun *absence* is always a negation cue that scopes over the prepositional phrase headed by *of* that is required by *absence*.

(6) 98570513 –05: Prominent appearing left pulmonary artery should be correlated with the presence or [<absence> of heart murmur].

2172166–05: Because insertion mutations in phoP are polar on phoQ, we constructed strains that expressed the phoQ protein in the [<absence> of PhoP] to test whether resistance to defensin requires only the phoQ gene product.

9442374–08: The [<absence> of slow migrating forms of I kappa B beta following stimulation] suggests that the phosphorylation does not necessarily constitute the signal-induced event which targets the molecule for proteolysis.

3.2. Absent

The adjective *absent* is always a negation cue. Its scope depends on the syntactic construction. If it modifies a noun as in (7), it scopes over the noun phrase. If it participates in a copulative construction as in (8), it scopes over the clause.

(7) 10415075–10: Finally, immunoblot analysis of nuclear extracts showed decreased or [<absent> p65 protein levels].

(8) 10415075–07: [NF-kappa B binding activity was <absent> in several SLE patients who were not receiving any medication, including corticosteroids].

A precise annotation of scope should allow discontinuous scope. In (9), *The OBF-1 mRNA* should be annotated as scope because it is the ellided subject of the copulative clause in which *absent* is the attribute.

(9) 785929–05: [The OBF-1 mRNA] is expressed in a highly cell–specific manner, being most abundant in B cells and essentially [<absent> in most of the other cells or tissues tested].

3.3. Cannot, could not

The modal *can* followed by the adverb *not* acts always as a negation cue. It scopes over the direct object of the main verb in active sentences (10) and over the subject, in passive sentences (11).

(10) 7958618–03: At low GSSG levels, T cells [<cannot> optimally activate the immunologically important transcription factor NF kappa B], whereas high GSSG levels inhibit the DNA binding activity of NF kappa B.

(11) 8662928–10: [This regulation <could not> be appreciably modified by enhanced expression of STAT proteins].

3.4. Either

The determiner *either* is used as a negation cue when it is under the scope of negation cues like *no* and *not*, as shown in (12). *Either* scopes over the noun phrase in which it occurs as a determiner of the head.

(12) 93061407–08: No age-related differences were found in the expression of [<either> of these two proto–oncogenes] by anti-CD2 activated T cells.

920124–05: No virus replication and no transcription of the Ad2 early genes was observed in [<either> of the cells].

93292612–14: The levels of GATA-1, Epo-R and globin mRNA expressed were not affected by a 24-hour incubation of [<either> cell line] with Epo, GM-CSF or interleukin-3 (IL-3).

3.5. Except

The preposition *except* is always a negation cue that scopes over the prepositional phrase that it heads, as shown in (13).

(13) 10203577 -15: Our results show that the tested agents ([<except> in domethacin]) are inhibitors of the T cell-mediated immune response, (...) .

92113075–12: There was a normal increase in ACTH, cortisol, and GH ([<except> in one obese patient]) in response to insulin-induced hypoglycemia, while cortisol production was elevated in three patients.

99734593–02: Normal kidneys bilaterally [<except> for minimal renal pelvic dilatation].

3.6. Exclude

In biomedical texts, the verb *exclude* is used as a negation cue with the meaning of 'ruling out, to prove to be unrelated or not for consideration' (14), or 'except for' (15). It scopes over its direct object in active sentences and over the subject, in passive sentences.

(14) 99669739–04: The purpose was to [<exclude> any renal disease].

(15) PMC1131882–213: BLASTP searches of the sea urchin sequences against al GenBank proteins, [<excluding> RAG1], detected only the ring finger domain of the sea urchin sequences.

Exclude, like *fail, lack*, and *deny* belong to a category called inherent negatives (Tottie, 1991), words that have a negative meaning and a positive form.

3.7. Fail

The verb *fail* is used always as a negation cue in the sense of 'perform ineffectively or inadequately'. It scopes over the infinitive clause.

