# An OWL Ontology for HPSG

**Graham Wilcock**
University of Helsinki
PO Box 9
00014 Helsinki, Finland
graham.wilcock@helsinki.fi

## Abstract

The paper presents an OWL ontology for HPSG. The HPSG ontology is integrated with an existing OWL ontology, GOLD, as a community of practice extension. The basic ideas are illustrated by visualizations of type hierarchies for parts of speech.

## 1 Introduction

The paper presents an OWL ontology for HPSG (Head-driven Phrase Structure Grammar) (Sag et al., 2003). OWL is the W3C Web Ontology Language (http://www.w3.org/2004/OWL). An existing ontology is used as a starting point: GOLD (Section 2) is a general ontology for linguistic description. As HPSG is a more specific linguistic theory, the HPSG ontology (Section 3) is integrated inside GOLD as a sub-ontology known as a community of practice extension (Section 4).

## 2 GOLD: A General Ontology for Linguistic Description

GOLD, a General Ontology for Linguistic Description (http://www.linguistics-ontology.org/) (Farrar and Langendoen, 2003) is an OWL ontology that aims to capture "the general knowledge of the field that is usually possessed by a well trained linguist. This includes knowledge that potentially forms the basis of any theoretical framework. In particular, GOLD captures the fundamentals of descriptive linguistics. Examples of such knowledge are 'a verb is a part of speech', 'gender can be semantically grounded', or 'linguistic expressions realize morphemes'." (Farrar and Lewis, 2005).

As far as possible GOLD uses language-neutral and theory-neutral terminology. For instance, parts of speech are subclasses of *gold:GrammaticalUnit* as shown in Figure 1. As GOLD is language-neutral, a wide range of parts of speech are included. For example, both Preposition and Postposition are included as subclasses of Adposition. The classes in the OWLViz graphical visualization (on the right in Figure 1) have been selected from the complete list in the Asserted Hierarchy (on the left).

Originally GOLD was intended to be neutral where linguistic theories had divergent views, but a recent development is the idea of supporting different sub-communities as communities of practice (Farrar and Lewis, 2005) within the GOLD framework. A community of practice may focus on developing a consensus in a specific area, for example in phonology or in Bantu languages. On the other hand, communities of practice may focus on competing theories, where each sub-community has its own distinctive terminology and divergent conceptualization. In this case, the aim is to capture explicitly the relationship between the sub-community view and the overall framework, in the form of a Community Of Practice Extension (COPE) (Farrar and Lewis, 2005). A COPE is a sub-ontology that inherits from, and extends, the overall GOLD ontology. Sub-ontology classes are distinguished from each other by different namespace prefixes, for example *gold:Noun* and *hpsg:noun*.

## 3 An OWL Ontology for HPSG

HPSG OWL is an OWL ontology for HPSG that is currently under development. As the aims of the first version of the ontology are clarity and acceptability,



*Proceedings of the ACL 2007 Demo and Poster Sessions*, pages 169–172,
Prague, June 2007. 

