



Будущие деятельности в 2021 г.



Онтология на ситуацията за
състояние – лингвистично моделране.
Съпоставително изследване за
български и руски

До конца года планируем:

- просмотреть и отсеять 1000+ предикатов только со стативным значением и представители наиболее типичных классов стативных предикатов;
- приписать и подтвердить фреймы тех самых предикатов.

Мы также планируем собрать больше параллельных корпусов текстов:

- из разных коллекций и репозиторий;
- болгарско – русский, русско – болгарский, но также и переводы с третьего языка.



Онтология на ситуациите за
състояние – лингвистично моделиране.
Съпоставително изследване за
български и руски

Благодарности

Второй болгарско-русский семинар проводится в рамках проекта „Онтология ситуаций состояния – лингвистическое моделирование. Сравнительное исследование болгарского и русского языков“.





Онтология на ситуациите за
състояние – лингвистично моделиране.
Съпоставително изследване за
български и руски



Проектные деятельности в 2021 г.

Светлозара Лесева, Христина Кукова, Ивелина Стоянова





Деятельность 1.2

Создание базы данных с эксцерпированным материалом



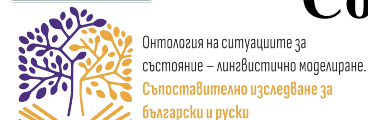
Онтология на ситуациите за
състояние – лингвистично моделране.
Съпоставително изследване за
български и руски

- Извлечены из Ворднета более 500 стативных синсетов.
- Извлечены 30 000+ параллельных пар предложений с потенциальными статическими предикатами.
- Фильтрация до 7 000 параллельных пар предложений.
- Ручная проверка кандидатов.
- Назначение соответствующих фреймов из Фреймнета и анализ **совместимости.**



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Ворднет:

- Семантическая сеть, в которой лексемы организованы в синонимические ряды (синсеты), состоящие из одного или нескольких синонимов (литералов) и связанные между собой концептуальными и лексическими отношениями (гиперонимия, меронимия, антонимия, сходство и т. д.).
- Принстонский Ворднет (PWN) содержит 120 000+ синонимических рядов, из которых 14 000+ глагольных синонимических рядов.
- БулНет содержит такое же количество синонимичных рядов, с более чем 85 000 синонимичных рядов, отредактированных экспертами, из которых 8 871 являются проверенными глагольными синонимическими рядами.
- Ворднет русского языка в свободном доступе: 48 354 синонимических рядов.



Търсене

принадлежа

Търси

Точно съвпадение:

Found: 10

1. **bg - v:** принадлежа:1; съм от:1; членувам:4
2. **bg - v:** принадлежа:2; отнасям се:2; спадам:1; числя се:1; съм от:2; класифицирам се:1
3. **bg - v:** свойствен съм:1; присъщ съм:1; принадлежа:3; наследявам се:1
4. **bg - v:** отнасям се:5; принадлежа:4
5. **bg - v:** принадлежа:5
6. **bg - v:** принадлежа:6
7. **bg - v:** състоя се:4; заключавам се:1; крия се:2; съм в:2; дължа се:2; лежа:8; принадлежа:7
8. **bg - v:** на мястото си съм:1; принадлежа:8
9. **bg - v:** мястото ми е:1; принадлежа:9
10. **bg - v:** принадлежа:10

▼ Synset: **bg - v:** принадлежа:10



дефиниция: представлявам част, член, привърженик, жител или под. на група, организация, населено място и др.

► **литерал:** принадлежа:10

▼ **пример:** Не е важен кой е човекът и към коя партия принадлежи в момента.

текст: Не е важен кой е човекът и към коя партия принадлежи в момента.

