

# AN ONTOLOGICAL ACCOUNT OF LINGUISTICS: EXTENDING SUMO WITH GOLD

*Scott Farrar*

Universität Bremen  
FB-10: Sprach- und Literaturwissenschaften  
28359 Bremen, Germany  
farrar@uni-bremen.de

## ABSTRACT

Recent advances in applied ontology and reusable linguistic resources such as WordNet have created exciting possibilities for synergy between knowledge engineering and linguistics. This paper details a General Ontology for Linguistic Description (GOLD), an effort to establish a common domain model for linguistics. GOLD extends the SUMO upper model and is inspired by data from a broad cross section of human languages. These two virtues of GOLD act to minimize language bias in the model. GOLD makes a clear distinction between grammar and semantics and, furthermore, clarifies the connection between linguistic and world knowledge. A detailed summary of GOLD is presented which includes the motivation for the conceptual model along with an explanation of the major entities and relations. By taking into account data from endangered languages, the GOLD model also suggests revisions and expansions to the SUMO category Process, in particular to account for verbal aspect in natural language.

## 1. INTRODUCTION

This paper addresses at least two topics which are of potential interest to the SUMO community: (1) eliminating language bias in ontologies and (2) formulating a general framework for relating the knowledge of language to that of broader world knowledge. As a part of the NSF-funded EMELD project [5], a General Ontology for Linguistic Description (GOLD) has been constructed. GOLD itself was introduced in [6]. The primary aim of the GOLD ontology is to provide a common framework for the comparison of linguistic data gathered from a broad cross section of human languages. By definition, then, GOLD seeks to provide a solution to the issue of language bias in ontology development. The final version of GOLD will provide enough explanatory adequacy to account for diverse linguistic phenomena found, and being discovered, in the world's approximately 6,000 languages.

The use of natural language is always situated within the context of the world around us. The first requirement, then, for a general ontology for linguistics is a theory-neutral model of the world. The creation of GOLD rests crucially on one such model, the Standard Upper Merged Ontology (SUMO) [13]. The current version of GOLD accounts for both language as a physical entity and language as part of psychologically reality. Accordingly, GOLD accounts for, or will account for, the subdomains

of phonology, morphology, syntax, semantics, and discourse. Whereas it has always been the occupation of linguists and philosophers to describe language as a system in its own right, this is the first systematic attempt to build an ontology for linguistics by taking into account a general world model such as SUMO. Other than being useful for language description, GOLD will aid in the creation of resources for natural language processing.

## 2. MOTIVATION FOR THE MAJOR CONCEPTUAL DISTINCTIONS

It has long been realized that a full account of language must take into account two fundamentally different, but related entities: form and meaning. Linguistic expressions are the physical manifestations of language—also called the morphs, or simply the forms. Linguistic expressions include the actual printed words you are reading on this page or the sounds produced when someone speaks. SUMO's existing model is adequate to describe linguistic expressions, with only a few revisions. A linguistic expression can be a PROCESS (signed or spoken) or an OBJECT (written). Therefore in GOLD, three modes of expression are distinguished: WRITTENEXPRESSION, UTTERANCE (the process of speaking), and GESTURE (the process of signing). Ontologically, a WRITTENEXPRESSION is a subclass of SUMO's CONTENTBEARINGOBJECT whereas the latter two are subclasses of PROCESS. For the sake of consistency, it is suggested that SUMO also include CONTENTBEARINGPROCESS as a subclass of PROCESS. The details of CONTENTBEARINGPROCESS are yet to be worked out, so the current discussion focuses on WRITTENEXPRESSION. The following subsumption hierarchy is given for WRITTENEXPRESSION. In Figure 1, Orth- stands for 'orthographic' to emphasize the separation between orthography and other modes of expression.

```

WRITTENEXPRESSION
  ORTHPART
    SIMPLEORTHPART
      ORTHROOT
      ORTHAFFIX
        ORTHPREFIX
        ORTHINFIX
        ORTHSUFFIX
        ORTHCIRCUMFIX
      ORTHSTEM
    ORTHWORD
      SIMPLEORTHWORD
      COMPLEXORTHWORD
      ORTHCOMPOUND
    ORTHPHRASE
  ...
Text

