

# BioNLP Shared Task 2013: Supporting Resources

Pontus Stenetorp,<sup>1</sup> Wiktor Gólik,<sup>2</sup> Thierry Hamon,<sup>3</sup> Donald C. Comeau,<sup>4</sup> Rezarta Islamaj Dogan,<sup>4</sup> Haibin Liu,<sup>4</sup> and W. John Wilbur<sup>4</sup>  
<sup>1</sup> National Institute of Informatics | <sup>2</sup> French National Institute for Agricultural Research (INRA) | <sup>3</sup> University Paris 13 | <sup>4</sup> National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health  
pontus@nii.ac.jp, wiktor.golik@jouy.inra.fr, thierry.hamon@univ-paris13.fr, {comeau,islamaj,liuh11,wilbur}@ncbi.nlm.nih.gov

## Supporting Resources

- Analyses that can be useful for shared task participants.
- Six participating groups cited usage in 2011.
- Shared effort between organisers and external groups.
- Follows the tradition of the 2009 and 2011 shared tasks.

1. Organisers issued an open call.
2. External groups submitted proposals.
3. Text-only annotation sub-sets made publicly available.

## Parsers and Formats

Name	Format(s)	Model
Enju	PTB, SD, SDC, CoNLL-X	Biomedical
McCCJ	PTB, SD, SDC, CoNLL-X	Biomedical
Stanford	PTB, SD, SDC, CoNLL-X	Combination

Table: Parsers and the model and formats used. PTB: PennTreebank-style, SD: Stanford Dependency-style and SDC: Collapsed SD.

- A variety of formats and parsers.
- Provided by the shared task organisers.

