BalkaNet Terms Glossary

This section contains a subset of terms used rather frequently throughout this volume. They represent part of the BalkaNet jargon which draws on the EuroWord-Net and Princeton WordNet terminology. Although many glosses are specific to the wordnet development terminology, being sometimes more specific than the general terms used in lexicography, we considered that this even partial term glossary would help the reader to better follow the individual presentations included herein. For each term we added a short definition (marked up with a **D:**) followed sometimes by an illustrative example (marked up with an **E:**).

adjective cluster

> **D:** A group of adjective synsets that are organized around antonymous pairs or triples of adjectives. An adjective cluster contains two or more head synsets which represent antonymous meanings. Each head synset has one or more satellite synsets.

aligned wordnets

> **D:** a set of monolingual wordnets with the property that their synsets are linked by complex equivalence relations to the records stored in an interlingual index; starting from one synset in a monolingual wordnet, via the interlingual index, one could reach the synsets in the other monolingual wordnets that lexicalize similar meanings. The quality of wordnets alignment is expressed (among the others) by the cross-lingual coverage.

also _see_ relation

> **D:** A semantic relation between related synsets.
>
> **E:** {`cold`} ALSO_SEE {`cool, frozen`}.

antonym

> **D:** A word that express an opposing meaning to that of another word. The *antonymy* lexical relation holds between two antonyms.
>
> **E:** wet-dry; man-woman; fall-rise.

BalkaNet

> **D:** An EC funded project (IST-2000-29388) that aims at developing a multilingual lexical resource representing semantic relations among basic concepts of the Balkan languages, more precisely aligned wordnets for Greek, Turkish, Romanian, Bulgarian, Czech and Serbian. Each of the BalkanNet wordnets is structured along the same lines as the EuroWordNet (`http://www.ceid.upatras.gr/ Balkanet/`).

BalkaNet Concept Set (BCS)

> **D:** A set of concepts commonly agreed to be implemented by all consortium members, in order to obtain a large-scale overlap between monolingual wordnets. Usually they correspond to meanings important in the languages involved

in the project. The projection of this set of concepts on PWN 2.0 is closed under *hyperonymy, holonymy* and *be_in_state* relations. It contains 8 516 concepts, including the EWN base concepts and common concepts proposed by each partner according to language specific criteria.

BCS1
>   **D:** The first BalkaNet Concept set, based on the EWN Base Concept set. Initially it contained 1 024 concepts, but due to refinements of PWN 1.5, PWN 1.6, PWN 1.7.1 and PWN 2.0, currently it contains 1 218 concepts. These concepts are lexicalized in all the languages represented in BalkaNet and each has a Top Ontology description attached to the concept.

BCS2, 3
>   **D:** The second and third BalkaNet Concept sets contain 3 510, respectively 3 788 concepts which are lexicalized in most languages represented in EWN. Also the concepts proposed by at least two partners were selected from a list of candidate concepts that would be relevant for the BalkaNet languages (according to monolingual selection criteria). The conceptual density criterion was also taken into consideration for the BCS selection.

BalkaNet ILI Datasource
>   **D:** A WMS component containing the ILI semantic data, along with a set of ontologies that are interconnected to ILI concepts.

Base Concept
>   **D:** A concept that is widely used, with its importance reflected in the ability to function as an anchor to attach other concepts; the anchoring capability is defined in terms of the number of relations (general or limited to hyponymy) and the high position of the concept in a hierarchy.

base form
>   **D:** The base form of a word or collocation is the form to which inflections are added.
>   **E:** `clean` $\rightarrow$ `cleanly, cleanliness, unclean`

***be_in_state* relation**
>   **D:** This relation specifies a value for an attribute; the values related by *be_in_state* are represented by descriptive adjectival synsets and the attributes by nominal sunsets.
>   **E:** `(tall:1)`$\rightarrow$ `be_in_state (stature:2 height:3)`.

