USGS
science for a changing world

National Geospatial Program

Grammar to Graph—An Approach for Semantic Transformation of Annotations to Triples

Scientific Investigations Report 2025–5064

U.S. Department of the Interior
U.S. Geological Survey

Grammar to Graph—An Approach for Semantic Transformation of Annotations to Triples

By Dalia E. Varanka and Emily Abbott

National Geospatial Program

Scientific Investigations Report 2025–5064

U.S. Department of the Interior
U.S. Geological Survey

U.S. Geological Survey, Reston, Virginia: 2025

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit https://www.usgs.gov or call 1–888–392–8545.

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov/ or contact the store at 1–888–275–8747.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:
Varanka, D.E., and Abbott, E., 2025, Grammar to graph—An approach for semantic transformation of annotations to triples: U.S. Geological Survey Scientific Investigations Report 2025–5064, 20 p., https://doi.org/10.3133/sir20255064.

Associated data for this publication:
Abbott, E.T., 2024, Grammar transformations of topographic feature type annotations of the U.S. to structured graph data: U.S. Geological Survey data release, https://doi.org/10.5066/P1BDPXKZ.

ISSN 2328-0328 (online)

Contents

Abstract....1
Introduction....1
    Purpose and Scope....2
Background....2
    Glossary Database....2
    Language Processing....3
        Dependency Grammar....3
        Part-of-Speech Classes....3
        Dependency Relation Properties....4
Methods....6
    Dependency Relation Property....6
        Part-of-Speech Rules....7
        Semantic Schema Mapping....9
        Directionality....9
    Property Patterns....10
Results....11
    Geographic Semantics....12
        Force Dynamics....12
        Social Space....12
        Subgraphs of Natural and Cultural Feature Properties....13
    Competency Questions....15
    Spatial, Topological, and Relative Relations....15
Discussion....16
Summary....16
Acknowledgments....16
References Cited....17
Glossary....19
Appendix 1. Topographic Property Patterns....20

Figures

1. Diagram of a dependency arc representing grammatical word dependencies among parts of speech categories and their abbreviations in an example phrase....3
2. Graphs showing the grammar classes and properties for OWL files using the Universal Dependencies prefix displayed on Protégé ontology design software interfaces....7
3. Diagram showing general part-of-speech dependencies and number of terms from glossary corpus used in this study....8
4. Part-of-speech dependency relations and property patterns for a subject-verb-object triple structure....9
5. Part-of-speech dependency relations and property patterns for a noun chunk triple structure....9
6. Verb modifiers and verb preposition subtypes and property patterns....10

7. Concept map for morphology feature types involving natural agents and cultural drivers 13
8. Diagram of bridge and arch semantics 14

Tables

1. Part of speech and dependency relations analyzed from the input phrase based on the gloss, “An aircraft facility is an area where aircraft can take off and land usually equipped with associated buildings and facilities” 5
2. Dependency arc formation from an input clause aligned in position with the head as subject, its parent part-of-speech category, the child as object, and the dependency relation of the child to the head 6
3. Number of unique terms from the input glossary and the frequencies of term occurrences grouped by part of speech 8
4. Software retrieval variables for simulating a clause using modifiers and function words 8
5. Head and token dependency relation direction in the positions of subject, property, and object 10
6. Examples of spatial property patterns—spatial conjunction and relative spatial position—for an example phrase 11
7. Example data showing topographic property patterns abstracted from multiple topography-related definitions 16

Abbreviations

ACL	adnominal clause
ADJ	adjective
ADP	adposition
ADV	adverb
ADVCL	adverbial clause
ADVMOD	adverbial modifier
AGENT	agent
AI	artificial intelligence
AMOD	adjectival modifier
AUX	auxiliary
CASE	case marker
CC	coordinating conjunction (dependency relation)
CCONJ	coordinating conjunction (part of speech)
CCOMP	clausal complement
COMPOUND	multiword expression
CONJ	conjunction
CSUBJ	clausal subject
DET	determiner
DOBJ	direct object
G2G	Grammar to Graph
GIS	Geographic Information System
NEG	negation
NLP	natural language processing
NMOD	nominal modifier
NOUN	noun
NSUBJ	nominal subject
NUM	numeral
NUMMOD	numeric modifier
OWL	Web Ontology Language
PART	particle
POBJ	object of preposition
POS	part of speech
PREP	preposition
PRON	pronoun

RDF	Resource Description Framework
ROOT	root
SCONJ	subordinating conjunction
SPARQL	recursive acronym for SPARQL Protocol and RDF Query Language
SPO	subject-property-object
SQL	structured query language
SVO	subject-verb-object
USGS	U.S. Geological Survey
VERB	verb

Grammar to Graph—An Approach for Semantic Transformation of Annotations to Triples

By Dalia E. Varanka[1] and Emily Abbott[2]

Abstract

Data annotation is the process of labeling data to show the outcome that a related data model should predict. In this study, annotation data were transformed into semantic graph triples, mainly for use with the Resource Description Framework (RDF), a type of entity-relationship-attribute data model for graph databases. The transformation of annotation data to semantic graph triples provides complex linguistic meaning with data handling advantages such as reduced data storage needs, improved logical specification of relations between objects, and reusable classes and properties that support logic and inference. A grammar-based framework in graph form supports user questions and queries.

The words defining approximately 334 topographic feature types compiled by the U.S. Geological Survey were tokenized as units of analysis and grouped by part of speech. Their dependency relations were identified for this study using natural language processing libraries. Dependency concepts are used as structured semantic relations among part-of-speech classes. Tokens, units equivalent to words, form instances of classes and were quantified within a tabular output format using PostgreSQL data storage software. Table data were logically aligned as triples following a mapping file and stored with an ontology file using Ontop virtual triplestore software. A grammar ontology schema for the data was synchronized to match queries whose results validated the graph's structure. The text analysis produced 8 part-of-speech classes of content words for object representations and 4 classes of function words for operational applications. Dependency relations formed 27 ontology properties for topographic subgraph structures. Token occurrences shaped overall ontology salience and formed a lexicon of syntactic terms for subgraph objects and properties. The schema ontology of class and property population shapes formed the lexicon of English terms. SPARQL Protocol and RDF Query Language (SPARQL) was used with the lexicon to conform data to RDF guidelines.

This study confirms the hypothesis that although linguistic logic varies from description logic, its approximation applies to ontology design. Property and query use case patterns extracted from the analysis support queries concerning complex topographic relations and patterns normally embedded within text definitions. The method used in this study could be applied to text forms in other domains, such as survey notes.

Introduction

This project, called Grammar to Graph (G2G), encompasses a data transformation process from natural language text strings to node-edge-node triples that **federate** to form a graph database. The transformation of annotation data to semantic **graph triples**, called "triples" throughout this text, provides complex linguistic meaning with data handling advantages such as reduced data storage needs, improved logical specification of relations between objects, and reusable classes and properties that support logic and inference. The transformation is based on English-language grammar rules. The scope of this report is to describe how geospatial feature-type definitions such as a bar (that is, a mound or ridge of naturally accumulated sand, gravel, or alluvium forming an underwater or exposed embankment or succession of ridges) can be analyzed using grammar rules that align with the Resource Description Framework (RDF) triples (Resource Description Framework Working Group, 2014). The RDF models statements of facts or web resources in expressions of the form subject-predicate-object, known as triples. For example, two nodes connected by an edge can describe a fact, represented as (subject, predicate, object) triples. Results were usable for developing topographic query types involving commonly used topographic domain properties.

This report is an accompanying publication to a publicly distributed U.S. Geological Survey data release (Abbott, 2024) that includes three tabular data files: 1) a glossary that forms the input data for a dependency analysis of definitions, 2) text strings tagged with their part of speech (POS) and the **dependency relation**, and 3) a quantified lexicon of each **token**, their occurrences, and the related POS categories of the input text corpus. The data release includes the ontology header and an approach for creating a subclasses query.

[1]U.S. Geological Survey.

[2]Colorado School of Mines, under contract to the U.S. Geological Survey.

Purpose and Scope

Natural language processing assisted by artificial intelligence (AI) is commonly composed of large language models using statistical approaches for word predictability and sense detection among large bodies of input text data. Applying an ontology as a schema structure with logic property relations improves the efficiency and accuracy of inferring information from AI technologies by vectorizing the graph and integrating it with multiple data models. The syntactic expression of words can vary, but the meanings of statements are reflected in grammar. The purpose of the G2G project is to study how grammar structures in graph databases may support natural language queries. This report describes the initial natural language processing stage of the G2G project as a prototype for procedures using geographical language data as an ontology and for further research.

This report describes the knowledge modeling scope of an ontology that approximates a range of targeted potential natural language queries by employing primitive entities to define basic concepts that can be used to build complex concepts. Topographic data share attributes of linguistic primitives learned early in life in creating categories as standard locality features and logical relations. Logic primitives are categories that cannot be defined in terms of other category criteria for inclusion in the same ontology. Primitives are compatible with disjoint properties whose classes are reused as separate elements for building flexible semantic (subgraph) models of more complex concepts by minimizing overlapping inferences.