(16) 9710600–05: In contrast, a mutant of Tax termed M22, which does not induce NK-kappaB, [<fails> to activate either IKKalpha or IKKbeta].

8622948–07: T cells, which [<fail> to proliferate], make little or no detectable cytokines.

3.8. Failure

The noun *failure* is a negation cue if it occurs with infinitive clauses, as in (17). It scopes over the noun phrase that it heads. If it occurs as a noun modifed by an adjective or by another noun as in (18), it is not considered to be a negation cue.

(17) 7479915–10: However, DCs lack Sp1, which may explain the [<failure> of HIV-1 to replicate in purified DCs].

10197731–07: Finally, we demonstrate that a [<failure> to degrade IkappaB-alpha in cytosols of TNF-treated T cells pretreated with PAO] is due to its interference ... .

99729302 –02: Nearly 6-year–old male with failure to toilet train.

(18) 9763613–02: Despite overwhelming evidence that enhanced production of the p75 tumor necrosis factor receptor (p75TNF-R) accompanies development of specific human inflammatory pathologies such as multi–organ *failure* during sepsis, ...

9843840-02: Positive-pressure mechanical ventilation supports gas exchange in patients with respiratory *failure*, but is also responsible for significant lung injury.

3.9. Favor

The verb *favor* in the sense of 'prefer' is considered to be a negation cue when it occurs with a preopositional phrase headed by the preposition *over*. It scopes over the prepositional phrase.

(19) 97740836–01: Right upper lobe linear density is [<favored>] to represent subsegmental atelectasis [<over> early infiltrate].

99598032 –02: With greater consolidation than volume loss, pneumonia is [<favored> over atelectasis].

If the verb does not occur with such a prepositional phrase (20), it is not considered to be a negation cue.

(20) 97686164 –02: Patchy right infrahilar infiltrate favored to represent pneumonia.

3.10. Impossible

The adjective *impossible* is a negation cue in copulative clauses, if it is followed by an infinitive construction. It scopes over the adjectival phrase that contains the infinitive construction in active sentences, and over the subject of the copulative clause and the adjectival phrase in passive sentences (21).

(21) 93141620 08–1: However, [these cells have been <impossible> to generate] given the requirement of GATA-1 for Epo receptor expression and red cell viability.

3.11. Instead of

The preposition *instead of* is always a negation cue. It scopes over the prepositional phrase that it heads:

(22) 1471-2105-8-239–442: Samples of the protein pair space were taken [<instead of> considering the whole space] as this was more computationally tractable.

1471-2105-8-225-97: [<Instead of> the whole length of the amino acid sequence], a window of 101 residues centered at the inframe stop codon was used as a BLAST query.

3.12. Lack (noun)

The noun *lack* acts always as a negation cue. It scopes over the prepositional phrase headed by *of*.

(23) 99586261 –01: The [<lack> of full inspiration] likely accounts for some crowding of bronchovascular markings.

99711562 –06: Given the [<lack> of increase in right hydronephrosis], the hydroureter may not be significant.

3.13. Lack (verb)

The verb *lack* is always a negation cue. It scopes over its object.

(24) 8183915–05: A 55-kDa isoform, I kappa B gamma-2, [<lacks> the 190 C-terminal amino acids of p70I kappa B gamma].

(25) 7659529–06: The GATA-1 and Sp1 binding sites in this promoter [<lacking> a TATA sequence] were necessary for a high level of transcription activation.

3.14. Loss

The noun *loss* is always as a negation cue that scopes over the full noun phrase that it heads. The noun phrase can have the structure: [loss + prepositional phrase headed by *of*] as in (26), or [adjective/noun + loss], as in (27). This cue is characteristic of clinical reports.

(26) 97649110–02: Cough for one week and [<loss> of appetite].

(27) 97655701–01: Pneumonia with [volume <loss> in the right upper lobe and right middle lobe].