```

Figure 1—Taxonomy of WrittenExpression

The subclasses of WRITTENEXPRESSION relate to one another via a straightforward mereology. One such axiom can be paraphrased as “an ORTHWORD must have a part which is an ORTHROOT”. The level of detail given in Figure 1 is needed to account for linguistic data from a variety of languages, though it is probably too detailed for ordinary NLP tasks. The orthography of Mandarin Chinese, for example, requires no ORTHAFFIX. Depending on the application, written expressions should be classed according to particular languages, e.g., *hat* is an instance of ENGLISHORTHWORD.

All three modes of linguistic expression (written, spoken, and signed) derive their essence from the fact that they are physical realizations of mental linguistic entities. The mental aspects of language are of primary interest to linguists. On the mental level, the units of language combine systematically according to the rules of grammar. The mental objects of interest here include the traditional notion of the morpheme and more complex linguistic constructions. They are subsumed in GOLD under the abstract concept LINGUISTICUNIT.

```

LINGUISTICUNIT
  SUBLEXICALUNIT (BOUND)
    INFLECTIONALUNIT
    DERIVATIONALUNIT
    CLITIC
    BOUNDROOT
    BOUNDSTEM
  LEXICALUNIT (FREE)
    FREEROOT
    COMPLEXLEXICALUNIT
      FREESTEM
      COMPOUND
  PHRASEUNIT
    NOUNPHRASE
    VERBPHRASE
  ...
  SENTENCE

```

Figure 2—Taxonomy of LinguisticUnit

For the moment, the concept LINGUISTICUNIT is subsumed directly under the very general SUMO category of ABSTRACT. This is not very satisfying ontologically, as there needs to be some category in

SUMO to account for mental units of all kinds, not just linguistic ones. The suggestion of a MENTALOBJECT as the superclass of mental entities such as LINGUISTICUNIT is given by the authors of DOLCE [11]. But the issue of mental category is far from being solved, and suggestions are welcome. Just as the subclasses of WRITTENEXPRESSION may be further classified into language specific entities, LINGUISTICUNIT also subsumes language specific classes. Examples of instances of LINGUISTICUNIT are given in (1).

- (1) (instance Un EnglishPrefix)  
(instance Kapu SwahiliRoot)

These instances should not be confused with their corresponding orthographic forms *un-* and *-kapu-*. One advantage of separating linguistic expressions from mental units is that it becomes possible to talk about the notion of an underlying form, e.g., *Simply* as an underlying form for *simplify*. Various theories approach the notion of an underlying form differently, but GOLD provides at least the basic constructs for dealing with them ontologically.

The taxonomy in Figure 2 is organized based on very general morphosyntactic and syntactic units of grammar. These are basic linguistic notions such as the morpheme, the word, and the phrase. The taxonomy of LINGUISTICUNIT is in the spirit of strongly typed theories such as Categorial Grammar or Head-Driven Phrase Structure Grammar. Individual applications of GOLD could be tailored to suit specific languages/theories. For example, a taxonomic hierarchy for the grammar of Mandarin Chinese might include the following.

```

MANDARINLEXICALUNIT
  NOUN
    TIAOCLASSNOUN
    BENCLASSNOUN
    ...
  VERB
    INTRANSITIVEVERB
    TRANSITIVEVERB

```

Figure 3—Possible Taxonomy for Mandarin Lexical Units

The point is that a language specific hierarchy like that in Figure 3 reflects language specific grammatical constructs, in this case, part of speech and valence. (Note, the current version of SUMO assumes that categories such as Noun, Verb, etc. are universal, an assumption which does not obtain empirically for all languages.) The concept TIAOCLASSNOUN refers to a formal category used specifically in Mandarin to classify nouns. So, for example, the classifier *tiao* is used as a measure word with nouns denoting long thin things. However, it is not immediately clear that members of the *tiao* class share a common semantics, as in (2).

