Translational morphosyntax: Distribution of negation in clinical records and biomedical journal articles

K. Bretonnel Cohen[a], Foster Goss[b], Pierre Zweigenbaum[c], Lawrence E. Hunter[a]

[a] *Computational Bioscience Program, University of Colorado School of Medicine, Aurora, CO, USA*
[b] *University of Colorado, Department of Emergency Medicine, Aurora, CO, USA*
[c] *LIMSI, CNRS, Université Paris-Saclay, Orsay, France*

Abstract

Prior knowledge of the distributional characteristics of linguistic phenomena can be useful for a variety of language processing tasks. This paper describes the distribution of negation in two types of biomedical texts: scientific journal articles, and physician progress notes from the intensive care unit. Two types of negation are examined: explicit negation at the syntactic level, and morphological negation at the sub-word level. The data show that the distribution of negation is quite different in the two document types at a high level of significance, with explicit negation much more frequent in the clinical documents than in the scientific publications, and with morphological negation much more frequent in the journal articles at the type level and more frequent in the journal articles at the token level. All code used in the study is freely available on GitHub[1].

Keywords:

Natural Language Processing, MeSH ID D009323; Text Mining, MeSH ID D057225; Linguistics, MeSH ID D008037

Introduction

Natural language processing (the treatment of human language by computers) is an increasingly-used tool in biomedical research and health care (Névéol & Zweigenbaum, 2015, 2016; Simpson & Demner-Fushman, 2012; Velupillai, Mowery, South, Kvist, & Dalianis, 2015). As the community has come to understand the challenges of this work, it has become clear that negation is a frequent cause of errors in language processing (W W Chapman, Bridewell, Hanbury, Cooper, & Buchanan, 2001a, 2001b; Goldin & Chapman, 2003; Grouin et al., 2011; Mutalik, Deshpande, & Nadkarni, 2001). This has stimulated a considerable amount of work on negation in clinical documents (e.g. the classic work cited above, and a considerable amount of more recent work summarized in (Uzuner, South, Shen, & DuVall, 2011)), and a lesser amount of work on negation in scientific literature (e.g. the work summarized in (Kim, Ohta, Pyysalo, Kano, & Tsujii, 2009; Morante, Liekens, & Daelemans, 2008; Nédellec, Bossy, Kim, Kim, & Ohta, 2013; Vincze, Szarvas, Farkas, Mora, & Csirik, 2008), as well as scientific literature corpus creation efforts (Agarwal, Yu, & Kohane, 2011; Pyysalo et al., 2007; Sanchez-Graillet & Poesio, 2007; Thompson, Iqbal, McNaught, & Ananiadou, 2009; Wilbur, Rzhetsky, & Shatkay, 2006)). However, while that work has made considerable progress, it has focussed almost exclusively on explicit negation by words with phrasal scope, such as *no* and *not.* In contrast, negation at a sub-word level—what has been called *morphological negation,* such as the *de-* in *dephosphorylate* or the *a-* in *afebrile,* meaning *without fever* —has received very little attention in the biomedical domain (or elsewhere). This is a gap in the literature, because this kind of negation has implications for many things in biomedical language processing and biomedical communication in general, ranging from lexicon/terminology design, to readability of and access to health care information by non-specialists, to the performance of natural language processing applications. Furthermore, the majority of work on negation in the biomedical domain has focussed on evaluation of system performance; very little of it has looked at distributional characteristics of negation in the relevant genres. This is a serious gap because it has implications for our understanding of the system performance that is the topic of most work on negation in the domain. This paper addresses both of those issues. In particular, we look at both clinical data and scientific literature and compare them with respect to their distributions of two kinds of negation: explicit negation (words such as *no* and *not*), and morphological negation (the *a-* in *afebrile*). Along the way we discuss a data set that we have prepared containing several thousand ambiguous words marked as to whether or not they begin with a negative prefix. The null hypotheses that we evaluate are that there are no differences in the distribution of negation between clinical and scientific biomedical texts at any of the levels to be examined; as will be shown, in fact there are such differences at the syntactic level and at the morphological level, and the differences are sometimes large.

