Theoretical Analyses and NLP for and via Language Learning and Teaching

Maria A. Todorova

Department of Computational Linguistics,
Institute for Bulgarian Language

Assessing reading literacy

Enriching the Semantic Network WordNet with Conceptual Frames

SIG webinar - November 13th 2024

Outline

Electronic language resources - multilingual and language specific for Bulgarian

Language tasks and language games for different scientific and educational purposes

- research on the basic vocabulary in Bulgarian
- language games for educational purposes

NLP tools

- a web interface for wordnets – Hydra for Web; supports two modes – a Single Wordnet mode and Parallel Wordnets mode
- Wordpress plugin for programming language games
- platform Readlet for investigation of the reading literacy and comprehension of early graders

Observations and analyses

Goals

1. To illustrate the way in which NLP resources like Wordnet and Framenet can be useful for language pedagogy by employing examples:
   - from an online language experiment with language tasks
   - and from a competition on Computational linguistics

2. To discuss and compare the results from different tasks.

3. To discuss the possibilities for further development of the language pedagogical potential of Frame semantics, lexical semantic networks and other types of resources and descriptive databases, like corpora and Framenets.

Language resources

Electronic language resources are (large in volume) collections of language data that have different purposes.

Language resources developed at the Institute for Bulgarian Language: http://ibl.bas.bg

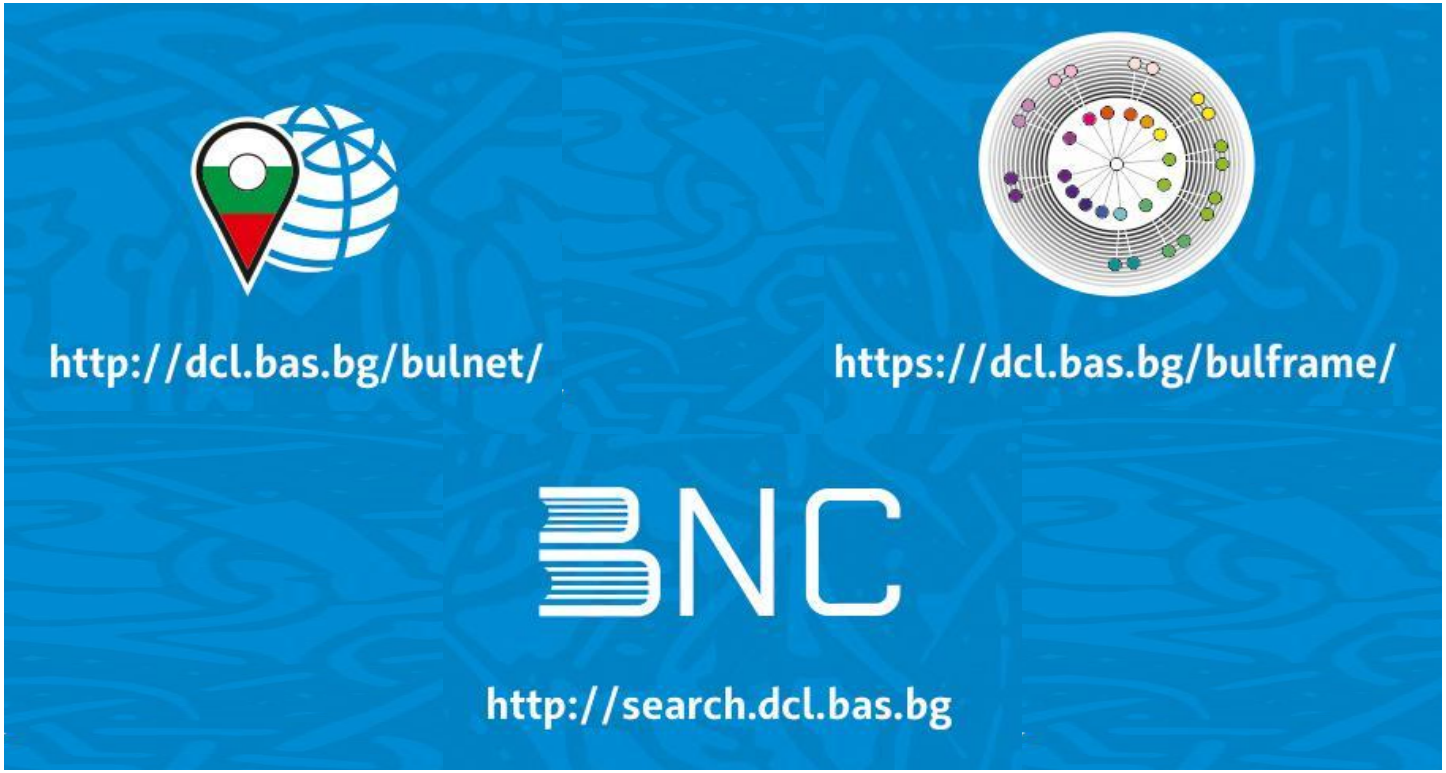
- Dictionaries
- BulNet – Bulgarian Wordnet
- Bulgarian National Corpus
- Language teaching resources

  and language games

  https://ibl.bas.bg/resursi/

https://www.youtube.com/watch?v=HOwPjyI-B_I

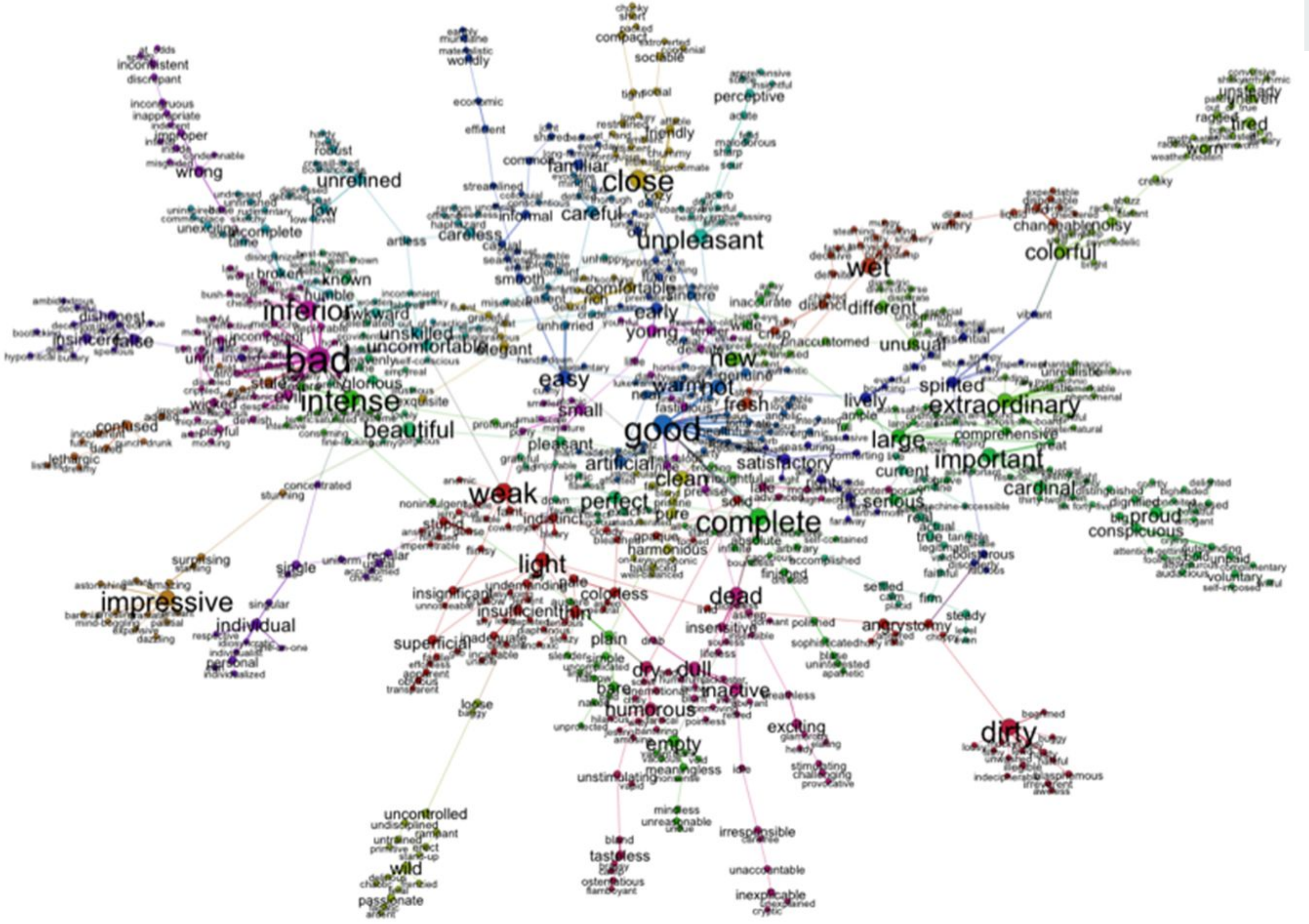
Resources of the Department of Computational Linguistics dcl.bas.bg