BalkaNet ILI (BILI)
>   **D:** The structured ILI used by the BalkaNet wordnets. It includes PWN (version 2.0), domain structuring (links to SUMO concepts), plus the records specific to the Balkan languages.

***category_domain* relation**
>   **D:** A semantic relation which allows topical classifications of the meanings represented by the synsets; the target synset of the relation is always a nominal

synset, but any synset, irrespective of its POS can be source of this relation, i.e. can be topical classified.

E: `(diplomatic immunity:1) CATEGORY_DOMAIN (law:2 jurisprudence:2)`

causes relation

D: A relation between an event (denoted by a verb or a noun) causing a resulting event, process or state referred to by another verb, noun or adjective.

E: `to kill (/a murder) CAUSES to die (/death)`.

Central WMS Server (CWMS Server)

D: The main server in WMS, serving requests for BalkaNet ILI data and domain information and acting as a directory service provider, registering Wordnet hosts and distributing this information to the other nodes of the network.

cognates

D: Two words from different languages, with similar orthographical form and sense; cognates can be similarly defined for languages that use different scripts, provided a common transcription is used.

E: `strophe (English) - strofă (Romanian) - strophe (French)`.

co-hyponym, coordinate term

D: A term at the same level of classification as another, sharing the same hypernym.

E: `cocktail party` is a co-hyponym of `birthday party`, both having `party` as the common hypernym.

collocation

D: A string of two or more words connected by spaces or hyphens, frequently co-occurring in natural language.

E: `man-eating shark, blue-collar, line of products`, etc.

complex equivalence relation

D: Equivalence relations used to map wordnet meanings to interlingual concepts and thus allowing inter-wordnets navigation.

E: EQ_SYNONYM: the synset in the starting wordnet has a meaning that is conceptually equivalent to an ILI-record;

EQ_NEAR_SYNONYM: the synset in the starting wordnet has a meaning that covers multiple ILI-records;

EQ_HAS_HYPERONYM: the synset in the starting wordnet has a more specific meaning than the ILI record (and an EQ_SYNONYM ILI-record is not defined);

EQ_HAS_HYPONYM: the synset in the starting wordnet has a more general meaning than the ILI-record (and an EQ_SYNONYM ILI-record is not defined);

Other complex equivalence relations, almost self-defined, are:

EQ_INVOLVED, EQ_ROLE, EQ_IS_CAUSED_BY, EQ_CAUSES, EQ_HAS_HOLONYM, EQ_HAS_MERONYM, EQ_HAS_SUBEVENT, EQ_IS_SUBEVENT OF, EQ_BE_IN STATE.

concept, interlingual concept

> **D:** A language independent semantic structure that represents equivalent meanings from more than two languages; in the ILI a concept is referred through a code, has an attached gloss and a set of links from and to the equivalent meanings from the multilingual dictionary.
>
> **E:** `ILI-code:  08232464-n`
> `GLOSS: a book in which names and transaction are listed`
> `LINKS: EN → (register_3)`
>
> > `RO → (registru_1, catastif_1, condică_1)`
> >
> > `SP → (registro_6)`
> >
> > `CA → (registre_3)`
> >
> > `BA → (erregistro_5, errejistro_1)`

conceptual density criterion

> **D:** This term assumes the structuring of the interlingual index (ILI). The criterion requires that once a concept (lexicalized in PWN as a nominal or verbal synset) was selected for being implemented in the monolingual wordnets, all its ancestors (up to the top level) were also selected; the concepts that correspond in PWN to adjectival synsets were selected so that they would be related (*be_in_state*) to nominal concepts in the selection.

conceptual domain

> **D:** A topical classification that groups related terms to the same topic. In BalkaNet some of the BILI nodes are linked to a conceptual domain.

conceptual distance

> **D:** The distance of conceptual nodes within the BILI taxonomy; it reflects semantic similarities between terms.