This paper takes a distributional and descriptive approach because in text mining and natural language processing, knowledge of the distribution of any linguistic phenomenon can help us predict the contribution of that phenomenon to error rates in our applications. Distribution of negation in particular is important both in natural language processing and in language science more generally. In a paper that we discuss in more detail below, (Wu et al., 2014) point out that distributional characteristics of linguistic phenomena can have deep implications for evaluating not just individual systems, but for evaluating the literature on a topic overall, where the performance that is reported in a paper may accurately describe the performance of a system when it is optimized for a specific dataset, but not be generalizable; this can lead to the conclusion that a particular problem is essentially solved,

[1] https://github.com/KevinBretonnelCohen/NegationDistribution

when in fact all that has been solved is dealing with a particular definition of that problem in a specific data set. In particular, Wu et al. point out that there has been considerable work on negation in biomedical text, particularly in clinical text, with a smaller body of work existing on negation in scientific journal articles. They describe a number of published solutions to negation in clinical texts, and observe that they are optimized to particular genres of text, and that those solutions do not necessarily generalize well at all.

From the point of view of system development and evaluation, knowledge of the distribution of a linguistic phenomenon can help us find suitable document sets to use for some specific task type that we want to work on (at the granularity of, say, parsing, coreference resolution, etc.). It can help us prioritize module development in system design, and in the case of negation, it may interact with usability of tools based on natural language processing, given what we know about human processing of negation from psycholinguistic studies: negated assertions are more difficult to process (Larrivée & Lee, 2015).

On a very practical level, the kind of distributional information that is studied in this paper can be used as a form of prior knowledge for setting parameters in machine learning applications that allow supervised under- and over-sampling (Cruz Díaz, Maña López, Mata Vázquez, & Pachón Álvarez, 2012).

Finally, to the authors' knowledge, there is a complete lack of literature that explores translational issues in negation. All work that we are aware of in the biomedical domain has looked either at clinical texts, or at published biomedical literature. To address that gap in the literature, we have compared two very different genres with relevance to translational research: on the one hand, scientific journal articles, and on the other hand, clinical documents.

Context of the present work

There has been a small amount of previous work on the distribution of negation. (Yaeger-Dror & Tottie, 1993) focussed on spoken versus written English and found differences in the distribution of morphological and explicit negation when comparing spoken versus written language. (W W Chapman et al., 2001b) examined the distribution of explicit negation *within* a genre, and found that it may be Zipfian, noting that "Negation phrases appear to comply qualitatively with Zipf's law," and that "The negation algorithm was triggered by sixty negation phrases with just seven of the phrases accounting for 90% of the negations." Subsequent work found this to be true across multiple Germanic languages, as well as the Romance language French (Wendy W. Chapman et al., 2013). This work can be summarized as showing that the distribution of negation is structured, and that it can be shown to vary in interesting ways both within and across genres. However, it remains the case that the literature on differences between morphological and explicit negation is very small, to the point that there have not been opportunities to evaluate the replicability of the associated findings, and the topic has not been touched on at all in the biomedical domain.

Wu et al. point out that one of the consequences of the sublanguage nature of clinical documents is that there is a limited number of ways to express negation; this is true, but previous studies of negation in clinical literature have focussed on negation at the syntactic level. Here we extend the domain of inquiry into a previously unstudied part of the grammar of biomedical text: the morphological level. Wu et al. pointed to the morphological differences in annotations as a possible explanatory factor that was uncharacterized. The work discussed here adds a considerable amount of data to that discussion, adding the ability to compare clinical data (the subject of the small amount of previous observations about distribution of negation in the biomedical domain) with data on scientific publications. The fact that we used the same processing on both data sources makes the results directly comparable, which has not been the case with any previous work.