Resources: The Bulgarian wordnet

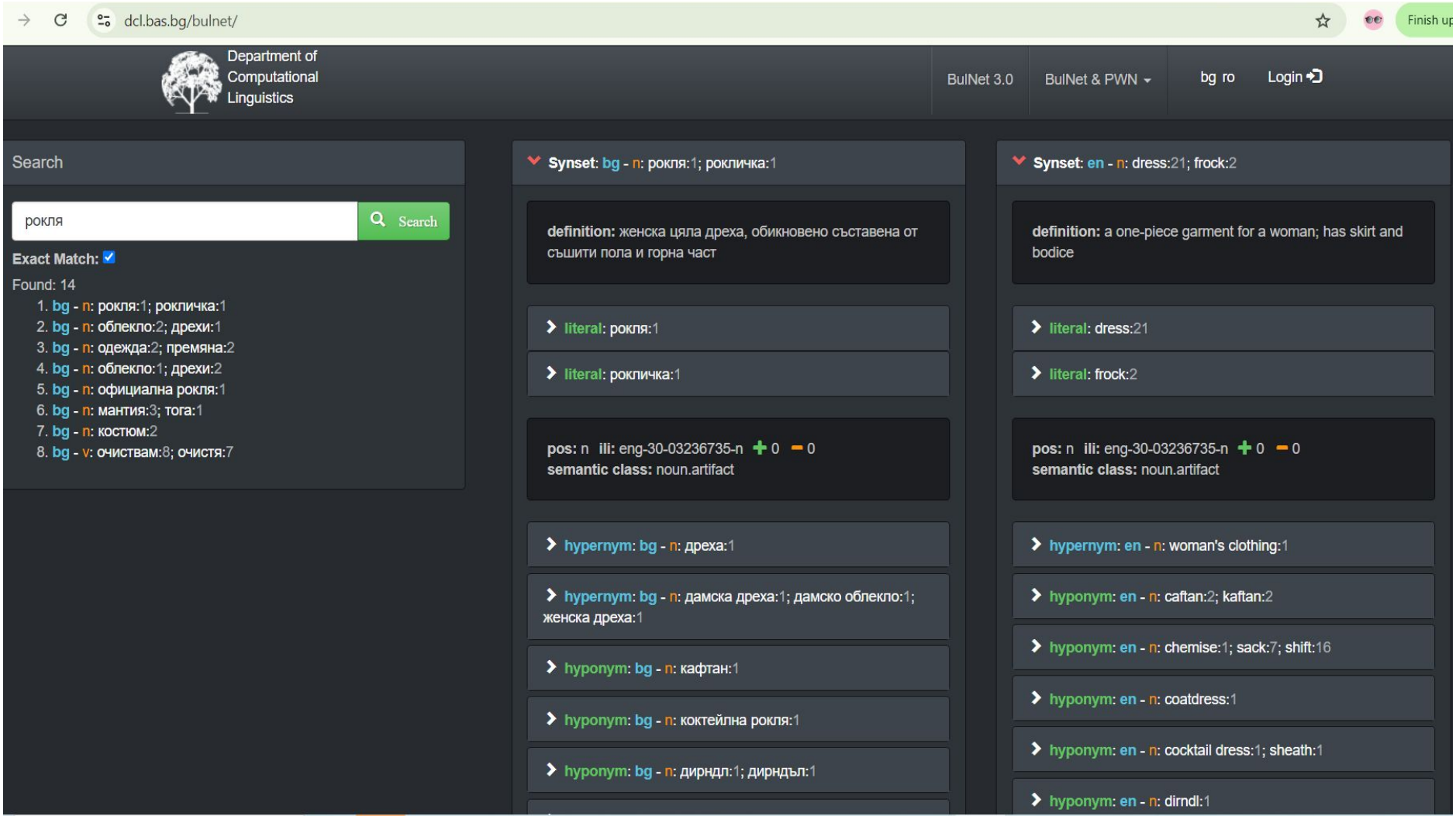
A lexical semantic network of synonym sets (synsets) interconnected by semantic, lexical and other relations.

- contains about 250,000 lexical units for Bulgarian, organized in more than 120,000 synonym sets.
- Synonym sets are interconnected with over 250,000 semantic relations: part – whole; antonymy; cause - effect, etc.
- the main structural relation is hypernymy (hyponymy).
- Words are classified in semantic classes (verbs of motion, emotion, communication, change …, nouns of body, person, artifact etc.)
- The synonym sets in the Bulgarian wordnet are linked to the equivalent synonym sets in 24 languages

