ISO/TC 37/SC4 N312
Date: 2006-08-21

ISO/TC 37/SC 4/WG 1/TermLR

List of Terms

1.1 **annotation**

1.1.1
binary value
There are two binary values: plus and minus. In xml, they may be represented as < plus/> and < minus/>.

1.1.2
boolean operator
Negation, conjunction, disjunction and conditional are each called boolean operator on truth values or propositions.

1.1.3
element
An element in xml normally consists of a start-tag, content and end-tag, as shown schematically by: <element type> content </element type>.

1.1.4
empty element
In xml, an empty element is an element that has no content. It is either represented as in the form of <element> </ element> or simply as <element />. Unlike the empty feature structure which is unique, there can be many different empty elements with differ

1.1.5
eXtensible Markup Language
XML. One of the conventional standards for a set of symbols and rules that are used for marking the structure of a text and the characteristics, properties and attributes of their components or basic elements.

1.1.6
feature specification
Assignment of a particular value to a feature in a feature structure

1.1.7
multiple inheritance
Some types may inherit properties from more than one supertypes. For example, a whale is both a mammal and a fish, thus inheriting properties from both of them.

1.1.8
multiset
A multiset, represented as f. .g or f gm, is a collection of objects that, unlike an ordinary set, allows the occurrence of identical elements in it. Hence, fa, a, bg is not the same as fa, bg. On the other hand, like an ordinary set, its members are not

1.1.9
unification
A binary operation on feature structures that combine two compatible feature structures into one representing exactly the information contained in the feature structures being unified.

1.1.10
collocation
Phrasal verbs, reflexive constructions and collocations are complex morphosyntactic structures but units at the semantic layer having one single lexicalised meaning.

1.1.11
facial expression
This new framework of course uses established methods from spoken dialogue evaluations but has to take into account new methods to handle multimodal characteristics like gestural input combined with speech input, graphical vs. speech output.

1.1.12
definite description
Referential readings of definite descriptions are handled by proposing referents from the external application context as well as the CLE context model.

1.1.13
annotation tool
To support manual annotation we developed an annotation tool for lexical semantic tagging (KiC) that allows for fast, and consistent manual tagging.

1.1.14
annotation process
Its main purpose is to speed up the annotation process through a user-friendly interface and by implementing several automatic ways to identify possible important/unimportant sentences.

1.1.15
anaphora resolution
All anaphora resolution methods make use of information about syntactic function, in rules such as: preference for a subject antecedent, syntactic parallelism, resolution of reflexives to the subject, Mitkov's collocations pattern rule.

1.1.16
common noun
In the case of antonomasia the proper name has polilexical units with common nouns among its variants (The fugitive Ghibelline); in the case of metaphor the common noun has a proper name among its variants, used frequently in a peculiar metaphorical manner.

1.1.17
co-occurrence
The tools for querying the corpus are based on Perl regular expressions and allow to check co-

occurrence of words or groups of words, specific morphological or lexical features of words.

1.1.18
constraint
Ensuring the validity of feature structures may require much more than simply specifying the range of allowed values for each feature. There may be constraints on the co-occurrence of one feature value with the value of another feature in the same feature

1.1.19
argument structure
As we know, it is the first research work to describe the argument structure through syntactic function

1.1.20
classifier
The classifier assigns classes to instances in an unlabeled data set, and then instances with top-ranking reliabilities of their labels are added to the labeled training set.

1.1.21
basic movement
The micro-level lexicon is used to translate the basic movements into animation, which is represented by a sequence of avatar postures.

1.1.22
aligner
First, an aligner performs a forced alignment of the speech with the string of phonemic symbols representing that signal.

1.1.23
context vector
Each training sample can be represented as context vectors with its contextual words.

1.1.24
Bitext
The result of an entire text alignment is also known as a bitext.

1.1.25
complex nominal
Accessing the lexical information of head nouns involved in some types of complex nominals and analyzing in particular their extended qualia structure, some clues to the interpretation of modifying PPs may be acquired.

1.1.26
feature set
The mechanism in Co-training assumes that two feature sets are independent.