3.15. Miss

The verb *miss* in the sense of 'lack' can act as a negation cue in transitive constructions (28), in adjectival participle clauses (29), and in unacusative constructions (30). In the first case, the cue scopes over the direct object of the verb. In the second case, it scopes over the noun phrase that is modified by the adjectival participle. In the third case, it scopes over the subject of the verb.

(28) 9154298–12: A deletion construct that contains a NF-kappa B site but is [<missing> the multiple response region] demonstrated a continued increase in IL-6 luciferase activity in LPS-stimulated CMV transfected cells.

(29) 93300824 –09: ... [an additional factor and/or post-translational modification of a factor, <missing>] in B cells, might be required for transactivation by NFAT.

(30) PMC1131882–170: ... (1) [their N–terminal portions are <missing>] and the RAG1–like sequences start at positions 17 or 18; ...

Miss is not considered to be a negation cue when it occurs as an attributive adjective preceding the noun, like in (31).

(31) 1471-2105-8-259-263–1: The total number of missing proteins will be lower.

3.16. Negative

The adjective *negative* is a negation cue when it refers to negative results of tests:

(32) 99802361–02: Otherwise [<negative>] renal ultrasound.

In most cases, like above, the cue and the scope coincide because it is not specified what was found to be negative in the test. In the example below, what is negative is specified. In this case it scopes over the prepositional phrase headed by *for*.

(33) 97080592 –12: Cells with nuclear p50 were [<negative> for the CD38, CD20, and CD2 markers].

In most occurrences, *negative* is not a negation cue (34), because it refers to negative regulations, negative functions, negative molecules, etc. In this case, it defines a type of entity or event, not the absence or unexistence of it.

(34) 9018153–01: Interaction of transcription factors RFX1 and MIBP1 with the gamma motif of the negative regulatory element of the hepatitis B virus core promoter.

3.17. Neither ... nor

The discontinuous conjunction *neither ... nor* acts always as a negation cue. It scopes over different parts of the sentence, depending on the syntactic structure of the clause and the type of phrase being coordinated.

If it coordinates phrases in a copulative clause (35) or in a passive clause (36), it scopes over the full clause. If it coordinates a verb complement of an active clause (37), it scopes over this complement, but if it coordinates the subject of an active clause (38), it scopes over the full clause. Finally, if it coordinates verb phrases, it scopes over the verb phrases.

(35) 2123553–10: [<Neither> FKBP binding <nor> inhibition of rotamase activity of FKBP alone is sufficient to explain the biologic actions of these drugs].

(36) 10072497–07: The inhibitory effects of NAC were not due to nonspecific toxicity as [<neither> the viability of DC <nor> their mannose receptor-mediated endocytosis were modified by NAC].

(37) 7641319–07: The thromboxane receptor gene has [<neither> TATA <nor> a CAAT consensus site].

(38) 9032271–08: [<Neither>a mutation in I kappaB alpha <nor> a mutation in p50 or relA, the two major subunits of NK-kappaB in this cell line, accounts for this phosphorylation defect].

(39) 9007200–08: V3-BH10 [<neither> blocked radiolabeled IL-2 binding to IL-2 receptors <nor> affected tyrosyl phosphorylation of several cellular proteins (p120, p98, p96, p54, and p38), which is immediately induced by IL-2 stimulation].

3.18. Neither (determiner)

The determiner *neither* acts always as a negation cue that scopes over the full clause.

(40) 7849291–06: Northern blot analysis showed that [<netiher> of these agents reduced LPS-stimulated TF mRNA accumulation], thereby suggesting a posttranscriptional mechanism for the effect.

3.19. Never

The adverb *never* is always a negation cue. It scopes over the clause, if it is a complement of a verb that is not the main verb in the sentence (41), and over the sentence, if it is a complement of the main verb (42).

(41) 7836389–05: At variance, the second peak [which has <never> been reported previously], lasted several hours.

