

Automatic Gapping Resolution for Russian AGRR

Ivan Smurov, Maria Ponomareva, Tatiana Shavrina, Kira Droганova

30.05.2019

Traditional approach

Gapping - omission of a repeating predicate in a non-initial coordinated clause where two remnants remain expressed.

Я принял её за итальянку, а его — за шведа.

I mistook her for Italian and him for Swede

'I mistook her for Italian and ~~I mistook~~ him for Swede'

Annotation

- **Remnants** **R1, R2** pronounced elements of the gapped clause, which have correlates
- **Remnant correlates** **cR1, cR2** parallel elements in the full clause similar to the remnants, both semantically and syntactically
- **Gap** **V** possible position of missing material
- **Gap controller** **cV** main predicate in the full clause

Мы [_{cV}вернем] [_{cR1}Вашим шинам] [_{cR2}былую хватку], а [_{R1}Вам] [_V] [_{R2}безопасность на дороге]!

We will.return your tires former grip and you safety on road

We will return former grip to you tires and ~~return~~ you safety on the road'



Gapping in AGRR corpus

Broader interpretation

Stripping vs. Gapping with one remnant

Stripping:

The man stole the car after midnight, **but not** the diamonds. [Merchant 2016]

Abby can speak passable Dutch, and Ben, **too**. [Wurmbrand 2013]

Все мы любим Мамбу и Сережа **тоже**.

All we love *Mamba* and Serezha too

'All of us love Mamba, and Serezha ~~loves it~~ too'

Stripping vs. Gapping with one remnant

Gapping with one remnant:

[_{cV} Добавляем] [_{cR1} муку, крахмал и разрыхлитель], а **в конце** [_v] [_{R1} сметану].
add flour starch and baking.powder and at.the.end sour.cream

'We add flour, starch and baking powder, and at the end ~~we add~~ sour cream.'

Рост цен [_{cV} составил] [_{cR1} 11,9%] (**за 2009 год** [_v] [_{R1} 4,4%]).
growth prices amounted.to 11.9% in 2009 year 4.4%

'The price growth amounted to 11.9 %(in 2009 ~~it amounted to~~ 4.4 %).'

Gapping in subordinate clauses

[_{cR1} Он] [_{cV} знает] [_{cR2} ее], как [_{R1} живописец] [_{R2} свою картину].
he knows her as painter his painting

'He knows her as painter ~~knows~~ his work'

Different types of remnants

- predicates

Одно может вдохновлять, а другое вгонять в тоску.

one can inspire and other put in melancholy

‘One thing can inspire and the other ~~can~~ put you in a melancholic mood.’

- preposition phrases

Советую вам поменьше думать о проблемах, и побольше — об их решении.

recommend you less think about problems and more - about their solution

‘I recommend you to think less about problems, and ~~think~~ more about solving them.’

Different types of remnants

- adverbs

Вначале они играли интересно, потом – прескучно.
at.first they played interesting.ADV after - boring.ADV.INT
'At first they played interesting, then ~~they played~~ extremely boring.'

- adjectives

Сердце ее было слишком чистым, чувства слишком искренними.
heart her was too pure feelings too sincere
'Her heart was too pure and her feelings ~~were~~ too sincere.'



Obtaining the data

Obtaining the data

- Validation of automatically obtained markup instead of annotating sentences from scratch
- Compreno template-based module for gapping detection was used
- Variety of genres: fiction, technical texts, news + social media (vk)
- Every sentence was evaluated by 2 assessors:
 - 0 - no gapping
 - 1 - markup is ok
 - 2 - wrong markup
 - 3 - hard to decide

Assessment analysis

On technical texts, fiction, news

	0	1	2	3
0	1533	138	129	136
1	240	5301	1021	237
2	213	451	1600	281
3	307	177	117	108

Out of 11989 sentences 44% were considered correctly annotated and 13% were unanimously considered to have no gapping.

Assessment analysis

On social media texts

	0	1	2	3
0	1817	232	174	118
1	154	1900	142	46
2	75	130	360	21
3	139	53	36	25

Out of 5422 sentences 35% were considered correctly annotated and 34% were unanimously considered to have no gapping.

Validation of the corpus

How well does it represent the
phenomenon?

