

An Ontology for Biological Reaction Events

Yuka Tateisi¹
yucca@is.s.u-tokyo.ac.jp

Tomoko Ohta¹
okap@is.s.u-tokyo.ac.jp

Takako Takai²
taka@nihs.go.jp

Jun'ichi Tsujii¹
tsujii@is.s.u-tokyo.ac.jp

¹ Department of Information Science, Graduate School of Science, University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

² Division of Chemi-Bio Informatics, National Institute of Health Science, Kamiyoga
1-18-1, Setagaya-ku, Tokyo 158, Japan

1 Introduction

Ontologies of biological terminology have been created to provide a model of biological concepts that can be used to integrate heterogeneous information sources such as multiple databases [1, 3, 5]. An ontology has several usage in computational linguistics such as building a large semantic dictionary of the field that can be used for query expansion in information retrieval and deep natural language understanding of the texts in the domain. We aim to construct an ontology to model biomolecular reactions in human. The ontology will be used by biological event information extraction systems from online research papers and documents.

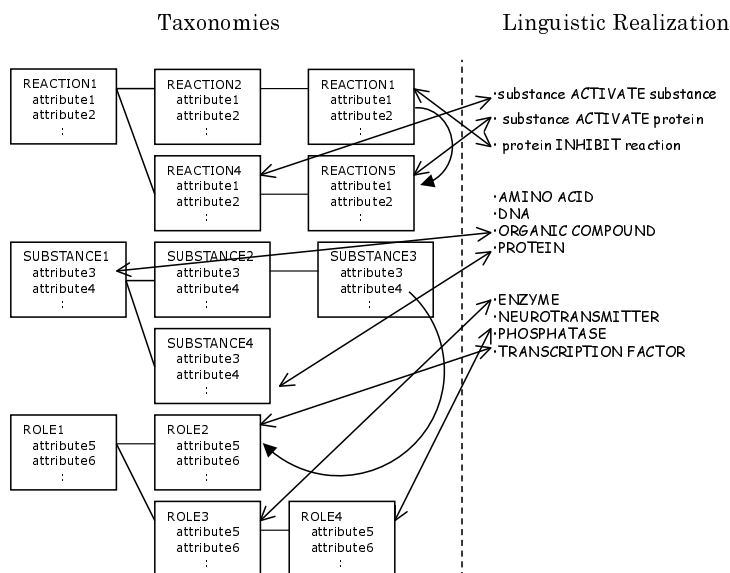


Figure 1: Elements of Ontology.

2 Method

An ontology is a description of concepts and their relations in a domain. Relations include generic relations such as taxonomy (*is-a* relation), synonymy, meronymy (*part-of* relation) and domain-specific

relations such as *is-a-substrate-of*, *is-coded-by*. We construct an ontology by building a set of taxonomies whose nodes are linked with each other by relations other than *is-a* relation. Such a system is naturally modeled by type hierarchies with attribute-value structures and represented by typed feature structures used in HPSG and other unification-based natural language processing systems [2]. We assume that there is a special attribute called *linguistic realization* that links a concept and the names it is denoted by. Our model of concepts and linguistic realizations are illustrated in Fig. 1.

The ontology is being implemented on a typed-feature manipulation language LiLFeS [4]. LiLFeS is a prolog-like programming language that can handle typed feature structures, on which various natural language processing programs are implemented. By using LiLFeS, we aim to seamlessly incorporate the ontology into natural language processing systems. A part of the ontology is shown in Figure 2.

```
name<-[bot]+[LING_REAL\string]. %LING_REAL: linguistic realization
substance<-[name].
role<-[name]+[EVENT\biological_event].
protein<-[substance]./. constr\(\LING_REAL\'protein\').

enzyme<-[protein,role] + [ACTIVE_CENTER\protein_substructure,SUBSTRATE\substance]./.
    constr\(\LING_REAL\'enzyme\' & ACTIVE_CENTER\WHOLE\%0 &
    EVENT\(\enzymic_event&ENZYME\%0)).
```

Figure 2: Ontology Representation in LiLFeS Language.

3 Conclusion

We have started to implement an ontology for biological reaction events. The ontology will be used in combination with natural language processing systems to extract knowledge on biological events and help make the entries of biological reaction databases.

Acknowledgments

This work is partially supported by Japan Society for Promotion of Science (JSPS-RFTF96P00502).

References

- [1] Baker, P.G., Goble, C.A., Bechhofer, S, Paton, N.W., Stevens, R., and Brass, A., An ontology for bioinformatics applications, *Bioinformatics*, 15:510–520, 1999.
- [2] Carpenter, B., *The Logic of Typed Feature Structures*, Cambridge University Press, 1992.
- [3] Kent R.E., Gene Ontology, located at <http://wave.eecs.wsu.edu/Bioinformatics/GeneOntology/GeneOntology.html>
- [4] Makino, T., Yoshida, M., Torisawa, K., and Tsujii, J.: LiLFeS – towards a practical HPSG parser, *Proc. COLING-ACL '98*, 807–811, 1998.
- [5] Schulze-Kremer, S., Ontologies for molecular biology, *Pacific Symposium on Biocomputing '98*, 695–706, 1998.