Background

Source data from a glossary of topographic features (Caro and Varanka, 2011) were normalized by removing punctuation and other characters or replacing these with different formatting for data processing consistency. Normalization allowed the syntax of the text to be examined and used flexibly for focusing on topographic vocabulary meaning and patterns of language use and their spatial nuances. The normalized data are defined as feature-type class semantic universals involving multiple entities and various relations for graph data.

The **dependency grammar** analysis provided semantically governed rules among object categories, such as structures and relations and their associated morphological processes to simulate the complexity of linguistic expression. Dependency grammar takes the verb as the central unit of the phrase. A phrase has a subject as the core argument and, secondarily, an object. These positions can be compiled as a three-part data structure based on the dependency relations between POS sets or lexical instances. An analogy is sometimes made that the subject-verb-object (SVO) languages, such as English, are intuitively related to subject-property-object (SPO) triple resources as a simple sentence form. This analogy is roughly correct, but linguistically, a POS serves as the subject, and a relation connects it to the verb. Some schemas of POS are associated with dependency relations for defining triples during postprocessing.

Data were tokenized and tagged with respective POS and dependency relations using spaCy pipelines for processing and analysis (ExplosionAI GmbH, 2024). The logic of each dependency relation among POS was semantically aligned as a triple pattern with property domain and range classes, and was manually mapped to structure the language processing data output as an RDF vocabulary of hierarchical types and other classes with word instances as set members and primarily object properties to create the grammar ontology files. Patterns in the semantic order of word associations were considered equivalent for SPARQL query matching against the database schema in the SPO forms derived from SVO clause structures. Subgraphs of complex relations consisted of object modifiers and other subproperties expanding upon the SPO form.

Glossary Database

The compiled glossary database includes approximately 334 topographic feature-type English-language labels and associated definitions whose original versions are provided in Caro and Varanka (2011). The input annotation data to study geospatial language were manually edited to ensure data quality. For example, nominative terms of glosses whose definitions are meant to persist with time allow the modality of verb forms (representing tense or mood) within the input definition to be omitted to focus on spatial aspects of the data. Subtypes of English, such as those expressing linguistic variety related to geographical areas, were not considered. Only active-voice phrases, such as “A zone where waves break into foam as they move towards the shore,” were included in the analysis.

All punctuation was removed and replaced according to the following rules:

- Where a slash appears in the topographic label (as in “sea/ocean”), it is replaced by a space.
- “And/or” is replaced with “or.”
- Commas are replaced with “and” or “or” in a list, often as a stylistic change.
- Periods are removed. Where there are two sentences, a natural language passage is written for each sentence individually.
- Brackets are removed if the bracketed phrase’s insertion makes grammatical sense.
- Contractions are broken out into separate words.
- Hyphenated words remain as compound words.

Punctuation rules consisted of:

- Terms appearing after “especially” are considered subclasses and are removed if the parent class encompasses them.
- Feature types with more than one definition are edited and either incorporated in a single definition or remain as multiple definitions where distinct characters are defined.
- The first determiner of the definition, such as “a” or “an” (as in “an area”), is assumed to be a subsumption relation of the feature label to its definition, such as “Feature [isA] Definition.” “The” is removed using the same rule, although this causes slight differences in semantics (that is, whether the referent is specified or unspecified). Determiners within the definition are kept.
- The linguistic and first-order logic concepts can be resolved for computation, although not perfectly. Some annotation of language relations reflects labels assigned to properties for ease of use.

Variations in writing style were interpreted following Modern Language Association formatting principles and guidelines for consistency (refer to Purdue Online Writing Lab, 2024).

Language Processing

Text data to instantiate user queries were organized among their POS category and dependencies on the basis of a predominant subject of a phrase called the **head**. Patterns among word dependencies reflect types of semantic roles such as object, modifier, or spatial preposition. These grammar rules form the domain and range classes for RDF properties aligned along equivalent grammar rules. Types of word dependency patterns among tokens are called **chunks** in the case of nouns and their modifiers.

Dependency Grammar

Dependency grammar organizes relations among linguistic units, such as words, according to their predominant or dependent roles to communicate meaning (fig. 1). The verb is taken to be the structural center of a clause and requires the semantic agreement of other words directly or indirectly connected to it according to its dependency direction, as in “is dependent on” or “has dependent.” The data structure is represented by the dependency arcs between the main word acting as the head and the head’s dependents. Dependency structures may be formed as chunks consisting of nouns with auxiliary relations to verbs, such as modifiers or other information, adding complexity, such as phrases.

Part-of-Speech Classes

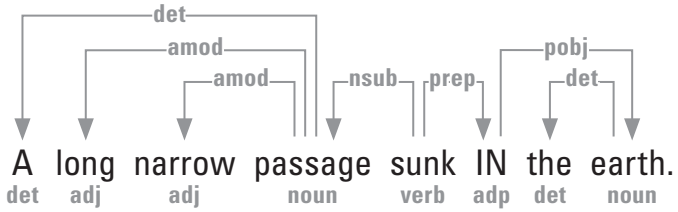
SpaCy natural language processing pipelines tokenize phrases and assign POS abbreviation tags to those tokens and tags for the dependency relations determined among tokens based on Universal Dependencies libraries, as presented by Marneffe and others (2024). Token texts and their associated tags compose the tabular data column headings of the USGS data release associated with this report (Abbott, 2024). The text analysis in this study produced eight POS classes of content words for object representations and four classes of function words for operational applications; these are listed below. Function words provide clarifying context to descriptions. For example, descriptions formed with the determiner “the” that appear to refer to exactly one object could indicate logical primitives.

The eight POS content words and their abbreviations are:

- Adjective (ADJ)
- Adverb (ADV)
- Auxiliary (AUX)
- Determiner (DET)
- Noun (NOUN)
- Numeral (NUM)
- Pronoun (PRON) and
- Verb (VERB)

The four POS function words are:

- Adposition (ADP)
- Coordinating conjunction (CCONJ)
- Particle (PART) and
- Subordinating conjunction (SCONJ)



Abbreviations

DET	determiner	**NOUN**	noun	**ADP**	adposition
ADJ	adjective	**NSUB**	nominal subject	**PREP**	preposition
AMOD	adjective modifier	**VERB**	verb	**POBJ**	object of preposition

Figure 1. Diagram of a dependency arc representing grammatical word dependencies among parts of speech categories and their abbreviations in an example phrase.

A token's **lemma** is an attribute that describes the root form or infinitive of that word. In the word's dictionary form, a lemma is used as a heading indicating the subject or argument of an annotation or a dictionary entry. Lemmas may be preferred forms for entity classes in an ontology because definitions usually reflect stable states such as roots or infinitives. However, linguistic class definitions include conditional properties indicated by the word token texts of their input clause. Word token texts vary to indicate semantic properties (such as person, number, tense, use, and the action role of subjects and objects) that may be transferable to graphs. POS tags, such as "noun," "verb," "adjective," and "adverb," indicate grammatical roles that help distinguish how a lemma may function within a graph, whether as a class, relationship, or property.

The POS sets transfer to the class-subclass hierarchy of taxonomy and tree-type databases to allow **Boolean operations** involving set members and inference. However, the relations and attributes may lack equivalent formal semantic axioms or formulas that enable some semantic similarity measures but limit logic applications. Some set relations are inherent in definitions of words, such as plurals. Different kinds of relations among objects, space, and time, such as agency, possession, or topology, are marked by prepositions or subordinate conjunctions. Verb relations and attributes can infer some processes. For example, in the meteorological phenomenon called the convection cycle, dynamic markers such as "cooling," "shrinking," and "sinking" can be used to define an entity such as a convective storm, "An atmospheric storm which is generated by the heating and convection of moist and unstable air masses" (The Ontology Lookup Service, 2025). Adverbs can reflect conditions of necessity, uncertainty, possibility, or permission, including time conditions. Some subordinate conjunctions beginning with "wh-," such as "where" and "when," provide indications of space and time.

Grouping POS with tags categorizes syntactic text terms into semantic roles and creates a lexicon. The lexicon offers quantitative summaries of the input data corpus. Each word has its lemma listed once in order of most frequent occurrence, with POS category tags included for grouping terms. Descriptive parts of speech, such as determiners, are removed from the lexicon. The same format is applied to chunks. Determiners are like noun modifiers when included in chunks, which are groupings of a noun and its modifiers. Chunking streamlines the processing workload if context is not a priority.

Dependency Relation Properties

Dependency relations are grammatically associated with certain parts of speech to form semantic patterns. The word token text has attributes of its POS and its lemma. The head, representing the predominant concept of the relationship, has attributes of the POS and the lemma. The dependency relation has one head and any number of children.

The dependency relation for the head and its children is similar to an ontology property's domain and range class. The following relations form the semantic transformation rules for the grammar ontology:

- Adnominal clause (ACL)
- Adverbial clause (ADVCL)
- Adverbial modifier (ADVMOD)
- Agent (AGENT)
- Adjectival modifier (AMOD)
- Auxiliary verb (AUX)
- Case marker (CASE)
- Coordinating conjunction (CC)
- Clausal complement (CCOMP)
- Multiword expression (COMPOUND)
- Conjunction (CONJ)
- Clausal subject (CSUBJ)
- Direct object (DOBJ)
- Negation (NEG)
- Nominal modifier (NMOD)
- Nominal subject (NSUBJ)
- Numeric modifier (NUMMOD)
- Object of preposition (POBJ)
- Preposition (PREP)
- Root (ROOT)

Table 1 shows an input clause, "An aircraft facility is an area where aircraft can take-off and land usually equipped with associated buildings and facilities," tokenized as words or as chunks and assigned their corresponding POS categories, and the dependency relation of the tokens to their arc heads of phrases within the clause.