1.1.27
TTS
TTS can support quality control in treebank and computational grammar development.

1.1.28

domain ontology
Mining terms and their relations from real-world free texts is attracting increasing attention, for example, the domain adaptation capability of IE systems relies on automatic acquisition of domain ontology and lexicosyntactic patterns for template filling.

1.1.29
context information
As a result of this experiment, we proved that context information throughout local density and selective sampling is more suitable and discriminative in WSD.

1.1.30
evaluation data
The corpus was divided into training data and evaluation data by extracting evaluation data randomly from the entire set of data.

1.1.31
manual annotation
The manual annotation of recorded multimodal behaviour on video may have several goals for researchers building multimodal systems: exploring the expected combination of input modalities, collecting training data for classifiers, extracting behaviour rule.

1.1.32
error rate
The reason for this restriction lies in the fact, that the word error rate for dictation for an untrained user and an open domain is still in the range of 10% which is far too high for wide spread use.

1.1.33
convolution
Results of the convolution and noise addition tests are presented for a speaker-dependent name recognition task.

1.1.34
annotation scheme
Contextual USAGE : Indeed, the fact that it contains the list of all data items that the annotation scheme can make use of, it is probably the best source of information for potential users or implementers who want to know whether a given item corresponds.

1.1.35
attribute
property of some object being described [ISO CD 2460-1]

1.1.36
boxed integer
integer in a box like 1 marking structure sharing in a feature structure [ISO CD 2460-1]

1.1.37
annotation
In this document the term annotation refers to the process of adding linguistic information to language data (“annotation of a corpus”) or the linguistic information itself (“an annotation”), independent of its representation.

1.1.38
Representation
The term representation refers to the format in which the annotation is rendered, e.g. XML, LISP, etc. independent of its content.

1.1.39
Stand-off annotation
Annotations layered over a given primary document and instantiated in a document separate from that containing the primary data.

1.2 data

1.2.1
collection
list, set or multiset of values

NOTE A list is an ordered collection of entities some of which may be identical, whereas a set or a multiset is an unordered collection of entities.

1.2.2
complex value
value represented either as a feature structure or as collection

1.2.3
data category selection
component of a TML's specification that constrains its informational content

USAGE The diagram exemplifies the various roles of a data category selection in the process of defining and using any linguistic annotation scheme

1.2.4
data element concept
concept that can be represented in the form of a Data Element, described independently of any particular representation

1.2.5
data object
discrete data, considered as a unit, representing an instance of a data structure that is known or assumed to be known [TERMIUM]

USAGE As such, at the base primary data objects are relatively simple in their structure; more complex data objects may consist of a list or set of contiguous or non-contiguous locations. [LREC 2002, 327.txt]

1.2.6
data set
collection of data in a major storage unit for which the system provides access facilities [TERMIUM]

NOTE that the training data which feeds the generation process is the very same dataset as that for which readings are generated. [LREC 2002, 313.txt]

1.2.7
lexical database
lexical resource. database collection consisting of individual data entries each of which documents a word and provides data pertinent to the senses associated with that word, as well as in some cases equivalent words in one or more languages

1.2.8

data category registry
data category specification used as a normative reference for the description of a TML [ISO 12620]

1.2.9
data element
a unit of data for which the definition, identification, representation and Permissible Values are specified by means of a set of attributes [ISO 12620]

1.2.10
empty path
path corresponding to the root node of a graph [ISO CD 2460-1]

1.2.11
data category
result of the specification of a given data field or the content of a closed data field

NOTE A data category is to be used as an elementary descriptor in a linguistic structure or an annotation scheme. Examples are: /term/, /definition/, /part of speech/ and /grammatical gender/.

1.2.12
directed acyclic graph
graph with directed edges and no cycle

NOTE DAG

1.2.13
form
any sequence of letters, pictograms and numerals used to write or pronounce a word

1.2.14
grammatical category
See also part of speech

1.2.15
natural language processing
the field of study covering knowledge and techniques which allow computerized processing of linguistic data. This field combines a variety of skills including linguistics, mathematical logic, statistics, and algorithms.

