




# Simulation of background knowledge and Bridging annotation in **Russian**

Dikonov V.G. <dikonov@iitp.ru>  
IITP RAS, Moscow, Russia

- 
- Introduction
  - Our approach
  - Annotation process
  - Evaluation / Problems

# Introduction: Theory

*Introduced by Herbert Clark in 1975*

The **Reader** builds a mental model of the world described in a text from pieces of information about **Entities** existing in that world.

- New Entities are found while reading
- The Reader relates each new entity with already known ones

This process is called **bridging**

- The relations are based on **background knowledge**
- Extra **associated** entities may be added to form the relations

*Bridging is an essential mechanism of understanding.*

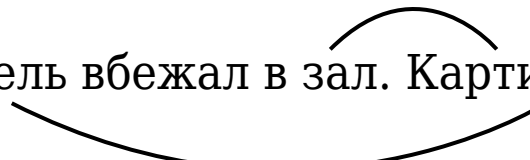
# Introduction: Bridging annotation

Bridging annotation creates anaphoric links between words. Words that denote new entities become **referents**, old – **antecedents**.

Bridging links:

- are not derived from syntax
- represent semantic relations between entities

Смотритель вбежал в зал. Картина была порвана.

A diagram illustrating bridging links in the Russian sentence "Смотритель вбежал в зал. Картина была порвана." Two curved lines connect the words "Смотритель" (viewer) and "Картина" (picture). One line is positioned above the words, and the other is below them, indicating a semantic relationship between the viewer and the picture.

Two groups of approaches rely on:

- Syntactic patterns (Roitberg & Khachko, 2017)
- Semantic resources (e.g. Princeton Wordnet)

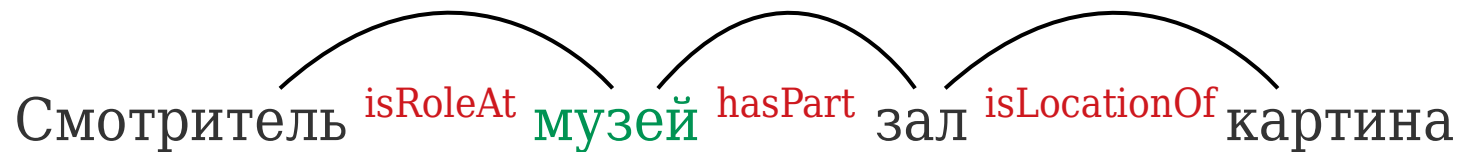
# Our approach

- Links meanings of words represented by ontology concepts

Картина – *PaintedPicture* «painting»,  
**not** *MotionPicture* «movie» / *PictureSituation* «situation in general»

- All relations have a semantic type
- Complex relation types

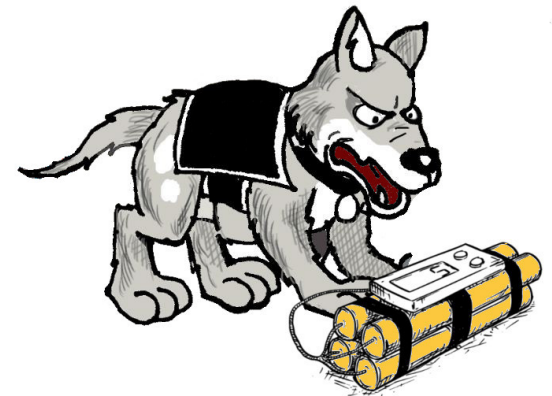
Смотритель *isRoleAt* музей *hasPart* зал *isLocationOf* картина



- Associations based on background knowledge  
Additional concepts, not mentioned in the text but derived from a model of domain, can be added to explain the relations.

# Associations

На станции метро "Владыкино" в Москве найдено взрывное устройство. Найденный предмет обследовали с использованием служебных **собак**.



# Associations

На станции метро "Владыкино" в Москве найдено взрывное устройство. Найденный предмет обследовали с использованием служебных собак.