An input clause can have multiple dependency relations that can form subgraphs within the same input sentence or across other definitions. Dependent children and their relation to the heads form multiple arcs within an input sentence to map to triples.

To simulate the node-edge-node pattern of triples, the dependency relations between words grouped by the POS of the data output by the natural language processing (NLP) software were directed edges, which connect the child to the head in the direction from child to head. However, directed edges can appear anywhere in the clause, so the meaning may be represented as pointing right or left. The object dependency represented by the triple property is assumed to be roughly the

Table 1. Part-of-speech and dependency relations analyzed from the input phrase based on the gloss, "An aircraft facility is an area where aircraft can take-off and land usually equipped with associated buildings and facilities."

[Input phrase, read top to bottom, is the tokenized gloss. Abbreviations for terms are provided in parentheses. Head, word that determines the nature of the phrase (Nordquist, 2019).]

Input phrase	Part-of-speech abbreviation	Dependency tag	Head
An aircraft facility	Noun (NOUN)	Nominal subject (NSUBJ)	is
is	Auxiliary (AUX)	Root (ROOT)	is
an area	Noun (NOUN)	Attributive (ATTR)	is
where	Subordinating conjunction (SCONJ)	Adverbial modifier (ADVMOD)	land
aircraft	Noun (NOUN)	Nominal subject (NSUBJ)	take-off
can	Auxiliary (AUX)	Auxiliary verb (AUX)	take-off
take-off	Verb (VERB)	Relative clause (RELCL)	an area
and	Coordinating conjunction (CCONJ)	Coordinating conjunction (CC)	take-off
land	Verb (VERB)	Conjunction (CONJ)	take-off
usually	Adverb (ADV)	Adverbial modifier (ADVMOD)	equipped
equipped	Verb (VERB)	Adnominal clause (ACL)	an area
with	Adposition (ADP)	Preposition (PREP)	equipped
associated buildings	Noun (NOUN)	Object of preposition (POBJ)	with
and	Coordinating conjunction (CCONJ)	Coordinating conjunction (CC)	associated buildings
facilities	Noun (NOUN)	Conjunction (CONJ)	associated buildings

same as in grammar in that the object is determined by, and secondary to, the subject. However, the triple syntax can be specified using an inverse property relation, creating a directed edge to the left or right of the specified nodes. Care is required to confirm directionality when mapping the dependency of the translation because some linguistic semantic dependencies are similar but not equivalent to the formal logic.

Some dependency relations between words of an input definition are related to a baseline triple based on a verb, its subject, and any object. Adding additional properties to these main parts of a triple creates chunks, which are nouns and their modifiers (usually adjectives). Chunks can be mapped to subgraphs in the positions of the subject and object nodes.

To produce the data for the topographic subject area study, NLP preprocessing was completed using two approaches. One approach was to analyze each input token to create chunks for the head of the dependency arc. The chunk pattern approach combines related dependent tokens to group commonly used POS, such as articles and determiners, or to document words commonly used together. The second approach, the token pattern approach, individualizes each unit by category to be examined by the category's clausal roles. Chunks are phrases formed as the head, dependency, and children of the subject and object relations of the verb of an input phrase. Modifiers and other dependents form clausal subjects with the head of the chunk, the core of the subject, or the object of the basic triple. In table 2, the arc heads "is," "in," and "of" are word tokens. "Any bowl-shaped depression" and "the surface" are chunks. The differences between these two approaches to information retrieval are the number and pattern of categories and properties.

Table 2. Dependency arc formation from an input clause ("A basin is any bowl-shaped depression in the surface of the land or ocean floor") aligned in position with the head as subject, its parent part-of-speech category, the child as object, and the dependency relation of the child to the head.

[Dependency arc formation between head, part-of-speech categories, and child is determined by Universal Dependencies software libraries. Abbreviations for terms are provided in parentheses.]

Head text	Head part of speech	Child	Child dependency to head
is	Auxiliary (AUX)	A basin	Nominal subject (NSUBJ)
is	Auxiliary (AUX)	any bowl-shaped depression	Attributive (ATTR)
any bowl-shaped depression	Noun (NOUN)	in	Preposition (PREP)
in	Adposition (ADP)	the surface	Object of a preposition (POBJ)
the surface	Noun (NOUN)	of	Preposition (PREP)
of	Adposition (ADP)	the land or ocean floor	Object of a preposition (POBJ)

Methods

This study's analysis aims to create schemas of POS associations with dependency relations for inferring triples. The tabular data results from NLP were integrated using Ontop, a virtual **knowledge graph** system, and SPARQL endpoint (Harris and Seaborne, 2013; Free University of Bozen-Bolzano, 2024). Data were stored using PostgreSQL software for the database (PostgreSQL Database Management System, 2024), including a relational data mapping file and ontology files with Ontop plugins for Protégé (Musen, 2015). Ontop converts SPARQL queries to SQL for PostgreSQL data retrieval. SPARQL was used to conform the RDF that connects instances in the database with International Resource Identifier prefixes from the furnished ontology, which is the Web Ontology Language standard vocabulary (W3C OWL Working Group, 2012). Postprocessing of the output structure for the ontology was required to restructure instances of dependency relations from the tabular column of the input data as properties, as were tokens for the lexicon, POS class hierarchy, and other Web Ontology Language (OWL) expressions (W3C OWL Working Group, 2012). This study included dependency relations for 27 ontology properties for topographic subgraph structures. Specific relations were differentiated manually as RDF subproperties using the Protégé software plugin for Ontop.

An initial competency query to test the general operation of the database design split the label and definition annotations of a linguistic concept, as is required by an ontology hierarchy. Subclasses of feature type for the ontology hierarchy are determined from the input glossary definitions by retrieving noun heads of the noun-subject relation combined with the auxiliary verb "is" or "are," indicating the subsumption relation represented by the W3 standard property rdfs:subClassOf property to build the hierarchy. Classes are associated with input data labels and definitions to be linked to the triples for reference. The approximate graph pattern of the SPARQL query, using prefixes for namespaces and triple syntax, is printed below.

```
SELECT ?FTI

WHERE { ?FT ud:hasPOS ud:NOUN .

?FT ud:dep ud:nsubj .

?FT rdfs:subClassOf g2g:FeatureType

?FT ud:hasToken ud:Token .

?FT rdfs:label ?Label .

rdfs:Label g2g:hasDescription g2g:Description . }
```

Dependency Relation Property

This section presents the four roles of chunk, property, object, and function, followed by grammar-semantic schema mapping. The subset of dependency relations identified from the analysis and query output comprises the subproperties of ud:dependencyRelation using the Universal Dependencies International Resource Identifier (fig. 2). The hierarchy of classes (fig. 2*A*) for subjects and objects and properties (fig. 2*B*) differs between the linguistic input clause and a comparable triple semantics. These differences represent the transformation of grammar semantics to the description logic of the Protégé software.

As shown in figure 2, the concept maps are expressed as RDF triples and arranged by their allowable relations to create OWL files. The columns on the left in figures 2*A* and 2*B* show linguistic classes and properties of the taxonomy as allowed by grammar rules. The transformation to description logic shown in the right column of figures 2*A* and 2*B* show changes in the hierarchy due to conforming from grammar to RDF standards. The issue of concern is the inheritance of attributes from parent classes. The inverted taxonomic hierarchy of grammar compared to triples shows that linguistic complexity predominates in gloss clauses and is greatly subsumed in graphs, where the core arguments, which are the primary drivers of the meaning, include complexity along with the simplified predominant triple.



Figure 2. Graphs showing the grammar classes and properties for Web Ontology Language (OWL) files using the Universal Dependencies prefix (ud:) displayed on Protégé ontology design software interfaces (Musen, 2015). *A*, a file version showing object classes abstracted from clauses, and *B*, a file version showing Resource Description Framework graph properties.

Part-of-Speech Rules

The number of POS terms output from the NLP was counted (table 3). Nouns predominate but include modifiers essential to the meaning of the nominal (the combined modifiers and noun). Other adjectives are descriptive. In addition to over 200 verbs, all other POS terms number fewer than 100, including modifiers and function words, which include prepositions that combine with nouns to suggest spatial relations and conjunctions that join more than one phrase that may indicate context.

Parts of speech other than nouns and verbs use dependency relations that add semantic specificity to the main parts of the triple through the data output from tokenization. Dependency relations analyses break up the input clause to be readable by logical relations (table 4). In this way, the data can be reused to connect phrases to each other in ways that support graph processing.

The twelve types of POSs are grouped into four roles (fig. 3). The chunk is based on the nominal consisting of a noun, adjectives, and other token dependencies of the noun. The property role includes the verb and adverb. The third position, object, includes the adposition or preposition and directly follows the property; the object has a chunk pattern.