► **is_usage_of:** **bg - v:** принадлежа:10

част на речта: v ili: eng-30-02756359-v +0 -0
семантичен клас: verb.stative

► **хипероним:** **bg - v:** съм:10; бе:2; бивам:1; няма:3

▼ Synset: **en - v:** belong:6

дефиниция: be a member, adherent, inhabitant, etc. (of a group, organization, or place)

► **литерал:** belong:6

► **пример:** They belong to the same political party

част на речта: v ili: eng-30-02756359-v +0 -0
семантичен клас: verb.stative

► **хипероним:** **en - v:** be:4



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделчране.
Съпоставително изследване за
български и руски

Список глаголов в Ворднете:

- Список основан на семантической классификации глаголов в Ворднете – стативные глаголы, когнитивные глаголы, глаголы эмоций, глаголы восприятия и т. д.
- Извлечены стативные глаголы, а также их гипонимы.



Пример

ID: eng-30-02756359-v

PWN Synset: {belong:5}

Сем. клас: **verb.stative**

Дефиниция: *be a member, adherent, inhabitant, etc. (of a group, organization, or place)*

Пример (англ.): *They belong to the same political party.*

Български: {принадлежа:10}

Руски: {принадлежать 02365119}



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделиране.
Съпоставително изследване за
български и руски

Пример

Гипероним ID: *eng-30-02604760-v*

PWN Synset: {*be:4*}

Сем. класс: **verb.stative**

Дефиниция: *have the quality of being; (copula, used with an adjective or a predicate noun)*

Гипоним 1 ID: *eng-30-01188342-v*

PWN Synset: {*be full:1*}

Сем. класс: **verb.consumption**

Деф: *be sated, have enough to eat*

Гипоним 2 ID: *eng-30-02604760-v*

PWN Synset: {*seethe:3; boil:4*}

Сем. класс: **verb.emotion**

Деф: *be in an agitated emotional state*

Гипероним – это глагол состояния. Мы также рассматриваем его гипонимы (наследующие его статическое значение), которые могут относиться к разным семантическим классам – консумация, эмоции, восприятие и так далее.



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделиране.
Съпоставително изследване за
български и руски

Semantic prime	Bulgarian WordNet		Russian WordNet	
	# synsets	# literals	# synsets	# literals
verb.stative	559	1725	392	641
verb.cognition	503	1776	389	605
verb.perception	342	1173	261	432
verb.emotion	264	1035	210	385
verb.change	27	29	6	10
verb.body	19	86	14	21
Others	72	252	56	96



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделчране.
Съпоставително изследване за
български и руски

➤ Фреймнет (<https://framenet.icsi.berkeley.edu/>)

- Лексическая база с 13000 значений из разных частей речи, каждое значение связано с определением и использованием, а также ему приписан семантический фрейм.
- Он также включает более 200000 аннотированных вручную примерных предложений, относящихся к более чем 1200 фреймам.



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделране.
Съпоставително изследване за
български и руски

- Мы приписываем синсетам фреймы из Фреймнета с соответствующими элементами фрейма.
- Фреймы также классифицируются как статические – фрейм State и связанные с ним фреймы при помощи иерархической релации Inherits from – Existence, Emotions, Being located, т.д.



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за състояние – лингвистично моделиране.
Съпоставително изследване за български и руски

PWN Synset: {seethe:3; boil:4} > Фрейм: Emotion heat

<p>Emotion [Emo] Эмоция</p>	<p>The Emotion experienced (usually restricted to words such as anger, fury, rage, passion):</p> <p>Rage BOILED inside him.</p> <p>He BOILED with rage.</p> <p>Pat's behavior made me SEETHE.</p>
<p>Experiencer [Expe] Экспириенцер Semantic Type: Sentient</p>	<p>The Experiencer is the individual that experiences the Emotion.</p> <p>Pat SEETHED with rage.</p>
<p>Seat_of_emotion [sea] Центр эмоции</p>	<p>The words in this frame can occur with an overtly expressed Seat of the Emotion. This frame element occurs as a PP Complement:</p> <p>Fury SEETHED within her.</p> <p>I was SEETHING inside.</p>

Синонимическому ряду из Ворднета приписывается фрейм Emotion heat на основе значения и принадлежности соответствующих глаголов *seethe, boil* к этому фрейму.