устройство *isObjectOf* обследовать  
*hasAgent* полицейский *isUserOf* собака



# Knowledge base

The background KB contains:

- Low-level **Ontology**

Provides 65031 concepts used in bridging relations

Based on the free UNLDC dictionary and SUMO, work-in-progress

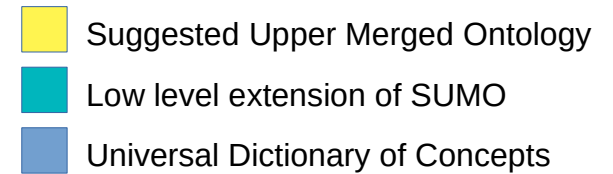
- **Domain descriptions**

Manually written brief descriptions of everyday-life domains in the form of RDF graphs (50-1000 triplets).

*Education, Shopping, Police, Theatre, Fishing, PostOffice, BankOffice...*



# Low-level Ontology

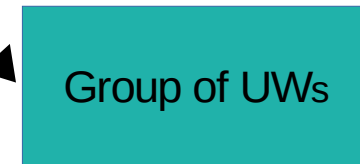
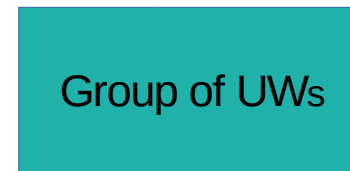
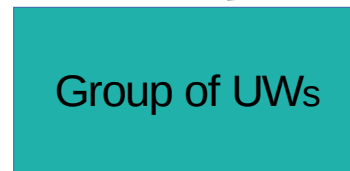


- Машинист паровоза  $\neq$  SocialRole Locomotive
- Режиссер ГАБТ  $\neq$  Position Organization  $\neq$  TheatreProfession EntertainmentCompany

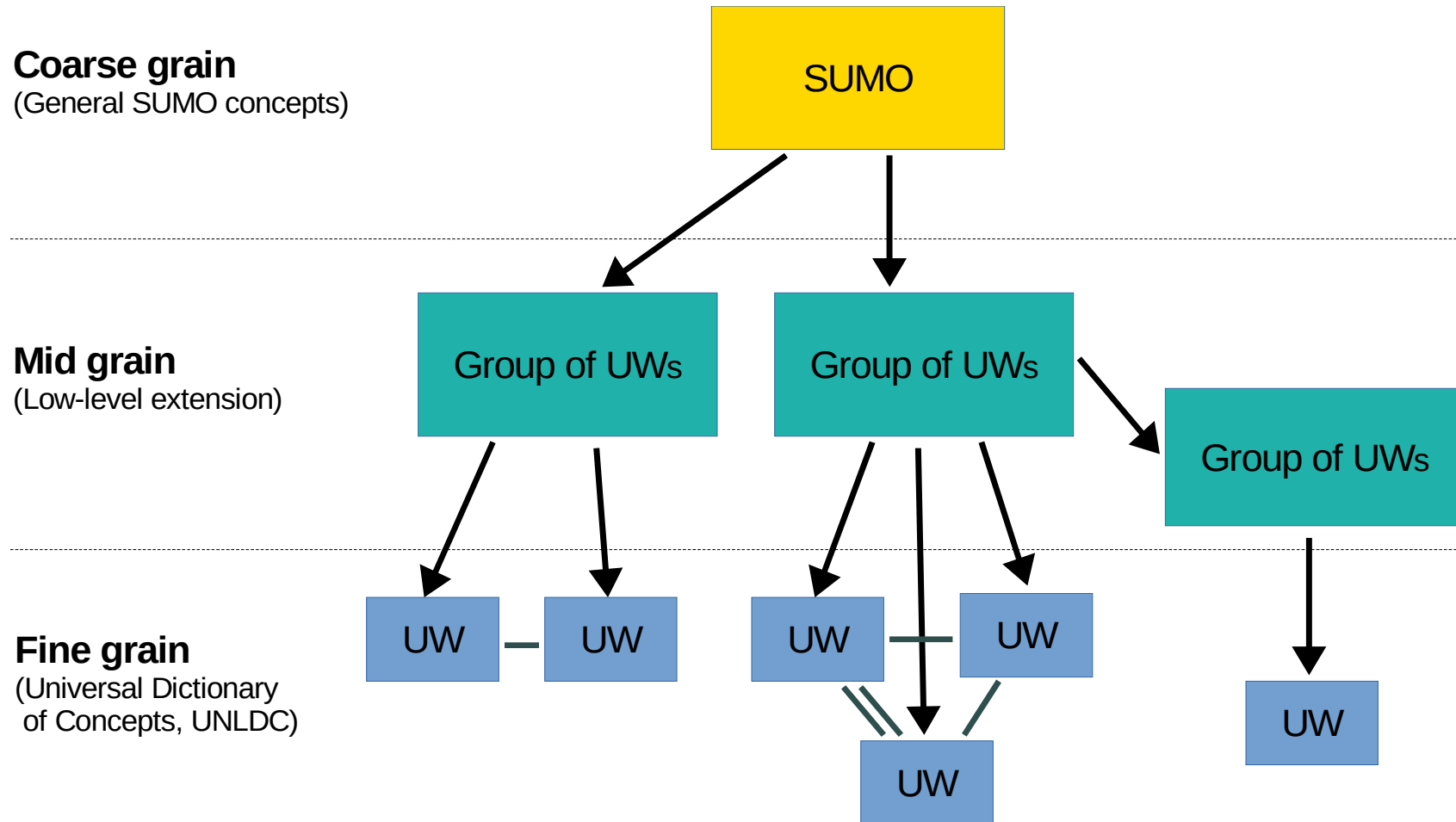
**Coarse grain**  
(General SUMO concepts)