Table 3. Number of unique terms from the input glossary and the frequencies of term occurrences grouped by part of speech.

[The input data corpus is the total set of tokens from the analyzed definitions. The number of term occurrences can indicate the frequency of reuse. There were 1,540 terms total, and 3,641 term occurrences.]

Part of speech	Abbreviation	Number of terms	Number of term occurrences
Proper noun	PROPN	11	16
Adjective	ADJ	31	51
Determiner	DET	4	11
Noun	NOUN	1,175	1,795
Numeral	NUM	1	2
Pronoun	PRON	11	93
Verb	VERB	215	482
Adverb	ADV	40	76
Adposition	ADP	36	730
Coordinating conjunction	CCONJ	5	324
Subordinating conjunction	SCONJ	11	61

Table 4. Software retrieval variables for simulating a clause using modifiers and function words.

[Properties forming a chunk pattern can include a function word such as a determiner (DET), adjective (ADJ) and numerical (NUM) modifiers, nominal noun (NOUN), and pronouns (PRON) referring to the objects of the chunk. Modifiers of the main verb include adverbs (ADV) and auxiliary (AUX) verbs. VERB, verb; NA, not applicable]

DET	PRON	ADJ	NUM	NOUN chunk	ADV	AUX	VERB
an	it	able	one	aborted take-off	alongside	will	accommodate
each	its	adjacent	NA	access	always	do	acquire
the	something	artificial	NA	access restriction	as	can	advance

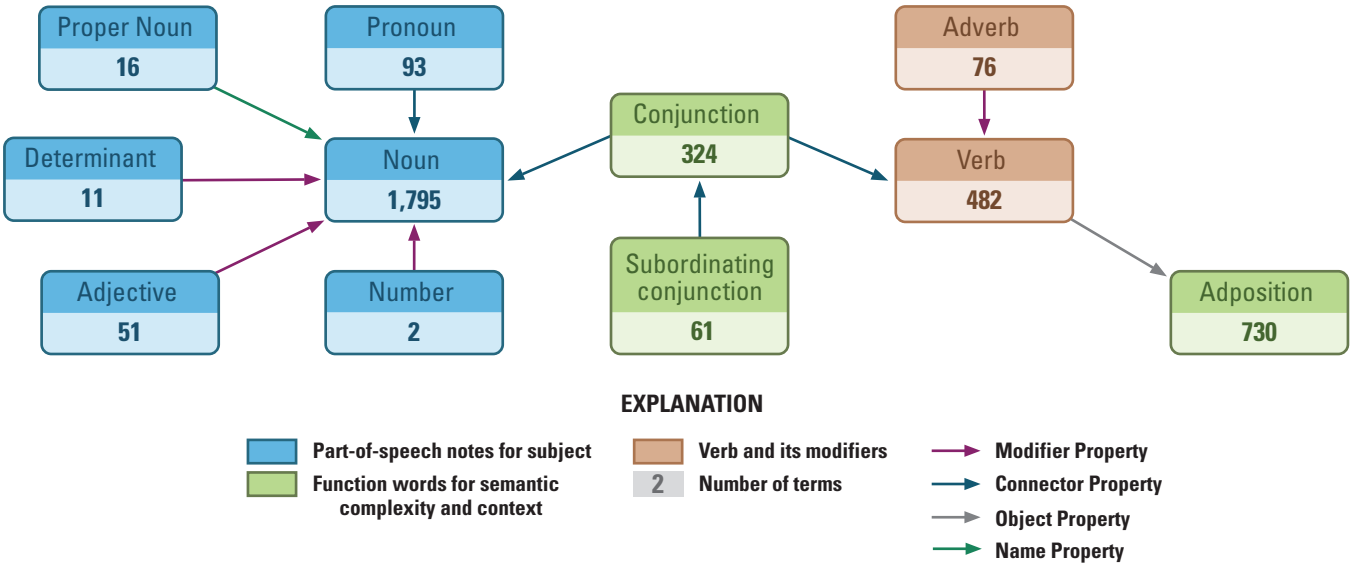


Figure 3. Diagram showing general part-of-speech dependencies and number of terms from glossary corpus used in this study.

The function role of conjunctions is to join or provide context for chunks. The semantic schema mapping rules to organize the chunk, property, and function are explained in the section of this report titled "Semantic Schema Mapping."

Semantic Schema Mapping

The semantic schema maps grammar rules to generate types of triple patterns, particularly RDF. The property establishes the primary semantic element with subject and object arguments. The subject is the head of the clause. The subject and property without an object fulfill the requirements of a clause. As a triple, the object can be implied, such as with a blank node. The subject and property have an object, and objects and subjects may have modifiers. Function words join or describe a context for multiple subjects and objects.

The dependency relations used as properties for triples have domain and range classes corresponding to grammar classes within the ontology structure. The SVO triple pattern has five subtypes of POS classes for the lexical graph data that will be queried using SPARQL (fig. 4).

Complexes of noun modifiers have four potential property patterns, as do verbs and their modifiers, to support property clusters. An object-like pattern called a chunk is used as the subject and reused as the object and as part of combined triples (fig. 5). Chunks have nominals such as modifiers and subtypes.

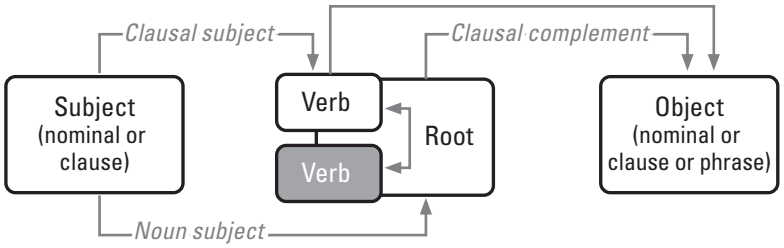
Parts of speech relating to the central grammatical verb combine to form the action property (fig. 6).

Retrieval variables for a data sample forming a verb property and its modifiers include adverbs (ADV), auxiliary (AUX) verbs, and the central verb (VERB). Each subproperty of the dependency relation property controls the appropriate structure and function relations of POS word instances to be represented in RDF.

Directionality

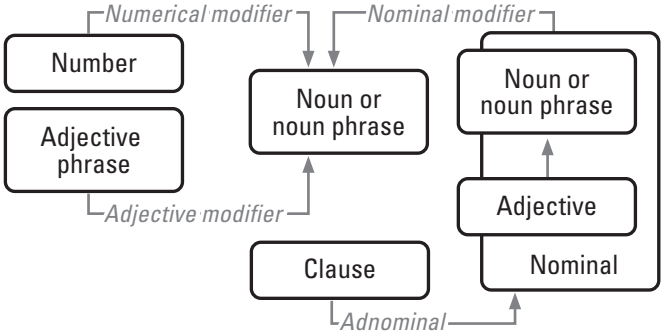
An object's dependency on a head preserves semantic coherence as a graph among tokens or POS categories to resemble the grammar of phrases and complex concepts. The head is always the subject, and the dependency direction is from children to head. The directionality of the RDF is in the inverse direction, from subject to object. A triple is read from left to right, although dependents can appear on either the left or right side in a dependency relation.

Data were formatted as a dependency arc starting at the head, shown in table 5 as HeadText, the dependency relation abbreviation, and a token representing the child. The children are more variable than the subject, so they are not grouped as often in a POS category, or the category is left blank (table 5). The arc head is analogous to the triple subject, and the token dependent aligns with the object as the argument of the verb.



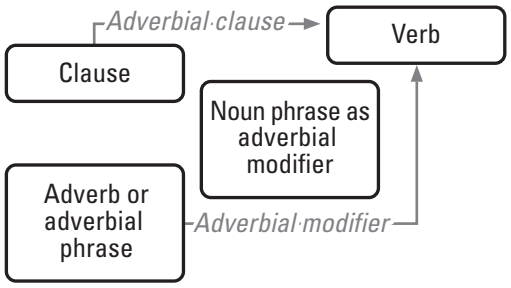
POS	Dependency	POS
Noun	nounSubject	Verb
Verb	directObject	Noun
Clause	clausalSubject	Verb
Verb	clausalCompliment	Clause
Verb	indirectObject	Noun

Figure 4. Part-of-speech (POS) dependency relations and property patterns for a subject-verb-object (SVO) triple structure. The SVO triple has five triple patterns used in this study, shown in the figure as the POS, Dep, POS columns of the data output.



POS	Dep	POS
Nominal	nominalModifier	Noun
Number	numericalModifier	Noun
Clause	adnominal	Nominal
AdjectivalPhrase	adjectiveModifier	Noun

Figure 5. Part-of-speech dependency relations and property patterns for a noun chunk triple structure.



POS	Dep	POS
Adverb	adverbialClause	Verb
Clause	adverbialModifier	Verb or Adverb

Figure 6. Verb modifiers and verb preposition subtypes and property patterns.

Table 5. Head and token dependency relation direction in the positions of subject, property, and object.