Methods

Materials

Since the goals of this study are translational in nature, the two sets of documents that were the materials for this work were drawn from the clinical domain and from the biomedical literature. These were MIMIC II progress notes on the one hand (Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, 2000; Saeed et al., 2011), and the CRAFT corpus on the other (Bada et al., 2009; Cohen et al., 2017; Verspoor et al., 2012).

MIMIC II progress notes: We worked with a sample of half a million words of progress notes from the MIMIC II corpus. These are free text notes written by physicians, reflecting the daily status of a patient in the Intensive Care Unit.

CRAFT corpus: This is an extensively annotated corpus of scientific journal articles in the domain of mouse genomics, broadly construed. It has been studied with respect to its similarity to PubMed/MEDLINE as a whole and has been found to be representative of the biomedical scientific literature (Verspoor, Cohen, & Hunter, 2009).

Methods: Explicit negation

We took samples of 10,000 consecutive words from both document types, for a total of 440,000 words each (the closest total sample size to the number of words in CRAFT, the smaller corpus). We counted the number of explicit negative words per 10,000-word sample. The set of explicit negatives that we counted was: *no|not|none|denies|nothing.* Details of the normalization can be found in the script, available at the GitHub site (see Footnote 1) and named *negativesEvery10KWords.pl.* (We note that one could argue about the completeness of the set of explicit negatives that was used in the experiment, but any omissions would affect both text types proportionally and would not be likely to change the overall conclusions of the study.)

Methods: Morphological negation

All word types in both corpora were collected. After normalizing for case and punctuation, we counted the number of tokens of each type. To clarify the intended meanings of the terms *type* and *token:* the word *denaturation* occurs 9 times in the CRAFT corpus. The word *ativan* appears 22 times in our sample of the MIMIC II corpus. These represent two *types* (*denaturation* and *ativan*), and a total of 31 *tokens.* (See the script *directoryToTypeTokenCounts.pl* on the GitHub site for details of the normalization.) Then we extracted all words beginning with any string that can be a negative prefix in English. To ensure objectivity, we obtained definitions of the set of negative prefixes in English from neutral third parties, listed below. This step produced a list of 5,196 words that can be thought of as ambiguous with respect to whether or not they begin with a negative morpheme. The extracted words were examined manually and classified as actually containing a negative prefix, or not.

All words from both document sets were presented as single words in isolation.

Then, with the judgements about which words did and did not begin with a negative prefix, along with the counts of each of those words, we calculated the total number of tokens beginning with an actual negative prefix in each document collection. We give the details of this process below.

Lists of English negative prefixes were collected from the following sources:

- lognlearn.jimdo.com/vocabulary/wordbuilding/negative-prefixes/
- http://www.englishhints.com/prefix-list.html
- http://grammarist.com/usage/negative-prefixes/

To build the set of words beginning with ambiguous strings, we first searched the two text collections for words beginning with the following character sequences, all of which are listed as negative prefixes in the 3rd-party sources described above: *un, no, a, de, dis, anti, il, im, in,* and *ir.*

The resulting set of words can be thought of as ambiguous with respect to whether or not they begin with a negative morpheme, since it contains both words such as *antiapoptotic* (CRAFT, 6 tokens), *immature* (MIMIC, 177 tokens), and *desaturation* (MIMIC, 72 tokens), which do begin with negative morphemes, and *anticipated* (CRAFT, 3 tokens), *improved* (MIMIC, 205 tokens), and *detailed* (MIMIC, 809 tokens), which do not. These are the words which were manually classified as either beginning with a negative morpheme, or not.

Guidelines

We developed the guidelines in three rounds, consisting of

1. a test of an initial set of guidelines on Amazon Mechanical Turk[2],
2. a subsequent test of a revision of the guidelines on Amazon Mechanical Turk,
3. and then group review of the guidelines by all of the authors.