conflict in sense assignment

> **D:** The situation generated by the existence of a literal tagged with the same sense identifier in more than one synset of the same wordnet.

content-word

> **D:** A word the grammatical category (part of speech) of which is noun, verb, adjective or adverb. In wordnets only synsets of content words are represented.

coordinate

> **D:** See **co-hyponym**.

cross-cluster pointer

> **D:** A semantic pointer from one adjective cluster to another.

cross-lingual coverage

> **D:** A quantitative figure that expresses the number of concepts in ILI/BILI which have been implemented in all wordnets of a system of aligned wordnets.

cross-POS (X-POS) relation

>**D:** A relation between synsets/words with different parts of speech.
>**E:** XPOS_synonymy, XPOS_near _synonymy, XPOS_hypernymy, XPOS_hyponymy, XPOS_antonymy, as well as the syntagmatic relations.

dangling node

>**D:** An existing synset which has no hypernym and is not mapped to a topmost node in ILI/BILI (such as unique beginners for the noun hierarchy in PWN).

dangling relation

>**D:** A relation in a certain wordnet supposed to connect two synsets that would correspond to two concepts in the commonly agreed sets, and of which one of the two synsets is missing.

definition

>**D:** A natural language description of a word meaning in a lexical resource.
>**E:** `noun_1 = a word that can be used to refer to a person or place or thing`.

definitional productiveness

>**D:** The number of sense definitions a noun participates in. It was used as on criterion for identifying the most relevant ILI concepts to be implemented in the Romanian wordnet.

***derived_from* relation**

>**D:** A lexical relation that links derivatives to their base forms.
>**E:** (adverb&adjective: `quickly` DERIVED_FROM `quick`),
>
>(adjective&adjective: `astomatal` DERIVED_FROM `stomatal`),
>
>(adjective&noun: `abbatical` DERIVED_FROM `abbey`).

direct antonyms

>**D:** A pair of words with opposed meanings between which there is an associative bond built up by co-occurrences. For the adjective clusters, direct antonymy is established only between head synsets.

domain

>**D:** A topical classification to which a synset has been linked with a CATEGORY, REGION or USAGE pointer.

domain concept

>**D:** A concept restricting a word meaning to a certain topic or domain.

domain ontology

>**D:** A set of ontologically structured domain concepts.

domain term

>**D:** A synset linked to a topical class. A domain term is further identified as being a CATEGORY_TERM, REGION_TERM or USAGE_TERM.

entailment

    D: A verb **X** entails **Y** if **X** cannot be done unless **Y** is, or has been, done.

    E: `eat_1` entails (`chew, masticate, mandicate`) and (`swallow, get down`).

equivalence relation

    D: A relation from a synset in a language-specific wordnet to an ILI-record in the ILI. Also see **complex equivalence relation**.

equivalent meanings

    D: The meanings (in more than two languages) that correspond to the same interlingual concept.

EuroWordNet (EWN)

    D: A multilingual lexical ontology with wordnets for several European languages (Dutch, Italian, Spanish, German, French, Czech and Estonian). The wordnets are structured in the same way as the Princeton WordNet, in terms of synsets (sets of synonymous words) with basic semantic relations between them. Each wordnet represents a unique language-internal system of lexicalizations. In addition, the wordnets are linked to an Inter-Lingual-Index, based on the Princeton wordnet. Via this index, the languages are interconnected so that it is possible to go from the words in one language to similar words in any other language. The index also gives access to a shared top-ontology of 63 semantic distinctions. This top-ontology provides a common semantic framework for all the languages, while language specific properties are maintained in the individual wordnets. (`http://www.illc.uva.nl/EuroWordNet/`)

expand model

    D: An approach to build a multilingual database based on the expansion of the PWN through the translation of a core set of synsets to other languages following their equivalence links.