[Results reflect the natural language phrases "support" as the clausal modifier of the noun "ability," "limited" as the modifier of "ability," and "able" having an open complement (XCOMP) for a core argument of the verb as HeadText. ACL, adnominal clause; AMOD, adjectival modifier]

HeadText	Dependency	Token
ability	ACL	support
ability	AMOD	limited
able	XCOMP	support

Property Patterns

The domain and range classes of properties offer a mixture of category tags, lexical instance texts, and lemmas in the database that reflect grammar for querying, competency questions, and determining vocabulary word sets. Table 6 shows examples of spatial dependency relations "above" and "below" using coordinating conjunctions for a phrase such as "A snag is a stem or trunk of a tree above or below the surface of the water" to characterize the concept using objects of a preposition to indicate relative spatial position.

The SPO patterns were manually interpreted using perception or experience of a linguistic relation of a token or POS class in addition to automated dependency relation arc creation. Manual testing, such as interchanging optional tokens and POS types to explore one-to-many relationships, was used. Some words served as semantic case markers, such as "by" to indicate agency and "to" to indicate functionality, spatial relation, or the receiver of the action.

The property patterns listed below in this section formulate variables for describing some semantic key concepts. These are 1) chunk subject, relating to objects and their attributes; 2) the SVO triple of the subject, verb, and object parts of predicates; and 3) conjunctions using grammatical rules for building Boolean-operation types of property patterns. A subset of the glossary definitions was written as subject-verb **predicate** clauses with no third object, such as when a chunk is a subject node of a predicate clause. The predicate consisting of a verb and an object was considered the SVO form. Conjunction relations have semantic functions comparable to labels for Boolean operations. Set theory properties applied to classes can operate on the graph model and connect datasets by linking namespaces of other common vocabularies, such as standards for cited references, data attribution, and other metadata.

The list below in this section provides a natural language question for formulating topographic database user queries. An informal representation of language formatted as a graph is included in the list for providing a semantic base for serializing SPARQL queries. Despite the similarities between dependency arc and triples, a logic notation that clearly aligns text data dependencies for conversion to triples was not widely available in related research literature, although language and logic notations can be used (Allwood and others, 1977; Sowa, 2010; Gregory, 2015; Wybraniec-Skardowska, 2022). Logic notation was excluded from the semantic transformation, and comparisons were made using the manual interpretation of data results from NLP as three parts of a single entity on one line, which organizes the data spatially (for database design) and is not meant to be interpreted as logical proofs. The style of arrows inserted for this function expresses directionality of text, category tags, or abbreviations. The semantic transformation is supported because, by understanding the rules of the property domains and ranges, direct string matching of classes, properties, and instances is entered in the ontology. Where concepts do not have exact matches, the owl:equivalentTo property connects the immediate triples. OWL provides a rich vocabulary for designing and structuring graphs. The RDF-based W3 standards assure the logic is correct when applied to data in compliant software, such as Protégé.

The three-column arrangement of output data from the NLP tabular results, as described in table 5, is expressed as RDF in the ontology or knowledge graph. The spaCy output tables were arranged as POS-dependency-POS to capture semantic relations. The data shown below is also arranged following a SPO triple convention basis, where the dependency property representation is simplified as a relation of the subject as an object and relation to the verb. The verb represents an action of relation between object sets or literals, whose dependency on the verb is subsumed to the property, and so are SPO subjects and objects classes. Members of those

Table 6. Examples of spatial property patterns—spatial conjunction and relative spatial position—for a phrase such as "A snag is a stem or trunk of a tree above or below the surface of the water."

[On the basis of data line extraction, patterns for a spatial conjunction concept, represented by a spatial relationship of prepositions "above" or "below," have a coordinating conjunction (CC) connector "or" and conjunction (CONJ) joining with "below." An object of a preposition (POBJ) is linked to a spatial preposition, "above."]

Property pattern	Head (subject)	Dependency (property)	Child (object)
Spatial conjunction	above	CC	or
Spatial conjunction	above	CONJ	below
Relative position	above	POBJ	deck
Relative position	above	POBJ	level

subject or object sets are included in the retrieval of data in the SPARQL triple patterns following this intermediary step between language and ontology formatted in a graph pattern using RDF. The arrows in the numbered examples below show the dependency of the verb in subject-object core arguments and correspond to applications in RDF. Potential dependencies of triple resources that can be modified by standard OWL properties are not included in arrow representations.

1. Chunk subject: property pattern for a competency question and data retrieval

   Question: What special types of objects are or have special properties?

   The head nominal of the chunk is identified first, followed by the subject dependency on the verb-related POS for structuring complex properties. Each line of the pattern inherits the data retrieved by the previous dependency arc. The final pronoun, nominal subject, and verb results approximate a chunk subject and dependency relations.

   Using NLP output tags, the variables to be retrieved are demonstrated below:

   ADJ→adjectival modifier→NOUN
   NOUN→auxiliary→AUX
   AUX→ADV→VERB
   PRON→NSUBJ→VERB

2. SVO or triple

   The SVO pattern includes an object with a subject and a verb. The structure of the object can be equivalent to the subject and modifiers, although the object appears after the adposition such as a preposition that indicates the receiver of the action.

   Question: What actions of subjects are associated with a spatial preposition?

   The auxiliary and active verbs with an adpositive are retrieved followed by the subjects of those actions or state of being. Adverb attributes of actions can specify the types of actions and their related adpositive.

   Linguistic order of variables to be retrieved:

   AUX→VERB→ADP
   NOUN→VERB→AUX
   ADV→VERB→ADP

3. Conjunctions

   Functions that join words that establish context for the main clause, including subordinating and coordinating conjunctions.

   Question: What related objects are likely to both be affected?

   Nouns are established as objects following the adposition of a verb. Multiple action nouns of chunks are related to each other, linked by conjunctions that also link multiple verbs and adpositions.

   Linguistic order of variables to be retrieved:

   VERB→ADP→NOUN
   NOUN→CCONJ→NOUN
   NOUN→SCONJ→NOUN
   VERB→CCONJ→ADP

The semantics of the notation described above were applied throughout this study based on topographic information and knowledge. The G2G approach's analysis of the translation among linguistic and graph semantic annotations was used to create some initial SPARQL graph patterns. The results described in this report's "Results" section focus on semantic translation.

Results

In this study, linguistic text-graph associations are common and can support structures for topographic information retrieval from the domain language, including the user perspective. Linguistic descriptions and words that support the primarily SVO form from spatial perspectives can be linguistic corollaries of logic concepts. Ontologists draw on linguistic schemas; for example, Casati and Varzi (2003) investigated the linguistic foundations of **mereotopologic** relations as a type of upper ontology. Linguistic labels were assigned to qualitative spatial relations functions that were later adapted as part of the GeoSPARQL standard. GeoSPARQL is an extension of SPARQL for determining topological relations by the presence or absence of interior, boundary, or exterior intersections among geospatial geometry objects, among other spatial data-handling functions (Car and others, 2024). Other tokens

have direct equivalencies with OWL, such as the determiner "any;" "one of several" (ROOT-PREP-AMOD) to indicate cardinality; "not" to indicate negation; and "other" to indicate the complement, as with owl:disjointFrom. Some property patterns have equivalencies with OWL property restrictions and specifications.

Asserted ontology subclasses and properties from the feature-type definitions can be validated through cross-thematic competency questions, which may also be a method for improving ontology logic organization for information retrieval by refining the queries and the queries' interactions within and external to the grammar ontology. The query use cases described in the "Geographic Semantics" section of this report match topographic knowledge questions to grammar organization to test the results of the approach. Results were iteratively evaluated for the geographic sense outcomes described in the next section, "Geographic Semantics." Linguistic expressions of topography are meant to be recognizable to public users sharing common concepts, whether those concepts are universal. Outcomes were manually evaluated to refine semantic pattern associations by examining the patterns of properties reflecting topographic vocabulary and then used for query use-case patterns.

Geographic Semantics

Spatial relation vocabularies commonly used for executable applications, such as those of the GeoSPARQL standard, can indicate concepts of location, feature geometry, and map algebra functions such as intersections, linear pathfinding, and buffering. A broader scope of geospatial data research addresses aspects of processes and temporal representation. The complexity of some geographical phenomena has been difficult to express using relational table technology. Complex vocabularies can be enabled by expanding the range of geospatial semantic representations within graph models that support data interoperability (Huang and others, 2019). Vocabularies for properties such as force dynamics and social space are not widely evident in research literature. The following two sections, "Force Dynamics" and "Social Space," discuss semantic aspects of a use-case study to articulate differences among categories of natural and cultural features based on the analysis of their subgraphs.

Force Dynamics

The topographic analysis of this study focuses on property patterns, including geographic process and force dynamic concepts, which include morphology and description, properties of human intention, and concepts of spatial data science. The property and lexical semantics of feature-type definitions build knowledge more flexibly than a literal clause by relating the data as triple propositions for graph processing. For example, a definition from the database for "arête," a landform type, would be "a sharp narrow ridge formed as a result of glacial erosion from both sides."

Changing the representation as an SPO pattern, in which case terms or labels take the three positions of a potential RDF serialized triple, creates the following series of ontology propositions to form the criteria of inclusion for a set of objects and attributes that comprise a dataset schema for information retrieval.

