

# Integrating Coreference Resolution for BEL Statement Generation

Miji Choi<sup>1,2</sup>, Haibin Liu<sup>3</sup>, William Baumgartner Jr.<sup>4</sup>, Justin Zobel<sup>1</sup>, and Karin Verspoor<sup>1</sup>

<sup>1</sup>University of Melbourne, Australia, <sup>2</sup>National ICT Australia (NICTA), Australia, <sup>3</sup>NCBI, Bethesda, MD, USA, <sup>4</sup>University of Colorado School of Medicine, USA  
joc1@student.unimelb.edu.au, {jzobel, karin.verspoor}@unimelb.edu.au  
haibin.liu@nih.gov, William.Baumgartner@ucdenver.edu

**Abstract.** We describe a pipeline system that automatically generates Biological Expression Language (BEL) statements from biomedical journal articles. The system incorporates existing systems for coreference resolution, event extraction, and BEL statement generation. While addressing the BEL track (Track 4) at BioCreative V (2015), we also investigate how incorporating coreference resolution might impact event extraction in the biomedical domain. In this paper, we report that our results evaluated on the training dataset show benefit from integrating a coreference resolution with event extraction.

## 1 Introduction

The BioCreative V (2015) Track 4 addressed the task of extraction of causal network information in terms of the Biological Expression Language (BEL). The BEL track aims to stimulate development of tools that extract events, and generate a corresponding BEL statement. To address the task, we developed a pipeline system which consists of the Turku event extraction system (TEES) [2], an automatic system for generating BEL statements [3], and a coreference resolution module. Coreference resolution has promised to improve system performance in event extraction [1, 4]. We incorporate a coreference resolution system developed for specifically biomedical texts using a simple rule-based approach. In this paper, we introduce the pipeline system in detail, and investigate how incorporating coreference resolution impacts event extraction performance on this task.

## 2 System description and methods

Our pipeline system consists of four components: coreference resolution, substitution, event extraction, and BEL statement generation. Input sentences are processed to identify coreference relations between anaphors and their referring mentions (antecedents). Those anaphors are replaced with their antecedents. Then, the coreference-substituted sentences are entered into an event extraction system, TEES [2], and results of the TEES system are post-processed and converted into BEL statements. In this way, we aim to identify events involving biological entities that are described linguistically using coreference.

**Coreference resolution.** The coreference resolution system identifies coreference relationships between anaphors and antecedents. Input sentences are tokenised and syntactically parsed using the Stanford parser [10]. Anaphors such as pronouns, e.g., *it* and *they*, and definite noun phrases which contain domain-specific nouns, such as *the protein* and *these genes* are identified. Noun phrases are selected as antecedent candidates, and the top ranked candidate is determined as the antecedent, with a set of syntactic and semantic rules.

**Event extraction.** We employ an existing event extraction system, TEES [2, 6], which was the best performing system in the BioNLP 2009 shared task. The system uses a Support Vector Machine (SVM), and classifies interactions between biological entities in the biomedical context. The input texts altered by the coreference substitution step are submitted to the TEES system, and events are detected by the GENIA GE11 event model trained with the BioNLP-ST'11 corpora [5].

**Generation of BEL Statements.** We utilise an existing automatic system for generating BEL statements [3]. The system post-processes and converts the extracted GENIA events into BEL statements as shown in Table 1. To address the BEL task, we added the BEL function *act* that represents molecular activities. For the function, a trigger word *activation* of Positive\_regulation was included for mapping.

**Table 1.** Mapping BioNLP event types into BEL functions

GENIA BioNLP-ST Event Type	BEL Function
binding	complex abundance; <i>p(..)</i>
gene expression	rna abundance; <i>r</i>
localization	translocation; <i>tloc</i>
phosphorylation	phosphorylation; <i>pmod(P)</i>
protein catabolism	degradation; <i>deg</i>
transcription	rna abundance; <i>r</i>
positive_regulation: activation	molecular activity; <i>act</i>

**BioEntity Normalisation.** In the process of generating BEL statements, an embedded protein normalisation component derived from the Liu et al. system [3] normalises protein mentions into concepts in the Protein Ontology. However, we found that this component failed to identify many protein mentions due to various naming conventions. For protein mentions not covered in the Protein Ontology, our system searches the mentions through the fields of symbol, synonyms, alternative names, and description in the resources of HGNC and MGI using exact string matching. Protein mentions not found in those resources, that might be related to other concepts such as MeSH Diseases and ChEBI, are ignored.