    From a technical point of view, the expand-model seems less complex, guaranteeing the highest degree of compatibility across the different wordnets. The problem for the expand-model is however that the multilingual system will be highly biased by the PWN. It will not only contain all the mistakes and gaps that are present in PWN (just like any other dictionary) but it will also be structured by the (American)-English lexicalization of Western concepts.

gloss

    D: A natural language definition and/or example sentences for a synset.

    E: `synset:1 - (a set of one or more synonyms)`

head synset

    D: Synset in an adjective cluster containing at least one word that has a direct antonym.

hierarchy preservation principle

    D: An assumption that gives the basic motivation for the automatic import of most of the semantic relations from PWN into another aligned wordnet: if two

source synsets $S_{1SOURCE}$ and $S_{2SOURCE}$ are linked by a semantic relation R and if the $S_{1TARGET}$ and $S_{2TARGET}$ are the correspondingly aligned synsets in the target wordnet, then they will be linked by the relation R. If in the source wordnet there are intervening synsets between $S_{1TARGET}$ and $S_{2TARGET}$, then the relation R is set between the corresponding target synsets only if R is declared as transitive (R^+, unlimited number of compositions, e.g. hypernym) or partially transitive relation (R^k with k a user-specified maximum number of compositions, larger than the number of intervening synsets between $S_{1TARGET}$ and $S_{2TARGET}$).

holonym

D: The name of the whole of which the meronym names a part. **Y** is a holonym of **X** if **X** is a part of **Y**. The holonymy relation holds between a meronym and its holonym.

E: `finger` HAS_HOLONYM `hand`.

hyper(o)nym, hyperonym

D: A synset Y is the hypernym of a synset X iff the meaning of Y is more general than the meaning of X (or the meaning of X is more specific than the meaning of Y). The hyper(o)nymy relation holds between a synset and its direct hypernym.

hyponym

D: A synset X is the hyponym of a synset Y iff the meaning of Y is more general than the meaning of X (or the meaning of X is more specific than the meaning of Y). The hyponymy relation holds between a synset and its direct hyponym.

ILI database or ILI

D: In EWN: an unstructured database of ILI records, inter-linking language-specific synsets from different wordnets.

In BalkaNet: see **BILI**.

ILI code

D: An unique alphanumeric identifier for an interlingual index record (ILIR). It is composed by a prefix that identifies the PWN version used as interlingual index (e.g. ENG20), followed a number (which represents the offset in the PWN of the respective synset) trailed by a letter identifying the part of speech of the literals in the synsets to which the ILI-code corresponds. Any synset in the monolingual wordnets which are mapped to a certain ILIR will have in their XML representation the same ID value as the ILI code of the respective ILIR.

E: In BILI: `ENG20-04295024-n`

ILI record (ILIR)

D: A piece of information indexed by an ILI code that helps one grasp a concept. In BalkaNet this piece of information is represented by a PWN synset, its attached gloss and all its direct relations (including domain/ontology specific relations). The concepts in BCS1 have also attached an ontological description

imported from EWN. In EWN an ILI record has the form
`(<ILI-code><ontological description><gloss>{<domain>})`

indirect antonym
> **D:** An adjective in a satellite synset that does not have a direct antonym has an indirect antonym via the direct antonym of the head synset.

inter-lingual alignment
> **D:** Explicit assigning of the same ID value (ILI code) in all languages to the XML representation of the synsets that are equivalent (EQ_SYNONYM) to the PWN synset represented by the ILIR with the same ILI code.

interlingual relation
> **D:** A relation holding between synsets in monolingual wordnets and ILI records.
> **E:** From a total of 20 relation types: EQ-SYNONYM, EQ-NEAR SYNONYM, EQ-HAS-HYPONYM, EQ-HAS-HYPERONYM etc.