- (2) yi tiao quan (one boat)  
yi tiao yu (one fish)

The problem is that noun classifiers, as pointed out by [1], may reflect purely formal criteria, rather than purely semantic; although grammars usually reflect a

combination thereof. It is an empirical question whether Mandarin nouns are classified via purely formal means, and the present work takes no particular stance. The example was merely given to illustrate that the GOLD gives the individual researcher the tools for encoding such a theoretical decision in a way such that disparate language descriptions may be rendered interoperable. So, another advantage of having the concept LINGUISTICUNIT is that organization based on purely formal criteria may be separated from organization based on semantics. Each LINGUISTICUNIT has a semantic pole, just as if has a phonological or orthographic pole. (Semantics will be the focus of Section 4).

A WRITTENEXPRESSION is related to LINGUISTICUNIT by the REALIZES relation. The various relations between subclasses of LINGUISTICUNIT itself concern the domain of syntax. The current version of GOLD does not include a treatment of syntax, although future research will focus on a general framework for syntactic relations and for representing grammars.

### 3. LINGUISTIC FEATURES IN GOLD

The organization of the categories subsumed by LINGUISTICUNIT pertains to a grammatical system. The individuals, therefore, are said to carry various linguistic attributes, or ‘features’, as they are commonly called in linguistics [15]. Thus, GOLD contains the concept LINGUISTICFEATURE which is subsumed by the SUMO category INTERNALATTRIBUTE. This category is in parallel with other attributes like SHAPEATTRIBUTE or PHYSICALSTATE.

```
INTERNALATTRIBUTE
  SHAPEATTRIBUTE
  ...
  LINGUISTICFEATURE
    PHONOLOGICALFEATURE
    ...
    MORPHOSYNTACTICFEATURE
      PARTOFSPEECHFEATURE
      LINGUISTICCLASSFEATURE
      AGREEMENTFEATURE
```

Figure 4—Taxonomy of LinguisticFeature

There are several varieties of features which linguists find useful for language description, including phonological, morphosyntactic and semantic. (See [5] for details.) A MORPHOSYNTACTICFEATURE is one which has the special status of applying most often to sublexical, or bound, linguistic units. Tense for example is a kind of MORPHOSYNTACTICFEATURE whose instances include PastTense, RelativeFutureTense, PostHodiernalTense, etc., each having a clearly defined semantics (discussed in the next section). For another example, take PARTOFSPEECHFEATURE. In English there are the instances Noun, Verb, Adjective, Adverb, etc. Though for some other language, only a subset of these will be appropriate, and the same feature will usually imply a different grammatical behavior of the LINGUISTICUNIT. Linguistic features cannot exist without the linguistic units to which they inhere. Features may be used to create a ‘feature system’, one way to represent a grammar [12].

The notion of a feature system is defined in terms of the GOLD ontology in [7].

### 4. SEMANTICS IN GOLD

In the previous sections, the notion of semantics has been alluded to in the context of linguistic units and various linguistic features. To begin the discussion of how GOLD treats semantics, the presentation of some basic philosophical assumptions concerning meaning is in order.

*Assumption 1: humans perceive, act in, and conceptualize their environment in the same way regardless of which language they happen to speak.*

That is, there is no linguistic relativism, at least in its strong form. Linguistic relativism, usually attributed to Sapir and Whorf, is the idea that a person’s linguistic system can influence their cognitive system and, therefore, the way in which they conceptualize reality. In terms of ontology, strong linguistic relativism would imply that there could be no common upper model to mediate between various languages, rendering machine translation impossible.

*Assumption 2: two humans must share a common linguistic system in order to communicate.*

Assumption 2 entails that two people must share the same language in order to understand each other: the same linguistic expressions (the same forms), the same mental linguistic units (roughly, the same grammar), and also a common semantics. That is, there is such a thing as semantic structure and it varies from language to language just as the other subsystems of language vary.