Фреймы приписаны автоматическим путем к PWN, а через него и к БулНет. 5000+ глагольных синсетов валидированы вручную

Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за състояние – лингвистично моделиране.
Съпоставително изследване за български и руски

Пример

SYNSET	SEM. CLASS	BG SYNSET	RU SYNSET	FRAME	FRAME Elements
eng-30-02603699-v	verb.stative	съществувам:3; съм:13; има:4; битувам:1	[быть, существовать]	Existence	Entity;;
eng-30-02614181-v	verb.stative	живея:2; жив съм:1; съм:18	[существовать]	Dead_or_alive	Protagonist:Sentient; Figure;;
eng-30-02614387-v	verb.stative	живея:8; поминувам:1; поминавам:1; помина:1	[жить]	Manner_of_life	Experiencer;; Lifestyle;; Manner;;
eng-30-02198234-v	verb.perception	струва ми се:2; сторва ми се:2; стори ми се:2	[казаться, представляться]	Unattributed_information	Unattributed_information 1 Reported_fact;;
eng-30-00077698-v	verb.body	задавам се:3; задавя се:3; давя се:3; задушавам се:1	[задыхаться]	Perception_body	Experiencer:Sentient; Body_part:Body_part;
eng-30-00623006-v	verb.cognition	поставям:4; поставя:4; задавам:1; задам:1	[говорить_загадками]	Stimulate_emotion	Experiencer:Sentient; Stimulus;;
eng-30-01188144-v	verb.consumption	умирам от глад:2; умра от глад:2; гладувам:3	[голодать, быть_голодным]	Perception_body	Experiencer:Sentient; Body_part:Body_part;
eng-30-02134927-v	verb.perception	звуча:1; прозвучавам:1; прозвуча:1; озвучавам:1	[издавать_звук, звучать]	Give_impression	Phenomenon;; Characterization;; Appraisal;; Inference;;



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделране.
Съпоставително изследване за
български и руски

- Автоматически извлечени непосредствените окружения глаголов состояния из Болгарско-русско-польского корпуса (<https://clarin-pl.eu/dspace/handle/11321/308>), что иллюстрирует реализацию предикатов в обоих языках (90 000 параллельные пары).
- Извлечены параллельные пары предложений, в которых встречаются глаголы из заранее определенного списка глаголов состояния.
 - Предстоит провести наблюдения и обобщения насчет частотности, дистрибуции и переводных соответствий в болгарском и русском языках.
 - Анализ может быть расширен другими параллельными и / или соотносимыми болгарско-русскими и русско-болгарскими корпусами.



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделране.
Съпоставително изследване за
български и руски

Пример

*БГ: Имайте предвид, че за това **съществува** и седмо доказателство, което е най-неопровержимото! На вас то съвсем след малко ще ви бъде демонстрирано!*

*РУ: Имейте в виду, что на это **существует** седьмое доказательство, и уж самое надежное! И вам оно сейчас будет предъявлено.*



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделиране.
Съпоставително изследване за
български и руски

Пример

*БГ: – Вашият разказ е изключително интересен, професоре, въпреки че далеч не **съвпада** с евангелските.*

*РУ: – Ваш рассказ чрезвычайно интересен, профессор, хотя он и совершенно не **совпадает** с евангельскими рассказами.*



Деятельность 1.2

Создание базы данных с эксцерпированным материалом



Онтология на ситуациите за
състояние – лингвистично моделране.
Съпоставително изследване за
български и руски

Пример

*БГ: Освен това хората **предпочитат** да омъжат дъщерите си по-скоро за продавачи на пуканки, отколкото за пастири.*

*РУ: И родители **предпочитают** брать в зятья торговцев, а не пастухов.*