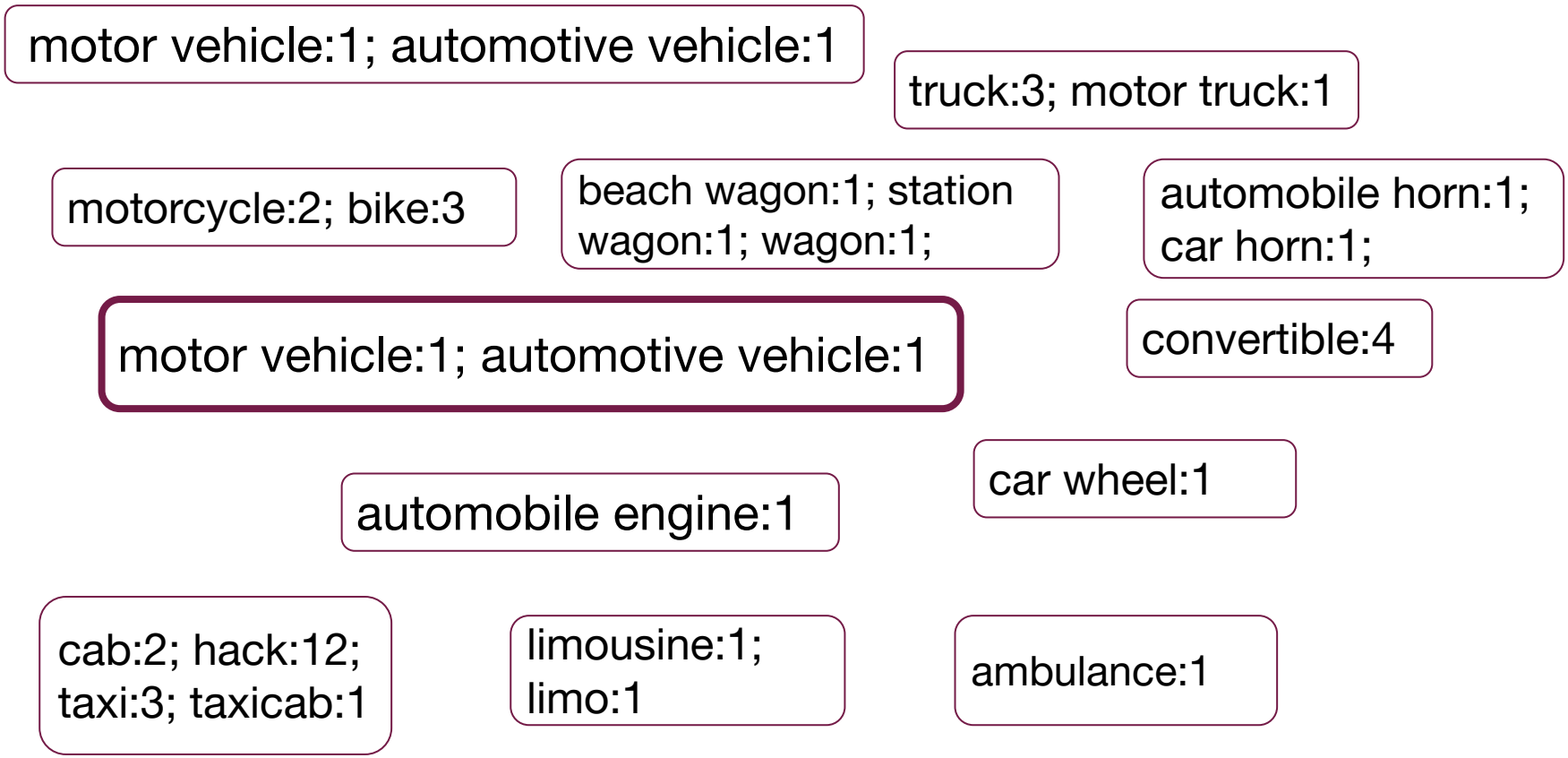


bad
good
inferior
intense
beautiful
close
unpleasant
careful
familiar
wet
colorful
changeable
noisy
weak
light
complete
dead
dirty
impressive
individual
extraordinary
important
large
new
fresh
easy
small
pleasant
artificial
clean
perfect
pure
harmonious
humorous
dry
dull
inactive
empty
meaningless
exciting
tasteless
wild
uncontrolled
unrefined
wrong
improper
inconsistent
perceptive
sociable
compact
friendly
worldly
economic
efficient
unskilled
uncomfortable
elegant
known
humble
broken
incomplete
careless
informal
artless
smooth
comfortable
sincere
early
young
different
unusual
inaccurate
distinct
unaccustomed
spirited
lively
comprehensive
cardinal
proud
conspicuous
voluntary
serious
satisfactory
current
steady
firm
insensitive
plain
simple
bare
superficial
insignificant
insufficient
thin
colorless
surprising
single
personal
confused
lethargic
playful
wicked
glorious
exquisite
insincere
false
dishonest
unsteady
tired
ragged
worn
creaky
vibrant
irresponsible
unaccountable
inexplicable
passionate
loose
bland
stimulating
unstimulating
mindless
unreasonable
idle
sophisticated
uninterested
polished
finished
accomplished
absolute
boisterous
faithful
settled
calm
concentrated
stunning
unhurried

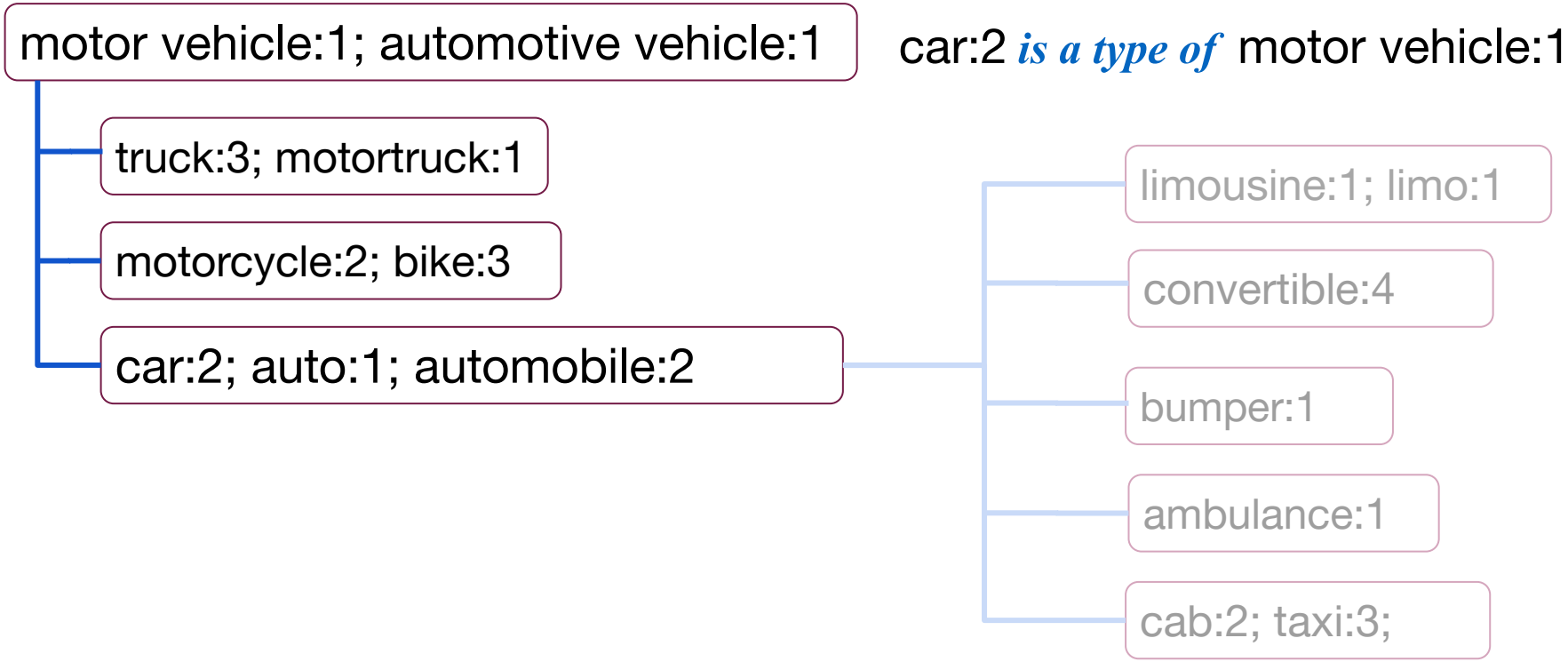
Hydra - an online tool for visualization of BulNet https://dcl.bas.bg/bulnet/



