

БУЛНЕТ (ЛЕКСИКАЛНО-СЕМАНТИЧНА МРЕЖА НА БЪЛГАРСКИЯ ЕЗИК) – ЧАСТ ОТ
СВЕТОВНАТА ЛЕКСИКАЛНО-СЕМАНТИЧНА МРЕЖА

SVETLA KOEVA

BULNET (BULGARIAN LEXICAL-SEMANTIC NETWORK) – PART OF THE GLOBAL
WORDNET

(Summary)

The brief overview of the Bulgarian wordnet (BulNet) as part of the global wordnet is presented in this paper. The (global) wordnet is an extensive lexical-semantic network that constitutes of synonymous sets (synsets) linked with the semantic relations existing between them. The cross-lingual nature of the global wordnet is provided by the establishing the relations of equivalence between synsets that express the same meaning in different languages. The special attention is given to the semantic, derivational and extralinguistic relations linking synonymous sets.

1. Въведение

Европейският проект БалкаНет (Стаму 2002) е насочен към създаването на синхронизирани лексикално-семантични мрежи за следните балкански езици – български, гръцки, румънски, сръбски, турски, допълване на чешката лексикално-семантична мрежа и приложението им за автоматична обработка на естествените езици. Целта на статията е да представи накратко основните резултати от работата по проекта и техните приложения.

2. Принстънският wordnet

Принстънският wordnet^[1] (PWN), създаден от Дж. Милър (Милър 1990) и неговите последователи (Фелбаум 1998), представлява лексикално-семантична мрежа – разновидност на традиционните семантични мрежи, чиято структура, най-общо казано, се състои от възли и връзки между тях. Възлите в лексикално-семантичната мрежа wordnet са синонимни множества (наречени *си́нсети*), които съдържат думи или съставни думи (наречени *литерали*). Дъгите в лексикално-семантичната мрежа wordnet изразяват семантични, словообразователни и екстралингвистични релации между обектите, разположени във възлите. Значението на литературите (sense), както и значението на синонимните множества (meaning) се разглежда като реализация на езиково независимите понятия (concepts). Семантиката на лексикалните възли в wordnet се изразява имплицитно чрез синонимната релация между литературите в синонимното множество и релациите към другите възли в мрежата и експлицитно чрез тълковното значение и примерите за употреба. Например лексикалните единици {*airplane:1, aeroplane:1, plane:1*} образуват едно синонимно множество, което се описва с тълковната дефиниция ‘*an aircraft that has a fixed wing and is powered by propellers or jets*’. Това синонимно множество е свързано с по-общо синонимно множество, наречено *хипероним*: {*heavier-than-air craft:1*} – ‘*a non-buoyant aircraft that requires a source of power to hold it aloft and to propel it*’; с няколко по-специфични, наречени *хипоними*: {*bomber:1*} – ‘*a military aircraft that drops bombs during flight*’, {*airliner:1*} – ‘*a commercial airplane that carries passengers*’ и т.н.; както и със синонимни множества, които изразяват неговите съставни части, наречени *мероними*: {*wing:2*} – ‘*one of the horizontal airfoils on either side of the fuselage of an airplane*’, {*accelerator:1, accelerator pedal:1, gas pedal:1, gas:5, throttle:2, gun:6*} – ‘*a pedal that controls the throttle valve*’ и т.н.

Думите в синонимните множества са подбрани така, че да бъдат взаимозаменяеми в даден контекст. В Принстънския wordnet при дефинирането на синонимите се прилага следният критерий за субституция:

Два израза са синоними в лингвистичен контекст К, ако субституцията на едната с другата (и обратно) в К не нарушава истинната стойност. (Милър 1990: 6).

Това означава, че в wordnet се включват не само пълните синоними, тъй като изискването не е ограничено до всеки контекст. Съществуват много примери за синоними (например *дъска* и *талпа*), които не могат да се заместят един с друг във всеки контекст по различни причини – синтактични, стилистични и др.

В Принстънския wordnet са включени два типа релации: между литературите (наречени лексикални) и между синонимни множества (наречени семантични). Семантичните релации се отнасят за всички литературите в двете синонимни множества, които свързват. Лексикалните релации свързват отделни думи в рамките на едно (по-рядко) или две синонимни множества. По-подходящо е противопоставянето: **литерална релация** ↔ **релация между синонимни множества**, тъй като литературните релации също могат да бъдат семантични. Релациите между синонимни множества са семантични: **хиперонимия** (Hypernym, Instance Hypernym), **хипонимия** (Hyponym, Instance

Hyponym), **меронимия** (Member meronym, Substance meronym, Part meronym), холонимия (Member holonym, Substance holonym, Part holonym), **субсъбитие** (Entailment), **каузация** (Cause), **глаголна група** (Verb group), **подобен на** (Similar to), **също така** (Also see) и **атрибут** (Attribute), и екстралингвистични: **тематична област** (Domain of synset – Topic), **регион на употреба** (Domain of synset – Region) и **сфера на употреба** (Domain of synset – Usage). Релациите между литерали са семантични: **синонимия** (Synonym) и **антонимия**, (Antonym), и словообразователни: **произхождащ** (Derived from adjective), **причастие** (Participle of verb), **произведен** (Derivationally related form) и **принадлежащ** (Pertainym).

Синонимните множества в Принстънския wordnet включват литерали, принадлежащи към четири части на речта: съществително, прилагателно, глагол и наречие, така наречените отворени класове думи (числителните са класифицирани като прилагателни). Фактически четирите части на речта образуват четири независими лексикално-семантични мрежи, между които няма семантични релации, а само словообразователни връзки. Съществителните имена се класифицирани в различни подкласове според това, дали означават действие, когнитивни процеси, чувства, дали са абстрактни или конкретни названия и т.н. Глаголите са разделени на подкласове в зависимост от това, дали са за възприятие, за движение, за допир, за изразяване на чувства, за мисловна дейност и т.н., и според синтактичната си рамка. Синтактичните рамки в wordnet имат следния формат: Somebody —s something to somebody, Somebody —s somebody, в който освен броят и видът на синтактично задължителните конституенти се включва и минимална информация за селективните ограничения при съчетаемостта. Прилагателните образуват подкласове според това, дали имат само предикативна употреба, само атрибутивна употреба, или и двете.

[i] <http://wordnet.princeton.edu//>