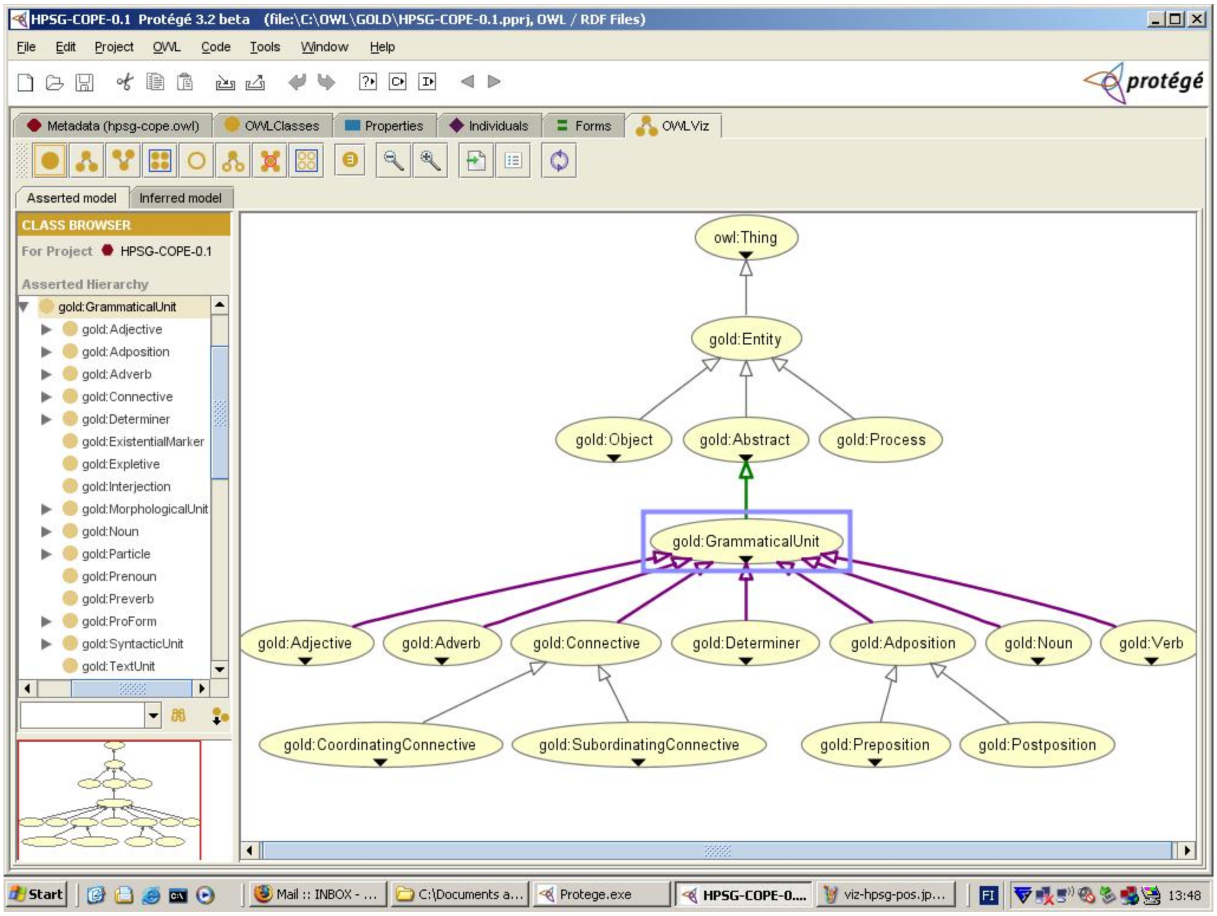


Figure 1: Parts of speech in GOLD

it carefully follows the standard textbook version of HPSG by Sag et al. (2003). This also means that the first version is English-specific, as the core grammars presented in the textbook are English-specific.

In HPSG OWL, parts of speech are subclasses of *hpsg:pos*, as shown in Figure 2. As this version is English-specific, it has prepositions (*hpsg:prep*) but not postpositions. Parts of speech that have agreement features (in English) form a distinct subclass *hpsg:agr-pos* including *hpsg:det* (determiner) and *hpsg:verb*. Within *hpsg:agr-pos*, *hpsg:comp* (complementizer) and *hpsg:noun* form a further subclass *hpsg:nominal*. This particular conceptualization of the type hierarchy is specific to (Sag et al., 2003).

The Protégé-OWL (http://protege.stanford.edu) ontology editor supports both visual construction and visual editing of the hierarchy. For example, if *hpsg:adj* had agreement features, it could be moved under *hpsg:agr-pos* by a simple drag-and-drop (in the Asserted Hierarchy pane on the left). Both the visualization (in the OWLViz pane on the right) and the underlying OWL statements (not shown) are automatically generated. The grammar writer does not edit OWL statements directly.

This is a significant advantage of the new technology over current grammar development tools. For example, LKB (Copestake, 2002) can produce a visualization of the type hierarchy from the underlying Type Definition Language (TDL) statements, but the hierarchy can only be modified by textually editing the TDL statements.

## 4 A Community of Practice Extension

HPSG COPE is a community of practice extension that integrates the HPSG ontology within GOLD. The COPE is an OWL ontology that imports both the GOLD and the HPSG ontologies. Apart from the import statements, the COPE consists entirely of