**Mid grain**  
(Low-level extension)



**Fine grain**  
(Universal Dictionary of Concepts, UNLDC)

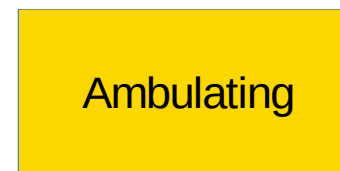


# Low-level Ontology

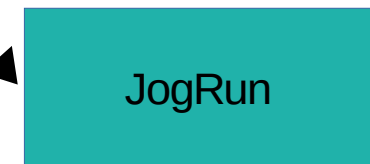
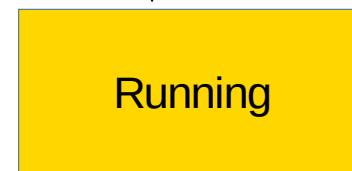
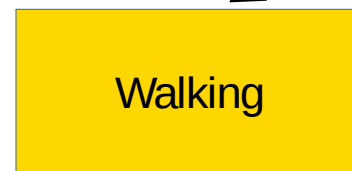
-  Suggested Upper Merged Ontology
-  Low level extension of SUMO
-  Universal Dictionary of Concepts

- Машинист паровоза  $\neq$  SocialRole Locomotive
- Режиссер ГАБТ  $\neq$  Position Organization  $\neq$  TheatreProfession EntertainmentCompany

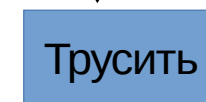
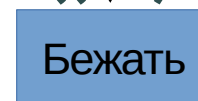
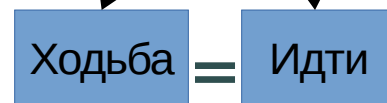
**Coarse grain**  
(General SUMO concepts)



**Mid grain**  
(Low-level extension)



**Fine grain**  
(Universal Dictionary of Concepts, UNLDC)



# Types of links

Group	Relations	* Each asymmetric type has a reverse type
Function	hasFunction , hasRoleAt , hasChief	
Part ↔ whole	hasPart , hasOptionalPart , hasDetachablePart , hasMember , hasSubEvent	
Object ↔ matter	hasIngredient , hasSubstance	
Event ↔ role	hasAgent , hasAgent2 , hasObject , hasObject2 , hasLocation , hasStartingPlacePoint , hasTerminalPlacePoint , hasInstrument , hasRecipient , hasBeneficiary , hasSource	
Cause ↔ result	hasResult , NewstatusOfAgent , NewstatusOfObject	
Temporal	before/after , during	
Misc. association	hasTime , hasResident , hasBeliever , hasAuthor , hasMaker , hasFrame , isUserOf , hasOwner , hasAttribute , cohyponym, SameAs	

- Open set of relation labels
- Complex relations are constructed from the basic types above.

