

Translational Morphosyntax: Distribution of Negation in Clinical Records and Biomedical Journal Articles

K. Bretonnel Cohen^{a,c}, Foster R. 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 School of Medicine, 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 progress notes. Two types of negation are examined: explicit negation at the syntactic level and affixal negation at the sub-word level. The data show that the distribution of negation is significantly different in the two document types, with explicit negation more frequent in the clinical documents than in the scientific publications and affixal negation more frequent in the journal articles at the type level and token levels. All code is available on GitHub¹.

Keywords:

Natural Language Processing; Data Mining; Linguistics

Introduction

Natural language processing (the treatment of human language by computers) is increasingly used in health care and in biomedical research [1–4]. Negation is a frequent cause of errors in language processing [5–9]. This has stimulated a considerable amount of work on negation in clinical documents (e.g. the classic work cited above and recent work summarized in [10]), and a lesser amount of work on negation in scientific literature (e.g. the work summarized in [11–14], as well as scientific literature corpus creation efforts [15–19]). However, while that work has made considerable progress, it has focused 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 *affixal* or *morphological negation*, such as the *de-* in *dephosphorylate* or the *a-* in *afebrile*, meaning *without fever*—has received very little attention in biomedical language processing. This is a gap in the literature, as 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 focused 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, [20] Wu et al. point out that distributional characteristics of linguistic phenomena can have deep implications for evaluating not just individual systems, but also 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, 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, as well as 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 select suitable document sets to use for some specific task type (at the granularity of, say, parsing, coreference resolution, etc.). It can help prioritize module development, 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 [21]. Distributional information can be used as a form of prior

¹<https://github.com/KevinBretonnelCohen/NegationDistribution>

knowledge in machine learning applications that allow supervised under- and over-sampling [22].

Finally, to the authors' knowledge, there is very little 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 and Tottie [23] focused on spoken versus written English and found differences in the distribution of affixal and explicit negation when comparing spoken versus written language. Chapman et al. [6] examined the distribution of explicit negation *within* a genre and found that it may be Zipfian, noting 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 [24]. Cohen et al. [25] compared the abstracts and bodies of full-text journal articles and found that article bodies had a higher percentage of explicit negatives, at 5.3/thousand words versus 3.8/thousand words in abstracts ($p < .01$ by Mann-Whitney-Wilcoxon), concluding that this was relevant to the relative difficulty of information extraction from the two text types. Verspoor et al. [26] made the same measurements in Open Access versus traditional journals, finding no significant difference between them and using that to argue that Open Access journals are representative of the biomedical literature as a whole. Kjellmer studied some of the interesting problems of affixal negation of adjectives in English, such as which adjectives can and cannot be negated (e.g., English has *unkind*, but not *uncruel*), and includes data on distribution of affixal negatives across types of adjectives [27]. Globally, this body of work can be summarized as showing that the distribution of negation is structured, and that it can be shown to vary in interesting ways (or not) 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 addressed at all in the biomedical domain.

Wu et al. [20] 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 focused 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. Using the same processing on both data sources makes the results directly comparable, which has not been the case with previous work.

Methods

Materials

Since the goals of this study are translational in nature, the materials for this work were drawn from the clinical domain and from the biomedical literature: *MIMIC II progress notes* on the one hand [28,29] and the *CRAFT corpus* on the other [30–32]. The rationale behind this choice of domains is that they are close to the opposite ends of the spectrum between bench (the mouse being a common model organism) and the bedside. Other choices could potentially be useful, e.g. journal articles with a clinical orientation, other kinds of clinical documents, etc.

MIMIC II progress notes: Half a million words of physician-written progress notes from the MIMIC II corpus. They reflect the status of patients in the Intensive Care Unit.

CRAFT corpus: A corpus of scientific journal articles in the domain of mouse genomics, previously shown to be representative of the biomedical scientific literature [26].

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 on GitHub and named *negativesEvery10KWords.pl*. (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.)

Affixal negation

All word types in both corpora were collected, and after normalizing for case and punctuation, the number of tokens of each type was counted. To clarify the 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 script *directoryToTypeTokenCounts.pl* on GitHub 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, named in a file on GitHub. 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.

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.

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 listed on the GitHub site: *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 words such as *antiapoptotic* (CRAFT, 6 tokens), *immature* (MIMIC, 177 tokens), and *desaturation* (MIMIC, 72 tokens), which 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 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 Turk²,
2. a subsequent test of a revision of the guidelines on Amazon Mechanical Turk,
3. ...and then review of the guidelines by all authors.