*Assumption 3: The limits of human cognition and the structure of the world itself may affect what kinds of structures are found in language, in particular what kinds of semantic structures can exist.*

In terms of human cognition, it is reasonable to expect that no language, for example, grammatically encodes an object’s infrared characteristics, at least partly due to the fact that the perceptual apparatus of humans does not perceive infrared light. As for the structure of the world itself, gravity may explain why there is a difference between ‘up’/‘down’ and ‘left’/‘right’ in various semantic systems.

*Assumption 4: there are no ineffable concepts.*

That is, there are no concepts present in commonsense reality (in an ontology) that cannot be put into words, somehow. Whereas some concepts are lexicalized, i.e., expressed as a single linguistic unit, others are grammaticalized, i.e., part of the closed class of linguistic units. There are other concepts for which there are neither words nor closed class items and, hence, must be expressed periphrastically.

The particular variety of semantics being pursued in GOLD is in agreement with the basic philosophy of Jackendoff [8] on semantic structure and the general framework of cognitive linguistics [16]. What then is a semantic system? A semantic system imposes a view of

the world on the grammatical system (not the speaker's cognition itself). A semantic system defines what aspects of a speaker's reality are expressed in the reporting of some conceptualized situation, or how the speaker represents the world. It seems to be a universal that a semantics consists of two subsystems, one lexical and the other grammatical [16]. Lexicalized concepts show up as lexical units. Grammaticalized concepts show up either as sublexical (bound) units or as grammatical constructions in the syntax.

Languages vary according to what concepts are lexicalized. Compare, for example, the word *wall* in English to the corresponding Italian words *muro* and *parete*. In the linguistic encoding of the four entities which hold up a house, Italian must indicate whether the outside part (*muro*) or the inside part (*parete*) is relevant. Also, consider that in many languages, there are only three words for the various colors: the equivalent of 'black', 'white', and 'everything else'. Perhaps an even more striking example of how the semantics of lexicalization may vary from language to language is found in Atsugewi. In Atsugewi objects—usually expressed as nouns in more familiar languages—are expressed as predicates. Thus, one cannot merely refer to 'a rock' or 'water', but a single lexical unit must package an object's type (in the sense of ontology) and how it happens to be moving (what kind of event it is participating in) [16]. Semantically, this means that the ontological type of an object is not referenced as is so common in Indo-European languages. Instead, the linguistic system is sensitive to such conceptual distinctions paraphrased, for example, as 'hard-thing-falling' and 'hard-thing-sliding'. The problem of lexicalization is well known in lexicography and machine translation, in particular that of how to map the lexical "space" of one language onto another. Another characteristic of lexical systems is that they are open ended with new items being lexicalized all the time.

There is even more variation in how languages grammaticalize certain concepts. Various candidates for what gets encoded grammatically include shape, possessability, how far an object is from the speaker, or when an event occurs in relation to some reference point. Whereas some grammars are neutral to whether a particular concept is expressed, others require it. For example, [17] analyze the differences in the semantic systems of Lakhota and English. Consider the scene expressed in English as: *A man broke a window with a rock*. In English, indicating whether the window was broken from a distance or up close is optional (cf. *smash* and *throw*). In Lakhota, such information is required in all sentences which describe actions, using either the verbal prefix *wo-* (action from a distance, as in shooting or throwing) or the prefix *ka-* (action performed in close contact). Similarly, whereas English expresses past tense by the single suffix *-ed*, a language like Yandruwandha has five kinds of past tense distinctions based on remoteness in time [4].