NOTE NLP

1.2.16
open data category
data category whose content cannot be fully enumerated due to the organic nature of language

EXAMPLE Typical open data categories might include /term/, /lemma/.

1.3 dialog

1.3.1
discourse model
model whose objective is to communicate with a user in a manner that is most natural for him or her [TERMIUM]

USAGE A template extraction algorithm was implemented that explores the discourse model , extracting and sequencing in temporal order instances of the 31 event types. [LREC 2002, 157.txt]

1.3.2
distributional analysis
mapping of how smaller linguistic units are distributed within larger ones [ASHER]

USAGE The distributional analysis gives a more accurate and more global view of the corpus, allowing to take better decisions. [LREC 2002, 90.txt]

1.3.3
imperative
verb mood used to express direct command. In English, the implied subject you is never expressed. [www.southwestern.edu/~carlg/Latin_Web/glossary.html]

EXAMPLE Do this

1.3.4
mood
mode or manner in which the action of a verb is represented

NOTE Latin has three moods: imperative, indicative, and subjunctive.
[www.southwestern.edu/~carlg/Latin_Web/glossary.html]

1.3.5
object language
language being described [ISO 12620]

1.3.6
order
state in which all components or elements are arranged logically, comprehensively, or naturally [www.wordreference.com/english/definition.asp?en=order]

1.3.7
passive voice
verb voice, in which the subject is the recipient of the action of the verb

EXAMPLE My lunch was eaten by them

1.3.8
typology
classification of languages according to structural similarities without reference to historical considerations [ASHER]

1.4 feature

1.4.1
feature
property of an entity

NOTE When provided with a value, it constitutes a feature specification, for example, .number is a feature, a pair .number singular is a feature specification, .singular is a value.

1.4.2
feature structure
set of feature specifications

NOTE The minimum feature structure is the empty feature structure

1.4.3
typed feature structure
feature structure sorted into a type which is labelled by the name of the type

NOTE In the graph notation, each node is labeled with the name of a type. In the matrix notation, a type name is ordinarily placed at the upper left corner of the inside of the pair of square brackets that represents a typed feature structure.

1.4.4
morphological feature
category induced from the inflected form of a word

NOTE: ISO 12620 provides a comprehensive list of values for European languages. An example of a morphological feature is: /grammatical gender/.

1.4.5
transcription
form resulting from a coherent method of writing down speech sounds transitive verb a verb which takes a direct object

1.4.6
attribute-value matrix
AVM. A very common notation in a matrix form which represents a feature structure consisting of pairs of an attribute, namely feature, and its value.

NOTE The acronym avm stands for \Attribute-Value Matrix where each row represents a pair of a feature and its value, separated by a colon (:), space ( ) or the equality sign (=).

1.4.7
feature structure declaration
sometimes called feature structure description. A feature structure may be described in a declarative manner through some description language. [ISO CD 2460-1]

1.4.8
reentrancy
phenomenon by which two paths point to the same node on a graph representing a feature structure. These paths are then called equivalent. As a result, the two paths leading to that intersecting node share its features or attribute values.

1.4.9
alternation
operation on feature values that returns one and only one value.

NOTE Given a feature specification F : a| b, where a| b denotes the alternation of a and b, F has either the value a or the value b, but not both.

1.4.10
empty feature structure
feature structure containing no feature specifications.

1.4.11
feature specification
assignment of a value to a feature

NOTE Formally, it is treated as a pair of a feature and its value.

1.4.12
graph notation
notation that uses a single rooted graph with a finite sequence of labelled and directed arcs to represent feature structures.

NOTE The name of each node including the root represents a type, the label of each arc a feature, and the name of the terminating node of each arc a value.

1.4.13
incompatibility
relation between two feature structures which have at least one common feature with a conflicting value.

NOTE Two feature structures that are not incompatible can be unified. The empty feature structure is compatible with any other feature structure.

1.4.14
matrix notation
attribute-value matrix AVM notation that uses square brackets to represent feature structures.

NOTE In a matrix notation, each row represents a feature specification, separated by a colon (:), space ( ) or the equality sign (=).

1.4.15
type
class of entities.

