Published as: Blake,C. (2014) Biological responses to chemical exposure: Case studies in how to manage ostensible inconsistencies using the Claim Framework, Inconsistency Robustness, Stanford, CA.

Biological responses to chemical exposure: Case studies in how to manage ostensible inconsistencies using the Claim Framework

Our daily decisions concerning which foods, drugs and substances to consume have a direct impact on level of exposure to chemicals. At the national level, policies set at by the Food and Drug Administration and the Environmental Protection Agency impact both our direct and indirect exposures. Although the latter is based on the best available scientific evidence, all of these decisions are made against a backdrop of uncertainty. Consider the phrase *median lethal dose (LD_{50})*, which is used in toxicology to capture the dose of a chemical that is required to kill half of the members in a given population. This measure embodies an intrinsic acceptance of inconsistencies in that it removes the need to explain why half the tested population dies at LD_{50} while the other half of the population survives. In addition to individual responses within the same study, differences between studies are inevitable when studying the highly complex and dynamic relationship between biological responses to chemical exposures, for example LD_{50} for acrylamide ranges between 85 and 1148. Our goal is to foreground this uncertainty, which can be hidden when taking only a global view of the biological responses to chemical exposure. To that end, we have developed the Claim Framework that captures how scientists communicate the results of their empirical study. The Claim Framework comprises four information facets (agent, object, change and dimension) that are pieced together to form five different claim types (explicit, implicit, comparison, correlation and observation). This paper provides case studies drawn from toxicology, medicine and epidemiology to illustrate how the level of abstraction provided by the Claim Framework is sufficient to capture experimental results from a variety of different study designs that are used to measure a biological response to chemical exposure.

1 Introduction

Scientists in medicine, toxicology and epidemiology spend much of their time predicting the biological response(s) that will ensue after chemical exposure. In medicine, the biological responses typically lessen the impact of a disease or injury, while in toxicology the biological response is typically harmful. Despite their differences, both these communities employ controlled experiments to reduce the inherent uncertainty in how an individual will respond to a given chemical; however, the very controls that limit variation within the lab contribute to uncertainty beyond the laboratory where the level and duration of exposure differ from those used in the control group. For example, scientists conducting a clinical trial deliberately select study subjects with the same (typically a single) medical condition, which can be problematic when the treatment is used by the general population who have multiple medical conditions, are of a different age, or have different behavioral factors.

In medicine, Phase Four trials occur after a treatment has been approved by the Federal Drug Administration (FDA). Such trials help to mitigate against the limitations of a laboratory study design by considering how a treatment works in the broader population. This drive to move beyond a laboratory setting is mirrored in toxicology where cumulative risk assessments that consider multiple chemicals are starting to be added to the more than 500+ risk assessments available in the Integrated Risk Information System (IRIS) database (a service provided by the Environmental Protection Agency (EPA)). Cumulative risk assessments can provide insight into the cancelling or synthesis (i.e. much more than additive) effects that multiple chemicals can have on biological responses. However, this change in focus drastically increases the search space of experimental conditions, which can lead to an increase in uncertainty. New ways to conduct experiments in-silica are starting to be explored to bridge this gap, but ultimately the quality of these computational models depends on the quality of the initial parameter settings.

In contrast to the controlled experiments that are employed in the medical and toxicology communities, epidemiologists focus on long term chemical exposures in a naturalistic setting where they can explore chronic health conditions and longer chemical exposure time frames (such as the nurses and teacher's

cohort) or opportunistic studies such as exposure during employment. However, observational study designs do not enable the researcher to control the amount of exposure for both ethical and pragmatic reasons and thus the types of claims that can be made from such studies will differ than those made in a controlled experiment.

It is against this backdrop of uncertainty that federal agencies make public policy that impact our direct and indirect exposure to chemicals. The information synthesis processes used in these agencies consider both acute and chronic levels of chemical exposure from a variety of study designs that differ with respect to the underlying population distributions, assumptions, exposures, and time frames and biological responses. Thus the manual processes employed by the FDA and EPA are the gold standard when it comes to biological responses to a chemical, but the effort required to synthesize new scientific findings into public policy can delay the potential health benefits of such findings in the general community. Moreover, many of the decisions about which foods to eat, drugs to take and toxins to consume, are made by individual consumers who do not have access to the resources or expertise in these agencies to synthesize evidence from scientific literature.