In developing the Mechanical Turk tasks, we took note of the ethical guidelines for crowdsourced linguistic data in [33].

Even after those three rounds of guideline development 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 GitHub.

To ensure the clarity and consistency of the guidelines and the neutrality of the resulting counts, the data was double-annotated. The annotators represent typical readers of the materials in question: an emergency room physician and 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 inter-annotator agreement was 0.94 before resolution, and the entire calculation of agreement is documented 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 affixal negation.

Replicability, repeatability, and reproducibility

All scripts, annotation guidelines for the affix study, and judgments of the individual annotators are on GitHub (see URL at the bottom of the first page). The annotation project and analysis were repeated by an independent third party to ensure that it was replicable. CRAFT is available at bionlp.sourceforge.net. MIMIC II requires a data use agreement, but is freely available.

Results

Explicit negation

The distribution of explicit negatives for the two document collections is shown 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. A Welch 2-sample t-test shows a statistically significant difference, $t = -27.092$, $df = 53.822$, $p\text{-value} < 2.2e-16$.

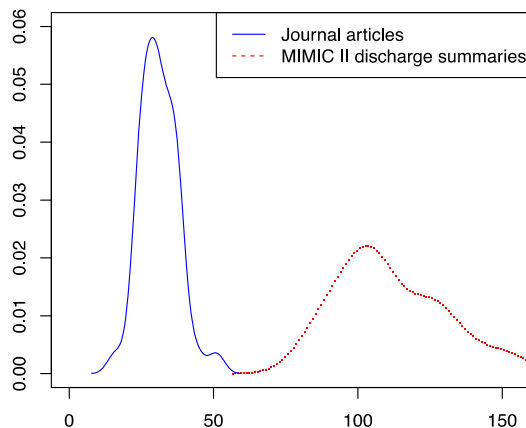


Figure 1 - Density distribution of frequency of explicit negation per 10,000 words, MIMIC II and the CRAFT corpus.

Affixal 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, $df = 1$, $p\text{-value} < 2.2e-16$, with the journal articles having a higher incidence of 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, $df = 2$, $p\text{-value} < 2.2e-16$, 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 1 - Counts and ratios of negated types and tokens.

Corpora and counts	CRAFT	MIMIC
Negated types	650	319
Ambiguous non-negated types	2,641	1,586
Non-ambiguous non-negated types	19,545	16,400
Ratio of negated types to non-negated types	0.028	0.017
Negated tokens	5,575	6,763
Ambiguous non-negated tokens	63,819	70,945
Non-ambiguous non-negated tokens	367,576	84,526
Ratio of negated tokens to non-negated tokens	0.013	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 affixal 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 affixal negation is also different at the

² Words from MIMIC II 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.

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 affixal level, 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

In a 2014 paper with far-reaching implications for the study of negation, and indeed for machine learning in natural language processing in general, Wu et al. [20] concluded 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 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 [34–36] 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. (Van Son et al. [37] gives an idea of what an affixal negation resource might look like, demonstrating the feasibility of the necessary annotation

tasks.) Thus, there are many potential applications for the distributionals that are reported on here.

Reproducibility

The code and data necessary to repeat /replicate this analysis are available for download at the GitHub repository named at the bottom of the first page of the paper. Future work that would be potentially revealing in terms of the reproducibility of the results reported here include at least the following, some of which are variations on the approach and some of which sample different populations:

- Using different sources of negation patterns, such as the most recent set of NegEx patterns, or those that could be mined from negation-annotated corpora, such as BioScope [12] and the BioNLP-ST shared task corpora [11,14].
- Expanding from the MIMIC II physicians' notes to other types of clinical data, from the CRAFT corpus to other scientific domains, and from English to other languages.

Acknowledgements

The work reported here was supported by NIH grants LM008111 and LM009254 to Lawrence E. Hunter, and NSF grant IIS-1207592 to Lawrence E. Hunter and Barbara Grimpe. Cohen's work was supported by Hunter's grants, by grant AHRQ R21HS024541-01 to Foster Goss (as was Goss's), and by generous funding from Labex DigiCosme (project ANR11LABEX0045 DIGICOSME), operated by ANR as part of the program Investissement d'Avenir IDEX ParisSaclay (ANR11 IDEX000302), as well as by a Jean d'Alembert fellowship from the Fondation Campus Paris-Saclay as part of the « Investissement d'Avenir » program operated by ANR. PZ received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant 676207.