Validation of the corpus

- test set from SynTagRus (v. 2015)
- all the sentences were manually verified and divided into three categories:
 - **1** - cases similar to the ones encountered in the AGRR corpus
 - **2** - cases of gapping not included in the AGRR corpus
 - **3** - cases considered other types of ellipsis rather than gapping

What does the AGRR corpus miss? (cat. 2)

- more than two remnants

[_{cR1} В Испании] [_{cR2} в 1923 году] [_{cV} установил] диктатуру [_{cR3} генерал Педро де Ривера],
. In S. . in 1923 year . established dictatorship . general Pedro de Rivera,
[_{R1} в Польше] [_{R2} в 1926-м] - [_{R3} Пилсудски].
. in Poland . in 1926 - . Pilsudski

‘In Spain, the dictatorship of General Pedro de Rivera was established in 1923, while in Poland ~~the dictatorship was established~~ by Pilsudski in 1926.’

- the order of remnants differs from the order of correlates

[_{cR1} Школа и уроки] [_{cV} принадлежали] [_{cR2} кругу мучительных обязанностей], а [_{R2}
. school and lessons . belonged.to . circle painful duties and .
душевному выбору] - [_{R1} зеленая птица с красной головой].
soul.ADJ choice - . green bird with red head

‘School and lessons belonged to the focus of painful duties, while a green bird with the red head ~~belonged to~~ the choice of the soul.’

What does the AGRR corpus miss? (cat. 2)

- The gap is controlled not by a verb, but by a noun

Бюрократизм привел к [cV отчуждению] [cR1 трудящихся] [cR2 от власти], [R1 крестьян]
red.tape led to alienation working.people from power peasants
[R2 от земли].
from land

‘Bureaucracy led to the alienation of working people from power, and alienation of peasants from the land’

- The correlate is higher syntactically than the elided predicate

Если [cR1 можно] [cV передать] [cR2 один университет], то почему [R1 нельзя] [R2
if is.possible transfer.INF one university, then why not.possible
другие]?!
others

‘If it is possible to transfer one university, then why can’t others be transferred?!’

What does the AGRR corpus miss? (cat. 2)

- Gapping “with generalization”

[_{cR1} Средства и способы] создаются талантливыми учеными, а [_{cV} реализуются]: [_{R1}
means and methods are.created talented scientists and are.realized
средства] - [_{R2} военной промышленностью], а [_{R1} способы] - [_{R2} военной наукой и
means - military industry and methods - military science and
опытом]
experience.

‘Means and methods are created by talented scientists, and are realized: the means are realized by the military industry, and the methods are realized by military science and experience.’

What the AGRR corpus does not include? (cat. 3)

- ellipsis in comparative constructions

От сна за рулем погибает столько же водителей, сколько от алкоголя
from sleeping behind wheel die as.many drivers how.many/as from alcohol
'As many drivers die from sleeping behind the wheel, as ~~many drivers die~~ from alcohol'

- stripping

В Сталингаде каждый сражается, причем как мужчины, так и женщины
in Stalingrad everyone fights and both men and women.'
'In Stalingrad, everyone continuously fights, both men and women ~~fight~~.'

What the AGRR corpus does not include? (cat. 3)

- sluicing

Медикам дается указание как-то бороться с этим явлением, а как – никому
doctors are.given instructions somehow cope with this phenomenon and how - no.one
не известно.

NEG knows

‘Doctors are instructed to somehow cope with this phenomenon, but no one knows, how to cope with it.’

- additional argument in elided clause

Правила меняются по ходу игры и всегда почему-то одним игроком
rules are.changed with progress game and always for.some.reason one player.INST

‘The rules are changed as the game progresses and for some reason ~~are changed~~ always by one player’

Category 2

- More than two remnants
- The order of remnants differs from the order of correlates
- The gap is controlled by a noun
- Gapping “with generalization”

Category 3

- Ellipsis in comparative constructions
- Stripping
- Sluicing
- Additional argument in elided clause

Shared Task Set-Up

Ellipsis resolution task decomposition

1. *Detect the clause with the gap*
2. *Find the controlling clause (i. e., find the head of the controlling clause)*
3. *Find the remnants and their correlates*
4. Replace the correlates with corresponding remnants
5. Consider agreement, coreference, word order and usage of quantifiers
6. Copy resulting structure to the second clause

Binary presence-absence classification

- **Definition:** for every sentence, participating systems must decide if there is a gapping construction in it.
- **Relevant metrics:** precision, recall and **f-measure**
- **SOTA on AGRR-2019:** 0.96
- **Desired behaviour:** since gapping is naturally rare precision is much more important than recall

Gap resolution

- **Definition:** for every sentence with gapping, participating systems must predict the position of the elided predicate and the pronounced predicate in the antecedent clause.
- **Relevant metrics:** symbolwise f-measure
- **SOTA on AGRR-2019:** 0.91

Full annotation

- **Definition:** for every sentence with gapping, participating systems must predict the linear position of the elided predicate and positions of its remnants in the clause with the gap, as well as the positions of remnant correlates and pronounced predicate in the antecedent clause.
- **Relevant metrics:** symbolwise f-measure
- **SOTA on AGRR-2019:** 0.89