The Bulgarian wordnet
PWN and BulNet

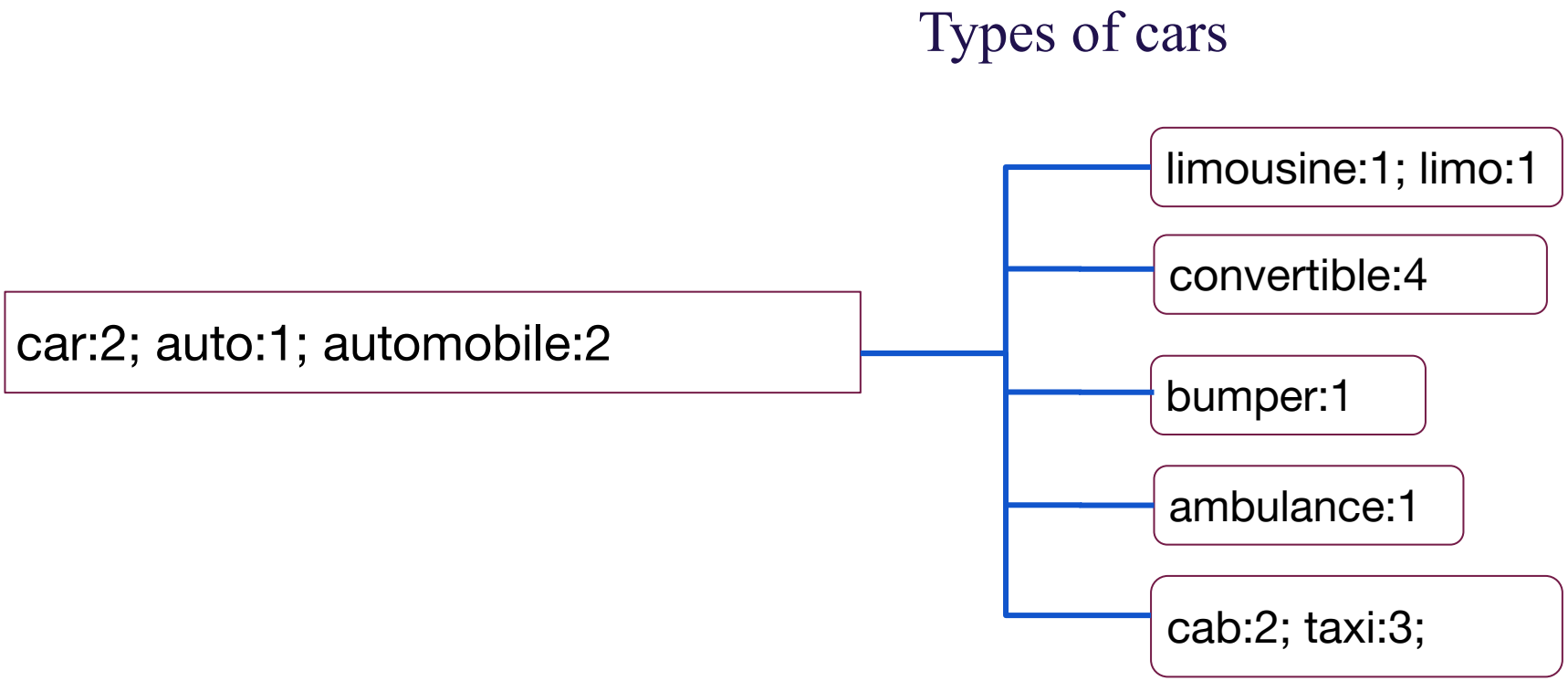


BulNet

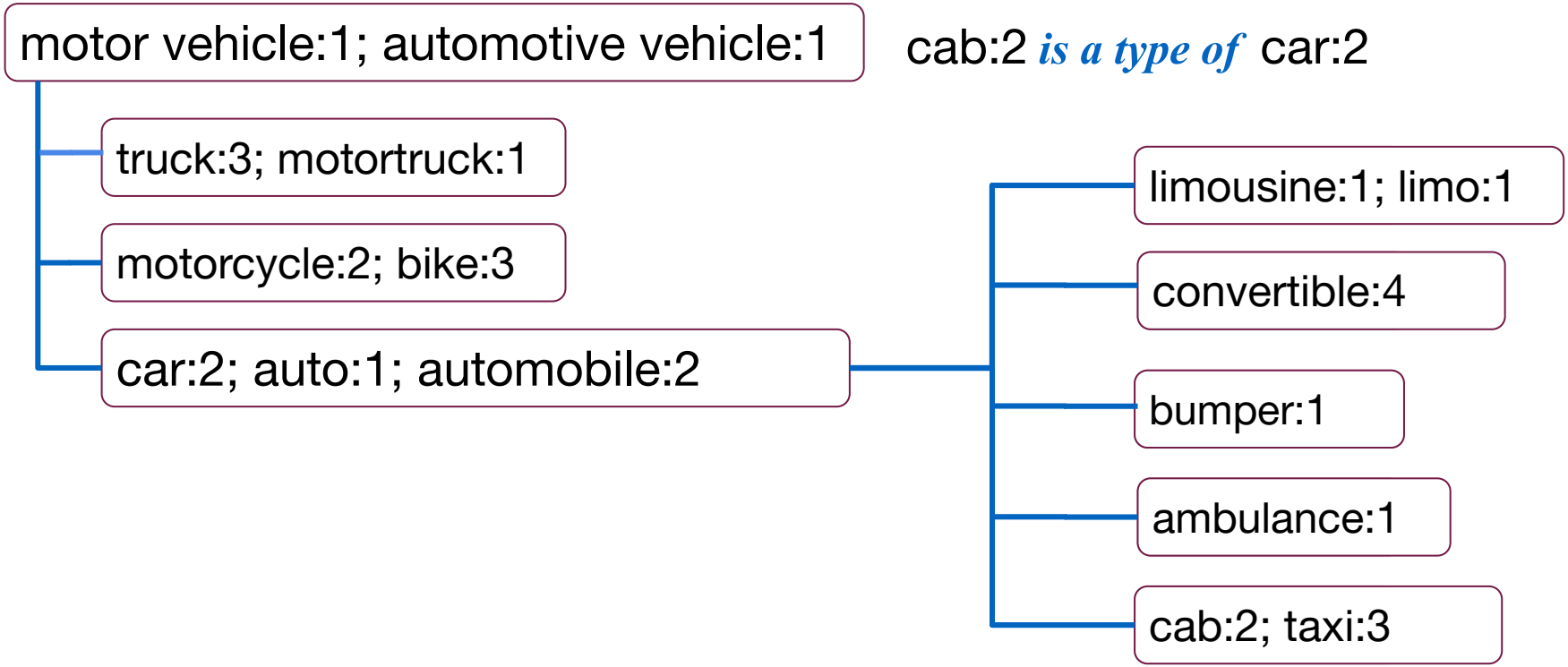


car:2 *is a type of* motor vehicle:1

PWN and BulNet

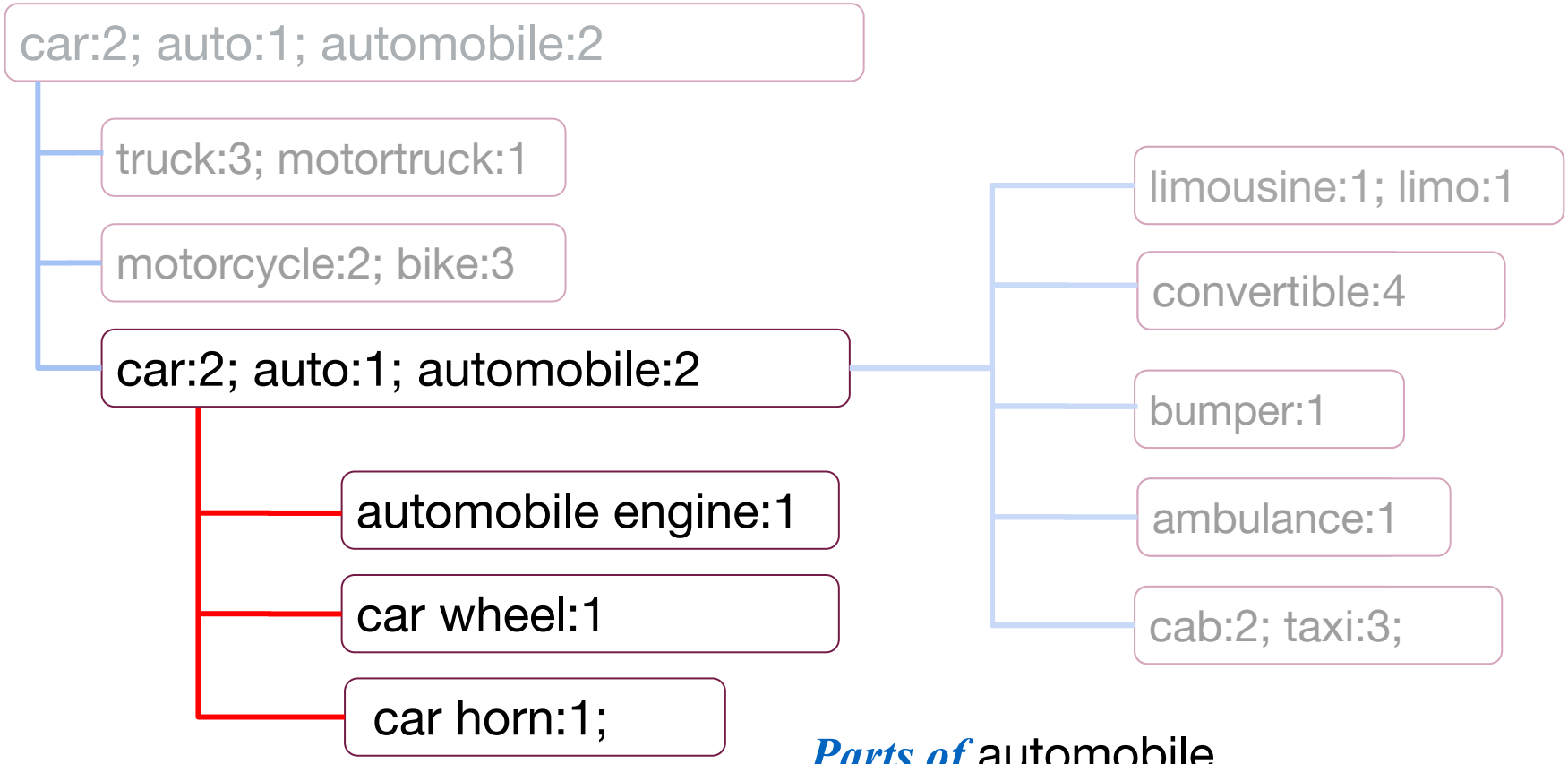


BulNet



cab:2 *is a type of* car:2

BulNet



Parts of automobile

Framenet

- FrameNet is a lexical database that focuses on the meaning and structure of words in relation to the frames or conceptual structures they evoke.

- FrameNet provides a rich network of semantic relations that connect frames, lexical units, frame elements, and other linguistic units. This allows for a detailed analysis of how words and concepts are related and how they contribute to the overall meaning of a sentence or discourse.

- FrameNet semantic relations are organized around frames, which represent specific concepts or scenarios.

Framenet

- Lexical units (LUs): LUs are words or multi-word expressions that evoke a specific frame. LUs are linked to FEs, indicating the role they play within the frame. Different LUs may have different syntactic realizations but evoke the same frame.

- Frame elements (FEs): FEs are the roles or semantic components associated with a frame. They represent the participants, attributes, or other semantic aspects of the frame. Each FE has a name and a description, and they are linked to lexical units (LUs) that evoke the frame.

- Frame-to-frame relations: Frames in FrameNet are related to each other through several relations in the frame hierarchy.

Framenet online https://framenet.icsi.berkeley.edu/frameIndex

---

Filling

Lexical Unit Inde

Definition:

These are words relating to filling containers and covering areas with some thing, things or substance, the Theme. The area or container can appear as the direct object with all these verbs, and is designated Goal because it is the goal of motion of the Theme. Corresponding to its nuclear argument status, it is also affected in some crucial way, unlike goals in other frames.

Lionel Hutz COATED the wall with paint.