Our goal is to enable scientists, decision makers and consumers to better understand the ostensible inconsistencies that surround a biological response to chemical exposure, which ultimately contributes to our health. We believe that understanding discourse in scientific literature holds the key to achieving this goal, as scientific articles provide an underutilized resource in how scientists conduct experiments and reconcile different results. Although scientist operate in a social context [1] which can manifest in biomedical literature as publication bias, articles have been used for centuries to document and extend findings and are thus critical in order to understand not just the current state of the scientific community, but also shed light into the argumentative structures that support a particular result. It is these structures on which the Claim Framework is based [2]. The Claim Framework draws from work in scientific rhetoric and follows a long tradition in sub-languages, which was first introduced by Harris, a linguist who characterized sentence structures used in a set of immunology articles [3]. The Claim Framework captures how scientists communicate the results of an empirical study and comprises four information facets (agent, object, change and dimension) that are pieced together to form five different claim types (explicit, implicit, comparison, correlation and observation). In this paper we draw examples from medicine, toxicology and epidemiology to illustrate how the semantic and quantitative facts captured in the Claim Framework can provide a way to build systems that are robust in the face of the ostensible inconsistencies involved with an individual's biological response to a given chemical.

2 Related work

Communication patterns used in scientific articles have been studied from several perspectives including computational linguistics, philosophy and have been particularly well-studied with biomedical literature.

2.1 Scientific sub-languages

One of the earliest attempts to characterize language used in scientific communication was conducted by Zellig Harris, who analyzed 14 full-text immunology articles that were published between 1935 and 1970 [4]. His manual analysis of each sentence within these articles revealed clear patterns between word classes and sentence constructions and caused to claim that "we have in science language something new: a number of different sentence types distinguished by their word classes, but all having the same operator argument, i.e. subject-verb-object structure" [4]. Fillmore also focused on enumerating argument types for a given set of verbs based on the premise that "The sentence in its basic structure consists of a verb and one or more noun phrases, each associated with the verb in a particular case relation" [5] and subsequent projects such as FrameNet extends this work.

Harris also observed that operators carry 'evidentiality meaning' [4]. The different claim types in the Claim Framework also capture different levels of evidence. Thus you would expect the study design to influence the way in which a scientist describes their results. For example, studies that use a controlled design (such as a randomized clinical trial) would contain more explicit claims, whereas observational studies (such as a longitudinal study) would contain more implicit or comparison claim types.

2.2 Scientific rhetoric

Scientific communication has also been explored from a rhetorical standpoint that situates a scientific article is within the research community. In particular the comparison claim type is similar to the Create a Research Space (CARS) model, which includes an 'establishing a niche' phase where an author counter-claims to establish a research gap [6], and the Rhetorical Structure Theory includes a contrast schema and antithesis relation that is used between different nucleus and satellite clauses [7]. Comparisons have been mentioned in these earlier study of physics articles where authors compare their results with previous experimental results (see sections 4.3 and 8.1 in [8]), which is also mirrored in Teufel and Moen's contrast category, which includes the action lexicon, better_solution, comparison and contrast [9]. In contrast to these models, we focus exclusively on the results of an empirical result.

2.3 Biomedical text mining

A large body of research has been conducted in biomedical text mining that relates to this work including the Gene Ontology (GO) Consortium (www.geneontology.org), a collaborative effort that provides manual annotations related to cellular components, biological processes and molecular functions of genes. Of the five relationships identified in GO three (regulates, positively_regulates and negatively_regulates) are most similar to the Claim Framework. As with the GO project, the Claim Framework also captures the level of evidence which is done via different claim types. The second area of related biomedical research identifies gene and protein relationships from biomedical literature, which is best summarized in [10]. Another rich resource for existing work is papers written for the Learning Language in Logic (LLL) Challenge, where natural language processing researchers compete for the best method to identify gene protein relationships from MEDLINE abstracts [11].