language internal relation (intra-lingual relation)
> **D:** A semantic or lexical relation that holds between synsets or words in the wordnet of a particular language. The language-specific relations are embedded into the monolingual wordnets.
> **E:** relations between nouns: synonymy, antonymy, hypernymy, etc;
>
> relations between verbs: synonymy, antonymy, hyperymy, entailment, etc;
>
> relations between adjectives: synonymy, antonymy;
>
> relations between adverbs: synonymy;
>
> syntagmatic (thematic) relations: Agent, Patient, Instrument, Location, etc.

language-specific concept
> **D:** A concept that is not lexicalized in English, represented in the ILI by a laguage-specific synset. It is mapped by an EO-SYNONYM relation to at least one synset (belonging to the language from which it originated)
> **E:** `BUL-907624109`

language-specific synset
> **D:** A synset that represents a concept not lexicalized in English. The language-specific synsets were manually added to the ILI by a local team. The new ILI record has an adequate prefix (identifying the language that generated it) and from there it could be linked to the synsets of other languages that have a similar lexicalized meaning.
> **E:** `kaşkaval:1 - kashkaval:1 - caşcaval:1` (in Turkish, Bulgarian, Romanian).

lemma
> **D:** Lower case ASCII text of word as found in the WordNet database index files. Usually the base form for a word or collocation.

lexical gap

    D: From a multilingual perspective, a concept that is not lexicalized in a given language, but it can be expressed by a free combination of words, whereas another language expresses the same concept with a lexical unit.

lexical pointer

    D: A lexical pointer indicates a relation between words in synsets (word forms).

lexical semantic network

    D: A special form of traditional semantic network, the nodes of which represent synsets – sets of actual words of a given language sharing (in certain contexts) a common meaning. PWN is a lexical semantic network.

lexical density of a wordnet

    D: The language dependent property of a wordnet to have implemented all the senses from a reference dictionary of the given language.

lexical/linguistic ontology

    D: An ontology in which exists only concepts that are lexicalized (therefore for which there exist literals, in at least one language, that have the corresponding senses). EWN and BalkaNet are lexical ontologies.

literal

    D: Any base form of a word or collocation appearing in one or more synsets of a monolingual wordnet.

location meronymy

    D: The semantic relation that stands between noun synsets for places or locations and the noun synsets of the larger places which include them.
    E: `city-N` HAS_MERO_LOCATION `center-N`.
      `oasis-N` HAS_HOLO_LOCATION `desert-N`.

machine readable dictionary (MRD)

    D: Explanatory dictionaries, synonym dictionaries, antonym dictionaries, phrasal dictionaries, valency dictionaries, etc. available in electronic format.

made_of meronymy

    D: The semantic relation which stands between noun synsets denoting things (concrete objects) and the stuff which compose them (concrete substances).
    E: `table-N` HAS_MERO_MADEOF `wood-N`.
      `paper-N` HAS_HOLO_MADEOF `book-N`.

meaning

    D: A set-theoretic equivalence relation (synonymy) over the set of senses in a given language. The meanings are related to language and are represented by the synsets in the wordnet of a language. The meaning is thus a language specific realization of a *conceptualization* which might be very similar to conceptualizations in several other languages. The gloss attached to one synset covers

all word senses in the synonymy set.

E: `politics:1, political relation:1` - (`social relations involving authority or power`).

member(-group) meronymy

D: The part-whole semantic relation between noun synsets denoting a set and its members.

E: `alphabet-N` HAS_MERO_MEMBER `character-N`.
    `deputy-N` HAS_HOLO_MEMBER `parliament-N`.

merge model

D: An approach to develop independently each wordnet on the basis of the available monolingual lexical resources available in machine readable format and the subsequent linking of the monolingual synsets to their semantic equivalents in other wordnets. The model is cost-effective since it uses existing resources. Because the different resources reflect the relations between words as separate language-internal systems, it is possible to maintain the language-dependent differences. A high degree of consistency can be achieved by combining resources because of a considerable variation in the way semantic information for equivalent words is coded within and across dictionaries. The model has to assure sufficient overlap in the coverage of the different wordnets and still maintaining language-specific properties of the relations and it has to interpret differences found across the different wordnets.

meronym

D: The name of a constituent part of, the substance of, or a member of something. **X** is a meronym of **Y** if **X** is a part of **Y**.