The point is that a given semantic system will emphasize particular aspects of a speaker's reality and

ignore others. The above examples generally concern two ontological parameters: 'granularity' and 'aspect'. Granularity refers to the amount of semantic detail encoded, as illustrated by the Italian example. Aspect on the other hand concerns the ontological vantage point from which the encoding is made, as in the Atsugewi example. Both granularity and ontological aspect create problems for ontological modelling. That is, to account for all the semantic distinctions which are found in natural language, an ontology should be highly granular. The problem of ontological aspect is more troubling. As reality can be viewed differently by various semantic systems, the ontology should perhaps consist of a lattice of upper models instead of one all-encompassing model. Neither of these solutions are possible from the standpoint of implementation. The remainder of this section offers a solution.

Considering the assumptions laid out in this section along with the above empirical observations, the category of SEMANTICUNIT is proposed as a kind of cognitive "container" for a linguistic unit's basic meaning. That is, a SEMANTICUNIT mediates between the linguistic form and the concept it designates. This conception of meaning is consistent with the various two-level approaches to meaning, as described in [2]. While this does not avoid the problem of granularity, having to proliferate categories with every new language that is taken into account, it does avoid the more serious problem of ontological aspect. That is, aspect may be represented at the level of SEMANTICUNIT instead of in the overall organization of the upper model. An instance of SEMANTICUNIT is information, so it is included in GOLD as a subclass of SUMO PROPOSITION.

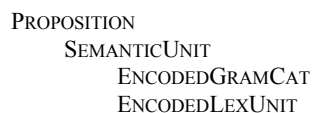


Figure 5—SemanticUnit in the Ontology

Distinctions in the taxonomy should be motivated from empirical evidence. The question that should be asked to motivate a semantic category is, how does such a distinction affect the grammar? The most obvious and universal distinction is the distinction between grammatical categories and lexical units. The former constitutes the meaning of the inflectional system of a language, while the latter constitutes the various lexical senses. Common grammatical categories found in the world's languages include tense, mood, aspect, case, etc. It is not the immediate aim of GOLD to characterize the concept TENSE or CASE. Rather, the meaning units corresponding to these general linguistic notions, e.g., PastTense or PerfectiveAspect, may be formulated in GOLD. Consider the various tense units. All such units encode a temporal relation between the time of utterance and some other reference point. Thus the meaning components of the various tenses are instances of ENCODEDTEMPORALRELATION, which is itself a subclass of ENCODEDGRAMCAT. This aspect of GOLD is the

subject of ongoing research within the EMELD project and the various instances have yet to be axiomatized.

As for ENCODEDLEXUNIT, a criterion for creating a taxonomy is to create distinctions only if it affects the grammar. Consider languages such as Ojibwa or Tamil which have lexical classification systems like Mandarin, but which are based solely on semantic grounds [1]. Such language specific hierarchies may be represented in the taxonomic structure of, for example, TAMILLEXUNIT. A taxonomic criterion for English noun senses could be the difference between mass and count items, since the mass/count distinction affects how lexical unit are used in the grammar, e.g., *some water*, but not *#some chair*. Granularity may be handled at the level of ENCODEDLEXUNIT. That is, given some basic upper model, such as SUMO, finer distinctions maybe specified in the semantics if they are not already present in the overall model. Consider the axiom in (3) which defines the sense for the Italian word *parete*, again meaning roughly ‘inside part of a wall’.

```
(3) (=>
      (instance Parete ItalianLexUnit)
      (exists ?X ?Y
        (and
          (part ?X ?Y)
          (instance ?Y Wall)
          (part ?Y Room))))
```

This axiom is probably incorrect, but it does illustrate how the level of SemanticUnit is mapped onto existing concepts in the SUMO. Instances of LINGUISTICUNIT relate to instances of SEMANTICUNIT via the HASSENSE relation.