FEs:

Core:

Agent [Agt] **Semantic Type:** Sentient **Excludes:** Cause	The Agent is the actor who instigates the filling.
Cause [cau]	An event which brings about the filling of the Goal.
Goal [Goal] **Semantic Type:** Goal	The Goal is the area or container being filled. Goal is generally the NP Object in this frame.
Theme [Thm] **Semantic Type:** Physical_object	The Theme is the physical object or substance which changes location. Rev. Lovejoy carefully BRUSHED the rolls with butter.

Language tasks – from theory to practice and vice versa

Experiment on verbs via Language games

Competition on Computational linguistics

Language tasks

Language games with verbs

A study aimed at the acquisition of basic verbs from pupils in primary and lower secondary education (7-10-year-olds and 11-14-year-olds).

It involves:

- 5 types of language tasks grouped into 4 variants, executed online
- A set of 188 verbs, of which 70% were verbs considered to be part of the core vocabulary (according to a set of criteria).

The study aims to test the basic competencies for recognising verbs' senses and their arguments (or participants in the situations referred to by the verbs – the semantic frame elements)

Programming of the tasks

- The tasks are programmed for online access through a plugin to the WordPress platform, developed and further improved by M. Yalamov with specific functionality for the purposes of some tasks.
- Selected verbs slide to specific places in a sentence or text.
- The answers are visualized when you go through all the tasks (and not after each task).
- The results are collected in tables from which data is extracted for analysis purposes.

Programming of the tasks

<quiz>
<qt> Задача 1. Избери подходящата дума от списъка под картинката, за да довършиш изречението.</qt>

<qd>**Хората...**</qd>

<ta>се прегръщат </ta>
<ta>се милват </ta>
<ta>се гушкат </ta>
<ta>се притискат </ta>
</quiz>

Language tasks – 5 types into 4 variants

http:// ibl.bas.bg/ igrasglagoli/

Задача 4. Кои от думите ще използваш, ако разказваш по картинката?

Избери глаголите, които според теб са подходящи за дадената картинка. Може да посочиш най-много пет глагола и най-малко един.

Самолетът _____.

Самолетът _____.

Самолетът _____.

Самолетът _____.

Самолетът _____.

прелита	лети	прегръща	кръжи	пътува	излита	бръмчи	мълчи	минава	лекува

НАПРЕД

(beforehand respondents fill in short questionnaires: age, grade level, sex)

Language tasks

Aims

- To test basic competencies for recognising verbs' senses and their arguments (or participants in the situations referred to by the verbs – the semantic frame elements)
- To confirm the target verbs as part of the set of basic vocabulary by:
  - measuring the frequency of the respondents' answers.
- To determine the semantic frames which may be evoked by the target verbs
  - via observations on the specific frame elements recognised by the children
- To confirm their understanding of the verbs' meaning

Hypothesis

- The basic vocabulary reflects the most essential phenomena such as concepts and situations related to nature, everyday life, spiritual and material culture, feelings, states, actions.
- The respondent's knowledge of the selected verbs is analysed through Frame Semantics (Fillmore 1982) and the semantic-syntactic description of the concrete frames in FrameNet.

The stimuli

- **Verbs** → refer to everyday activities and situations such as eating and drinking, movement, perception, speech, weather, etc.
- **Picture stimuli** → selected from datasets of free images (such as MultiPic (Duñabeitia et al. 2018), etc.); illustrating mostly participants in the situations defined by the verb (agent, theme/patient, instrument, location, etc.).
- **Context** →verbs are used in sentences (and in short texts).

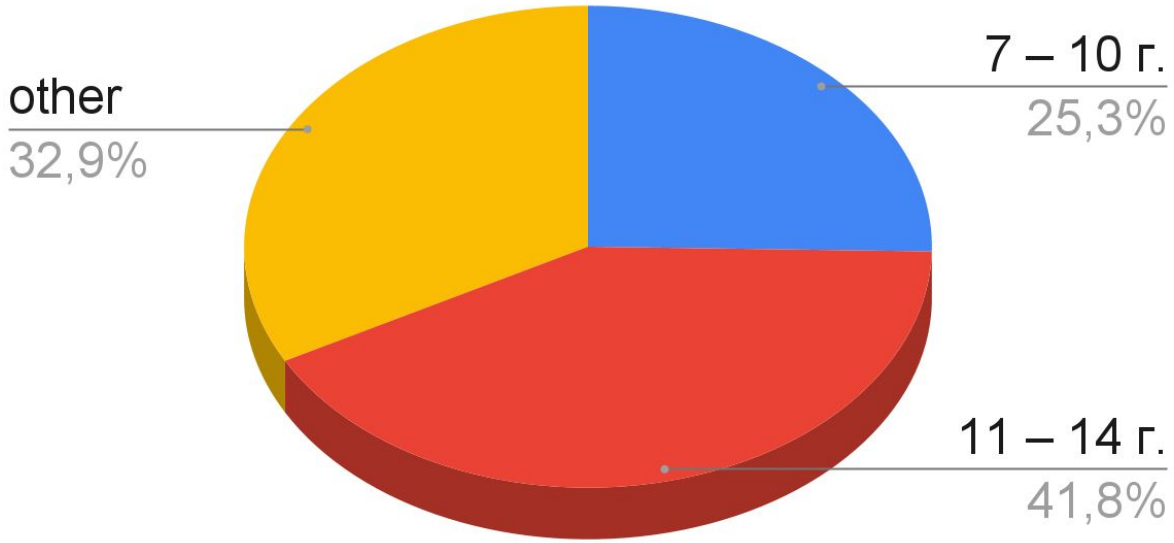
Selection of verbs

Verb	AoA	1-4 grades	BSC	BulNC	semantic class	Synset in BulNet	Definition in BulNet
пиша	write 'to put on paper' 2;	195	1	200.57	verb.communication	пиша:5	communicate or express by writing
играя	play 'to have fun' 2;	184	0	313	verb.social	играя:5	be at play; be engaged in playful activity; amuse oneself in a way characteristic of children
чета	read 'to learn what printed words mean' 2;	85	2	75.97	verb.cognition	чета:1; прочитам:1; прочета:1	interpret something that is written or printed

The database is created by Ivelina Stoyanova.

Respondents

557 respondents of whom 374 are 7- to 14-years-old.



Association Language tasks

- Association tasks involving picture stimuli representing participants in the situation referred to by the verb.

- Respondents might choose any verb from a set of 4 verbs where at least 1 refers to the main sense and is assumed to be part of the basic vocabulary, without additional encoding of the manner of action (which may be encoded by prefixes, suffixes, etc.).

- Results: respondents preferred verbs from the core vocabulary.

Language tasks 1

С ухото... .

- ☐ подслушвам
- ☐ чувам
- ☐ подочувам
- ☐ слушам

Frequency				
	1 - 4 grades	BCS	WN	BulNC
eavesdrop	0	0	3	18,4
hear	**47**	**1**	**1**	**304,87**
overhear	0	2	2	45,42
listen	**63**	**2**	**1**	**66,32**

Result		
	7 – 10	11 – 14
eavesdrop	4,8%	3,4%
hear	**73,8%**	**79.3%**
overhear	2.4%	0%
listen	**19%**	**17.2%**

Language tasks 2

Корабът _____

Корабът _____

Корабът _____

Корабът _____

Корабът _____

спира	плува	мълчи	акостира	продължава	плава	писука	износва	потегля	превозва

	7 – 10	11 – 14
stop	*9.1%*	*7.8%*
swim	**8%**	**7.5%**
hush	0%	1%
shore	**12.5%**	**11.4%**
continue	*4.5%*	*7.25%*
sail	**22.7%**	**20.3%**
tweet (!!!)	6.8%	5.9%
wear off	0%	0.7%
depart	**17%**	**18.3%**
transport	**19.3%**	**19.9%**

Language tasks 3 - verbs from domain

Изберете тези глаголи, които според Вас

са свързани с **ядене и пиене.**

- ☐ горя
- ☐ дъвча
- ☐ гриза
- ☐ бръсна
- ☐ бия
- ☐ мия
- ☐ бягам
- ☐ пия
- ☐ хапвам
- ☐ гълтам

Semantically related verbs associated with a simple general description involving information about their semantic class.
"Choose verbs that are associated with **eating and drinking**."
(verb.consumption)

	7 – 10	11 – 14
burn	0%	0.7%
chew	**19.2%**	**21.1%**
nibble	**19.2%**	**17.1%**
shave	0%	0%
beat	1%	0.3%
wash	1%	0.3%
run	1%	0%
drink	**20.2%**	**20.7%**
snack	**19.2%**	**20.7%**
swallow	**19.2%**	**19.1%**

Contextual language tasks 4

Изберете подходящия глагол.