E: Some meronyms for `car`: `accelerator, wheel, automobile engine`, etc.

meronymy

D: The semantic relation that holds between a part and the whole; the reverse relation is holonymy.

monosemous

D: Having only one **sense** for a grammatical category.

E: The noun `synonym`.

near antonymy relation

D: A relation of antonymy between two whole synsets.

E: {`sell, exchange for money`} NEAR_ANTONYM {`buy,purchase,take`}.

non-lexicalized concept

D: A concept that is not lexicalized in a certain language (it can be expressed by a free combination of words). Non-lexicalized concepts are represented by non-lexicalized synsets.

non-lexicalized synset

D: An empty synset in the wordnet for the language that does not lexicalize a given concept. A non-lexicalized synset has a gloss. The non-lexicalized synsets

are apparently redundantly preserved in the hierarchy but their purpose is to reflect the proper interlingual relation between the concept and the closest lexicalized synsets in the wordnet.

parallel corpus

D: A collection of texts in several languages, where one text in a hub language is the original and all the other texts represent its translations. Usually, in a parallel corpus are included professional human translations. A parallel corpus is frequently tagged and aligned at different levels (paragraph, sentence, word). **E:** George Orwell's *Nineteen Eighty Four* (10 languages, sentence aligned, POS tagged).

part of speech (POS)

D: Same as grammatical category. WordNet and the wordnets consider only nouns, verbs, adjectives, and adverbs.

part meronymy

D: The relation stands for the typical component-whole relation between noun synsets, namely, something which is either topologically or temporally included in a larger entity and which as well bears some kind of autonomy, non-arbitrary boundaries, and a definite function with respect to the whole.
E: `car-N` HAS_MERO_PART `engine-N`.
     `keyboard-N` HAS_HOLO_PART `computer-N`.

phraset

D: A multiword expression which is used to express a concept.
E: `ENG-synset  {cornfield}`
     `ITA-synset  {GAP}`
     `ITA-phraset {campo_di_grano}`

pertainym

D: A relational adjective. Adjectives that are pertainyms are usually defined by such phrases as "of or pertaining to" and do not have antonyms. A pertainym can point to a noun or another pertainym.
E: `musical instrument, nervous disease`

polysemous

D: Having more than one sense for one grammatical category.

polysemy count/degree

D: The number of senses of a word in a grammatical category.

Princeton WordNet (PWN)

D: An online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory, developed by the Cognitive Science Laboratory at Princeton University under the direction of Professor George A. Miller. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations

(both semantic and lexical) link the synonym sets. The Princeton WordNet (`http://www.cogsci.princeton.edu/~wn/`) is the standard model for the development of more than 50 monolingual wordnets for more than 40 languages all over the world. A reference to this global endeavour can be found on the Global Wordnet Association home page (`www.globalwordnet.org/gwa/`).

quality control
> **D:** A major task in the BalkaNet project concerning two main issues: validating the quality of the contents and structure of each monolingual wordnet on the one hand, and validating the quality and contents of each wordnet in comparison to the other wordnets within BalkaNet.

recursive definition
> **D:** A gloss definition that uses one of the entry word from the synset it defines.
> **E:** <u>fall</u>(20), light(5) - (<u>fall</u> to somebody by assignment or lot; "The task fell to me"; "It fell to me to notify the parents of the victims").

satellite synset
> **D:** Synset in an adjective cluster representing a concept that is similar in meaning to the concept represented by its head synset.

semantic interlingual validation
> **D:** A validation process that checks, against a multilingual parallel corpus, how the synsets of each monolingual wordnet cover actual use of language and to what degree the established interlingual equivalences among synsets of different wordnets are supported by parallel human translations.