In developing the Amazon Mechanical Turk tasks, we took note of the ethical guidelines for crowdsourced linguistic data discussed in (Sagot, Fort, Adda, Mariani, & Lang, 2011).

We were edified to find that even after those three rounds of guideline development, even our domain-expert annotators---one of whom had written the guidelines—still had questions about some specific cases. The final set of guidelines is available on the GitHub site in the file *Instructions for real experiment.rtf*.

To ensure the clarity and consistency of the guidelines and the neutrality of the resulting counts, the data was independently double-annotated by two of the authors. They represent typical readers of the materials in question:

- An emergency room physician.
- A former registered cardiovascular technologist with a PhD in linguistics and a specialty in biomedical language.

For the final annotation step, the word types from CRAFT and MIMIC II were combined into a single file and randomized (both with respect to corpus and with respect to ranking within each corpus). The shell script that shows the data flow, along with the randomization script, can be found on the GitHub site in the files *createExperimentDataSet.sh* and *randomizeTypeTokenFiles.pl.* The inter-annotator agreement was 0.94 before resolution, and the entire calculation of agreement is documented in the file *InterannotatorAgreementNegation.Rmd* on GitHub.

Finally, we used a two-tailed t-test to assess the statistical significance of the observed differences in explicit negation, and the chi square test for the data on morphological negation.

Replicability and reproducibilty

All scripts, as well as annotation guidelines for the morphology study and the judgements of the individual annotators, are available at the GitHub site in Footnote 1. The CRAFT corpus can be downloaded at http://bionlp.sourceforge.net/. The MIMIC II corpus requires signing a data use agreement, but it can be found at https://mimic.physionet.org/. To explore reproducibility of the results, some avenues suggested by the larger context reviewed above would be to repeat the experiments on other MIMIC note types; on other subject-matter domains of journal articles; and in other languages.

Results

Explicit negation

The distribution of explicit negatives for the two document collections is show in Figure 1. The distributions are quite different, with a mean of 111 per 10,000-word sample for the MIMIC II progress notes, and a mean of 31 per 10,000-word sample for the CRAFT corpus. Welch 2-sample t-test shows a statistically significant difference, t = -27.092, df = 53.822, p-value < 2.2e-16.

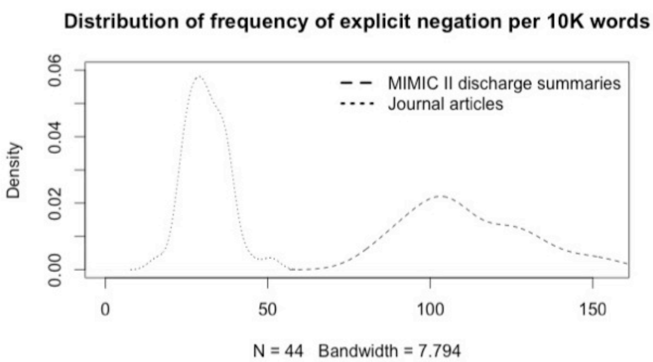

Figure 1—Density distribution of frequency of explicit negation per 10,000 words, MIMIC II and the CRAFT corpus.

Morphological negation

The distribution of morphologically negated and non-negated words is shown in Table 1, along with their ratios on the type level and on the token level. The distribution of morphological negation is different at the type level in the two genres, chi square = 8866.8, df = 1, p-value < 2.2e-16, with the journal articles having a higher incidence of

[2] *Words from MIMIC II that were used in the AMT tests of the guidelines were manually screened by an author with HIPAA and human subjects training to ensure that those words did not contain any identifying or potentially identifying information. This was in addition to the screening that has already been done by the MIMIC Consortium.*

morphologically negated types (0.028) than the clinical documents (0.017). Additionally, the distribution of morphological negation is also different at the token level in the two genres, chi square = 14338, df = 2, p-value < 2.2e-16, with the journal articles having a higher incidence of morphologically negated tokens, although the magnitude of the difference is smaller than that at the level of types (0.013 for CRAFT, 0.012 for MIMIC II).