The authors thank Tiffany Callahan, who repeated the analysis to ensure that it was replicable, finding a bug in the process. We assure the excellent MEDINFO reviewers that the next paper will be titled *Of mice and men*, as suggested. We also thank Stephen Wu, Wendy Chapman, and the members 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 discussion. Mayla Boguslav, Harrison Pielke-Lombardo, Jingbo Xia, and Elizabeth White gave helpful comments on the manuscript.

References

1. Névéal A, Zweigenbaum P. Clinical Natural Language Processing in 2014: Foundational Methods Supporting Efficient Healthcare. *Yearb Med Inform* (2015) **10**:194–8. doi:10.15265/IY-2015-035
2. Névéal A, Zweigenbaum P. Clinical Natural Language Processing in 2015: Leveraging the Variety of Texts of Clinical Interest. *IMIA Yearb* (2016) 234–239. doi:10.15265/IY-2016-049
3. Simpson MS, Demner-Fushman D. “Biomedical Text Mining: A Survey of Recent Progress,” in *Mining Text Data* (Boston, MA: Springer US), 465–517. doi:10.1007/978-1-4614-3223-4_14
4. Velupillai S, Mowery D, South BR, Kvist M, Dalianis H. Recent Advances in Clinical Natural Language Processing in Support of Semantic Analysis. *Yearb Med Inform* (2015) **10**:183–93. doi:10.15265/IY-2015-009
5. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan B. Evaluation of negation phrases in narrative clinical reports. *Proc AMIA Symp 2001* (2001) 105–109.
6. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A