semantic network
> **D:** A graph-like knowledge representation in which nodes represent concepts, while arcs between these nodes represent relations between concepts and are labeled so as to indicate the semantics of the relationships.

semantic validation
> **D:** The validation process by which someone assures that all the literals in a synset represent the same meaning and that the synset is complete.

sense
> **D:** One of the possible interpretations of a literal in a wordnet. Each sense of a word is in a different synset. It is represented by a number that follows the given literal.
> **E:** `market:3`

sense identifier
> **D:** The label appended to the literal representing a polysemous lexical item in order to identify different senses of the same item.

sentence frame

    D: A list of generic sentence frames, attached to each verb synset, that illustrates the types of simple sentences in which the verbs can be used.

    E: `give` - `{transfer possession of something concrete or abstract to somebody; "I gave her my money"; "Can you give me lessons?"; "She gave the children lots of love and tender loving care"}`

                `*> Somebody -- s somebody something`

                `*> Somebody -- s something to somebody`

***similar_to* relation**

    D: Relation of semantic similarity between a focal synset and peripheral or satellite synsets that have close referential meaning.

    E: `{damp, dampish, moist}` SIMILAR_TO `{wet}`.

***subevent* relation**

    D: An entailment relation between two verbs: the activity denoted by one argument of the relation is temporally properly included in the activity denoted by the other argument

    E: `(snore:1)`SUBEVENT`(sleep:1)`.

subordinate

    D: Same as **hyponym.**

subsense

    D: A sense of a word that is slightly different from one main sense. Usually, in a reference dictionary a subsense is marked by appending to the main sense identifier a number or a letter.

SUMO – Suggested Upper Merged Ontology

    This ontology is being created with the goal of developing a standard upper ontology that will promote data interoperability, information search and retrieval, automated inferencing, and natural language processing. The SUMO has been translated into various representation formats, but the language of development is a variant of KIF (a version of the first-order predicate calculus). An upper ontology is limited to concepts that are meta, generic, abstract or philosophical, and hence are general enough to address (at a high level) a broad range of domain areas. Concepts specific to particular domains are not included in an upper ontology, but such an ontology does provide a structure upon which ontologies for specific domains (e.g. medicine, finance, engineering, etc.) can be constructed. (`http://ontology.teknowledge.com/`)

superordinate

    D: Same as **hypernym.**

synonym

    D: a word that has the same meaning with another one and can be interchanged with the latter in a given context; one or more synonyms forms a synset.

E: `form:1, word form:1, signifier:1, descriptor:1` are synonyms when used with the specified PWN sense.

synset

  D: A synonym set; a set of words that lexicalize the same meaning and are interchangeable in some context.

synset ID

  D: An identifier comprising of the specific version of the PWN used and a sequence of digits representing the offset in the original database of the respective synsets, followed by a tag denoting the part of speech of the literals in the encoded synsets.

  E: `ENG20-04295024-n`

syntactic validation (of a wordnet)

  D: The validation process by which someone assures that a wordnet is well formed.

  E: compliance with the DTD for the VisDic editor, no duplicate literals are in the same synset, no sense duplication (literal & sense label), no dangling nodes (conceptual density), no loops, valid set of semantic relations, valid synsets identifiers and many others.

Top Concept

  D: A language-independent concept representing a fundamental semantic distinction.

Top Concept Ontology (EWN)

  D: A hierarchy of language-independent concepts (Top Concepts), reflecting important semantic distinctions, e.g. Object and Substance, Dynamic and Static. The most important purpose of the top-ontology is to provide a common starting point and high degree of compatibility across the wordnets as far as the Base Concepts are interpreted. The Top Ontology is linked to the ILI/BILI, thus providing additional language-independent structuring of the interlingual index, together with the Domain Ontology. The ontology consists of 63 higher-level concepts that classify a set of 1 218 BILI-records representing the most important concepts. At the first level there are 3 types of entities: **1stOrder-Entity** (any concrete entity (publicly) perceivable by the senses and located at any point in time, in a three-dimensional space), **2ndOrderEntity** (any Static Situation (property, relation) or Dynamic Situation, which cannot be grasped, heart, seen, felt as an independent physical thing; they can be located in time and occur or take place rather than exist; e.g. continue, occur, apply), **3rdOrderEntity** (an unobservable proposition which exists independently of time and space; they can be true or false rather than real; they can be asserted or denied, remembered or forgotten; e.g. idea, though, information, theory, plan).