Table Number 1–Counts and ratios of negated types and tokens in CRAFT and in MIMIC II progress notes.

Corpora and counts	Count or ratio
CRAFT negated types	650
CRAFT ambiguous non-negated types	2,641
CRAFT non-ambiguous non-negated types	19,545
CRAFT ratio of negated types to all non-negated types	**0.028**
CRAFT negated tokens	5,575
CRAFT ambiguous non-negated tokens	63,819
CRAFT non-ambiguous non-negated tokens	367,576
CRAFT ratio of negated tokens to all non-negated tokens	**0.013**
MIMIC negated types	319
MIMIC ambiguous non-negated types	1,586
MIMIC non-ambiguous non-negated types	16,400
MIMIC ratio of negated types to all non-negated types	**0.017**
MIMIC negated tokens	6,763
MIMIC ambiguous non-negated tokens	70,945
MIMIC non-ambiguous non-negated tokens	484,526
MIMIC ratio of negated tokens to all non-negated tokens	**0.012**

Discussion

The results of the hypothesis tests can be summarized thus:

The distributions of explicit negation are different between the two genres by two-tailed t-test, with the clinical notes having a much higher incidence of negation than the journal articles.

The distribution of morphological negation is different at the type level in the two genres by chi square, with the journal articles having a higher incidence of morphologically negated types.

The distribution of morphological negation is also different at the token level in the two genres by chi square, with the journal articles having a higher incidence of morphologically negated tokens, although the magnitude of the difference is smaller than that at the level of types.

Thus, there are differences in the distribution of negation between the two genres at both levels: explicit negation, and morphological negation. Furthermore, the difference is in different directions at the two levels. At the level of explicit negation, there is more negation in the clinical texts. In contrast, at the level of morphological negation, there is more negation in the journal articles.

These findings are especially relevant to a translational perspective on biomedical natural language processing, since the experiments reported here compared scientific literature to a clinical textual genre. From that perspective, the implications of the findings are that language processing systems that target the mapping of findings from the scientific literature to electronic health records will need to take a nuanced approach to handling negation, taking into account the different distributional characteristics of negational phenomena in the two genres.

Conclusions

Wu et al. conclude from their analysis of generalization versus optimization in clinical-domain negation detection systems that the best way to improve performance in negation detection is to manually annotate more data; in particular, they refer here not to increasing the sizes of the corpora that we already have, but to annotating negation in data drawn from other distributions besides the corpora that are already available. The work reported here is a contribution in that direction, as one of the results of the work is a large set of words from clinical records and scientific journal articles, available at the GitHub repository, that have been annotated for the presence of a derivational, prefixal negation morpheme. Because the methodology that we describe here can yield relatively rapid judgements with good inter-annotator agreement, this two-corpus study can be rapidly extended to additional scientific and clinical genres.

In addition to the relevance of these findings to biomedical language processing, there are also implications for the construction of semantic resources for the domain. The community's investment in lexical, terminological, and ontological resources continues to be strong. The findings that we report here have implications for the approach to building those resources. Modern lexical-semantic resources such as PropBank and VerbNet (Kipper, Korhonen, Ryant, & Palmer, 2008; Palmer, Gildea, & Xue, 2010; Palmer, Kingsbury, & Gildea, 2005) include separate entries for predicates that are related by the negative prefixes that have been studied in this paper. The Open Biomedical Ontologies seem to be following this strategy. However, since they have large numbers of "reversible" state-changing predicates, they do not seem to be keeping up, and if they can, may find the explosion in the number of terms to be overwhelming. For example, the Gene Ontology currently (file go-basic, version releases/2016-12-24) contains 8 terms that begin with *phosphorylation* (up from 6 in 2014), but only has 3 of the corresponding terms beginning with *dephosphorylation* (unchanged from 2014). A mechanism for dealing procedurally with this kind of prefixation could considerably reduce the maintenance load of biomedical resources like the Open Biomedical Ontologies. Thus, there are considerable potential applications for the kind of insights into the distributional characteristics of the phenomena that are reported on in this paper.