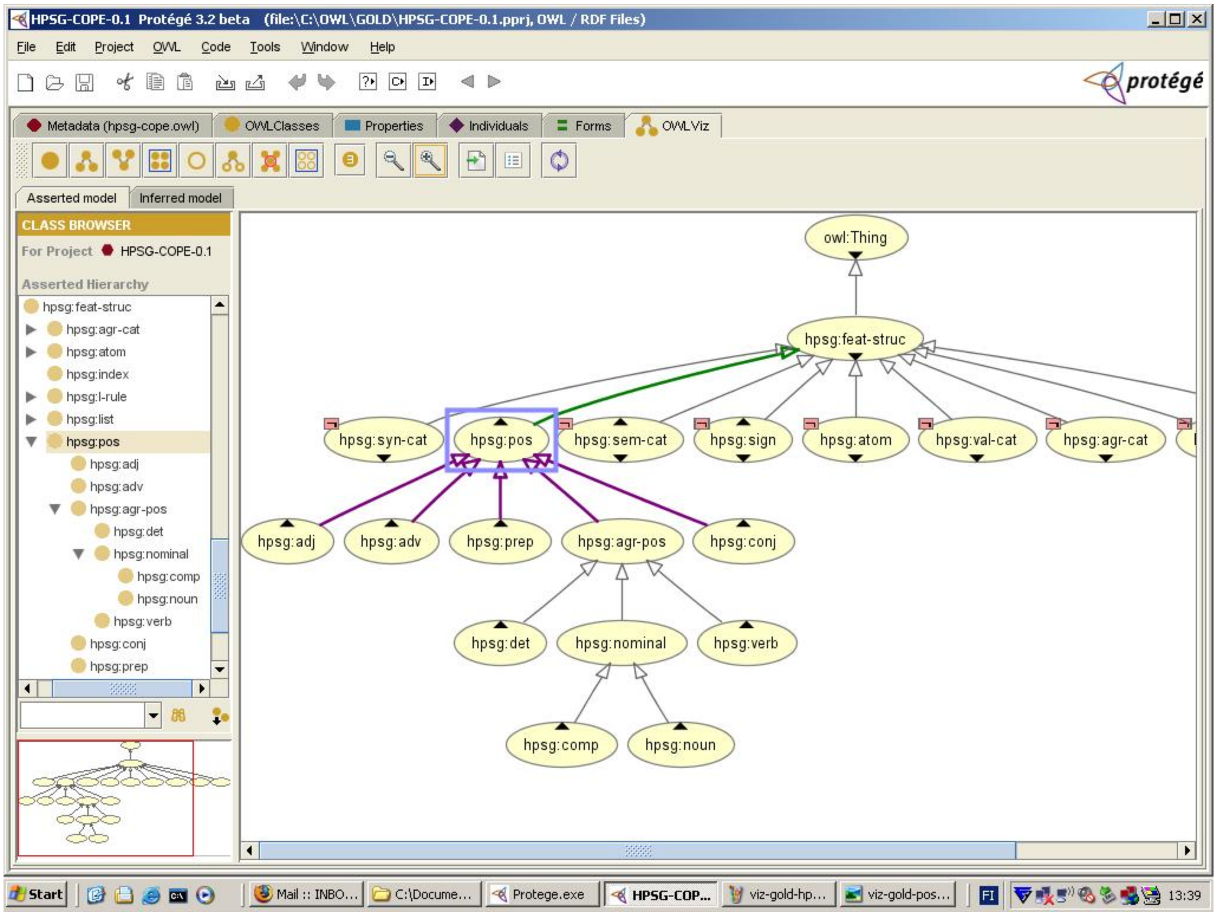


Figure 2: Parts of speech in HPSG

*rdfs:subClassOf* and *rdfs:subPropertyOf* statements. HPSG COPE defines HPSG classes as subclasses of GOLD classes and HPSG properties as subproperties of GOLD properties.

In the COPE, parts of speech in HPSG are subsumed by appropriate parts of speech in GOLD, as shown in Figure 3. In some cases this is straightforward, for example *hpsg:adj* is mapped to *gold:Adjective*. In other cases, the HPSG theory-specific terminology differs significantly from the theory-neutral terminology in GOLD. Some of the mappings are based on definitions of the HPSG terms given in a glossary in (Sag et al., 2003), for example the mapping of *hpsg:conj* (conjunction) to *gold:CoordinatingConnective* and the mapping of *hpsg:comp* (complementizer) to *gold:SubordinatingConnective*.

Properties in HPSG OWL are defined by HPSG COPE as subproperties of GOLD properties. For example, the HPSG OWL class *hpsg:sign* (Sag et al., 2003) (p. 475) properties:

| | |
|---|---|
| PHON | type: list (a sequence of word forms) |
| SYN | type: gram-cat (a grammatical category) |
| SEM | type: sem-struc (a semantic structure) |

are mapped to the GOLD class *gold:LinguisticSign* properties:

| | |
|---|---|
| hasForm | Range: PhonologicalUnit |
| hasGrammar | Range: GrammaticalUnit |
| hasMeaning | Range: SemanticUnit |

by the HPSG COPE *rdfs:subPropertyOf* definitions:

| | | |
|---|---|---|
| hpsg:PHON | subproperty of | gold:hasForm |
| hpsg:SYN | subproperty of | gold:hasGrammar |
| hpsg:SEM | subproperty of | gold:hasMeaning |

## 5 Conclusion

The paper has described an initial version of an OWL ontology for HPSG, together with an approach to integrating it with GOLD as a community of prac-