2.4 Automating the claim framework

At this point, we have developed automated methods to identify explicit and comparison claim types. The approach used to populate explicit claims is strongly influenced by Fillmore's emphasis on the implicit and pre-suppositional levels of communication associated with verbs [5]. Specifically, a combination of semantic features related to verbs and syntactic features were used to identify explicit claims (see [2] for details). Verb categories were generated from the full text biomedicine articles from the Genomics Track of the Text Retrieval Conference (TREC) [12]. An important difference between the general verb-argument structures explored by Harris and Fillmore is that explicit must only consider the findings reported in an article.

In addition to semantic features (i.e. the verbs), the system developed to identify explicit claims employs syntactic features which are similar to those used in the RelEx system, which identifies genes and proteins [13] relationships; however, our approach does not apply the apply tight constraints to terminal nodes used by RelEx where a terminal node must be a gene or a protein or in Rosario and Hearst where terminal nodes must be a treatment or disease [14, 15]. Thus, our automated approach is more similar to ARBITER, a computer program that identifies binding relationships from text [16], which was developed as part of the Semantic Knowledge Representation Project [17] and subsequently extended [18].

We have also developed methods to identify comparison claims automatically. In the initial experiment a binary classifier was used with semantic and syntactic features (similar to the explicit claims) to discern a comparison from non-comparison sentence [19]. Our subsequent work focuses on identifying the specific

agent, object and dimension of change from a given sentence [20]. As with the earlier work semantic and syntactic features are employed, but rather than discerning a comparison from a non-comparison sentence, the system classifies noun phrases as an agent, object or dimension of change.

At this point we have not developed automated methods to identify implicit, observations and correlations. The current implementations do not contain domain specific terms, but experimental results thus far have focused on biomedical texts, with limited work on social science literature[21].

3 Case studies

The best way to explain how the how biological responses to chemicals are captured using the Claim Framework is to provide real examples from actual papers in toxicology, medicine and epidemiology. The case studies below illustrate how scientists describe grapple with uncertainty and why a representational language that reflects that uncertainty must include semantic and quantitative components.

3.1 Toxicology

The toxicology phrase *median lethal dose (LD_{50})*, which describes the dose of a chemical at which half of a tested population dies. An article may provide the dose-response curve (usually as a graphic), and the time-frame in which death occurred, but LD_{50} alone enables a toxicologist to convey to a colleague that a chemical is acutely toxic. This term, which appears in nearly ten thousand MEDLINE sentences, is an interesting example of the toxicology sublanguage because the term does not appear in our general vernacular and because it embodies the inherent uncertainty associated with a biological response to a chemical. Specifically the acceptance that although half of a population will die, the other half will survive and that science is not able to provide the biological responses for an individual (usually mice or rats), but rather can only provide a summary for the overall population.

At first glance LD_{50} values can appear to be inconsistent. Consider acrylamide, which is used in a range of products such as paper and dyes and water treatment plans and is produced naturally when cooking at high temperatures. Acrylamide levels are currently regulated in the US under the Safe Drinking Water Act, which requires that water treatment plants not exceed 0.05 percent dosed at $1mg/L^1$ and reported under California's Proposition 65, which doesn't limit chemical exposure per se, but provides consumers with a list about chemicals that have been deemed carcinogenic (acrylamide was added in 1990) or cause reproductive toxicity (acrylamide was added in 2011). Other exposures to acrylamide, such as the levels in foods are not currently regulated in the US.

Table 1. Ranges of LD_{50} Reported from Acrylamide Exposures (see [22])

Species (route)	LD_{50} in mg/kg bw	References
Rat (oral)	107 – 251	IPCS (3), EU (5), NTP (54)
Rat (dermal)	400	IPCS (3), NTP (54)
Rat (i.p.)	90 – 120	IPCS (3), NTP (54)
Mouse (oral)	107 – 170	IPCS (3), NTP (54)
Guinea pig (oral)	150 – 180	IPCS (3), EU (5)
Rabbit (dermal)	1,148	EU (5)
Cat (i.v.)	85	IPCS (3)

Table 1 shows LD_{50} values for acrylamide ranges, which between 85 and 1,148. More acrylamide is required to induce toxicity in rats through their skin (dermal) than by ingesting the chemical (oral) and toxicity also depends on the species, where rabbits require a concentration that is almost 3 times higher than the rate reported for rats given the same delivery mechanism (dermal). Note also that LD_{50} is reported as a rate (in

this case mg/kg) to control for members of a population who are of a different size. Some might argue that identifying factors that influence LD_{50} is a core mission of toxicology research.