(42) 7759875–09: [Yet, the levels of the activating NF-AT complex <never> reach those observed in similarly stimulated normal T cells].

3.20. No:DT

The determiner *No* is a negation cue except when it belongs to multiword hedge cues like *no signs of, no evidence of, no proof, no guarantee that*, as exemplified in (43). However, these cases are listed as negation triggers in the list of negations used by the system NegEx (see Subsection 3.30.).

(43) 97666653 –01: *No evidence of* focal pneumonia.
8622883–04: Although a role of p16 in this regulation has been presumed, there is *no proof* so far that loss of this tumor suppressor gene really affects E2F-mediated regulations.
1471-2105-8-225–184: This is practically indispensable for objective classification of candidates, but there is *no guarantee that* unknown proteins with the 23rd amino acid will score higher than the thresholds.

Most cases of *no* as negation cue occur in clinical reports. *No* occupies the first position of a nominal sentence. The full noun phrase in which *no* is a determiner is under the scope of the negation (44).

(44) 97636245–01: [<No> acute cardiopulmonary].
97640573–02: [<No> features of bacterial pneumonia].

If *no* modifies a noun, it scopes over the noun phrase (45), and if it modifies an adjective, it scopes over the adjectival phrase (46).

(45) 93171603–12: In contrast, protease treatment and analysis of SJO extracts showed [<no> detectable levels of the band 1pk activity].

(46) 93153207–09: This low frequency of response was [<no> different from that previously detected using cultures of T cells and autologous].

3.21. No longer

The adverb *no longer* acts always as a negation cue. If it modifes an adjectival phrase in a copulative clause (47), it scopes over the adjectival phrase. If it modifies a verb, it scopes over the verb phrase (48).

(47) 99585768 –02: The previously noted opacity seen in the upper lobe on 1/2 is [<no longer> present].

(48) 7523507–08: Upon stimulation of the cells, this protein [<no longer> prevents binding of DNA by Band A], and supression of binding is restored within 30 min.

3.22. None

The pronoun *none* is always a negation cue. If it has a subject function (49) it scopes over all the sentence, whereas if it has an object function (50) it scopes over the object noun phrase.

(49) 10403270–11: All mAECA possessed high activity against macrovascular EC, but [<none> had significant antimicrovascular EC activity].

(50) 9808586–05: Normal RBC show [<none> of these phenomena].

3.23. Not:RB

The adverb *not* is not always a negation cue. When it is a cue, the scope depends on the syntactic construction and on the type of phrase modified by *not*. If it modifies a verb, it scopes over the verb phrase in active sentences (51) and over the clause, in passive sentences (52). If it modifies other phrases, it scopes over the phrase (53).

(51) 10358154 –07: We show that PNU156804 does [<not> inhibit c-myc and bcl-2 mRNA induction].

(52) 10233882–03: However, [a role for NF-kappaB in human CD34 (+) bone marrow cells has <not> been described].

(53) 10329625–01: Disruption of alpha beta but [<not> of gamma delta] T cell development by overexpression of the helix-loop-helix protein Id3 in committed T cell progenitors.

It is not a cue in the following cases:

- When it is preceded by *can* or *could*, the negation cue is *cannot*.

    (54) 97652710–02: Because the trachea is straight we [<cannot> fully rule out double aortic arch].

- When it is followed by *only, the only*.

    (55) 2172166 –09: Defensin resistance is not the only virulence characteristic ...

- In hedge cues like *it is not clear, it is not evident*.

    (56) 7585505–04: Because it is not clear at which point in cell cycle the peripheral B lymphocites are arrested ...
    9070319–02: ... it is not evident whether NF-kappa B acts as a messenger system for germline C epsilon transcription.
    8622883–02: The role of alterations of the MTS1 tumor suppressor gene on chromosome 9o21 ... is not yet clear.

- When it modifies an epistemic verb like in *it is not known*. In these cases it is a hedge cue.