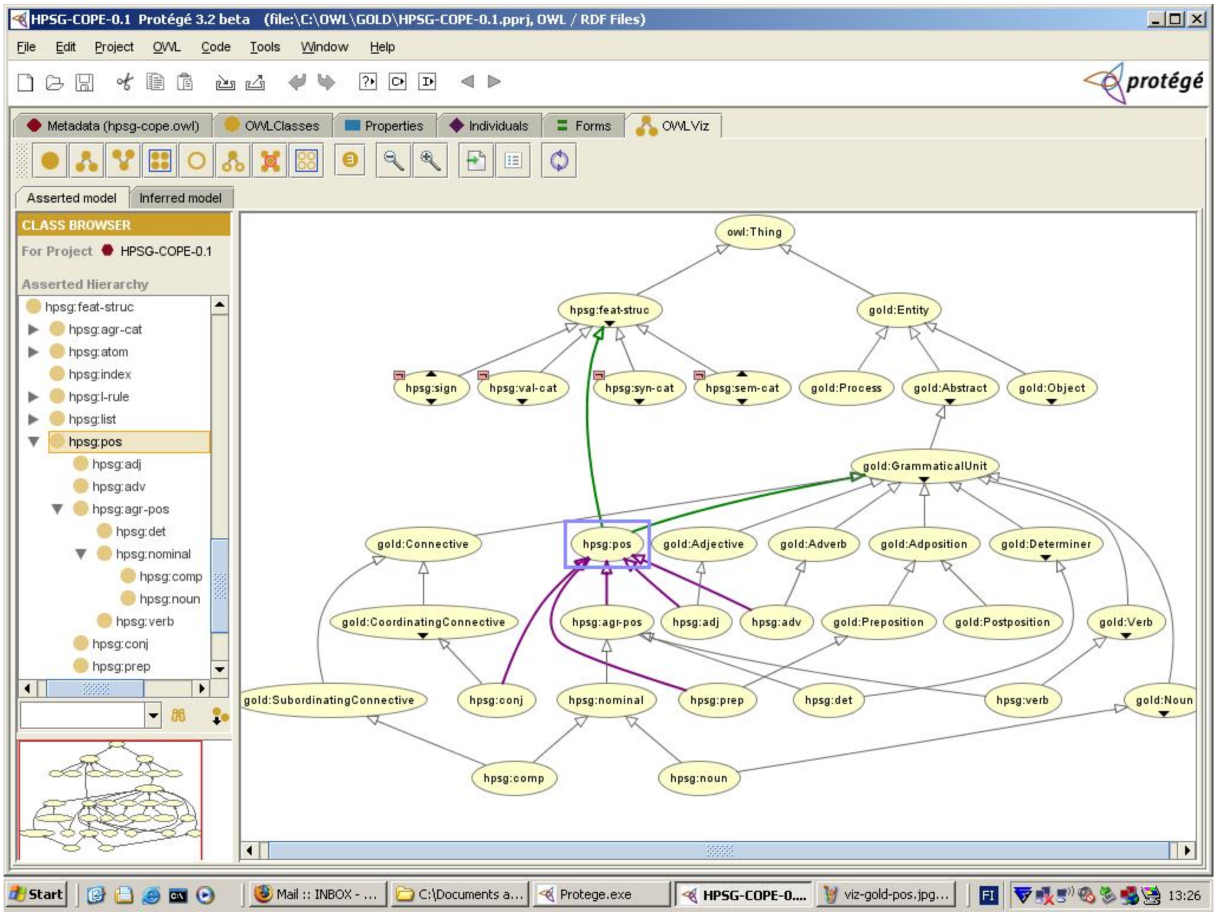


Figure 3: Parts of speech in the Community of Practice Extension

tice extension. Perhaps a rigorous foundation of typed feature structures and a clear type hierarchy makes HPSG more amenable to expression as an ontology than other linguistic theories.

Protégé-OWL supports visual development and visual editing of the ontology. This is a significant practical advantage over existing grammar development tools. OWLViz provides graphical visualizations of any part of the ontology.

OWL DL (Description Logic) reasoners can be run inside Protégé to check consistency and to do cross-classification. One current research topic is how to exploit reasoners to perform automatically the kind of cross-classification that is widely used in HPSG linguistic analyses.

Another current topic is how to implement HPSG lexical rules and grammar rules in the ontology. An interesting possibility is to use the W3C Semantic Web Rule Language, SWRL (Wilcock, 2006).

## References

Ann Copestake. 2002. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford, CA.

Scott Farrar and D. Terence Langendoen. 2003. A linguistic ontology for the semantic web. *GLOT International*, 7.3:97–100.

Scott Farrar and William D. Lewis. 2005. The GOLD Community of Practice: An infrastructure for linguistic data on the web. http://www.u.arizona.edu/~farrar/.

Ivan A. Sag, Thomas Wasow, and Emily Bender. 2003. *Syntactic Theory: A Formal Introduction*. CSLI Publications, Stanford, CA.

Graham Wilcock. 2006. Natural language parsing with GOLD and SWRL. In *RuleML-2006, Rules and Rule Markup Languages for the Semantic Web (Online Proceedings)*, Athens, GA.