# Domain descriptions

- Included statements are expected to be true only within the domain.
- Common knowledge of an average Russian/USSR person.

Domain	Subdomain	S	V	O	
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasPart СЦЕНА,театральная сцена	TheaterStage
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasOptionalPart ложа	BoxAuditoriumSeat
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasPart ПАРТЕР	OrchestraSeating
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasOptionalPart АМФИТЕАТР	Amphitheatre
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasOptionalPart БАЛКОН	BalconyStructure
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasOptionalPart ГАЛЕРКА	PeanutGallery
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasPart проход	AislePassageway
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasPart СИДЕНЬЕ	Seat
StagePlace	Opera	AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasOptionalPart ЯМА,оркестровая яма	PitArea
StagePlace	Ballet	AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasOptionalPart ЯМА,оркестровая яма	PitArea
StagePlace		AuditoriumArea	АУДИТОРИЯ,ЗАЛ	hasPart ЛЮСТРА	Chandelier
StagePlace		TheaterStage	СЦЕНА,театральная сцена	hasOptionalPart суфлёрская будка	PromptBox
StagePlace		TheaterStage	СЦЕНА,театральная сцена	hasPart театральный свет	TheaterLight
StagePlace		TheaterStage	СЦЕНА,театральная сцена	hasOptionalPart ПРОЖЕКТОР,СОФИТ	SpotlightLamp
StagePlace		TheaterStage	СЦЕНА,театральная сцена	hasOptionalPart ПРОЖЕКТОР	Projector
StagePlace		TheaterStage	СЦЕНА,театральная сцена	hasOptionalPart РАМПА	Footlights
StagePlace		TheaterStage	СЦЕНА,театральная сцена	hasPart ЗАНАВЕС	TheaterCurtain
StagePlace		TheaterStage	СЦЕНА,театральная сцена	hasPart КУЛИСА	Coullisse
StagePlace		PerformanceStageWing	КУЛИСЫ	hasPart ГРИМЕРКА	MakeUpRoom
StagePlace		PerformanceStageWing	КУЛИСЫ	hasPart СКЛАД1	Stockroom
StagePlace		MakeUpRoom	ГРИМЕРКА	hasPart ЗЕРКАЛО	Mirror
StagePlace		SeatSpace	МЕСТО	hasLocation АУДИТОРИЯ,ЗАЛ	AuditoriumArea
StagePlace		Scenery	ДЕКОРАЦИЯ	hasLocation СЦЕНА,театральная сцена	TheaterStage
StagePlace		FlatScenery	плоскостная декорация	hasLocation СЦЕНА,театральная сцена	TheaterStage
StagePlace		PerformanceProgram	ПРОГРАММА,ПРОГРАММА	hasTopic КОНЦЕРТ1	ConcertPerformance
StagePlace		PerformanceProgram	ПРОГРАММА,ПРОГРАММА	hasTopic СПЕКТАКЛЬ,представление	DramaticPerformance
StagePlace	Opera	Libretto	ЛИБРЕТТО	hasTopic ОПЕРА	Opera
StagePlace	Ballet	Libretto	ЛИБРЕТТО	hasTopic БАЛЕТ	Ballet

# Domain descriptions

- Only the most important concepts for a domain
- Augmented with:
  - ontology class relations, e.g. food → bread, milk...
  - semantic relations from UNLDC
    - inherited from PWN 3.0, (meronymy, derivation links, not “topic terms”)
    - additional

airplane *isAgentOf* fly, *isUsedBy* pilot, *hasDevicePurpose* transportation

Equal or richer link data than in projects based on English Wordnet.  
Less lexical coverage. (37733 Russian words, excluding MWEs)