The entries in the WordNet lexical database most closely correspond to GOLD instances of ENCODEDLEXUNIT. It is at this level of the ontology where the familiar WordNet-style relations are defined, for example, antonym, synonym, troponym, etc.

```
BINARYPREDICATE
LINGUISTICRELATION
  ANTONYM
  SYNONYM
  TROPONYM
  HYPERNYM
  HYPONYM
```

Figure 6—Instances of LINGUISTICRELATION

Notice that HYPERNYM is defined separately from SUBCLASS which gives the developer the power to separate the strictly semantic notion of ‘is-a’ from the formal SUBCLASS relation. In this way language-specific hierarchies like the one in WordNet can be preserved while still maintaining ontological consistency via the link from a linguist unit to its semantics. Whereas linguistically, it makes sense to encode that “A dog is a pet.” and “A dog is an animal”, ontologically only the latter makes sense. It is not required to include the hypernym/hyponym relations for a specific language, though it may be useful for NLP.

## 5. SUGGESTIONS FOR SUMO

When considered against the enormous variety of linguistic data, the SUMO upper model provides a solid conceptual basis for a notional account of grammatical categories. However, based on linguistic evidence, SUMO is deficient in at least one domain, that of PROCESS. The following section offers a general critique of PROCESS and suggests revisions based on linguistic evidence.

To begin, the term *process* is misleading to most linguists. A process, at least in English, refers to some kind of state of affairs which is ongoing. The category label is used in SUMO, however, to refer to all kinds of states of affairs, not just those which are ongoing. Therefore, some general term to replace PROCESS is needed. The term *event* has similar misleading connotations, though it is often used by linguists to refer in general to things which “takes place”. The term *perdurant* is suggested as a neutral alternative, after [11]. Terminological issues aside, the notion of what distinguishes a SUMO PROCESS from a SUMO OBJECT is not at all clear. There needs to be some ontological assurance in the form of an axiom, or axioms, that a PROCESS is a temporal entity, one whose parts are not present all at once but unfold in temporal phases. This is especially important linguistically as will be discussed below. Finally, the criteria for organizing the various subclasses of PROCESS is unclear. On the one hand, they are organized based on the number of participants. But this is not always the case. For example, some subclasses are distinguished by what kind of change takes place. A case in point is DAMAGING which is classified as INTERNALCHANGE. DAMAGING usually, however, involves two participants, and should at least cross classify as a DUALOBJECTPROCESS. The following linguistic observations may help solve these and other problems related to the SUMO category PROCESS.

The grammatical category ‘aspect’ (not to be confused with ontological aspect) relates most directly to PROCESS. Aspect captures “the different ways of viewing the internal temporal constituency of a situation” [3]. In other words aspect describes the “internal temporal shape of events or states” [14]. Aspect as such often shows up formally as attached to the verb and sometimes has formal interaction with the tense system of a language. The most common contrast in an aspectual system is the perfective/imperfective distinction, which is also referred to as viewpoint aspect. Various Slavic languages are prototypical cases of grammaticalized aspect, however, English examples will be given for brevity.

- (4) a. John read the book (and finished it).  
 b. John was reading the book when the phone rang.

That is, the reading PROCESS is either viewed as a whole as in (4a) or as ongoing as in (4b). Depending on the type of PROCESS being described, there is an interaction with certain time expressions.

- (5) #The was breaking when the phone rang.  
 #John reached the finish line for an hour.

To account for the type of ungrammaticality in (5), linguists often categorize predicates according to their ‘aktionsarten’. This term essentially refers to the PROCESS type, using the language of SUMO. Linguists usually credit Zeno Vendler for the various action types, thus they are known as the Vendler.

```

EVENT
  ACHIEVEMENT
  ACCOMPLISHMENT
STATIVE
  STATE
  PROCESS

```

Figure 7—The Vendler classes

For grammatical reasons as shown in (5), one needs to classify predicates according to the type of PROCESS they denote. The domain of PROCESS in the current version of SUMO is partially inspired by the verb classification of Beth Levin [10]. Levin’s classification is the result of a very thorough examination of the semantic and syntactic properties of English verbs. However, Levin’s efforts are ultimately specific to English and should be reserved for an organization of English verbal semantic units. Recent work in the EMELD project suggests that the SUMO PROCESS hierarchy be revised according to the following scheme, here called PERDURANT hierarchy.

```

PERDURANT
  STATE (most time stable)
    e.g., KNOWING
  OCCURRENCE (less time stable)
    PROCESS
      e.g., RUNNING
      e.g., FIXING
    EVENT
      CULMINATEDEVENT
        e.g., RIPENING
      PUNCTUALEVENT
        e.g., BREAKING