- GlacialErosion formed ridge
- GlacialErosion formed (sides exactly 2)
- Both sides form ridge
- Ridge was formed from both sides by glacial erosion
- Both sides were formed by glacial erosion

The style of writing reflects the normalization of the input data adapted for the ontology taxonomy. Spaces between complex terms are omitted to indicate a single-class label. Object classes with multiple members are written in single numbers.

The propositions listed above build a pattern that fits a subgroup of the dataset, defining natural features of a morphological process, which entails the determination of a shape that is described verbally (like a metaphor) or geometrically. The shape helps define the feature-type class along with supporting details. The dependency of subjects, properties, and objects in the pattern connects the feature type to additional data such as its identity or context.

Social Space

The language analysis in this study inferred human intention and the agency of socially constructed properties. For example, a park is a place or area that, set aside by human agents, will allow for specific attributes while suppressing other competing attributes. Socially constructed spatial relations, such as "set aside" or "preserve," may seem more like environmental attributes and not specific spatial relations with physical manifestations. The distance and implicit nature of human influence control the environment surrounding or setting conditions within which an entity operates. Spatial relations are required for entities to exist and operate in their environments. Socially constructed spatial relations can be understood as informal logic relations between sets of entities.

Cultural geographers have analyzed socially constructed spatial relations and found that they were not adequately modeled in Geographic Information Systems (GIS). Massey (2004) found that humans create and control spaces of interrelations. In those relations, temporal change differs from natural change, and socially constructed relations involve quicker observable temporal change in general. The outcomes of morphologic processes are less determinant and more varied than those of long-lasting natural processes.

The G2G approach of listing natural language propositions to align with semantic technology standards, as was shown for an arête in the "Force Dynamics" section of this report, was applied to a group of definitions involving socially constructed spatial relations, also known as cultural features. The key aspects of cultural-features morphology were agents, objectives, and technology that broaden the scope of topographical shapes. An example of statements defining cropland as "land that is plowed or cultivated for crop production" appears below as a list of necessary classes and properties forming the criteria of inclusion.

- Plow: Overturn an area of land with a plow
- Overturn: Temporal change of geometric shape
- Cultivate: Break up the soil in preparation for sowing or planting; to raise or grow plants

As demonstrated by the semantic analysis above, topographic morphology is driven by agents with objectives using tools, based on the natural environment. Cropland varies in shape based on this pattern of interaction. The shape of the cropland area is determined by the interactions as, for example, mixed-use cultivation such as a vineyard designed for small-scale production on a larger farm, cropland driven by circular irrigation systems, or cultivated terraces adjusted to contoured terrain.

The semantic patterns of cultural features were initially similar to the natural feature patterns, but some differences were evident. Morphologic processes included physical force dynamics and human-driven, tool-based forces. The "shape" descriptor, important for natural features, is replaced by a term of human interaction, such as "cultivated." In the natural feature example in the "Force Dynamics" section of this report, all arêtes would look similar based on their definition and generalized geometric shape. However, this is not so for all croplands. Socio-spatial relations are affected by environmental conditions and the cumulative history of human modifications; the abandonment of cropland to physical forces also affects relations. These relations are often observable in the topography of places and regions.

Subgraphs of Natural and Cultural Feature Properties

A use case to guide queries along semantically specific property paths differentiates natural and cultural features using force dynamics and social space patterns, beginning with a general morphology class whose criteria are inherited by both natural and cultural features. The use case concept relies on common understandings about topography for testing this graph-based querying approach and is easily altered for other general-purpose uses.

Two knowledge concepts to disambiguate features using semantic graph searches are commonality, for which we heuristically assume a natural feature is the entailment or predecessor of cultural features, and cause and action. Shown in figure 7, commonality is tested if there is a missing edge between object classes or if there are duplicate nodes. The cause and action pattern tests for a common ancestor node or alternative action pathway along edges.

Commonality is tested if an edge between object classes is missing or if there are duplicate nodes. The cause and action test is done for a common ancestor node or an alternative action pathway along edges. Three definitions from the glossary database for the feature types "bridge" and "arch" that pose a logical contradiction are used for semantically differentiating the data for this use case:

- An arch is a natural opening in a rock mass.
- A bridge is a human-made structure carrying across a body of water or depression.
- A bridge is the same as an arch.

The tokenized definitions of arch and bridge were mapped for commonality and cause and action patterns using the natural or cultural feature properties (fig. 8). The feature-type patterns shown in blue in figure 8 are the same pattern as in figure 7, with the definitions for arch and bridge in black print (fig. 8). The semantic paths of term definitions are linked by properties. Topological relations to geospatial features are included in figure 8.

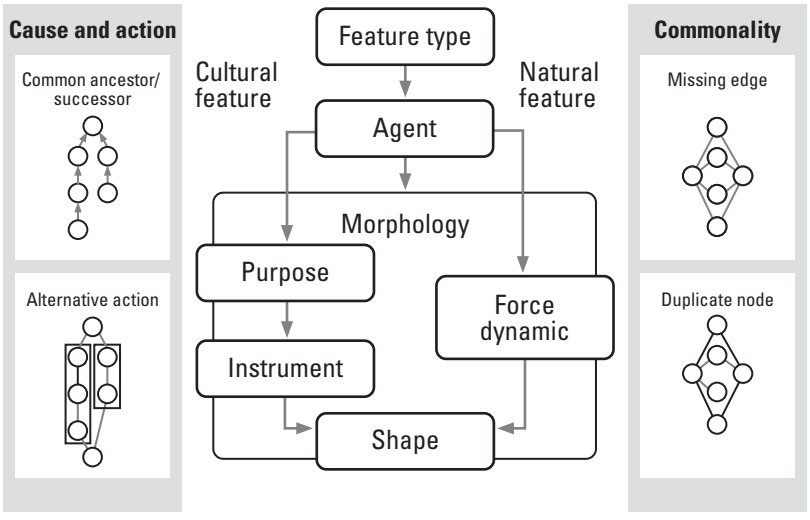


Figure 7. Concept map for morphology feature types involving natural agents (force dynamics) and cultural drivers (purpose and instrument). Graph property paths are diagrammed for "Cause and action" and "Commonality" concepts.

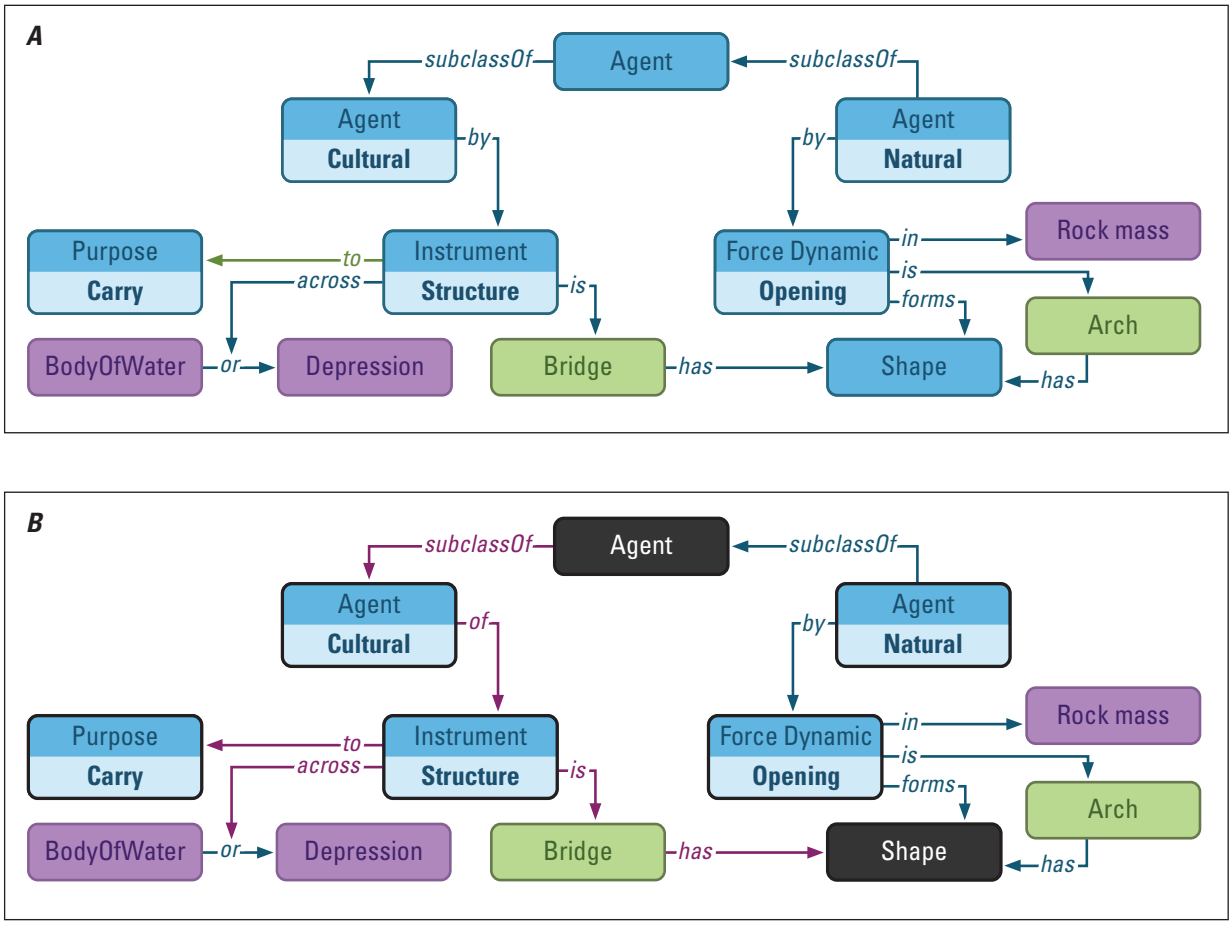


Figure 8. Diagram of bridge and arch semantics. *A*, the semantics for a bridge as a natural feature resemble those for an arch, but with the added property "for" indicating a purpose, as shown by the green arrow. *B*, "Agent" and "Shape" are common ancestors of a concept hierarchy and indicate data inheritance along graph paths between the "cultural" and "natural" subclasses.