- simple algorithm for identifying negated findings and diseases in discharge summaries. *J Biomed Inform* (2001) **34**:301–310.
7. Goldin I, Chapman WW. Learning to detect negation with “not” in medical texts. *Proc ACM SIGIR* (2003)
 8. Mutalik PG, Deshpande A, Nadkarni PM. Use of general-purpose negation detection to augment concept indexing of medical documents. *J Am Med Informatics Assoc* (2001) **8**:598–609.
 9. Grouin C, Deléger L, Rosier A, Temal L, Dameron O, Van Hille P, Burgun A, Zweigenbaum P. Automatic computation of CHA2DS2-VASc score: information extraction from clinical texts for thromboembolism risk assessment. *AMIA . Annu Symp proceedings AMIA Symp* (2011) **2011**:501–10. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22195104> [Accessed December 24, 2016]
 10. Uzuner O, South BR, Shen S, DuVall SL. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J Am Med Inform Assoc* (2011) **18**:552–556. doi:10.1136/amiajnl-2011-000203
 11. Kim J-D, Ohta T, Pyysalo S, Kano Y, Tsujii J. Overview of BioNLP’09 shared task on event extraction. *Proc Nat Lang Process Biomed NAACL 2009 Work Shar task* (2009)
 12. Vincze V, Szarvas G, Farkas R, Mora G, Csirik J. The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. *BMC Bioinformatics* (2008) **9**:
 13. Morante R, Liekens A, Daelemans W. Learning the scope of negation in biomedical texts. *Proc Conf Empir Methods Nat Lang Process* (2008) **715–724**.
 14. Nédellec C, Bossy R, Kim J-D, Kim JJ, Ohta T. Overview of BioNLP shared task 2013. in *BioNLP Shared Task 2013 Workshop*, 1–7.
 15. Sanchez-Grailllet O, Poesio M. Negation of protein-protein interactions: analysis and extraction. *Bioinformatics* (2007) **23**:i424-32. doi:10.1093/bioinformatics/btm184
 16. Thompson P, Iqbal SA, McNaught J, Ananiadou S. Construction of an annotated corpus to support biomedical information extraction. *BMC Bioinformatics* (2009) **10**:349.
 17. Pyysalo S, Ginter F, Heimonen J, Bjorne J, Boberg J, Jarvinen J, Salakoski T. BioInfer: a corpus for information extraction in the biomedical domain. *BMC Bioinformatics* (2007) **8**:50.
 18. Wilbur JW, Rzhetsky A, Shatky H. New directions in biomedical text annotation: definitions, guidelines and corpus construction. *BMC Bioinformatics* (2006) **7**:
 19. Agarwal S, Yu H, Kohane I. BioNOT: a searchable database of biomedical negated sentences. *BMC Bioinformatics* (2011) **12**:420. doi:10.1186/1471-2105-12-420
 20. Wu S, Miller T, Masanz J, Coarr M, Halgrim S, Carrell D, Clark C, Chapman W, Bridewell W, Hanbury P, et al. Negation’s Not Solved: Generalizability Versus Optimizability in Clinical Natural Language Processing. *PLoS One* (2014) **9**:e112774. doi:10.1371/journal.pone.0112774
 21. Larrivé P, Lee C. *Negation and polarity: Experimental perspectives*. Springer International Publishing (2015).
 22. Cruz Diaz N, Maña López MJ, Mata Vázquez J, Pachón Álvarez V. A Machine-Learning Approach to Negation and Speculation Detection in Clinical Texts. *J Am Soc Inf Sci Technol* (2012) **63**:1398–1410.
 23. Yaeger-Dror M, Tottie G. Negation in English speech and writing: A study in variation. (1993).
 24. Chapman WW, Hillert D, Velupillai S, Kvist M, Skeppstedt M, Chapman BE, Conway M, Tharp M, Mowery DL, Deleger L. Extending the NegEx lexicon for multiple languages. in *Studies in Health Technology and Informatics*, 677–681.
 25. Cohen KB, Johnson HL, Verspoor K, Roeder C, Hunter LE. The structural and content aspects of abstracts versus bodies of full text journal articles are different. *BMC Bioinformatics* (2010) **11**:492. doi:10.1186/1471-2105-11-492
 26. Verspoor K, Cohen KB, Hunter L. The textual characteristics of traditional and Open Access scientific journals are similar. *BMC Bioinformatics* (2009) **10**:183. Available at: <http://www.biomedcentral.com/1471-2105/10/183>
 27. Kjellmer G. Negated Adjectives in Modern English: A corpus-based study. *Stud Neophilol* (2005) **77**:156–170.
 28. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K SH. PhysioBank, PhysioToolkit, and PhysioNet Components of a New Research Resource for Complex Physiologic Signals. *Circulation* (2000) **101**:215–220.
 29. Saeed M, Villarreal M, Reisner AT, Clifford G, Lehman L-W, Moody G, Heldt T, Kyaw TH, Moody B, Mark RG. Multiparameter Intelligent Monitoring in Intensive Care II (MIMIC-II): A public-access intensive care unit database. *Crit Care Med* (2011) **39**:952–960.
 30. Bada M, Eckhart M, Garcia K, Evans D, Sitnikov D, Baumgartner WA, Ogren P V, Lanfranchi A, Howard A, Corvey W, et al. The Colorado Richly Annotated Full-Text (CRAFT) Corpus: A Resource for BioNLP Research. *Proc Seventh Rocky Mt Bioinforma Conf* (2009)35.
 31. Verspoor K, Cohen KB, Lanfranchi A, Warner C, Johnson HL, Roeder C, Choi JD, Funk C, Malenkiy Y, Eckert M, et al. 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* (2012) **13**:
 32. Cohen KB, Verspoor K, Fort K, Funk C, Bada M, Palmer M, Hunter LE. “The {Colorado Richly Annotated Full Text (CRAFT)} Corpus: Multi-Model Annotation In The Biomedical Domain,” in *Handbook of Linguistic Annotation*, eds. N. Ide, J. Pustejovsky (Springer).
 33. Sagot B, Fort K, Adda G, Mariani J, Lang B. Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé. in *TALN’2011 - Traitement Automatique des Langues Naturelles* Available at: <http://hal.inria.fr/inria-00617067>
 34. Palmer M, Kingsbury P, Gildea D. The {P}roposition {B}ank: an annotated corpus of semantic roles. *Comput Linguist* (2005) **31**:71–106.
 35. Kipper K, Korhonen A, Ryant N, Palmer M. A large-scale classification of English verbs. *Lang Resour Eval* (2008) **42**:21–40.
 36. Palmer M, Gildea D, Xue N. *Semantic role labeling*. Morgan and Claypool (2010).
 37. van Son CM, van Miltenburg CWJ, Morante-Vallejo R. Building a Dictionary of Affixal Negations.

Address for correspondence:

Kevin Bretonnel Cohen, kevin.cohen@gmail.com