```

Figure 8—Proposed Perdurant taxonomy

Thus, PERDURANT is organized by the particular time stability criteria of its subclasses. This characterization deviates from the traditional Vendler classes as it takes into account more than just linguistic predicates. Note that here PROCESS refers only to on-going activities such as those expressed by ‘run’ or ‘fix’. The class EVENT is reserved for entities which entail a change of state.

Also found in the aspectual systems of many languages are grammatical categories which refer to specific temporal parts or ‘phases’ of a PERDURANT. Consider the following English examples.

- (6) a. John started to read the book.  
 b. John finished reading the book.

(6a) exemplifies the inceptive and (6b) the completive aspect. What is required is the ontological machinery for talking about parts of processes, a mereology for PERDURANT. [9] suggest such a mereology by introducing the notion of a ‘fiat part’ of PERDURANT. This

way various linguistic expressions, such as in (6), could more easily be defined according to the ontology.

## 6. REFERENCES

- [1] Aikhenvald, A. (2000) *Classifiers: A Typology of Noun Categorization Devices*. Oxford: Oxford University Press.
- [2] Bateman, J. A. (1997) The theoretical status of ontologies in natural language processing. In *Proceedings of the Workshop on Text Representation and Domain Modeling—Ideas from Linguistics and AI*. (9-11 October, 1991:Berlin, Germany).
- [3] Comrie, B. (1976) *Aspect*. Cambridge, UK: Cambridge University Press.
- [4] Comrie, B. (1985) *Tense*. Cambridge, UK: Cambridge University Press.
- [5] EMELD (Electronic Metadata for Endangered Languages Data). See <http://emeld.org/>.
- [6] Farrar, S. and D. T. Langendoen (2003a) A linguistic ontology for the Semantic Web. *GLOT International* 7 (3), 97-100.
- [7] Farrar, S. and D. T. Langendoen (2003b) Markup and the GOLD ontology. Presented at the EMELD Workshop on Interlinear Text, E. Lansing, MI, USA.
- [8] Jackendoff, R. (1990) *Semantic Structures*. Cambridge, MA: The MIT Press.
- [9] Grenon, P. and B. Smith (2003) SPIN and SPAN: Prolegomenon to geodynamic ontology.
- [10] Levin, B. (1993) *English Verb Classes and Alternations*. Chicago: University of Chicago Press.
- [11] Masolo, C., S. Borgo, A. Gangemi, N. Guarino, A. Oltramari, and L. Schneider (2002) WonderWeb deliverable D17 (version 2.0). The WonderWeb Library of Foundational Ontologies and the DOLCE ontology.
- [12] Maxwell, M., Simons, G. and Hayashi, L. (2002) A Morphological Glossing Assistant. Presented at the LREC Workshop on Resources and Tools in Field Linguistics, Las Palmas, Spain.
- [13] Niles, I. and A. Pease (2001) Toward a Standard Upper Ontology. In *Proceedings of the 2<sup>nd</sup> International conference on formal Ontology in Information Systems (FOIS-2001)*.
- [14] Payne, T.E. (1997). *Describing Morphosyntax: A Guide for Field Linguists*. Cambridge: Cambridge University Press.
- [15] Shieber, S. M. (1986) *An Introduction to Unification-Based Approaches to Grammar*. Chicago: University of Chicago Press.
- [16] Talmy, L. (2000). *Toward a cognitive semantics. Volume I-II: Typology and process in concept structuring*. Cambridge, MA: The MIT Press.
- [17] van Valin, R. D. and R. J. LaPolla (1997) *Syntax: Structure, Meaning, and Function*. New York: Cambridge University Press.