Acknowledgements

The work reported here was supported by NIH grant R21 HS024541-01A1 to Foster Goss, NIH grants LM008111 and LM009254 to Lawrence E. Hunter, and NSF grant IIS-1207592 to Lawrence E. Hunter and Barbara Grimpe. The authors thank Stephen Wu, Wendy Chapman, and the seminar series of the group Information Langue Écrite et Signée of the Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur, Centre National de la Recherche Scientifique for discussions of the work reported here. Cohen's work was supported by Hunter's grants and by generous funding from Labex DigiCosme (project ANR11LABEX0045DIGICOSME) operated by ANR as part of the program « Investissement d'Avenir» Idex ParisSaclay (ANR11IDEX000302).

References

Agarwal, S., Yu, H., & Kohane, I. (2011). BioNØT: a searchable database of biomedical negated sentences. *BMC Bioinformatics*, *12*(1), 420.

Bada, M., Eckhert, M., Garcia, K., Evans, D., Sitnikov, D., Baumgartner, W. A., … Hunter, L. (2009). The Colorado Richly Annotated Full-Text (CRAFT) Corpus: A Resource for BioNLP Research. *Proceedings of the Seventh Rocky Mountain Bioinformatics Conference*.

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. (2001a). Evaluation of negation phrases in narrative clinical reports. *Proceedings of AMIA 2001*.

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001b). A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of Biomedical Informatics*, *34*, 301–310.

Chapman, W. W., Hillert, D., Velupillai, S., Kvist, M., Skeppstedt, M., Chapman, B. E., … Deleger, L. (2013). Extending the NegEx lexicon for multiple languages. In *Studies in Health Technology and Informatics* (Vol. 192, pp. 677–681).

Cohen, K. B., Verspoor, K., Fort, K., Funk, C., Bada, M., Palmer, M., & Hunter, L. E. (2017). The {Colorado Richly Annotated Full Text (CRAFT)} Corpus: Multi-Model Annotation In The Biomedical Domain. In N. Ide & J. Pustejovsky (Eds.), *Handbook of Linguistic Annotation*. Springer.

Cruz Díaz, N., Maña López, M. J., Mata Vázquez, J., & Pachón Álvarez, V. (2012). A Machine-Learning Approach to Negation and Speculation Detection in Clinical Texts. *Journal of the American Society for Information Science and Technology*, *63*(7), 1398–1410.

Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, S. H. (2000). PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, *101*(23), 215–220.

Goldin, I., & Chapman, W. W. (2003). Learning to detect negation with “not” in medical texts. *Proceedings of the ACM SIGIR*.

Grouin, C., Deléger, L., Rosier, A., Temal, L., Dameron, O., Van Hille, P., … Zweigenbaum, P. (2011). Automatic computation of CHA2DS2-VASc score: information extraction from clinical texts for thromboembolism risk assessment. *AMIA ... Annual Symposium Proceedings. AMIA Symposium*, *2011*, 501–10.

Kim, J.-D., Ohta, T., Pyysalo, S., Kano, Y., & Tsujii, J. (2009). Overview of BioNLP’09 shared task on event extraction. *Natural Language Processing in Biomedicine (BioNLP) NAACL 2009 Workshop: Shared Task*.

Kipper, K., Korhonen, A., Ryant, N., & Palmer, M. (2008). A large-scale classification of English verbs. *Language Resources and Evaluation*, *42*(1), 21–40.

Larrivée, P., & Lee, C. (2015). *Negation and polarity: Experimental perspectives*. *Negation and Polarity: Experimental Perspectives* (Vol. 1). Springer International.

Morante, R., Liekens, A., & Daelemans, W. (2008). Learning the scope of negation in biomedical texts. *Empirical Methods in Natural Language Processing*.