Commonality is tested by missing edges or duplicate nodes in the subgraph of the input clause token nodes. "Bridge" includes a property indicating a purpose, which is missing for "arch." The commonality (fig. 8*A*) of the two feature type classes is changed by "carry," which disambiguates the cultural feature-type bridge from a natural bridge.

Differences in cause and effect are matched against the subgraphs. "Cause and action" is defined as "common ancestor or succession" on nodes and "alternative action" on edges. Arch and bridge have a common ancestor, Agent, and can have a common succession to the class Shape, which is shown in figure 8*B* by blue arrows, and as an alternative action differentiating artificial and natural semantic paths, which are shown by purple arrows.

The semantics of shape may explain that a bridge is the same thing as an arch if a bridge includes an overhead linear connection from point to point. Roughly speaking, bridge and arch share the shape of an overhead point-to-point stationary connection. A bridge with the same natural features as an arch, without a purpose, could be the same thing as an arch. When the purpose "carrying across" is added, the bridge shape differs from an arch.

The use case demonstrates how graph properties can support semantic specification to resolve potential issues of ambiguity, differentiation, and relative meaning. Graph data processing depends on patterns for these functions. The "Results" section clarifies the property patterns for some topographic semantic patterns.

Competency Questions

The following are competency questions used in this study to query and validate basic semantic roles of properties linking classes, subjects, actions, and receivers. The first set of competency questions focus on predication:

- What do subjects do to objects? If an object is missing, a blank node can be used to expand the predicate through reification.
- What subjects spatially relate to other objects? Spatial relations are represented using the POS preposition.
- What types of subject act in particular ways to objects? Adverbs and prepositions indicate actions to an indirect object.
- What types of objects have properties reflecting actions of the subject on them? Nominals are objects whose adjectives and other modifiers represent attributes as receivers of actions (that is, verbs).

Other questions and related queries match complex aspects of topographic-feature queries using a graph pattern structure. Examples given below are followed by data retrieval variables that could be used to represent semantic data retrieval patterns:

- What attributes differentiate feature type subclasses? Token AMOD ADJ
- What features have spatial relations attributed to rivers? River PREP ADP or token PREP rivers

For information about actions, agents, and affordances, some examples of semantic transformation from question to computational query values are listed below:

- What multiple actions are involved in processes? VERB CC and
- What were agents of actions? VERB AGENT by
- What conditional activities are possible? VERB AUX to, affords DOBJ entry

Spatial, Topological, and Relative Relations

This section lists data variables organized by topographical relation property patterns for knowledge concepts (spatial prepositions, relative position, multiple relations, feature topology, topological relations, and joins), which are listed on the left of the colon in the list below. Listed to the right of the colon are related property patterns of those knowledge concepts to retrieve data that is not specified in advance, yet that should be acceptable as geographic sense (as described in the section of this report titled "Geographic Semantics"). These example patterns are not definitive; they are flexible and are included as a proof-of-concept.

- Spatial prepositions: access PREP between
- Spatially relative to: above POBJ level
- Spatial conjunction: above CONJ below
- Feature topology: access compound traffic
- Topological relation to: across POBJ bottom
- Topological joins: across CC or

We designed and adapted topographic property patterns for competency questions that were intended to validate the ontology for retrieving information for spatial, topological, and relative relation concepts. The G2G approach proposes a way to represent complexity for a place or locality study that involves multiple perceptible and experiential factors by representing the complex semantics of attributes and properties. The property patterns retrieved from the analyzed data constitute topographic knowledge patterns to use as ontology properties. The property patterns are used to create query patterns to retrieve data for competency questions and information retrieval queries.

All pattern dependency relation arcs emerge from and maintain associated text that forms the scope of the lexicon and syntax of the pattern schemas from property types determined from data examples such as those shown in table 7. Referring to the table, the head, the relation, and a dependent token chain together as sets of triples form a graph to query. Classes allow multiple token instances, but a particular token may mark a certain property type or an exemplar of a meaning, enabling the token to function as part of the pattern. Tokens can be indicators of a way the object is used by including the grammar case, such as the genitive case of possession indicated by an apostrophe, or can be interchangeable within a POS category if the semantic role is consistent for the set of instances or among more than one general POS if the property is semantically rich in meaning. The selected model for tokens and POS categories must be recognizable as ontologically believable to the topographic data user.

In this study, patterns were detected that relate to spatial or thematic topographical content whose semantic meaning is related by a dependency arc based on Head.text/dep and children position open links for variable values such as tokens or POS. Examples are expanded as patterns in appendix 1. The objective of this study was to retrieve node and property values providing geographic sense from a single queryable question, but a single line of a triple-type pattern is most often insufficient for complex questions. Additional lines must be added to create a subgraph query pattern for complex questions.

Table 7. Example data showing topographic property patterns abstracted from multiple topography-related definitions. The head, dependency, and child are read on each line from left to right.

[A property pattern for a concept listed in the left column follows the dependency of children to a head shown in the next three columns. MARK, case marker; AMOD, adjectival modifier; DOBJ, direct object; AGENT, agent; COMPOUND, multiword expression. Other abbreviations defined in table 1.]

Concept	Head	Dependency arc	Child
Multiples	VERB	CC	and
Subordinating joiner	Acquired	MARK	after
Ordinal data level	NUM	AMOD	more
Ordinal data level	NUM	ADVMOD	than
Modifier: verb	VERB	ADVMOD	naturally
Modifier: noun	NOUN	AMOD	major
Affordances	VERB	AUX	to
Affordances	Affords	DOBJ	entry
Agent	VERB	AGENT	by
Spatial preposition	Access	PREP	between
Spatial preposition	Access	PREP	to
Spatial conjunction	Above	CC	or
Spatial conjunction	Above	CONJ	below
Spatially relative to something	Above	POBJ	deck
Spatially relative to something	Above	POBJ	level
Feature topology	Access	COMPOUND	traffic
Topological conjunctions	Across	CC	or
Topological conjunctions	Across	CONJ	into
Topological relation to feature	Across	POBJ	bottom
Topological relation to feature	Across	POBJ	course
Topological relation to feature	Across	POBJ	terrain

Discussion

The G2G study demonstrates that, although grammar rules differ from description logic axioms used by RDF and SPARQL, semantic transference approximations can enable ontology design for querying the text of long-tailed datasets with data engineering requirements that do not require machine learning methods typical of large language models. The method used in this study is reproducible specifically to support queries concerning topographic relations and patterns normally embedded within text statements, such as definitions. The G2G method may be applied to other domain text forms, such as land survey notes, social media sources, or chat box output, but these text forms were not tested. The approach is meant to be open. From any perspective, queries should produce results evaluated for geographic sense to refine ontological commitments of semantic patterns. The resulting topographic data representations create a basic semantic framework based on a natural language base map similar to topographic base maps used for thematic cartography.

Summary

The graph-to-graph technique transforms dependency relations between words transformed to triples that behave as properties connecting types of nodes. Nouns with dependents can create triple-type structures for graph software processing. The grammar rules of dependency relations function as schemas for triples.

This study demonstrated the transformation of a feature-type definition to subject-predicate-object formations typical of triple resources. Dependency relation arcs form data structure edges to create semantic template patterns. The dependency relations of subjects and objects create a semantic framework that is instantiated with vocabulary syntax, creating a lexicon. The semantics of graphs derived from natural language using grammar ontology contributes to knowledge properties and inference among triples. Grammar creates a discourse of queries that help refine patterns to update the ontology.

A grammar ontology for topographical language queries shows that the processing pipeline is a repeatable approach that supports the geographical sense of data retrieval. The graph design was tested as a basic application that is useful for limited text corpora, such as annotations or notes. Statistical accuracy measurements of the triple schemas populated with word or phrase instances could be improved by vectorizing the graph with other artificial intelligence-assisted processing.

Acknowledgments

Thanks are extended to Richard Brown and other U.S. Geological Survey system administrators for the software referenced in this report. Emily Abbott's work was done while she was a Student Contractor with the U.S. Geological Survey. Samantha Arundel and Ethan Shavers reviewed the manuscript and offered important suggestions. Evelyn Hampton edited the manuscript with insight and rigor. This work was conducted using Protégé.