NOTE Primitive entities that features take as their value are considered types, too, namely unit

classes consisting of themselves. For example, words can be sorted into types such as `verbs', `nouns', `adjectives', and soon.

1.4.16
unification
operation that combines two compatible feature structures into one.

NOTE Feature structures can be unified only if they do not have an incompatibility. Feature structures carry partial information and, by unification, their information is incremented.

1.4.17
union
operation that combines two sets, or multisets, into one.

NOTE The equivalent operation for lists is concatenation.

1.4.18
value
information about an entity.

NOTE There are two types of feature values: atomic value and complex value.

1.4.19
Feature structure description
A feature structure is often described in a declarative manner through some description language. This should not be confused with feature system declaration fsd which describes the set of all valid feature structures.

1.4.20
shared value
Feature value shared by two or more features in a feature structure. In graph notation, a node to which two or more paths merge represents the value shared by the paths.

1.4.21
identity element
The empty feature structure is an identity element of the operation called unification on feature structures, since it yields the identical result when unified with any other feature structure just as the number 0 is an identity element for the algebraic operation.

1.4.22
markup
The process of adding formatting or other processing commands to a text in order to provide information about the logical as well as physical structure of its content.

1.4.23
multivalued feature
Some features, especially used in linguistics, take sets, multisets or lists as values. These are called multivalued features. The feature comps, for instance, takes a list of complements like object and indirect object as value.

NOTE they do not violate.

1.4.24
negation
In this standard, negation applies to feature values only and is not understood as a truth function as in ordinary bivalent logics. It is understood more or less in a set-theoretic sense.

1.4.25
path
Sequence of feature names which, in the graph notation, label each of the arcs, starting from the root. The notion of path can also be extended in the same manner to other notations.

1.4.26
subsumption
A reflexive, anti-symmetric and transitive relation between two feature structures.

1.4.27
structure sharing
A feature structure in which one or more feature-values are shared or re-used. See reentrancy.

1.4.28
tag
The feature structure in xml notation is surrounded in <fs> tags, and each of its features with <f> tags.

1.4.29
type, feature structure

Elements of a domain can be sorted into classes in a structured way, based on similarities of properties. These classes are called types.

NOTE In linguistics, for instance, class names like phrase, word, pos (parts of speech), noun, and verb are often taken as types.

1.4.30
typing
By typing, each feature structure is assigned a particular type. Each feature specification with a particular value is then constrained by this typing. A feature structure of the type noun, for instance, would not allow a feature like tense in it or a spec.

1.4.31
boxed label
Label in a box like 1 or A which is used for marking structure sharing in an avm.

NOTE The index is normally marked with an integer, but it can be any alpha-numeric symbol that can be used as a coreferential index.

1.4.32
type constraint
The construction of well-formed feature structures are constrained by some type inheritance hierarchy. A feature structure of some subtype must inherit all the constrains laid on its supertypes.

1.4.33
atomic value

In a feature structure, the value of each feature is either atomic or complex. An atomic value is some primitive type of object that has no internal feature specification or value structure. See complex value.

1.4.34
bag
See the term multiset.

1.4.35
generalization
While unification enriches information, generalization captures common features from various feature structures. It is a binary total operation on two feature structures.

1.4.36
boolean value
In bivalent logic, two boolean values, truth and falsity, are admitted. They are often represented as 1 and 0, + and -, or positive and negative as polarity values.

1.4.37
distinctive feature
Feature that distinguishes an object from others. The sound segments /p/ and /b/ in English, for instance, are distinguished from each other in the specification of a feature voicing: one is voiceless and the other voiced.

1.4.38
concatenation
Two feature list values may be concatenated into a single list, represented by the concatenation operator.

1.4.39
compatibility
Two feature structures are compatible if and only if none of the features that they have in common has a conflicting value. On the other hand, two incompatible feature structures contain at least one identical feature which has a conflicting value.

1.5 relation

1.5.1
relation
Contextual Usage: This attribute may be used for specifying which of various logical relations the given value has to the actual value of the feature.