# Annotation

Domains: Shopping, Police

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation

StateTradeInspection

В конце августа **Госторгинспекция** проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation

StateTradeInspection      Verify      (if X is TRUE)  
CheckAnalyze      (conformity)

В конце августа **Госторгинспекция** **проверила** ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

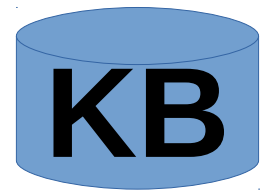
Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...



# Annotation

?



StateTradeInspection

Verify  
CheckAnalyze

В конце августа **Госторгинспекция** **проверила** ряд сетевых супермаркетов экономкласса в Москве.

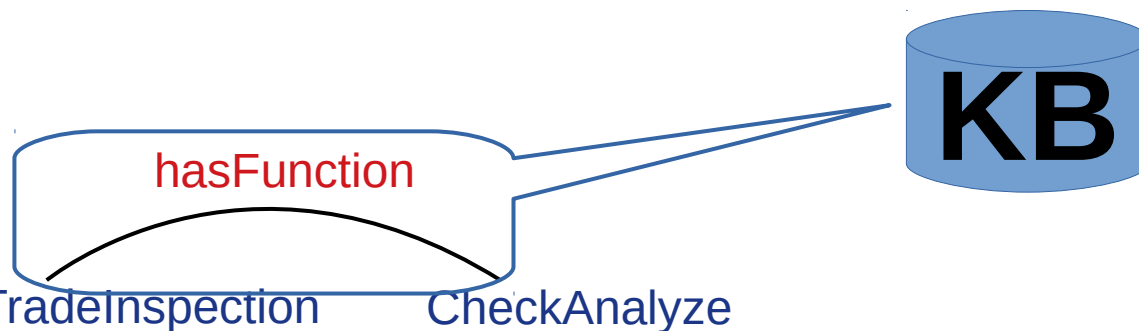
И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation



В конце августа **Госторгинспекция** **проверила** ряд сетевых супермаркетов экономкласса в Москве.

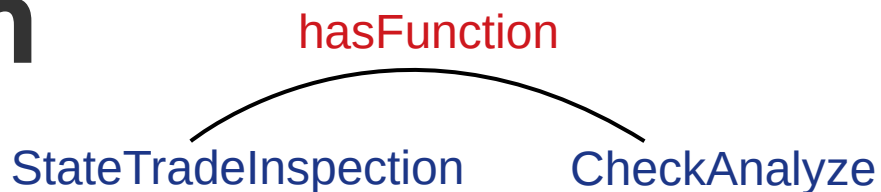
И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation



В конце августа **Госторгинспекция** **проверила** ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation

IsObjectOf ПРОВЕРКА hasAgent

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation

hasAgent

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве. isResultOf

hasLocation

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

IsObjectOf ПРОДАЖА hasAgent

# Annotation

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве. И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

hasLocation

synonym

hasLocation

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

IsObjectOf ПРОДАЖА hasAgent / hasLocation



# Annotation

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

IsIngredientOf

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

IsMakerOf

IsAgentOf ПРОДАЖА isAgent2

# Annotation

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation – Words

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Annotation - Entities

В конце августа Госторгинспекция проверила ряд сетевых супермаркетов экономкласса в Москве.

И обнаружила, что доля поддельной сгущенки в ряде магазинов доходила до 65 процентов.

Настоящая сгущенка состоит из трех элементов: сахара, молока, воды.

Но многие предприятия разработали для сгущенки свои технические условия (ТУ), которые допускают замещение молочной основы дешевыми маслами растительного происхождения.

...

# Manual corpus annotation attempt

- 2 short texts
- 6 annotators
- Chosen approach – Entities based

## Result:

**197** bridging relations marked in total.

Very low inter-annotator agreement ! Unlabeled relations !