Какво правя сутрин?

Сутрин обичам да _____ вкусна и здравословна храна. 

Ето сега _____ портокали за любимия сок. 

Преди това _____ едно яйце в тенджера. 

В момента _____ домата на парчета. 

Взех филийки хляб, за да ги _____ в тостера. 

Когато филийките са готови, ще ги _____ с масло. 

След това _____ сол върху филийките. 

Върху филийките _____ и малко кашкавал. 

Накрая _____ портокалов сок в голяма чаша и _____ с наслада вкусната напитка.

препека	сварих	ще настържа	ще изпия	намажа	ще налея	ще поръся	изцеждам	ям	режа

“What do I do in the morning?”

In the morning I like to.....(eat) delicious and healthy food.
Now I..... (am squeezing) my favorite orange juice.
Before that, I .....(boiled) an egg in a pot.
I .....(am cutting) the tomato into pieces.
I took some slices of bread to..... (toast) them in the toaster.
When the slices are ready, I will ..... (spread) some butter on them.
Then I ..... (will sprinkle) salt on the slices.
I .....(will grate) some cheese on top of the slices.
Finally, I .....(will pour) orange juice into a large glass and ..... (will drink) the delicious drink.

Contextual language tasks

Aim at gathering information about the respondents' ability to acquire knowledge in a more complex situation.

These tasks are the most difficult ones and combine a complex of stimuli – the respondents have to take into account lexical, grammatical and morpho-semantic specifics of the verbs (of concrete and abstract meanings, from all semantic classes, i.e., cognitive verbs, verbs of emotions, stative verbs, motion verbs, etc.). A choice may involve:

- a paronym of the correct verb
- a verb that does not meet the requirements for the form
- a verb that does not meet the semantic requirements of the context
- a verb similar in meaning but with a syntactic realization that is incompatible with the context

Language tasks 5

Алиса **скучаеше (страдаше, доскучаваше (b), нуждаеше)** и **си мислеше (приспиваше, успиваше, колебаеше)** дали да **набере (прибере, отнесе, обере (a)** един букет от маргаритки в тежката следобедна горещина.

Alice was beginning to ***get bored (hurt, dreary, needy)*** *and she was* ***considering in her own mind (sleeping, getting sleepy, hesitating)*** *whether to* ***pick (gather, carry, reap)*** *a branch of daisies in the hot afternoon.*

През това време един Бял Заек със светлочервени очи **подскочи (посочи, поклати (b), изсмя)** край нея.

At the same time a White Rabbit with pink eyes ***ran (pointed, swayed, laughed)*** *close by her.*

Evaluating the readability of texts

The Flesch Reading Ease Index for evaluating the readability of a text is calculated using the following formula:

206.835 – 1.015 x (average sentence length) – 84.6 x (average number of words per sentence)

Where:

The average sentence length is calculated from the total number of words in the text divided by the number of sentences in the text.

The average number of words in a sentence is calculated from the total number of characters in the text divided by the number of words in the text.

The result of this formula is a numerical value, higher values indicate easier to read text, while lower values are associated with more difficult to read texts.

Evaluating the readability of texts

*Алиса **скучаеше** и си **мислеше** дали да **набере** един букет от маргаритки в тежката следобедна горещина. През това време един Бял Заек със светлочервени очи **подскочи** край нея. Това не се **стори** необикновено на Алиса и тя не се **изненада** дори когато **чу** как Заека си **говори** „О, божичко, божичко!“. По-късно, като **размисли**, **реши**, че това е доста необичайно.*	readability: 147, FRE = (206.835 - (1.015 * (59 words / 1 sentence)) level: Extremely Easy age: (below) 4th Grade

Language tasks 5

A total of 260 respondents filled in at least one verb position, with only 5% of respondents correctly filling in all positions in the task given, while 68% of responses were incomplete or incorrect, or possibly arbitrary.

Difficulties: reflexive verbs (esp. obligatory reflexive verbs), causation (in a sequence of actions):
Most errors were made with polysemous verbs, verbs of perception and cognition such as ’to think’, ’to look’, or abstract verbs such as ’to pull out’, свия ’to shrink’, вися ’to hang’, as well as verbs with low frequency of use such as здрача се ’to dusk’, тъмнея ’to get dark’, some of which are not exactly core vocabulary

Assessing the understanding of words in context

Analysis of respondents' knowledge of lexical elements through the theory of frame semantics, using information from Framenet.

The basic concept of frame semantics, developed by C. Fillmore (1972), is based on the fact that by assimilating a word, a person acquires information about the situation in which the word is used, about the participants in the situation and about the relationships between them: someone does something to someone under certain accompanying circumstances (where, when, how, for what purpose, for what reason, etc.).

Semantic frames represent situations from the world around us and their participants.

Assessing the understanding of words in context

Verb	Semantic class	7-10 olds	11-14 olds	Frame	Context
чета 'read.v'	verb. cognition	32%	0.3%	Reading perception: The **Reader1** attends to a **Text2** to process its Information	**книгата2**, която **сестра1** ѝ чете. (The book which her sister was reading)
набера 'pick.v'	verb. contact	38%	41.9%	Food gathering: A **Gatherer1** removes **Crop2** ripe and ready to an accepted degree	**Алиса1**... да набере един букет от **маргаритки2** ... (Example 11)
подскоча 'jump.v'	verb. motion	11.2%	70%	Self motion: The **Self mover1**, a living being, moves under its own direction along a **Path2**.	През това време един **Бял Заек1** ... подскочи край **нея2** (Example 12)

Language tasks – theoretical models and education

- Competition on Computational Linguistics
- Language games for Sofia Festival of Sciences
- Language games in schools

Competition on CL– from theory to practice

СЪСТЕЗАНИЕ
по компютърна
лингвистика

2024

Do linguistics and mathematics have points of contact? Can language tasks be solved?

Be sure to answer these questions yourself by participating in the Computational Linguistics Competition. Participating in the competition opens up new horizons in which you discover a small fraction of the ways in which computers "understand" language.

The Competition on Computational Linguistics is an open contest for junior students or high school students conducted through the Educational Platform of the Institute for the Bulgarian Language. *https://ibl.bas.bg/moodle/course/view.php?id=3*

No prior knowledge of linguistics or logic is necessary.

Състезание по компютърна лингвистика

2025

Организатор на Състезанието по компютърна лингвистика е Секцията по компютърна лингвистика към Института за български език „Проф. Любомир Андрейчин" към Българската академия на науките.

Състезанието по компютърна лингвистика е предназначено за ученици в прогимназиален и гимназиален етап на обучение и се провежда онлайн чрез Образователната платформа на Института за български език към БАН.

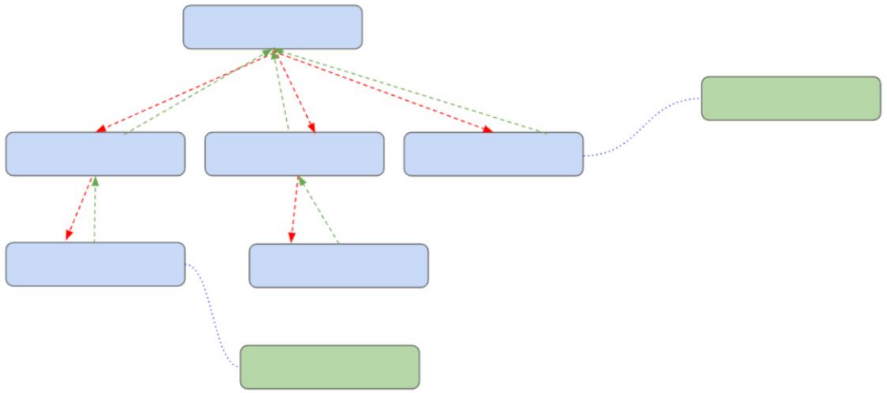
Competition on CL

In solving problems on CL, students learn about the diversity and consistency of language and exercise their logic skills.