1.5.2
type inheritance hierarchy
types are ordered in some hierarchical order so that objects of a lower type inherit properties of their super-types. In linguistics, these hierarchies are often used to organize linguistic descriptions, especially lexical information. [ISO CD 2460-1]

1.5.3
validation
value of a feature in a feature structure may either be atomic or complex. A value is complex if it is a feature structure itself or a list of values, again either atomic or complex. [ISO CD 2460-1]

1.5.4
working language
language used to describe objects [ISO 12620]

1.5.5
conceptual domain
finite list of simple data categories that may be the values of a complex data category

1.6 word

1.6.1
unique
value which is applied to categories with a unique or very small membership, such as negative particle, which are `unassigned' to any of the standard part-of-speech categories. The value unique cannot always be strictly applied.

1.6.2
homophone
word that sounds like another word but is different in writing or meaning

NOTE An example of difference in spelling is “pair” compared to “pear” or “pare” in “The cook used a knife to pare the pair of pears”.

1.6.3
inflected form
form that a word can take when used in a sentence or a phrase

NOTE An inflected form of a word is associated with a combination of morphological features, such as grammatical number or case.

1.6.4
part of speech
category assigned to a word based on its grammatical and semantic properties See also grammatical category

NOTE ISO 12620 provides a comprehensive list of values for European languages. Examples of such values are: /noun/ and /verb/.

1.6.5
romanization
transliteration from a non-Latin script into a Latin script

1.6.6
script
set of graphic characters used for the written form of one or more languages (ISO/IEC 10646-1, 4.14)

1.6.7
simple data category
data category that may be the possible content of a closed data category, but that cannot itself be further sub-divided

EXAMPLE /masculine/, /feminine/, and /neuter/ are possible simple data categories

associated with the conceptual domain of the closed data category /grammatical gender/ as it is associated with the German language.

1.6.8
transliteration
form resulting from the conversion of one writing system into another.

1.6.9
word
in the context of a given language, is a description composed of at least a part of speech and a lemmatized form

NOTE The description can include more morphological information and/or syntactic and semantic information. A word is either a single word or a multi-word expression.

1.6.10
word class
See also part of speech

1.7 others

1.7.1
base document
document containing data to be captured in order to be processed by a data processing system [TERMIUM]

1.7.2

cardinal numeral
numeral of the class whose members : - are considered basic in form - are used in counting, and - are used in expressing how many objects are referred to [www.sil.org/linguistics/GlossaryOfLinguisticTerms/WhatIsACardinalNumeral.htm]

1.7.3
communication
transfer of data among functional units according to sets of rules governing data transmission and the coordination of the exchange [TERMIUM]

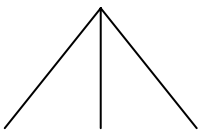
USAGE Thirdly, a major characteristic of human communication is behavioural coordination. [LREC 2002, 214.txt]

1.7.4
context
text which illustrates a concept or the use of a designation [ISO 12620]

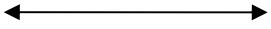
Annex A
(informative)