Number of annotators	1	2	3	4	5	6
% relations tagged	75,1%	13,2%	7,1%	3%	1%	0,5%

Every annotator recorded his personal associations.

- background knowledge ( Education, life experience ... )
- intentions ( What the Reader wants to learn ? )

Is it possible / SAFE to train annotators to make identical associations ???

# Manual corpus annotation attempt

- 2 short texts
- 6 annotators
- Chosen approach – Entities based

## Result:

**197** bridging relations marked in total.

No inter-annotator agreement ! Labeled relations !

Number of annotators	1	2	3	4	5	6
% relations tagged	<b>89,15%</b>	<b>8,4%</b>	<b>2%</b>	0,4%	0%	0%

Every annotator recorded his personal associations.

- background knowledge ( Education, life experience ... )
- intentions ( What the Reader wants to learn ? )

Is it possible / SAFE to train annotators to make identical associations ???

# Variability of complex relations

- CondensedMilk SameAs YieldIndefiniteQuantity hasFrame Market hasMember RetailStore  
Сгущенка — продукция на рынке, который включает магазины
- CondensedMilk SameAs YieldIndefiniteQuantity isObjectOf Selling hasAgent RetailStore  
Сгущенка — продукция, которую продают магазины
- CondensedMilk hasIngredient Milk hasFrame Market hasMember RetailStore  
В сгущенке есть молоко, которое существует на рынке, который включает магазины
- CondensedMilk hasFrame Market isFrameOf Selling hasAgent RetailStore  
Сгущенка существует на рынке, где продают, а делают это магазины
- CondensedMilk hasFrame Market isFrameOf CounterfeitGoods hasLocation Shop  
Сгущенка существует на рынке, где есть фальсификат, находящийся в магазинах
- CondensedMilk hasMaker Manufacturer isMakerOf CounterfeitGoods hasLocation Shop  
Сгущенку делают производители, делающие фальсификат в магазинах
- CondensedMilk isObjectOf Selling hasAgent RetailStore isOwnerOf Shop  
Сгущенку продает магазин (организация), который владеет зданием магазина.
- CondensedMilk isObjectOf Selling hasAgent ChainBusiness hasMember RetailStore  
Сгущенку продает торговая сеть, части которой магазины
- CondensedMilk isObjectOf Selling hasAgent RetailStore  
Сгущенку продают магазины



# Manual vs Computer

*Entity-based*

*Word-based*

**197** relations tagged manually, **532** tagged automatically

**78 (39,5%) overlap** between the sets (Unlabeled comparison)

Computer outperformed most human annotators!

Annotators	A	B	C	D	E	F	Computer
Relations	22 11,1%	23 11,6%	35 17,7%	46 23,3%	72 36,5%	<b>84</b> <b>42,6%</b>	<b>78</b> <b>39,5%</b>

- Out of 454 relations tagged only by computer
  - 50 relations are duplicates of human ones
  - 45 relations are wrong because the entities are different

предприятие *isMakerOf* сгущенка (поддельная)

предприятие *isMakerOf* сгущенка (настоящая)