So how might an information system capture the data reflected in Table 1? One strategy is to encode the known factors into Resource Description Framework (RDF) triples, which have become a staple for representing knowledge in the semantic web. The RDF triple shown in (1) is problematic because acrylamide only caused some of the population to die. A better RDF triple would employ the toxicology sublanguage of LD_{50} as a predicate and the chemical and amount as arguments (see 2 and 3).

causes (Acrylamide, death)		(1)
General form:	LD_{50} (Chemical, Amount mg/kg body weight)	(2)
For this instance:	LD_{50} (Acrylamide, 85)	(3)

Additional RDF triples could be added to capture the species and method of delivery from Table 1 that influence LD_{50}, however trying to define all the factors apriori would be perpetually incomplete as factors continue to unfold as new experiments are conducted. For example, sentence (4) below reveals a temporal component to LD_{50} that is not reflected in Table 1.

> The LD_{50} (and 95% CI) were estimated to be 251 mg/kg (203-300 mg/kg) of acrylamide at 24hr post-dosing and 175 mg/kg at 168 hr post dosing. PMID=452021 (4)

Rather than trying to identify every feature that influences LD_{50}, the claim framework focuses on how a scientist, in any discipline, describes the result of their study. For example the results from sentence 4 would be captured as the two separate observational claims:

Claim 1: Claim Type: Observation	Claim 2: Claim Type: Observation
Object: LD_{50}	Object: LD_{50}
$Object_{Modifier}$: acrylamide, 251 mg/kg, 203-300	$Object_{Modifier}$: acrylamide,175 mg/kg, 159-191
$Change_{Modifier}$: 24 hr postdosing	$Change_{Modifier}$: 168 hr postdosing
Source: PMID=452021, section= Results	Source: PMID=452021, section= Results

The claim framework is not the first (and we daresay not the last) effort to annotate rhetoric from scientific articles. The difference is that the claim framework focuses on how scientists communicate results rather than an array of rhetoric types [9, 23] which is critical when working with full-text articles. Several other efforts have taken a Message Understanding Conference approach were the system first identifies entities, which are then used to identify relationships [24-26]. We propose an approach that focuses on the general sentence structures and leaves the unification of noun phrases to subsequent processing steps. From a public policy standpoint, the most important gap in the RDF triples in 1-3 is that the results are de-coupled from the article, which is a sharp contrast to how scientific results are currently used to inform public policy where a direct connection between the claim and the original source document is critical to maintain transparency (see the right hand column of Table 1). A knowledge representation that does not include this link is unlikely to be adopted by agencies who conduct risk assessments in toxicology.

3.2 Medicine

Biomedicine has a long history of honing study designs that mitigate against bias, such as the triple-blind randomized cross-over design where the patient, principal investigator and statistician are all blinded to the treatment and where patients change treatments within the study. Several study designs have become so formulaic that the data that a scientist should report has been operationalized, such as in the CONSORT[2] group who created a 25 point checklist of information that should be reported when publishing a randomized clinical trial [27], and CARE for case reports [28].

Despite agreement on study design and formalization of the data that should be reported, the lack of head-to-head drug comparisons has recently received attention to the point where editors of several leading medical

journals developed a set of recommendations for Comparative Effectiveness Research (CER) who collectively stated that a key challenges was that "CER should directly compare tests or active treatments - so called head-to-head comparisons - of viable clinical alternatives within the current stand or practice (which in some cases may be no intervention)" [29]. Their argument was based in part on an analysis of rheumatoid arthritis articles, where only 5 of 91 trials included a head to head comparison [30]. They go on to say that "Under the current incentives for drug development, we have more medicines and more choices, but we often lack the scientific evidence to make choices among them. The use of "placebo-only practice" reduces risk for the pharmaceutical industry, but clearly it does not translate into good medicine or good public policy." [31]. This discussion underscores the limitations of using science articles that mirrors observations made in the social sciences[1].