Concept diagrams

A.1 Graphic representation used in the concept diagrams

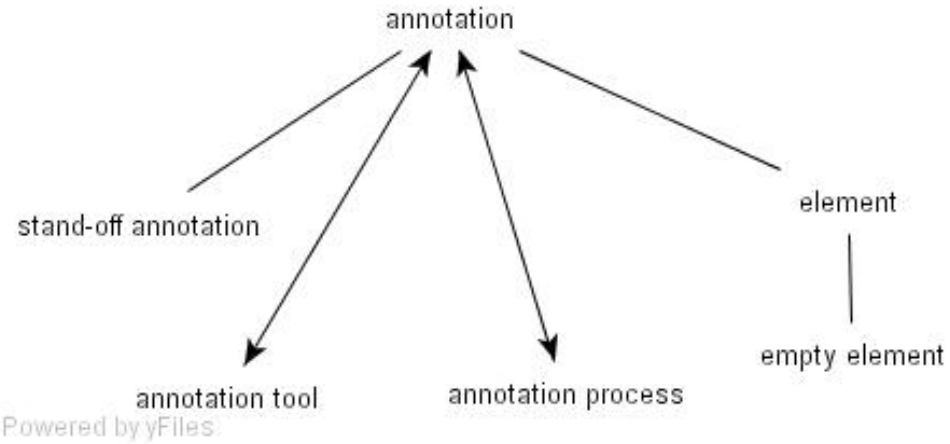


General relations
Are represented by tree diagram

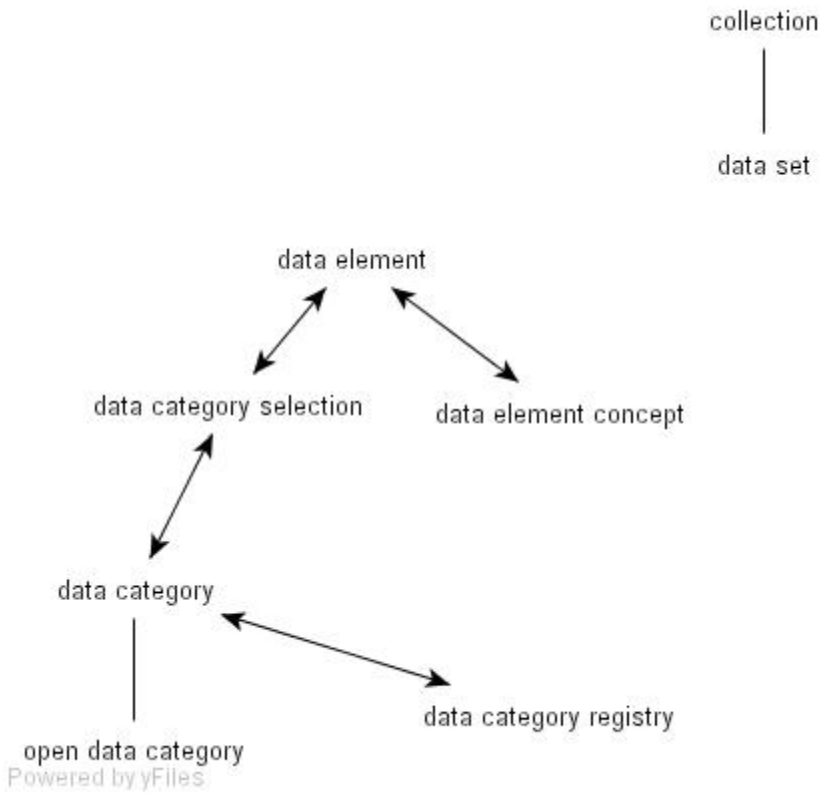


Associative relations
Are represented by arrow diagrams

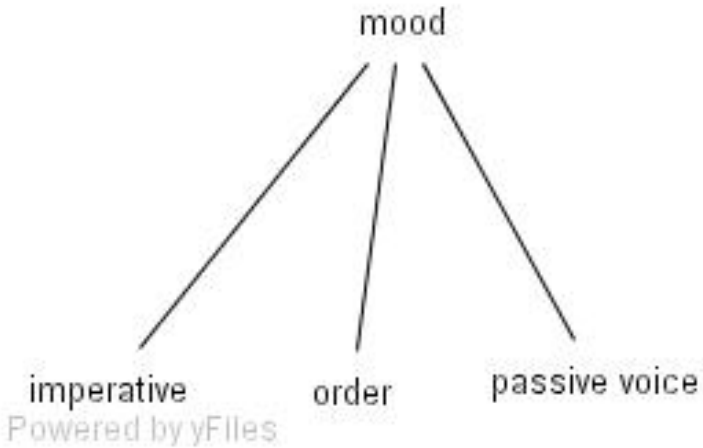