Mutalik, P. G., Deshpande, A., & Nadkarni, P. M. (2001). Use of general-purpose negation detection to augment concept indexing of medical documents. *Journal of the American Medical Informatics Association*, *8*, 598–609.

Nédellec, C., Bossy, R., Kim, J.-D., Kim, J. J., & Ohta, T. (2013). Overview of BioNLP shared task 2013. In *BioNLP Shared Task 2013 Workshop* (pp. 1–7).

Névéol, A., & Zweigenbaum, P. (2015). Clinical Natural Language Processing in 2014: Foundational Methods Supporting Efficient Healthcare. *Yearbook of Medical Informatics*, *10*(1), 194–8.

Névéol, A., & Zweigenbaum, P. (2016). Clinical Natural Language Processing in 2015: Leveraging the Variety of Texts of Clinical Interest. *IMIA Yearbook*, (1), 234–239.

Palmer, M., Gildea, D., & Xue, N. (2010). *Semantic role labeling*. Morgan and Claypool.

Palmer, M., Kingsbury, P., & Gildea, D. (2005). The Proposition Bank: an annotated corpus of semantic roles. *Computational Linguistics*, *31*(1), 71–106.

Pyysalo, S., Ginter, F., Heimonen, J., Bjorne, J., Boberg, J., Jarvinen, J., & Salakoski, T. (2007). BioInfer: a corpus for information extraction in the biomedical domain. *BMC Bioinformatics*, *8*(1), 50.

Saeed, M., Villarroel, M., Reisner, A. T., Clifford, G., Lehman, L.-W., Moody, G., … Mark, R. G. (2011). Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database. *Crit Care Med*, *39*(5), 952–960.

Sagot, B., Fort, K., Adda, G., Mariani, J., & Lang, B. (2011). Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé. In *Traitement Automatique des Langues Naturelles*.

Sanchez-Graillet, O., & Poesio, M. (2007). Negation of protein-protein interactions: analysis and extraction. *Bioinformatics (Oxford, England)*, *23*(13), i424-32.

Simpson, M. S., & Demner-Fushman, D. (2012). Biomedical Text Mining: A Survey of Recent Progress. In *Mining Text Data* (pp. 465–517). Boston, MA: Springer US.

Thompson, P., Iqbal, S. A., McNaught, J., & Ananiadou, S. (2009). Construction of an annotated corpus to support biomedical information extraction. *BMC Bioinformatics*, *10*(1), 349.

Uzuner, O., South, B. R., Shen, S., & DuVall, S. L. (2011). 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *Journal of the American Medical Informatics Association : JAMIA*, *18*(5), 552–556.

Velupillai, S., Mowery, D., South, B. R., Kvist, M., & Dalianis, H. (2015). Recent Advances in Clinical Natural Language Processing in Support of Semantic Analysis. *Yearbook of Medical Informatics*, *10*(1), 183–93.

Verspoor, K., Cohen, K. B., & Hunter, L. (2009). The textual characteristics of traditional and Open Access scientific journals are similar. *BMC Bioinformatics*, *10*(1), 183.

Verspoor, K., Cohen, K. B., Lanfranchi, A., Warner, C., Johnson, H. L., Roeder, C., … Hunter, L. E. (2012). A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools. *BMC Bioinformatics*, *13*(207).

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

Wilbur, J. W., Rzhetsky, A., & Shatkay, H. (2006). New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC Bioinformatics*, *7*.

Wu, S., Miller, T., Masanz, J., Coarr, M., Halgrim, S., Carrell, D., … Sohn, S. (2014). Negation’s Not Solved: Generalizability Versus Optimizability in Clinical Natural Language Processing. *PLoS ONE*, *9*(11), e112774.

Yaeger-Dror, M., & Tottie, G. (1993). *Negation in English speech and writing: A study in variation*.

Address for correspondence: Kevin Bretonnel Cohen, kevin.cohen@gmail.com