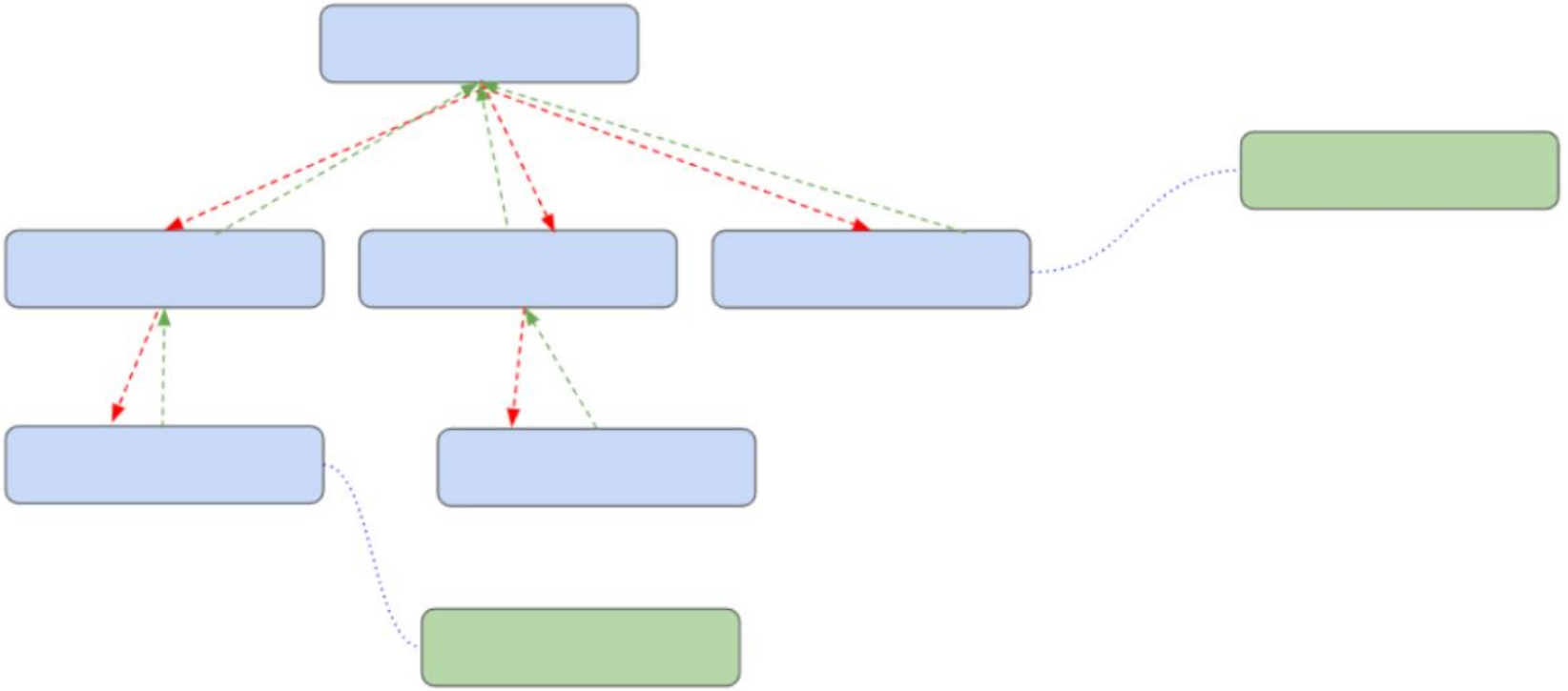
- **Tasks on relations between words and their meanings 2018**
- **Tasks on metaphorical meanings 2024**

Tasks on relations between words and their meanings

The diagram below (also called a graph) represents relations between concepts and their meanings. The nodes of the graph are filled with words that denote given concepts, and the arcs that connect them express semantic relations or word-forming connections between them. The different colors of the nodes represent different parts of speech. The arrows indicate direction and different colors mean different semantic relations or word-forming connections. For example, body, head, and eye are related by the relation *is part of* and the relation *has par*t as follows: body has part head, and head has part eye; a head is a part of a body, and an eye is a part of a head

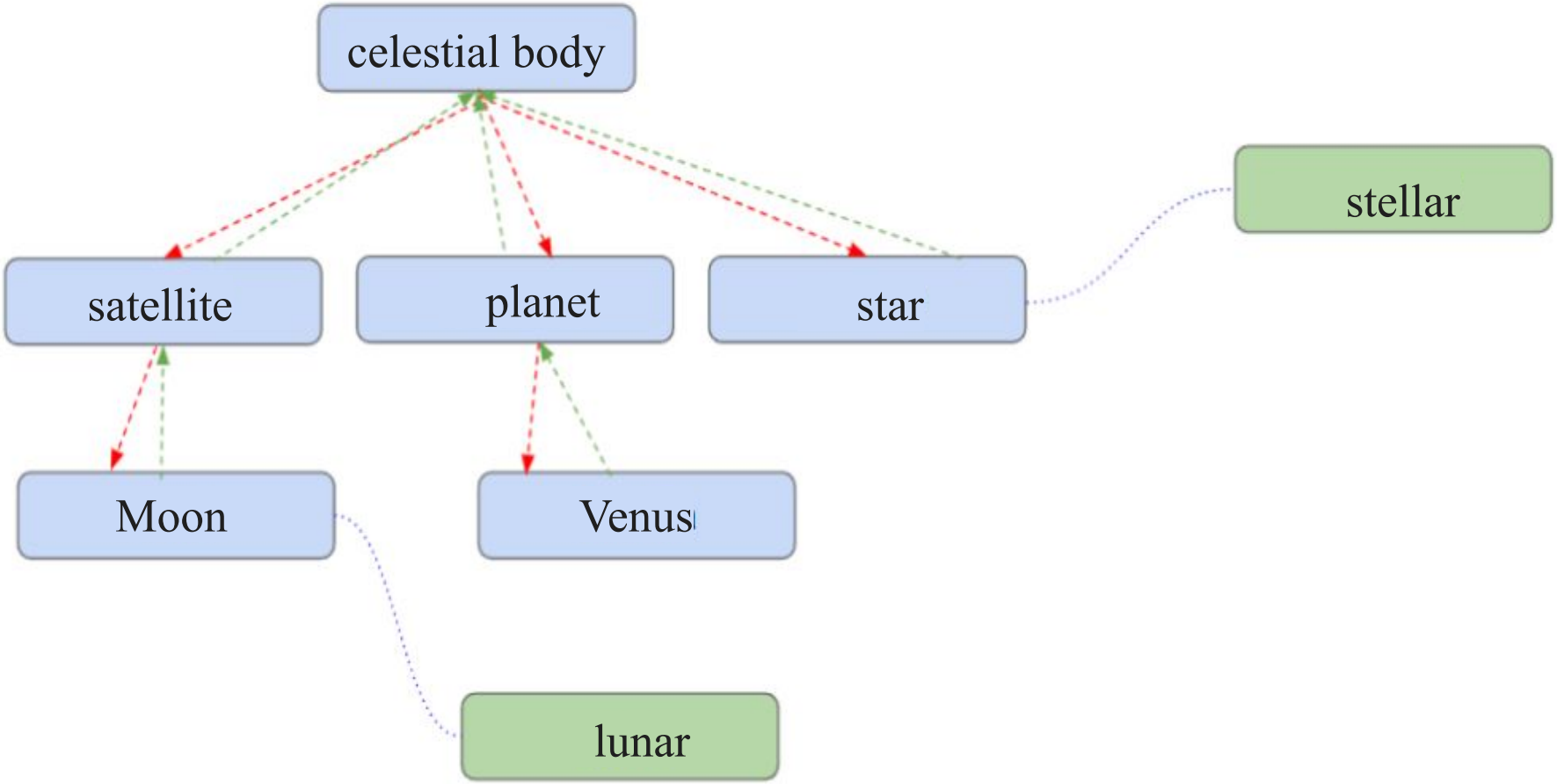


Fill in the graph so that it correctly illustrates the relationships between the concepts that are denoted by them. In one node, you can fill in only one word or compound word (for example, neutron star). Select one suggestion from the list below that is suitable for node 1 and drag it to the specified location.



sunset | celestial body | space |planetary | sunrise | star | planet | moon | universe | stellar | astronomy | companion | galaxy | lunar | neutron | Venus

Tasks on relations between words and their meanings



Tasks on metaphorical meanings

A metaphor is a way of connecting concepts that follows the pattern We talk about X as if it were Y, where X is a word naming an abstract concept from one subject area and Y is a word or phrase naming a concept from another subject area.

For example, the metaphorical linking pattern We talk about TIME[X] as if it were MONEY[Y] can be illustrated by the following sample sentences:

- This rule will save us [I SAVE MONEY] two hours [TIME].
- This mistake will cost us [COST MONEY] a whole week of delay[TIME].
- We wasted [WASTE MONEY] a whole month [TIME].

The words *save, cost and waste* in these sentences are connected metaphorically with the word denoting the abstract concept time[X], and in their direct meaning, they are usually combined with words or expressions naming a certain amount of money[Y].

Tasks on metaphorical meaning

You have given 20 sentences. They are distributed equally in 5 groups so as to illustrate exactly 5 patterns of metaphors. For each sentence, choose the appropriate metaphorical pattern.