A.2.1 annotation



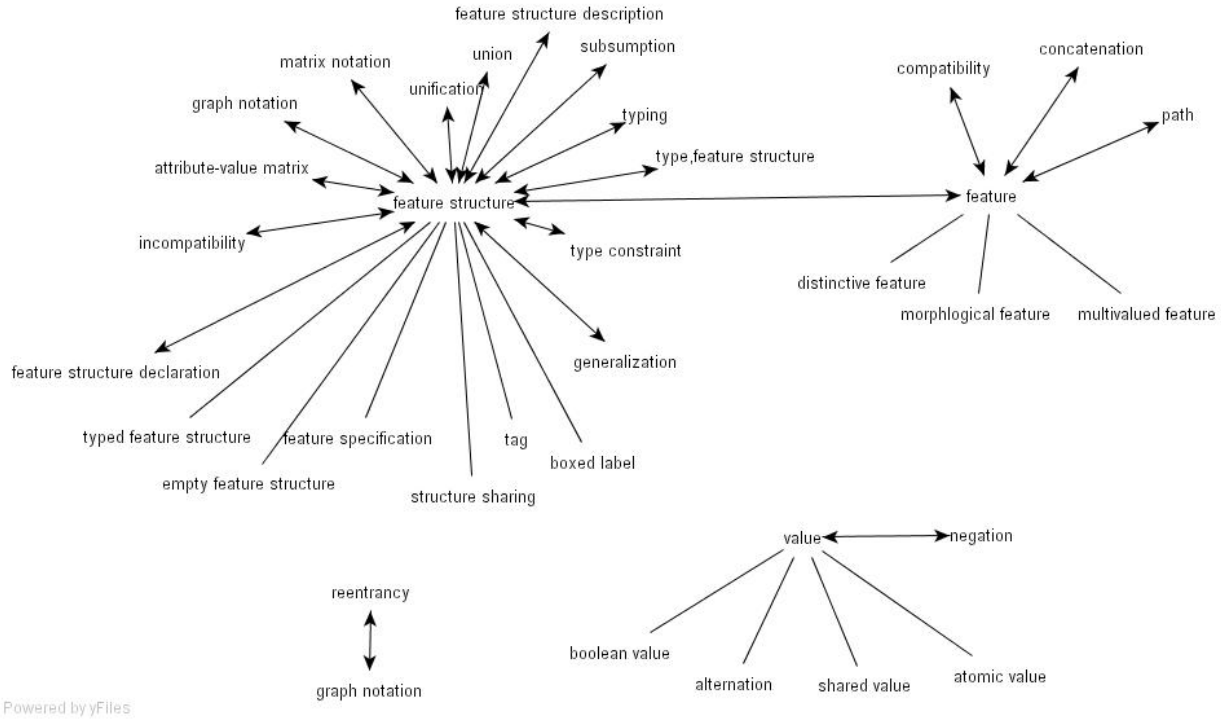
A.2.2 data



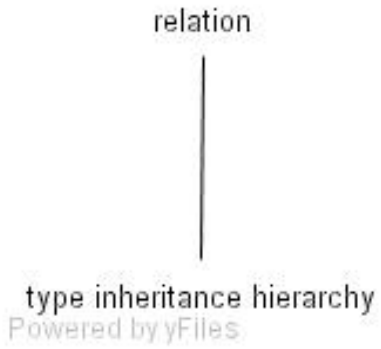
A.2.3 dialog



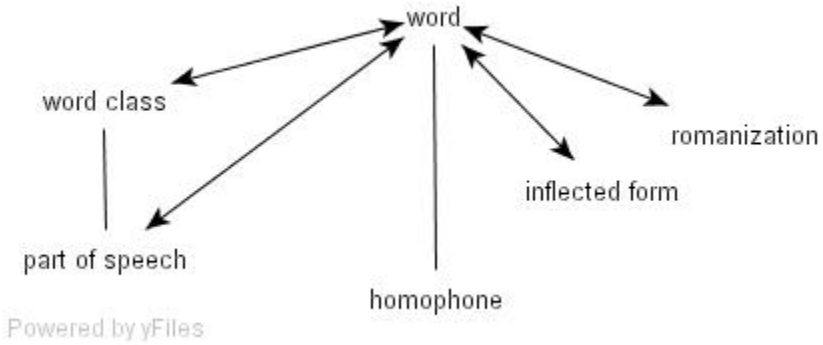
A.2.4 feature



A.2.5 relation



A.2.6 word


word
word class
romanization
inflected form
part of speech
homophone
Powered by yFiles