translation equivalents

  Two wordforms in a parallel corpus that appear in the same translation unit and represent reciprocal translations are called translation equivalents. Translation equivalence is a context dependent relation.

translation equivalence pair

   D: A pair of words (w_{TL}, w_{SL}) in which w_{SL} and w_{TL} are translation equivalents.

translation unit

   D: In a sentence aligned parallel corpus, containing texts in N languages, a translation unit is a set of N text segments, one for each language, that are reciprocal translations. Such a text segment usually contains a single sentence, but not always.

   E: From the *1984* parallel corpus:

```
<tu id="0zz.113">
   <seg lang="en">
     <s id="0en.1.1.24.2">
       <w lemma="Winston" ana="Np">Winston</w>
       <w lemma="be" ana="Vais3s">was</w>
       ...
     </s>
   </seg>
   <seg lang="ro">
     <s id="0ro.1.2.23.2">
       <w lemma="Winston" ana="Np">Winston</w>
       <w lemma="fi" ana="Vmii3s">era</w>
       ...
     </s>
   </seg>
   <seg lang="cs">
     <s id="0cs.1.1.24.2">
       <w lemma="Winston" ana="Np">Winston</w>
       <w lemma="se" ana="Px--d-ypn-n">si</w>
       ...
     </s>
     ...
   </seg>
  ...
   </tu>
```

troponym

   D: A verb expressing a specific manner elaboration of action expressed by another verb. **X** is a troponym of **Y** if **X** is **Y** in some manner.

   E: `chant`, `intone`, `vocalize` and `croon` are troponyms of `sing`.

unique beginner

   D: A noun synset with no superordinate.

   E: `entity`, `state`, `action`, `event`, `psychological feature`, `be`, `become`, `cause to be perceived`, `know`, `like`, `take`, etc.

verb_ group **relation**

    D: The relation groups several similar overlapping meanings of the verbs.
    E: `(act:2 behave:1 do:9)`VERB_GROUP`((act:5 play:8 act-as:2)`
    `(dissemble:3 pretend:2 act:9))`

VisDic

    D: A graphical application for viewing and editing MACHINE READABLE dictionaries. It is the standard browser of the BalkaNet wordnets (as such it uses a specific DTD for the XML encoding). Most of the program behavior can be configured. There are 6 types of views (XML, user defined, tree, edit, words, selected entries) adaptable for each dictionary independently. VisDic was primarily developed for browsing and editing wordnets, but the tool can be configured for any type of dictionary – monolingual, translational, thesaurus or just a plain corpora. VisDic allows working with up to 10 dictionaries at the same time (`http://nlp.fi.muni.cz/projekty/visdic/`).

wordnet

    D: A lexical knowledge base for one given language, modeled after the principles of PWN.

WordNet

    D: See Princeton WordNet.

wordnet datasource

    D: A **WMS** component containing the semantic data that form a monolingual Wordnet and provides basic semantic operations.

word-level alignment

    D: In a parallel text: the process of making the correspondence between tuples of words that are reciprocal translations.

Wordnet Management System (WMS)

    D: An infrastructure for storage, querying and browsing within and across wordnets. This large scale, distributed, service-oriented system is essentially a distributed network of servers, each one hosting a monolingual wordnet. A central server (CWMS server) is responsible for establishing a coherent communication among the peripheral servers and it also holds responsibility for handling multilingual search requests.