# Computer-only relations

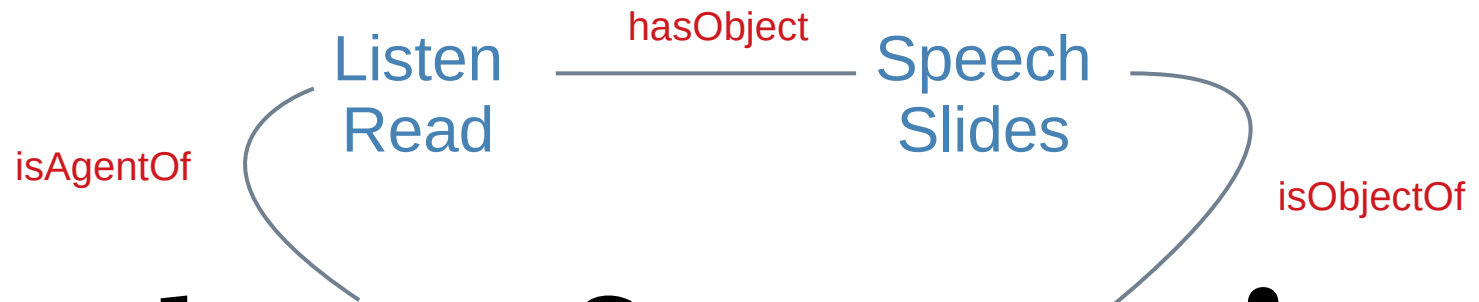
- продукцией *isObjectOf* торгуют **Doubles syntax**
- производителей *isMakerOf* ПРОДУКЦИЯ *isObjectOf* продажу **OK**
- производителей *isMakerOf* этикетку **Irrelevant**
- производителей *isMakerOf* КОНТРАФАКТ *isObjectOf* ОБНАРУЖИВАТЬ *hasAgent* Контролеры **Same**
- производителей *isMakerOf* КОНТРАФАКТ *isObjectOf* ОБНАРУЖИВАТЬ *hasAgent* эксперты **Entity**
- производителей *isMakerOf* ПРОДУКЦИЯ *SameAs* КОНТРАФАКТ *isObjectOf* толкают
- производителей *isMakerOf* ПРОДУКЦИЯ *SameAs* МОЛОКО *hasLocation* полках
- производителей *isMakerOf* ПРОДУКЦИЯ *isObjectOf* ПРОДАВАТЬ *hasAgent2* покупатель
- покупатель *isAgent2Of* ПРОДАВАТЬ *hasAgent* СУПЕРМАРКЕТ *isLocationOf* толкают
- покупатель *isAgent2Of* продажу **OK**
- покупатель *isAgentOf* ИСКАТЬ *hasObject* МОЛОКО *isLocationOf* этикетку
- покупатель *isAgentOf* ИСКАТЬ *hasObject* МОЛОКО *hasLocation* полках
- покупатель *isAgent2Of* ПРОДАВАТЬ *hasAgent* СУПЕРМАРКЕТ *isObjectOf* инспекции
- покупатель *isAgent2Of* ПРОДАВАТЬ *hasAgent* сети **OK**
- покупатель *isAgentOf* ИСКАТЬ *hasObject* МОЛОКО *SameAs* производства **BAD**

# Associations

List of associated concepts generated from the example text.

Not present in the text, but linked in the KB.

Concept	Rank	Upper Class	Translation
PayGiveMoney	113	Giving	платить
Commodity	81	Product	товар
Buying	63	FinancialTransaction	покупка,покупать
WebStore	60	RetailStore	интернет-магазин
ApparelStore	38	RetailStore	магазин одежды
Bakeshop	38	GroceryStore	булочная
Barbershop	38	RetailStore	цирюльня
BazaarShop	38	RetailStore	базар
BookStore	38	RetailStore	книжный магазин
BoothShop	38	RetailStore	палатка
...			
BuyIn	37	Buying	закупать
Keeping	35	IntentionalProcess	хранение
Carrying	33	Transfer	нести
PaymentMoney	33	price	плата,выплата (деньги)
HaulCarry	31	LandTransportation,Guiding	везти
BlackMarket	29	FinancialTransaction	[черный рынок]
MonopolyMarket	29	FinancialTransaction	монополия
Kidnapping	24	Robbing	похищение
PoliceDetective	22	PoliceOfficer	сыщик,детектив
SamplePurchase	22	Experimenting,Buying	контрольная закупка
monetaryValue	22	SingleValuedRelation	цена



# Thank you for attention!

Public resources

Universal Dictionary of Concepts (UNLDC):  
<https://github.com/dikonov/Universal-Dictionary-of-Concepts>

SUGGESTED UPPER MERGED ONTOLOGY (SUMO):  
<http://www.adampease.org/OP/>