### 3 Results

We compare system performance with and without coreference resolution on the training dataset. Since our approach does not specifically require training new models, we use all training data only as test data. Overall results when incorporating coreference resolution are slightly higher than without coreference. We observe an increase in F-score from 17.0 to 27.2, when considering the resolution of personal pronouns, as shown in Table 2. However, Recall and F-score were negatively impacted by the inclusion of coreference resolution for possessive pronouns and definite noun phrases.

**Table 2.** Comparison of performance between an approach with coreference resolution and an approach without it on the training dataset

		Pers. Pronoun			Poss. Pronoun			Definite NP			ALL		
		P	R	F	P	R	F	P	R	F	P	R	F
Without Coref.	Term (T)	51.9	55.4	53.6	62.6	59.2	60.8	69.4	65.4	67.3	60.4	59.1	59.7
	Function-Secondary (Fs)	33.3	7.1	11.8	62.5	27.4	38.1	75.0	40.9	52.9	62.5	24.8	35.5
	Function	33.3	5.4	9.3	46.0	17.5	25.4	60.0	32.1	41.9	48.2	16.9	25.0
	Relation-Secondary (Rs)	58.3	63.6	60.9	78.2	71.7	74.8	90.6	87.9	89.2	75.1	72.8	74.0
	Relation	16.0	18.2	17.0	30.8	34.6	32.6	50.0	45.5	47.6	31.7	30.3	31.0
	Statement	8.2	9.1	8.6	17.7	14.2	15.7	20.8	15.2	17.5	14.9	12.8	13.8
With Coref.	Term (T)	59.2	56.3	57.7	64.4	58.4	61.4	67.7	58.3	62.7	63.1	57.3	60.3
	Function-Secondary (Fs)	42.9	15.8	23.1	71.9	25.6	37.7	78.6	35.5	48.9	67.2	24.8	36.3
	Function	37.5	12.0	18.2	57.1	16.4	25.5	73.3	26.2	38.6	56.3	17.0	36.1
	Relation-Secondary (Rs)	64.7	69.8	67.2	83.1	69.5	75.7	82.9	69.4	75.6	77.4	69.7	73.4
	Relation	27.4	27.0	27.2	38.9	31.2	34.7	45.2	38.8	41.8	36.6	31.5	33.8
	Statement	10.0	9.5	10.0	13.8	9.2	11.1	23.5	16.3	19.3	13.9	10.4	11.9

We report our final results on the test dataset (Table 3). Our system achieved higher precision and lower recall overall as compared to performance on the training dataset. In contrast with the results on the training dataset, the coreference approach reduced system performance except for at the secondary function (Fs) level, since the coreference approach produced more false positives. In a second test phase (Stage 2), gold standard entities for the test dataset were given to focus the analysis on event extraction rather than entity recognition. This produced improvement from roughly 20 F-score to 35 for Statements using coreference.

**Table 3.** Results for the task 1 at the levels of Term, Fs, Function, Rs, Relation and Statement

	Run 1			Run 2			Run 3			Stage 2		
	P	R	F	P	R	F	P	R	F	P	R	F
Term (T)	84.21	21.33	34.04	81.01	21.33	33.77	81.01	21.33	33.77	97.41	37.67	54.33
Function-Secondary (Fs)	75.00	5.36	10.00	80.00	7.14	13.11	83.33	8.93	16.13	69.23	16.07	26.09
Function	75.00	4.55	8.57	60.00	4.55	8.45	42.86	4.55	8.22	72.73	12.12	20.78
Relation-Secondary (Rs)	91.53	26.73	41.38	87.10	26.73	40.91	87.10	26.73	40.91	96.81	45.05	61.49
Relation	60.38	15.84	25.10	57.14	15.84	24.81	55.17	15.84	24.62	75.61	30.69	43.66
Statement	54.35	12.38	20.16	51.02	12.38	19.92	49.02	12.38	19.76	67.61	23.76	35.16

## 4 Discussion

There is a performance difference between the training and the test dataset. The result on the training dataset with the coreference approach shows system performance improvement compared to the results using an approach without coreference resolution. In contrast, the coreference approach slightly reduced the performance on the test dataset, returning additional false positives. There are 10 sentences including 11 anaphors in the test dataset. Our system produced four additional BEL statements associated to 2 sentences among them, and these statements are all false positives due to errors in normalisation of entity mentions to IDs. Our system failed to identify correct identifiers of entities, even though it successfully identified namespaces of entities and a function. We also found that 2 of the 11 anaphors do not involve proteins; these were not identified in the event extraction process due to the limitations of our system. We explore this further below.