References Cited

Abbott, E.T., 2024, Grammar transformations of topographic feature type annotations of the U.S. to structured graph data: U.S. Geological Survey data release, https://doi.org/10.5066/P1BDPXKZ.

Allwood, J., Andersson, L.-G., and Dahl, O., eds., 1977, Logic in linguistics: Cambridge, Mass., Cambridge University Press. [Also available at https://doi.org/10.1017/CBO9780511621017.]

Car, N.J., Homburg, T., Perry, M., Knibbe, F., Cox, S.J.D., Abhayaratna, J., Bonduel, M., Cripps, P.J., and Janowicz, K., eds., 2024, OGC GeoSPARQL—A geographic query language for RDF data (version 1.1): Open Geospatial Consortium website, accessed March 3, 2025, at https://docs.ogc.org/is/22-047r1/22-047r1.html.

Caro, H.K., and Varanka, D.E., 2011, An analysis of spatial relation predicates in U.S. Geological Survey feature definitions: U.S. Geological Survey Open-File Report 2011–1235, 37 p., accessed March 3, 2025, at https://pubs.usgs.gov/of/2011/1235/.

Casati, R., and Varzi, A., 2003, Parts and places—The structures of spatial representation: Cambridge, Mass., The MIT Press, 248 p.

ExplosionAI GmbH, 2024, spaCy—Industrial-strength natural language processing (ver. 3.0): Explosion AI software release, accessed March 19, 2025, at https://spacy.io.

Free University of Bozen-Bolzano, 2024, Ontop—A virtual knowledge graph system Ontop website, accessed June 8, 2025, at https://ontop-vkg.org/.

Gregory, H., 2015, Language and logics—An introduction to the logical foundations of language: Edinburgh, U.K., Edinburgh University Press, 328 p., accessed July 8, 2025, at http://www.jstor.org/stable/10.3366/j.ctt1g09w89.

Harris, S., and Seaborne, A., eds., 2013, SPARQL 1.1 query language—W3C recommendation 21 March 2013, W3C web page, accessed March 19, 2025, at https://w3.org/TR/sparql11-query/.

Huang, Y., Yuan, M., Sheng, Y., Min, X., and Cao, Y., 2019, Using geographic ontologies and geo-characterization to represent geographic scenarios: ISPRS International Journal of Geo-Information, v. 8, no. 12, 19 p. [Also available at https://doi.org/10.3390/ijgi8120566.]

Marneffe, M. de, Ginter, F., Goldberg, Y., Hajič, J., Manning, C., McDonald, R., Levin, L., Nivre, J., Petrov, S., Pyysalo, S., Schneider, N., Schuster, S., Silveira, N., Tsarfaty, R., Tyers, F., Zeldes, A., and Zeman, D., 2024, Universal Dependencies guidelines (ver. 2): Universal Dependencies website, accessed March 19, 2025, at https://universaldependencies.org/#language-.

Massey, D., 2004, Space-time, 'science' and the relationship between physical geography and human geography: Transactions of the Institute of British Geographers, v. 24, no. 3, p. 261–276. [Also available at http://doi.org/10.1111/j.0020-2754.1999.00261.x.]

Musen, M.A., 2015, The protégé project—A look back and a look forward: AI Matters, v. 1, no. 4, p. 4–12, accessed March 19, 2025, at https://doi.org/10.1145/2757001.2757003.

Nordquist, R., 2019, Head (words)—Glossary of grammatical and rhetorical terms: ThoughtCo. website, accessed February 1, 2025, at https://www.thoughtco.com/head-words-tern-1690922.

PostgreSQL Database Management System, 2024, PostgreSQL 16: PostgreSQL Global Development Group website, accessed June 8, 2025, at https://www.postgresql.org/.

Purdue Online Writing Lab, 2024, MLA formatting and style guide: Purdue University website, accessed June 1, 2024, at https://owl.purdue.edu/owl/research_and_citation/mla_style/mla_formatting_and_style_guide/mla_formatting_and_style_guide.html.

Resource Description Framework Working Group, 2014, Resource Description Framework (RDF): W3C Semantic Web Standards, accessed March 19, 2025, at https://www.w3.org/RDF/.

Sowa, J.F., 2010, The role of logic and ontology in language and reasoning, *in* Poli, R., and Seibt, J., eds., Theory and applications of ontology—Philosophical perspectives: Dordrecht, Netherlands, Springer Dordrecht, p. 231–263, accessed July 8, 2025, at https://doi.org/10.1007/978-90-481-8845-1_11.

The Ontology Lookup Service, 2025, Convective storm: European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI) web page, accessed July 8, 2025, at http://purl.obolibrary.org/obo/ENVO_01001294.

W3C OWL Working Group, 2012, OWL 2 Web Ontology Language document overview (second edition)—W3C recommendation 11 December 2012: W3C web page, accessed January 16, 2025, at https://www.w3.org/TR/2012/REC-owl2-overview-20121211/.

Wybraniec-Skardowska, U., 2022, Logic - Language - Ontology—Studies in universal logic [chap. 12]: Basel, Switzerland, Birkhäuser, p. 257–278, accessed July 8, 2025, at https://doi.org/10.1007/978-3-031-22330-3_14.

Glossary

Boolean operations A branch of algebra involving conjunction, disjunction, and negation operations expressed and corresponding with binary operators AND, OR, and NOT.

children Words or other tokens with dependency relations to the arc head.

chunk Nominals or other nouns with grammatical relationships to other parts of speech.

dependency relation The dependency of the meanings of words and other units of texts among each other and aligning with grammar rules.

dependency grammar Linguistic relations that organize units, such as words, based on the completeness of meaning according to the units' predominant or dependent roles.

head The word that determines the nature of the phrase (Nordquist, 2019).

federate A way to integrate multiple decentralized data as sources to independently query using semantic technology query systems.

graph triples A data model following graph theory structured as a three-part node-edge-node semantic structure. Nodes are represented as object classes, member sets, and data literals. Edges are represented as relationships or properties linking nodes.

knowledge graph A type of database that maintains knowledge or information in a graph form. A typical knowledge graph includes a plurality of nodes representing objects (also referred to as entities) and a plurality of edges connecting the nodes, where the edges represent properties as relationships between the objects (for example, is a parent of, is located in, and so on). One common type of knowledge graph is based on the Resource Description Framework (RDF), which models statements of facts or web resources in expressions of the form subject-property-object, known as "triples," also referred to here as "graph triples." For example, two nodes connected by an edge can describe a fact, which can be represented as (subject, property, object) triples.

lemma An attribute of a word that describes its root form or infinitive.

mereotopology A theory (using mereological and topological concepts) of the relations among wholes, parts, and the boundaries among parts.

predicate The subject of the statement performs an action, and the verb or predicate describes what the subject does. If there is a receiver of the action, the verb followed by an object is a predicate.

token An instance of a linguistic symbol or expression.

Appendix 1. Topographic Property Patterns

Each entry in this appendix describes a concept demonstrated by data variables structured by the logical order of semantic roles.

Affordances: Patterns that indicate affordances include conditional information

Language pattern: token auxiliary (AUX) Head.text

Agency: The volitional or primary cause of the action expressed by a verb and the agentive case. The head of the pattern "verb and agent" is the dependent marked with the word "by."

Language pattern: verb (VERB) agent (AGENT) by

Multiple actions: Actions occurring concurrently

Language pattern: VERB coordinating conjunction (CC) and

Ordinal property: Data whose values are ordered

Language pattern: numeral (NUM) adjectival modifier (AMOD) more

Spatial conjunction: Related spatial positions

Language pattern: above CC or; above conjunction (CONJ) below

Spatially relative: The relative relation of one thing to another

Language pattern: above object of preposition (POBJ) deck; above POBJ level

Feature topology: Topological relations among features

Language pattern: access compound traffic

Spatial prepositions: A spatial relationship indicated by a preposition

Language pattern: access preposition (PREP) between; access PREP to

Modifier—verb: A term that describes or adds detail to a verb similar to an attribute

Language pattern: VERB adverbial modifier (ADVMOD) naturally

Modifier—noun: A term that describes or adds detail to an object similar to an attribute

Language pattern: noun (NOUN) AMOD major

Subordinating conjunction: Word connecting the dependent clause to a main clause

Language pattern: acquired mark after; token AUX Head.text

Topological conjunctions: a word joining together words, phrases, or clauses using topological relation indicated by a preposition or as an adverb.

Language pattern: across CC or; across CONJ into

Topological relation to feature: Relation indicated by a

Language pattern: across POBJ bottom; across POBJ course; across POBJ terrain

National Geospatial Technical Operations Center
U.S. Geological Survey
Box 25046, Mail Stop 510
[or 6th and Kipling, MS 510]
Denver Federal Center
Denver, Colorado 80225-0046
Or visit the National Geospatial Technical Operations Center website at
https://www.usgs.gov/national-geospatial-technical-operations-center
Publishing support provided by the Science Publishing Network,
Denver and Reston Publishing Service Centers

ISSN 2328-0328 (online)
https://doi.org/10.3133/sir20255064