The call to action by journal editors underscores the importance of direct comparisons between different drugs, but we have observed that scientists often report their results by comparing subjects who were given a particular drug with a placebo or control group. The claim framework includes a comparison claim type that includes at the two entities being compared and the way in which the entities were compared, which we call the basis of the comparison. Comparison sentences can be further decomposed into gradable comparatives, which enable us to order the entities, for example in sentence (10) tamoxifen (TAM) is less than the control animals with respect to uterine weights. In addition to the entities (depicted with bold and underline), comparison sentences must provide a comparison basis. In sentence (10) and (11) the basis of the comparison is uterine weights (depicted with an underline). Gradable comparative sentences also include a change term (depicted as italic and underline) and may include a set of change modifiers (depicted in italics).

> In the present study, uterine weights of intact animals treated with **TAM**[Agent] was *decreased* as compared with **controls**, although *not significantly*. PMID=12189200 (10)

Non-gradable comparison sentences do not provide the information necessary to order entities with respect to the comparison basis. For example, sentence (11) is a non-gradable comparison because we cannot rank tamoxifen and 4-hydroxytamoxifen with respect to uterine weight. Non-gradable sentences can be further characterized as similar and different, where sentence (11) is a similar non-gradable comparison.

> Since **tamoxifen**[Agent] and **4-hydroxytamoxifen**[Object] had *nearly identical* [Basis Modifier] effects on uterine weight[Comparison Basis], … PMID=10190564 (11)

> … rats treated with **tamoxifen**[Agent] and 4-hydroxytamoxifen, the uterine weights[Comparison Basis], were *decreased*[Change] by *25% (P < 0.05)* [ChangeModifier] compared with the **solvent control group**[Object]. PMID=10190564 (12, claim 1)

> … rats treated with tamoxifen and **4-hydroxytamoxifen**[Agent], the uterine weights[Comparison Basis], were *decreased*[Change] by *25% (P < 0.05)* [ChangeModifier] compared with the **solvent control group**[Object]. PMID=10190564 (12, claim 2)

> **E2 treatment** *increased*[Change] the uterine weight[Comparison Basis], compared with **control animals**[Object] *(P < 0.001)* [ChangeModifier], whereas raloxifene did not significantly affect the uterine weight. PMID=12639932 (13)

One reason that comparisons are not considered is because from a natural language processing perspective comparative structures have earned a reputation of being "notorious for its syntactic complexity" [32] and as a "very difficult structure to process by computer" [33]. However, despite that difficulty from a processing perspective recent efforts have been successful in identifying comparison sentences [19, 34] and our own work is extending these methods to identify the specific agent, object and basis of change [20].

Preliminary results suggest that the basis of the comparison can be a powerful way to summarize research results. For example, sentences 10-13 describe the difference between breast cancer treatments and their effect on uterine weight, which is measured because the uterus is very sensitive to estrogen, and some of the most prescribed breast cancer treatments operate by blocking estrogen. Using these automated methods would enable help to explain the inconsistent results in treatments by providing users with a detailed summary of the specific items that were measured during the controlled study, for example breast cancer

drugs may be similar with respect to their impact on reducing the uterine weight, but very different with respect to actually treating cancer.

3.3 Epidemiology

Epidemiologists employ both descriptive and analytical methods to "study of the distribution and determinants of disease frequency" [35]. In contrast to typical study designs used in medicine and toxicology that allow the researcher to control the amount and duration of chemical exposure, most epidemiologists employ opportunistic strategies to collect data. The field has also developed a sub-language to communicate key findings, such as reporting odds ratios for case-controlled study designs and relative risk or standardized mortality rates for cohort study designs [36] and such data is beginning to be formalized such as in the STROBE standard [37] and meta-analysis [38] for cohort studies.

An observational study design better captures the naturalistic setting in which chemical exposure occurs and can thus be used to mitigate against the limited number of randomized head to head trials [31] and the tight constraints that are imposed during subject selection in controlled trials. However, because the researcher does not specifically control the amount or duration of chemical exposure, nor do they control the range of factors that influence exposure, identifying the characteristics that will be studied apriori is even more unrealistic than in a medical setting. Consider the following results of a study that used a population of men who were exposed to styrene (STY) and butadiene (BD) as part of their employment at a styrene-butadiene rubber plant [39].