**Error analysis.** Our system is limited to identifying events involving gene and protein entities only, due to the reliance on BANNER for entity recognition. There is a notable difference in the results between Stage 1 and Stage 2. With the gold standard entities, the overall performance in Stage 2 improved comparing to Stage 1 (Run 1 in Table 3).

There is also a limitation stemming from which trigger words are used to produce BEL statements. Low Recall at the Function and Function-Secondary Levels in particular shows that our system missed to capture event trigger mentions associated to many BEL functions. When the original BEL generation component was developed, the trigger words were derived from a limited dataset. This list could be greatly enriched by taking advantage of the BioNLP Shared Task 2011 and 2013 data. We will explore expansion of this list in future work.

**Impact of coreference resolution.** Even though the process of coreference resolution resulted in a slight performance reduction in the final result on the test dataset, the coreference resolution approach has the potential for discovery of hidden and complex biological events. For example, the following passage contains an interacting relationship between the personal pronoun *It* and the gene *TIMP-1* implied in the text.

*Interestingly, IL-13 did cause an approximately 80% decrease in pulmonary  $\alpha$ 1-AT expression (Figure 13). It also caused a significant increase in TIMP-1 expression that was seen after as little as 1 day and was readily apparent...*

Our system identifies the coreference relationship between the anaphor *It* and the gene *IL-13* previously mentioned in the text, and automatically substitutes the pronoun with its antecedent. Consequently, the event  $p(HGNC:IL13)$  increases  $p(HGNC:TIMP1)$  is successfully extracted. This could not be extracted without coreference resolution.

**Acknowledgments.** This work was supported by the University of Melbourne, and by the Australian Federal and Victorian State governments and the Australian Research Council through the ICT Centre of Excellence program, National ICT Australia (NICTA). The project receives funding from the Australian Research Council through a Discovery Project grant, DP150101550.

## References

1. Kim, J. D., Ohta, T., Pyysalo, S., Kano, Y., Tsujii, J. I.: EXTRACTING BIOMOLECULAR EVENTS FROM LITERATURE THE BIONLP09 SHARED TASK. *Computational Intelligence*, 27(4), 513-540 (2011)
2. Bjrne, J., Salakoski, T.: Generalizing biomedical event extraction. In *Proceedings of the BioNLP Shared Task 2011 Workshop* (pp. 183-191). Association for Computational Linguistics (2011)
3. Liu, H., Baumgartner Jr. W., Catlett, N., Matthews, A., Roberts, P., Roeder, C., Hooser, A., Ziemek, D., Hunter, L., Cohen, K. B.: Automatic Generation of BEL Statements from Text-mined Biological Events. In *Proceedings of the BioLINK SIG*, 2312-671X, pp. 58, (2013) [http://biolinksig.org/proceedings/2013/biolinksig2013\\_Liu\\_Baumgartner\\_etal.pdf](http://biolinksig.org/proceedings/2013/biolinksig2013_Liu_Baumgartner_etal.pdf)
4. Miwa, M., Thompson, P., Ananiadou, S.: Boosting automatic event extraction from the literature using domain adaptation and coreference resolution. *Bioinformatics*, 28(13), 1759-1765 (2012)
5. Kim, J. D., Pyysalo, S., Ohta, T., Bossy, R., Nguyen, N., Tsujii, J. I.: Overview of BioNLP shared task 2011. In *Proceedings of the BioNLP Shared Task 2011 Workshop* (pp. 1-6). Association for Computational Linguistics (2011)
6. Bjrne, J., Heimonen, J., Ginter, F., Airola, A., Pahikkala, T., Salakoski, T.: Extracting complex biological events with rich graph-based feature sets. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing: Shared Task* (pp. 10-18). Association for Computational Linguistics (2009)
7. Lee, H., Chang, A., Peirsman, Y., Chambers, N., Surdeanu, M., Jurafsky, D.: Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, 39(4), 885-916 (2013)
8. Nguyen, N., Kim, J. D., Miwa, M., Matsuzaki, T., Tsujii, J.: Improving protein coreference resolution by simple semantic classification. *BMC bioinformatics*, 13(1), 304 (2012)
9. Leaman, R., Gonzalez, G.: BANNER: an executable survey of advances in biomedical named entity recognition. In *Pacific Symposium on Biocomputing* (Vol. 13, pp. 652-663) (2008)
10. Chen, D., Manning, C. D.: A fast and accurate dependency parser using neural networks. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (Vol. 1, pp. 740-750) (2014)