    (57) 7645208–04: However, it is not known whether this mechanism alone is sufficient to explain the block to HIV replication ....
    747417–05: The mechanism of generation of p50 is not known.
    7925300–02: The activation of nuclear factor kappa B (NF-kappa B) in intact cells is mechanstically not well understood.

- In the construction *, if not + adverb,*:

    (58) 7650486–09: ... NK cell activation early after CD16 ligand binding involves primarily, if not exclusively, NFATp, ...

- As part of the conjunction *whether or not*:

    (59) 8151786–04: Stimulation of T cells ... resulted in the recruitment of transcriptional factors to a similar level whether or not the cells expressed the nef gene.

- Followed by certain adverbs like *necessarily, simply*

    (60) 8626528–07: Induction of NF–MATp35 was shown to depend on de novo protein synthesis and was restricted to T cells that received a mitogenic combination of T cell stimuli, not necessarily including CD28 signaling.
    8977228–12: Hence, these cytokines do not simply enhance monocyte differentiation, but have complex and slightly divergent effects ...

In certain cases, either *not* or the verb that it modifies are followed by an adverb. and does not express a categorical negation. In these cases it is not annotated as negation cue:

(61) 8645086–09
Although the mechanism of HIV-1 activation with heat shock has not been fully elucidated yet, it is presumed PKC plays an important role in HIV-1 activation.

(62) 97733696–05
Since the aortic arch and the descending thoracic aorta are also not clearly identified ...

(63) 2172166–10
The virulence defect ... is not completely explained ...

(64) item 99571347–02
This is not significantly changed allowing for better aeration on today's ....

(65) 7524762–11
SCL mRNA levels...did not increase significantly in ....

3.24. Or

The conjunction *or* belongs to a discontinuous negation cue if it is preceded by *neither*, as explained in Section 3.18.:

(66) 9887050–09: However, [<neither> TNF <or> LPS stimulated VCAM-1 expression in HUAECs].

3.25. Rather than

The conjunction *rather than* acts always as a negation cue that scopes over the phrase that it modifies.

(67) 7718519–05: Gel shift analyses demonstrate that NK-kappa B nuclear translocation is stimulated primarily by IL-1 [<rather than> by antigen receptor signals].

(68) 97025418–04: The proliferative responses of purified CD4+ or CD8+ T cells are suppressed more strongly by TPCK when anti-CD28 [<rather than> the phorbol ester PMA] is used as the mitogenic coactivator.

3.26. Rule out

The verb *rule out* occurs in clinical reports. Like other verbs, it scopes over the object in active constructions and over the subject in passive constructions (69). When the verb is used in imperative (70), it does not act as a negation cue.

(69) [Pneumothorax was <ruled out>].

(70) 97702062–03: Rule out pneumothorax].

3.27. Unable

The adjective *unable* is always a negation cue that scopes over the the full clause in copulative sentences

(71) 9341193–08: Analysis of Tax mutants showed that [two mutants, IEXC29S and IEXL320G, were <unable> to significantly transactivate the c-sis/PDGF-B promoter].

3.28. With the exception of

The prepositional multiword *with the exception of* occurs always as a negation cue that scope over the prepositional phrase.

(72) 7478534–09: Mutation of the core ETS binding site from –GGAA– to –GGAT– prevents the binding of ETS–like factors [<with the exception of> ETS1].

97088634 –03: The gene is expressed in all hematopoietic cells [<with the exception of> T-cells and plasma cells].

This cue can have variants, if an adjective modifies the noun *exception* like in *with the notable exception of*.

3.29. Without

The preposition *without* acts always as a negation cue that scopes over the prepositional phrase that it heads.

(73) 10352279 –01: CD28 costimulation augments IL-2 secretion of activated lamina propria T cells by increasing mRNA stability [<without> enhancing IL-2 gene transactivation].

3.30. Additional cues

In this subsection we list negation cues that are not found in the BioScope corpus, but that are reported in previous work.