Full annotation

- **Definition:** for every sentence with gapping, participating systems must predict the linear position of the elided predicate and positions of its remnants in the clause with the gap, as well as the positions of remnant correlates and pronounced predicate in the antecedent clause.
- **Relevant metrics:** symbolwise f-measure
- **SOTA on AGRR-2019:** 0.89

Symbolwise f-measure example

- Text: Утопающий, чтобы спасатель мог подтянуть его к судну, должен взяться за петлю руками, а в идеальном варианте — продеть петлю себе подмышки.
- True annotation: cV: 54:60 cR1: 61:84 cR2: V: 111:111 R1: 111:138 R2:
- Test annotation: cV: 54:60 cR1: 60:81 84:85 cR2: 0:9 V: 110:110 R1: R2:
- F-measures: cV: 1.0, cR1: 0.89, cR2: 0.0, V: 0.0, R1: 0.0, R2: 1.0
- Total f-measure: 0.482

SynTagRus results compared to AGRR test

Corpus	Participant	Binary classification	Gap resolution	Full annotation
<u>AGRR2019</u>	Winner	0.96	0.90	0.89
	Second best	0.95	0.86	0.84
SynTagRus gapping test set	Winner	0.91	0.76	0.77
	Second best	0.88	0.67	0.64

SynTagRus results by class

Category	Total number of samples	Participant	Number of sentences classified as gapping	Fraction of sentences classified as gapping
0	1166	Winner	8	0.007
		Second best	30	0.026
1	507	Winner	433	0.854
		Second best	420	0.828
2	75	Winner	26	0.35
		Second best	37	0.49
3	100	Winner	6	0.06
		Second best	13	0.13

Track organization

Shared task:

1. Closed track – an open-source track, convenient for research groups and student teams.
2. Open track - no restriction on data and systems used

3 different gapping tasks:

1. Binary presence-absence classification
2. Gap resolution
3. Full annotation

Timeline:

- January 26th 2019 - data release
- February 2nd - noisy data release
- February 23rd - submission deadline
- March 5th - results

AGRR Data

		No gapping		Has gapping		sum
dev	vk	670	2760	326	1382	20548
	other	2090		1056		
train	vk	2860	10864	1366	5542	
	other	8004		4176		
test	vk	343	1365	185	680	2045
	other	1022		495		
sum			14989		7604	22593

100k+ sentences with noisy data - with automatic gapping annotation

Questions

Closed Track:

- 9 teams - 5 universities, 2 IT-companies, 2 independent researchers

Open Track:

- 0 participants

Closed track:

- Should universal pretrained models be allowed?

Deadlines:

- Some participants were late!

Results

team	binary			gap resolution	full
	precision	recall	f-measure	f-measure	f-measure
fit_predict	0.969	0.95	0.959	0.905	0.892
EXO	0.899	0.964	0.931	0.815	0.786
Koziev Ilya	0.774	0.903	0.834	0.677	0.647
Derise	0.801	0.906	0.850	0.665	0.622
Meanotek	0.891	0.781	0.832	0.635	0.514
MGY-DeepPavlov	0.934	0.644	0.762	0.601	0.587
vlad	0.778	0.915	0.841	0.574	
MorphoBabushka	0.763	0.619	0.683	0.466	0.440
nsu-ai	0.485	0.123	0.196	0.037	0.036

Late Submissions

team	results				
	binary			gap resolution	full
	precision	recall	f-measure	f-measure	f-measure
MГУ-DeepPavlov	0.973	0.646	0.776	0.617	0.599
MГУ-DeepPavlov	0.898	0.934	0.916		
MГУ-DeepPavlov	0.97	0.712	0.821	0.658	0.653
EXO	0.946	0.946	0.946	0.859	0.836

Results

team	binary			gap resolution	full
	precision	recall	f-measure	f-measure	f-measure
fit_predict	0.969	0.95	0.959	0.905	0.892
EXO	0.899	0.964	0.931	0.815	0.786
Koziev Ilya	0.774	0.903	0.834	0.677	0.647
Derise	0.801	0.906	0.850	0.665	0.622
Meanotek	0.891	0.781	0.832	0.635	0.514
MGY-DeepPavlov	0.934	0.644	0.762	0.601	0.587
vlad	0.778	0.915	0.841	0.574	
MorphoBabushka	0.763	0.619	0.683	0.466	0.440
nsu-ai	0.485	0.123	0.196	0.037	0.036

github.com/dialogue-evaluation/AGRR-2019