> During 1943-1991, the cohort had a total of 386172 and an average of 25 person-years of follow-up, with **3976 deaths observed**[Agent] compared to **4553 deaths expected**[Object] based on *general population mortality rates* [Comparison Basis] (*standardized mortality ratio (SMR) = 87, 95% confidence interval (CI) = 85-90* [Basis Modifier]) PMID= 8901897 (14)
>
> *More*[ChangeDirection] than **expected**[Agent] *leukemia deaths*[Comparison Basis] occurred in the **overall cohort**[Object] (**48**[Agent] **observed/37**[Agent/Object] **expected**[Object], *SMR = 131*[ChangeMod], *CI = 97-174*[ChangeMod]) and among ever hourly subjects (45/32, SMR = 143, CI = 104-191). PMID= 8901897 (15, claim 1)
>
> *More*[ChangeDirection] than **expected**[Agent] *leukemia deaths*[Basis of Comparison] occurred in the overall cohort (48 observed/37 expected, SMR = 131 CI = 97-174[ChangeMod]) and among **ever hourly subjects**[Object] (**45/32**[Agent/Object], *SMR = 143*[ChangeMod], *CI = 104-191*[ChangeMod],). PMID= 8901897 (15, claim 2)

Authors use the standardized mortality rates to report their findings such as SMR=87 (see sentence 14), which means that the number of deaths was lower than the expected. One factor that appears to play a role is where the study subject worked which would influence the amount of chemical exposure, however the researchers had no control over how and when people transitioned between different jobs or how the work environment changed. The highest SMR value of 431 was reported for men who had worked in the laboratories, had worked at the plant for more than 10 years and who were hired more than 20 years ago (sentence not shown).

From a claim framework perspective the claims in sentences 14 and 15 are captured in the claim framework as comparison claim type, but we also see claims in this article reported as an as observations and as an implicit claims where factors such as age of subjects, role at the rubber plant, race, time worked are typically used as agents, the disease is the object and the SMR values including the confidence intervals are captured as change modifiers. Interestingly the results of the study (see 16 and 17) use the explicit claim type:

> … **mortality patterns**[Object] by **race, years worked and process group within the SBR industry**[Agent] did *not* indicate a *causal association*[Change] with occupational exposures. PMID= 8901897 (16)
>
> These results indicate that exposures in the **SBR industry**[Agent] *cause*[Change] **leukemia**[Object]. (17)

Observational studies are particularly prone to publication bias that occurs when the results when an article that finds a relationship between the disease and potential factors are more likely to be published than

studies that do not detect a relationship. Negation, which is captured as a modifier in the claim framework could play a critical role in mitigating against this type of publication bias, as would considering the main body of a full text article. For example the rubber plant study reported the 22 different SMR values but only 7 of those appear in the abstract. Prior work shows that fewer than 8% of the claims appear in the abstract [2], and more importantly that there are systematic differences in information that is reported in the abstract and the full text [40].

4 Closing comments

Biological responses to chemical exposure play an important role in human health. Government agencies such as the FDA and EPA use the results reported in scientific literature to determine if there is an association between individual chemicals and harmful responses and if such an association exists, they work with law-makers to establish limits to protect human health. In addition, our daily decisions about foods, drugs and substances to consume also has a direct impact on our chemical exposure.

We have developed a bottom-up approach to identifying claims from scientific articles that assumes that there exists a sub-language that scientists use to convey their experimental results. The Claim Framework employs four information facets (agent, object, change and dimension) that are pieced together to form five different claim types (explicit, implicit, comparison, correlation and observation) to capture results. In contrast to systems that first identify entities and then try to find relationships that include those entities, we focus first on the claim type, and leave noun phrase unification to latter steps of the process. This late binding approach is well suited for scientific literature where new factors are established with each new experiment. Moreover, maintaining fidelity to the original terms used from an article (or at the very least mapping back to the original terms) is a critical component of this approach, in much the same way that risk assessment in toxicology and meta-analyses in medicine must cite the exact text and maintain the link back to the original source document.

The case studies from toxicology, medicine, and epidemiology presented in this paper illustrate how the claim framework can accurately reflect the information necessary to remove ostensible inconsistencies when working with biological responses to chemical exposure. There is still much work to be done with respect to unifying noun phrases (in particular new kinds of anaphoric references), but systems that employ the claim framework will be able to foreground the factors that influence biological responses to chemical exposure. More importantly, such as system could identify areas where inconsistencies are not explained. These edges of scientific knowledge are precisely the areas in which both scientists and policy makers should focus.