The noun *inability* is reported in Sanchez-Graillet and Poesio (2007) with the following example:

(74) The selectivity of this interactions was demonstrated by the [<inability> of hTR to interact with GST] ... .

Mutalik et al. (2001) report on the following additional cues: *no history of*, *deny* (75). They also report on cases in which the adverb *not* negates a verb phrase, whereas the negated concept is the subject of the verb. It concerns mostly perception verbs like *see* in passive constructions (76).

(75) Patient [<denies> shortness of breath, chest pain, fever, chills, nausea, vomiting diarrhea, and abdominal pain].

(76) [Dyspnea is <not> seen].

Additionally, a list of negation triggers can be found in the web page of NegExp mentioned above. The differences with the negation cues listed for the BioScope corpus are the following:

- Half of the NegEx triggers are patterns related to the verb *rule out*, like: *ruled him out for, ruled her out for, ruled the patient out for, ruled out against, ruled him out against*.

- Patterns with the triggers *no signs of, no evidence of* and similar: *no findings of, no findings to indicate, no mammographic evidence of, no radiographic evidence of, no new evidence, no other evidence, no evidence to suggest, no sign of, no signs of, no suggestion of*.

- Patterns with the trigger *without*: *without, without any evidence of, without evidence, without indication of, without sign of*.

- Patterns involving *not*: *not appear, not appreciate, not associated with, not complain of, not demonstrate, not exhibit, not feel, not had, not have, not know of, not known to have, not reveal, not see, not to be*.

- Patterns with the trigger *no*: *no abnormal, no cause of, no complaints of, no new, no significant, no suspicious*.

- Patterns with the trigger *never*: *never developed, never had*.

- Other triggers: *checked for, declined, declines, denied, denies, denying, evaluate for, free of, resolved, test for, unremarkable for, with no*

In future work we will integrate these patterns with the cues of the BioScope corpus.

4. Conclusions and further research

In this paper we presented a list of negation cues and the description of their scopes in biomedical texts. The list is based on the negation cues that occur in the BioScope corpus. It has been shown that the scope of the cues depends mostly on the part-of-speech of the cue and on the syntactic construction in which it occurs. A descriptive summary per cue is presented in Table 1

The list of cues is not exhaustive, since it is based on one corpus. As future research, we would like to continue enriching this list with cases from other corpora and we would like to analyse the negation cues of general domain corpora in order to compare cross-domain behaviour of the cues and produce annotation guidelines to annotate a general domain corpus.

Additionally, we would like to analyse the content of the words that fall under the scope, in order to define an algorithm that determines what information is negated.

Acknowledgments

5. References

M. Averbuch, T. Karson, B. Ben-Ami, O. Maimon, and L. Rokach. 2004. Context-sensitive medical information retrieval. In *Proc. of the 11th World Congress on Medical Informatics (MEDINFO-2004)*, pages 1–8, San Francisco, CA. IOS Press.

S. Boytcheva, A. Strupchanska, E. Paskaleva, and D. Tcharaktchiev. 2005. Some aspects of negation processing in electronic health records. In *Proc. of International Workshop Language and Speech Infrastructure for Information Access in the Balkan Countries*, pages 1–8, Borovets, Bulgaria.

W. W. Chapman, W. Bridewell, P. Hanbury, G. F. Cooper, and B.G. Buchanan. 2001a. A simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform*, 34:301–310.

W. W. Chapman, P. Hanbury, G. F. Cooper, and B.G. Buchanan. 2001b. Evaluation of negation phrases in narrative clinical reports. In *Proc AMIA Symp. 2001*, pages 105–109.