Tasks on metaphorical meaning

The metaphorical pattern in the following 4 sentences is:
We talk about FEELING as if it were a LIQUID.

He finally poured out his grief in front of her.
His soul overflows with happiness.
His soul is flooded with the warm wave of resignation and hope.
He swallowed his disappointment and said nothing.

Tasks on metaphorical meaning explanation

The words *poured, spilled, poured, swallowed* refer to situations related to LIQUID: 'a substance in a liquid state at room temperature and pressure' (https://dcl.bas.bg/bulnet/).
Although the verb *swallow* can be associated with both FOOD and LIQUID, it should be noted here that the four sentences should illustrate the same pattern, and therefore the most appropriate general pattern for the group should be chosen.
The words *sorrow, happiness, resignation, hope and disappointment* are associated with the abstract idea of FEELING; 'the experiencing of affective and emotional states' (https://dcl.bas.bg/bulnet/)


ОПОЗНАЙ
ЧЕРНАТА МУХА - ВОЙНИК


КАМПАНИЯ НА
НАПИСАНОТО ОСТАВА
ПИШИ! ПРАВИЛНО!

Assessing reading literacy

Investigation of the reading literacy and comprehension of early graders via the platform Readlet, developed by the Comphys Lab at CNR ILC in Pisa.

Project leader of the Bulgarian team: prof. Svetla Koeva

Bulgarian team members: Ivelina Stoyanova, Valentina Stefanova, Tsvetana Dimitrova, Maria Todorova, Hristina Kukova

Istituto di Linguistica Computazionale "Antonio Zampolli"
Consiglio Nazionale delle Ricerche

Project leader of the Italian team: prof. Vito Pirrelli

Italian team members: Claudia Marzi, Marcello Ferro, Andrea Nadalini, Alessandro Lento

Main goal: To increase the level of literacy and reading abilities of primary school children through education. To contribute to the achievement of this goal, a thorough investigation focused on assessing reading literacy and comprehension in children of early school age in Bulgaria and Italy will be carried out.

Funding: the Bulgarian Academy of Sciences and the National Research Council of Italy

---

The project is performed in collaboration with 21st Hristo Botev Secondary School in Sofia.

Assessment of reading skills via Readlet

The Institute of Computational Linguistics in Pisa has developed a platform for automatic analysis and assessment of reading skills using a tablet.

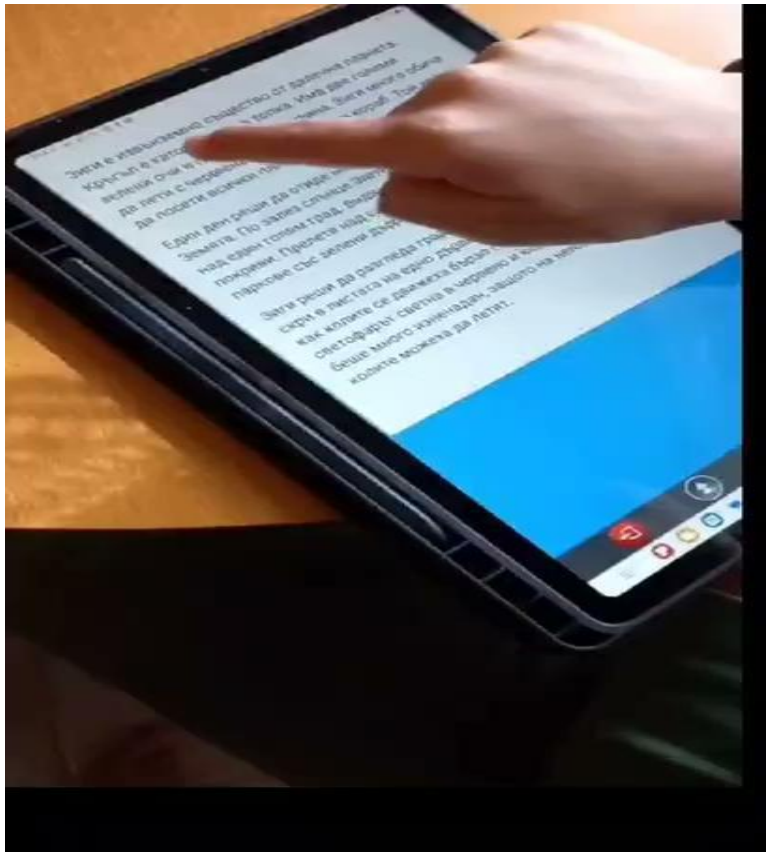
The platform Readlet provides an accurate assessment of reading skills based on empirical material.

The platform is used for automatic collection, pre-processing, and analysis of synchronized multimodal reading data, which includes:

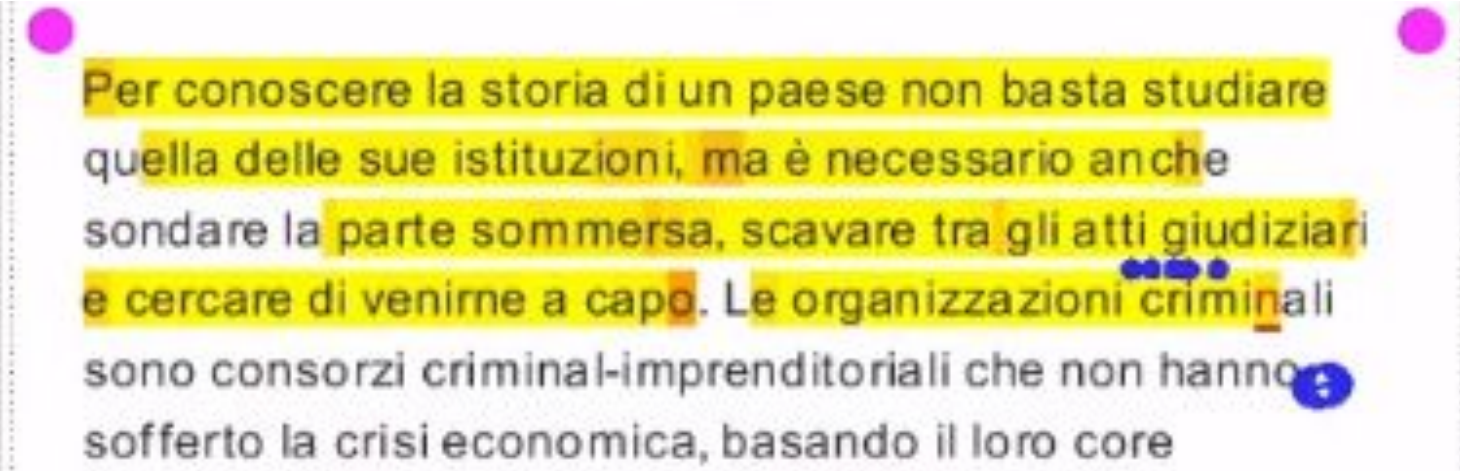
- voice recording when reading;
- record the time of swipe of the finger under the text when reading;
- the time to answer questions that test the comprehension of the text;
- and the number of correct answers.

https://www.readlet.it/apps/readlet/?lang=en

During a tablet reading session, the student is offered a short one-page text on the tablet's touch screen.
When reading on a tablet, the short text is displayed on its screen, expecting the student to point to each word in the text while reading it aloud or in his mind, and to slide his index finger across the screen.



Analyses of the results



As a result of the respondent's the programme is recording the information about the movement of the finger on the text and synchronizes it with the recording of the voice.
In the illustration above, the yellow color indicates that the text has been traversed with a finger, the orange places indicate where the finger has stopped or slowed down.

Analyses of the results

By correlating the data of the finger trajectory, the audio recording and the annotated reading text, an analysis with specific linguistic features of the reading skills of children of different ages are made, thus achieving a better understanding of:

- the main mechanisms in children's reading strategies
- which features of the text make it difficult to read
- opportunities to improve reading skills and overcome some deficits and difficulties.

For more information:

Claudia Marzi, Anna Rodella, Andrea Nadalini, Loukia Taxitari, Vito Pirrelli. Does finger-tracking point to child reading strategies?

Plenary Talkof Prof. Vito Pirrelli on CLIB 2024: Written Text Processing and the Adaptive Reading Hypothesis

Conclusions

I shared some observations on the application of some NLP tools and electronic language resources on language learning and teaching.

I presented research on the basic vocabulary in Bulgarian for students in the initial stage of education through language games online

I gave basic idea of the approach for investigation of the reading literacy and comprehension of early graders via the platform Readlet, developed by the Comphys Lab at CNR ILC in Pisa.

Thank you for your attention!

maria@dcl.bas.bg