References cited

[1] B. Latour, *Science in action : how to follow scientists and engineers through society*. Cambridge, Mass.: Harvard University Press, 1987.

[2] C. Blake, "Beyond genes, proteins, and abstracts: Identifying scientific claims from full-text biomedical articles," *Journal of Biomedical Informatics,* vol. 43, pp. 173-189, 2010.

[3] Z. S. Harris, *Mathematical structures of language*: Wiley, 1968.

[4] Z. S. Harris, M. Gottfried, T. Ryckman, P. Mattick, A. Daladier, T. N. Harris*, et al.*, *The Form of information in science : analysis of an immunology sublanguage*. Dordrecht Netherlands ; Boston: Kluwer Academic Publishers, 1989.

[5] C. J. Fillmore, "The case for case," in *Universals in Linguistic Theory*, E. Bach and R. T. Harms, Eds., ed New York: Holt, Rinehart and Winston Inc, 1968, pp. 1-90.

[6] J. Swales, *Genre Analysis: English in Academic and Research Settings*: Cambridge Applied Linguistics, 1990.

[7] W. C. Mann and S. A. Thompson, "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization," *Text,* vol. 8, pp. 243-281, 1988.

[8] J. G. Kircz, "Rhetorical structure of scientific articles: the case for argumentational analysis in information retrieval.," *Journal of Documentation,* vol. 47, pp. 354-372, 1991.

[9] S. Teufel and M. Moens, "Summarizing Scientific Articles -- Experiments with Relevance and Rhetorical Status," *Computational Linguistics,* vol. 28, pp. 409-445, December 2002 2002.

[10] D. Zhou and Y. He, "Extracting interactions between proteins from the literature," *Journal of Biomedical Informatics,* vol. 41, pp. 393-407, 2008.

[11] J. Cussens and C. Nédellec. Learning Language in Logic Workshop (LLL05) at the *22nd International Conference on Machine Learning,* Bonn, Germany, 2005.

[12] W. Hersh, A. Cohen, L. Ruslen, and P. Roberts, "Genomics track overview," in *TREC 2007 Working Notes*, 2007.

[13] K. Fundel, R. Kuffner, and R. Zimmer, "RelEx - Relation extreaction using dependency parse trees," *Bioinformatics,* vol. 23, pp. 365-371, 2007.

[14] B. Rosario and M. A. Hearst, "Classifying Semantic Relations in Bioscience Texts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Barcelona, Spain, 2004, pp. 430-8.

[15] B. Rosario and M. Hearst, "Multi-way relation classification: application to protein-protein interaction," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing* Vancouver, British Columbia, Canada 2005, pp. 732-9.

[16] T. Rindflesch, C., J. V. Rajan, and L. Hunter, "Extracting Molecular Binding Relationships from Biomedical Text," in *Proceedings of the 6th Applied Natural Language Processing Conference*, 2000, pp. 188-95.

[17] T. C. Rindflesch, M. Fiszman, H. Kilicoglu, and B. Libbus, "Semantic Knowledge Representation Project," Lister Hill National Center for Biomedical Communications, National Library of Medicine, National Institutes of Health, Department of Health and Human Services, September 25, 2003.

[18] M. Fiszman, T. C. Rindflesch, and H. Kilicoglu, "Summarizing Drug Information in Medline Citations," in *AMIA Annu Symp Proc*, 2006, pp. 254-8

[19] D. Hoon Park and C. Blake, "Identifying comparative sentences in full-text scientific articles," presented at the Workshop on Detecting Structure in Scholarly Discourse at Association of Computational Linguistics, Jeju, South Korea., 2012.

[20] A. Lucic and C. Blake, "Automatically Summarizing Medical Literature," presented at the GSLIS Research Showcase, University of Illinois, 2014.

[21] S. Ahmed, C. Blake, K. Williams, N. Lenstra, and Q. Liu, "Identifying Claims In Social Science Literature," in *iConference*, Fort Worth TX, 2013.

[22] "NTP-CERHR Monograph on the Potential Human Reproductive and Developmental Effects of Acrylamide," Center for the Evaluation and Risks to Human Reproduction, National Toxicology Program, Ed., ed: NIH Publication No 05-4472, 2005.

[23] A. de Waard and J. Kircz, "Modeling Scientific Research Articles - shifting perspectives and persistent issues," in *Conference on Electronic Publishing*, Toronto, Canada, 2008.

[24] T. C. Rindflesch and M. Fiszman, "The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text," *Journal of Biomedical Informatics,* vol. 36, pp. 462-77, December 2003 2003.

[25] J.-D. Kim, T. Ohta, Y. Teteisi, and J. i. Tsujii., "GENIA corpus - a semantically annotated corpus for bio-textmining," *Bioinformatics,* vol. 19, pp. i180-i182, 2003.

[26] S. Van Landeghem, J. Bjorne, C. H. Wei, K. Hakala, S. Pyysalo, S. Ananiadou*, et al.*, "Large-scale event extraction from literature with multi-level gene normalization," *PLoS One,* vol. 8, p. e55814, 2013.

[27] D. Moher, K. F. Schulz, and D. G. Altman, "The CONSORT Statement : Revised Recommendations for Improving the Quality of Reports of parallel-group randomized trials," *Journal of the American Medical Association,* vol. 285, pp. 1987-91, 2001.

[28] J. J. Gagnier, G. Kienle, D. G. Altman, D. Moher, H. Sox, D. Riley*, et al.*, "The CARE guidelines: consensus-based clinical case report guideline development," *J Clin Epidemiol,* vol. 67, pp. 46-51, Jan 2014.

[29] H. C. Sox, M. Helfand, J. Grimshaw, K. Dickersin, P. L. M. Editors, D. Tovey*, et al.*, "Comparative effectiveness research: challenges for medical journals," *PLoS Med,* vol. 7, p. e1000269, Apr 2010.

[30] C. Estellat and P. Ravaud, "Lack of head-to-head trials and fair control arms: randomized controlled trials of biologic treatment for rheumatoid arthritis," *Arch Intern Med,* vol. 172, pp. 237-44, Feb 13 2012.

[31] S. Malozowski, "Comparative efficacy: what we know, what we need to know, and how we can get there," *Ann Intern Med,* vol. 148, pp. 702-3, May 6 2008.

[32] J. W. Bresnan, "Syntax of the Comparative Clause Construction in English," *Linguistic Inquiry,* vol. 4, pp. 275-343, 1973.

[33] C. Friedman, "A General Computational Treatment Of The Comparative," presented at the Association of Computational Linguistics, Stoudsburg, PA, 1989.

[34] M. Fiszman, D. Demner-Fushman, F. M. Lang, P. Goetz, and T. C. Rindflesch, " Interpreting Comparative Constructons in Biomedical Text," in *Proc. 2007 Workshop on Biomedical Natural Language Processing (BioNL'07)*, Prague, Czech Republic., 2007, pp. 37-144.

[35] B. MacMahon and T. F. Pugh, *Epidemiology; principles and methods*, 1st ed. Boston,: Little, 1970.

[36] C. H. Hennekens, J. E. Buring, and S. L. Mayrent, *Epidemiology in medicine*, 1st ed. Boston: Little, Brown, 1987.

[37] J. Poorolajal, Z. Cheraghi, A. D. Irani, and S. Rezaeian, "Quality of Cohort Studies Reporting Post the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement," *Epidemiol Health,* vol. 33, p. e2011005, 2011.

[38] D. F. Stroup, J. A. Berlin, S. C. Morton, I. Olkin, G. D. Williamson, D. Rennie*, et al.*, "Meta-analyses of Observational Studies in Epidemiology," *Journal of the American Medical Association,* vol. 283, pp. 2008-12, 2000.

[39] E. Delzell, N. Sathiakumar, M. Hovinga, M. Macaluso, J. Julian, R. Larson*, et al.*, "A follow-up study of synthetic rubber workers," *Toxicology,* vol. 113, pp. 182-9, Oct 28 1996.

[40] C. Blake, "Using Secondary Information to inform Public Policy," presented at the Internet, Politics, Policy 2012, Oxford Internet Institute, University of Oxford, 2012.

---

[1] See http://www.epa.gov/safewater/consumer/pdf/mcl.pdf

[2] See http://www.consort-statement.org/home/