Cue	POS	Ambig	Scopes over:
absence	N	N	PP headed by *of*
absent	ADJ	N	NP, if attributive; clause, if in copulative clause
cannot	MV	N	object of verb in active clause; subject of verb in passive clause
could not	MV	N	object of verb in active clause; subject of verb in passive clause
either	DET	Y	NP that *either* modifies
except	PRP	N	PP headed by *except*
exclude	V	N	object in active clauses; subject in passive clauses
fail	V	N	infinitive clause
failure	N	N	NP
favor over	V	Y	PP headed by *over*
impossible	ADJ	Y	AP in active clauses; subject and AP in passive clauses
instead of	PRP	N	PP headed by *instead of*
lack	N	N	PP headed by *of*
lack	V	N	object of *lack*
loss	N	N	NP
miss	V	Y	object in transitive constructions; NP modified by adjectival participle clauses; subject in unacusative clause
negative	ADJ	Y	PP headed by *for*
neither ... nor	CNJ	N	clause, verb prhases or verb complement depending on type of coordinated phrases and syntactic structure of clause
neither	DET	N	clause
never	ADV	N	sentence or clause, depending on the type of verb it modifies (main or subordinated)
no	DET	Y	the phrase that it modifies, NP or AP
no longer	ADV	N	the phrase that it modifies, AP or VP
none	N	N	clause, if subject; object noun phrase, if object
not	ADV	Y	clause or VP, if it modifies a verb; else, modified phrase
rather than	CNJ	N	phrase being modified
rule out	V	Y	object of verb in active clause; subject of verb in passive clause
unable	ADJ	N	full clause
with the exception of	PRP	N	PP headed by *with the exception of*
without	PRP	N	PP headed by *without*

Table 1: List of negation cues

N. Collier, H.S. Park, N. Ogata, Y. Tateisi, C. Nobata, T. Sekimizu, H. Imai, and J. Tsujii. 1999. The GENIA project: corpus-based knowledge acquisition and information extraction from genome research papers. In *Proceedings of EACL-99*.

P. L. Elkin, S. H. Brown, B. A. Bauer, C.S. Husser, W. Carruth, L.R. Bergstrom, and D. L. Wahner-Roedler. 2005. A controlled trial of automated classification of negation from clinical notes. *BMC Medical Informatics and Decision Making*, 5(13).

l. M. Goldin and W.W. Chapman. 2003. Learning to detect negation with 'Not' in medical texts. In *Proceedings of ACM-SIGIR 2003*.

S. Goryachev, M. Sordo, Q.T. Zeng, and L. Ngo. 2006. Implementation and evaluation of four different methods of negation detection. Technical report, DSG.

H. Harkema, J. N. Dowling, T. Thornblade, and W. C. Chapman. 2009. ConText: An algorithm for determining negation, experiencer, and temporal status from clinical reports. *Journal of Biomedical Informatics*, 42:839–851.

L. R. Horn. 2001. *A natural history of negation*. CSLI, Stanford, CA.

Y. Huang and H.J. Lowe. 2007. A novel hybrid approach to automated negation detection in clinical radiology reports. *J Am Med Inform Assoc*, 14(3):304–311.

R. Morante and W. Daelemans. 2009. A metalearning approach to processing the scope of negation. In *Proc. of CoNLL 2009*, pages 28–36, Boulder, Colorado.

A.G. Mutalik, A. Deshpande, and P.M. Nadkarni. 2001. Use of general-purpose negation detection to augment concept indexing of medical documents. a quantitative study using the UMLS. *J Am Med Inform Assoc*, 8(6):598–609.

R. Quirk, S. Greenbaum, and G. Leech. 2000. *A comprehensive grammar of the English language*. Longman, London.

L. Rokach, R.Romano, and O. Maimon. 2008. Negation recognition in medical narrative reports. *Information Retrieval Online.*

O. Sanchez-Graillet and M. Poesio. 2007. Negation of protein-protein interactions: analysis and extraction. *Bioinformatics*, 23(13):424–432.

G. Tottie. 1991. *Negation in English speech and writing: a study in variation.* Academic Press, New York.

V. Vincze, G. Szarvas, R. Farkas, G. Móra, and J. Csirik. 2008. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics*, 9((Suppl 11)):S9.