## Parsing Pipeline

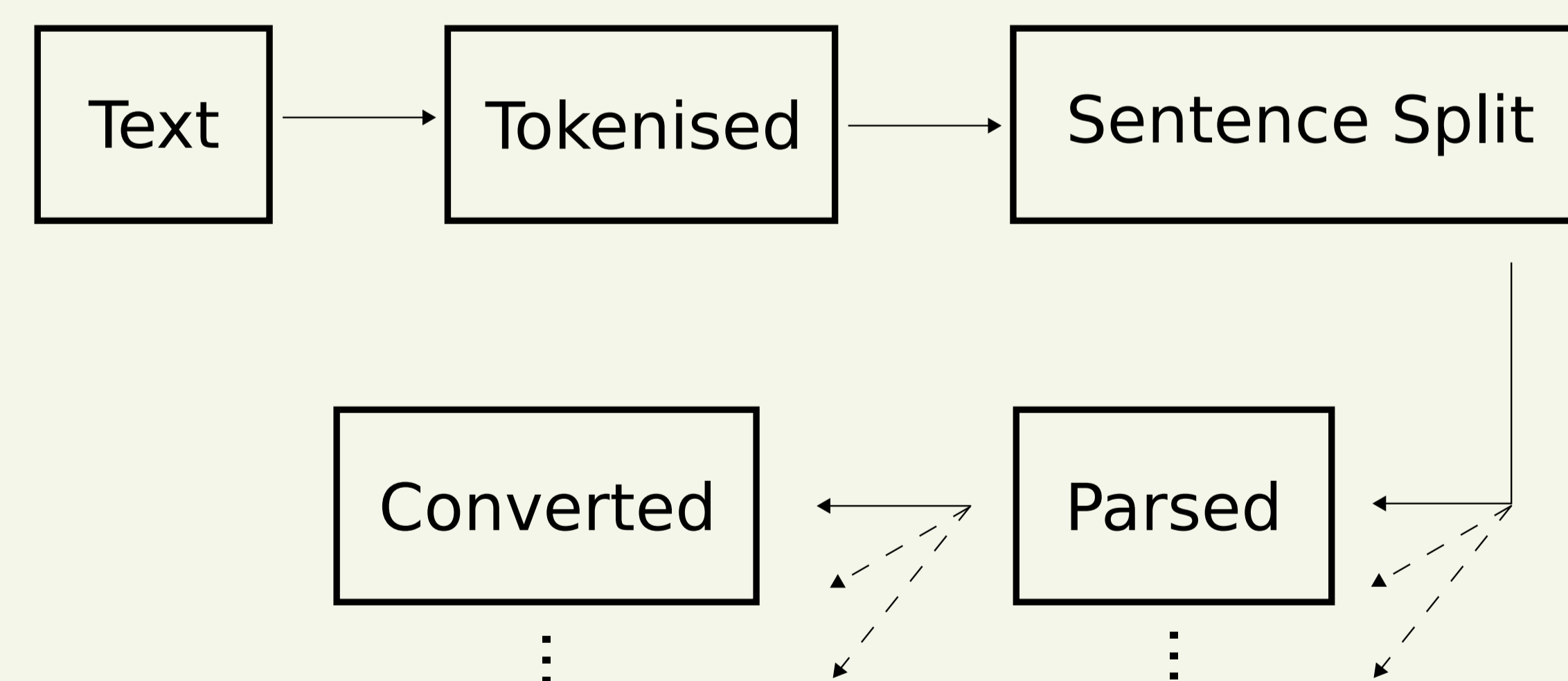


Figure: Parsing pipeline architecture.

- Single shared pipeline between the parsers.
- Allow for interoperability.

Pipeline and documentation available at:

[github.com/ninjin/bionlp\\_st\\_2013\\_supporting](https://github.com/ninjin/bionlp_st_2013_supporting)

## BioC

Biodefense : it 's more than duct tape .  
[NN] [PN] [GE] [RR] [CSN] [NN] [NN]

Figure: Example of annotations induced by MedPost.

Constraint-induced movement therapy is a commonly used intervention to improve upper limb function after stroke .

constraint-induced movement therapy be a common used intervention to improve upper limb function after stroke .  
[JJ] [NN] [NN] [VBZ] [DT] [RB] [JJ] [NN] [TO] [VB]

Figure: Example of annotations induced by BioLemmatizer.

- **BioC**: XML-format addressing tool/platform interoperability.
- Output of two systems supplied in this format.
- **MedPost**: shallow parser.
- **BioLemmatizer**: lemmatiser.
- Provided by members of the National Center for Biotechnology Information.

Tool and documentation available at:

[bioc.sourceforge.net](http://bioc.sourceforge.net)

## BioYaTeA

Behaviour influences cholesterol plasma levels in a pig model .  
behaviour influence cholesterol plasma level in a pig model .  
[Term] [Term] [Term]  
[NN] [VVZ] [NN] [NN] [NNS] [IN] [DT] [NN] [NN] [SENT]

Figure: Example of annotations induced by BioYaTeA.

- Modified version of the YaTeA term extractor.
- Terms, lemmas, parts-of-speech and constituent analysis.
- Two output formats: TSV-based and XML-based.
- Provided by members of the French National Institute for Agricultural Research (INRA) and University Paris 13.

Tool and documentation available at:

[search.cpan.org/~bibliome/Lingua-BioYaTeA/](http://search.cpan.org/~bibliome/Lingua-BioYaTeA/)

## Cocoa

Derivation of normal macrophages from human embryonic stem (hES) cells  
[Process] [Body Part] [Body Part]

for applications in HIV gene therapy.  
[Process] [Organism] [Procedure]

Figure: Example of annotations induced by Cocoa.

- Noun phrase-level entity annotations.
- More than 20 different semantic categories.
- Format compatible with the shared task.
- Provided by S. V. Ramanan of RelAgent Private Ltd.

Web API and documentation available at:

[npjoint.com](http://npjoint.com)

## Availability

Output of all the systems and additional links at:

[2013.bionlp-st.org/supporting-resources](http://2013.bionlp-st.org/supporting-resources)

## Acknowledgements

Special thanks to Richard Johansson for providing his updated *pennconverter* ([github.com/ninjin/pennconverter](https://github.com/ninjin/pennconverter)) and to the Natural Language Processing community at large for maintaining the continuing trend of making tools publicly available.

This work was in part supported by the Quaero programme funded by OSEO (the French agency for innovation). The research of Donald C. Comeau, Rezarta Islamaj Dogan, Haibin Liu and W. John Wilbur was supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM).