The Database of Constructions with Lexical Repetitions RepLeCon and Inter-Annotator Agreement

Olga Blinova (o.blinova@spbu.ru), Elena Vilinbakhova (e.vilinbakhova@spbu.ru)

CompLing2022, ITMO University, St. Petersburg, Russia, 24.06.2022

Report Structure

1. RepLeCon: a general description
2. Rhetorical relations
3. Inter-Annotator agreement
4. Results

RepLeCon

RepLeCon – a database of **CON**structions with **LE**xical **REP**etitions

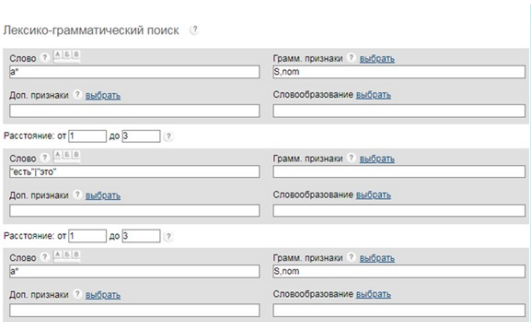
Five language pairs:
Russian ⇔ *English*
Russian ⇔ *French*
Russian ⇔ *German*
Russian ⇔ *Italian*
Russian ⇔ *Spanish*

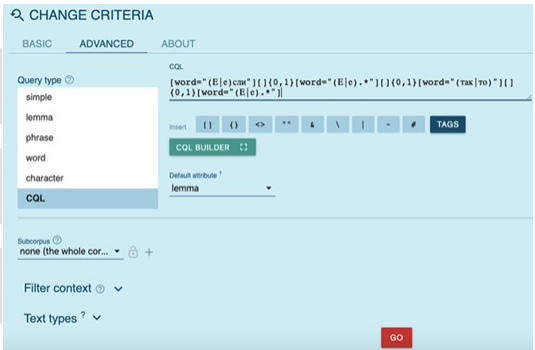
Goal

- to provide a deeply annotated dataset of Russian constructions with lexical repetitions and their equivalents in translation;
- to contribute to theoretical linguistic research on repetitions;
- to elaborate on its practical applications.

Corpora

Russian National Corpus, OPUS parallel corpus





Constructions

Tautologies:

- Identity-establishing tautologies: X есть X (война есть война)
- Negative tautologies: X (это) не X (у нас зима это не зима)
- Conditional tautologies: если P, то P (если он придёт, то придёт)
- (Cor)relative tautologies: что P, то P (кто знает, тот знает)
- Comparative tautologies: X как X (люди как люди)
- Disjunctive tautologies: P или не P (быть или не быть)

Constructions

Tautological infinitives:

- with negation: Vinf не Vfin (видеть не видел)
- without negation: Vinf Vfin (звонить звонил)

Temporal constructions with repetitions:

- NP-Nom в NP-Acc (минута в минуту)
- NP-Nom за NP-Ins (неделя за неделей)
- NP-Gen в NP-Ins (из года в год)
- NP-Gen на NP-Acc (со дня на день)
- NP-Nom от NP-Gen (день ото дня)

Constructions

And more constructions with repetitions:

- X так X (гулять так гулять)
- NOM INS: X X-ом (работа работой)
- X не X, а (актёр не актёр, а так)
- Vfin не Vfin (ищи не ищи)
- Constructions with reduplicated pronouns: PRN-PRN, а (кто-кто, а Петя придёт)
- "Grammaticised" repetitions: ни X, ни Y (ни дождь, ни снег)

Annotation

Five general categories:

1. Context description
2. Structure
3. Semantics
4. Pragmatics
5. Characteristics of translations

CASE STUDY: RHETORICAL RELATIONS

Rhetorical Structure Theory, Mann and Thompson (1988)

- describes the diversity of structures observed to occur in texts
- posits various relations between text parts
- identifies both the transition point of a relation and the extent of the items related

RHETORICAL RELATIONS IN MONOLOGUES

Jasinskaja and Karagjosova's proposal: (Jasinskaja and Zeevat 2008); (Jasinskaja 2009); (Jasinskaja and Karagjosova 2021)

Six rhetorical relations

- Resemblance relations: *Elaboration*, *Parallel*, 'semantic opposition' type of *Contrast*
- Contiguity in space and time relations: *Narration*
- Causal relations: *Explanation*, *Result*, 'denial of expectation' type of *Contrast*.

Taxonomy supported by real data in the domains of connectives and anaphora.

\+ **Condition**

RHETORICAL RELATIONS IN MONOLOGUES

Condition recognizes how the realization of one situation depends on the realization of another situation.
(1) You said that every woman is special in her own way. That there's no two alike. So if that's true.....then **Dee Dee is Dee Dee**. Let her be who she is. [OpenSubtitles2011, Dr. T the Women, 2000]

In the RR of *Explanation* two discourse units describe two state of affairs, and one of them is presented as the reason for the other.
(2) I'm afraid I have no choice. **Policy is policy**. [OpenSubtitles2011, Lexx. Season 2, Episode 2, 1997-2002]

RHETORICAL RELATIONS IN DIALOGUES

Asher and Lascarides' proposal, Asher and Lascarides (2003; 2013) Segmented Discourse Representation Theory

- *Question-Answer Pair* suggests that the contribution is a direct response to the question
- *Indirect Question-Answer Pair* holds when the contribution is not a direct response to the question, but the questioner can infer the necessary information from it
- *Partial Question-Answer Pair* implies that the contribution is may rule out some true answers, but is not sufficiently informative that the questioner can compute a direct answer from it
- *Correction* holds when the contribution is aimed at correcting the previous utterance
- *Denial* is employed to rebut the previous contribution
- *Acknowledgement* is used to express the acceptance of the previous utterance

RHETORICAL RELATIONS IN DIALOGUES

Question Answer Pair suggests that the contribution is a direct response to the question.

(3) — What's your thesis?
— All I'm saying is **people are people**. We do what we do...
[OpenSubtitles2011, United States of Tara, 2010]

Indirect Question Answer Pair holds when the contribution is not a direct response to the question, but the questioner can infer the necessary information from it.

(4) — With all respect, doctor, I'm counting on Excelsior.
— Excelsior? **A ship is a ship** Why would you want that bucket of bolts?
[OpenSubtitles2011, Star Trek IV: The Voyage Home, 1986]

Inter-annotator agreement (IAA)

"Inter-annotator (or inter-coder) agreement has become the quasi-standard procedure for testing the accuracy of manual annotations. This process is based on the assumption that if multiple coders agree in their coding decisions of the same material we can be certain that — at least for this set of data and this set of coders — annotations are free of unsystematic and distorting variations" (Bayerl, Paul 2011).

Sources of disagreement:

- gaps in the annotation guide,
- difficulty in interpreting a markable item,
- annotators' carelessness.

IAA calculation scheme

Agreement measures:

- **Percentage agreement** (number of items on which annotators agree / total number of items)
- **Chance-corrected agreement measures (Bennett, Alpert Goldstein's S, Scott's pi, Fleiss' kappa etc.)**

IAA calculation steps:

1. The fields of the RepLeCon database were divided into "technical" fields and "interpretable" ones.
2. All "interpretable" fields were annotated by **two coders with year-long annotation experience**.
3. The received data was analyzed using percentage agreement and chance-corrected agreement measures.

Number of agreed responses

130 constructions with lexical repetitions

The fields where information about the **rhetorical relations** of constructions with the preceding (prec) and following (foll) discourse.

Role = core vs satellite

RR = rhetorical relations

field	N of consistent annotation
role-prec	128
role-foll	127
rr-prec	126
rr-foll	104

IAA calculation

R "irrCAC" package

`https://cran.r-project.org/web/packages/irrCAC/index.html`

The coefficients were calculated for each of the 4 fields ("rr-prec", "role-prec", "rr-foll", "role-foll") separately.

Raw dataset ("rr-prec"):

Annotator1	**Annotator2**
PQAP	PQAP
PQAP	PQAP
E	E
P	P

IAA calculation results

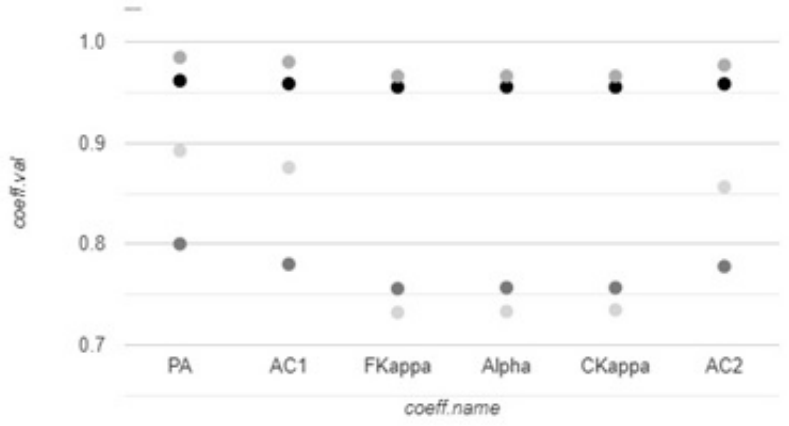
Six coefficients of agreement: Percent Agreement, Gwet's AC (AC1), Fleiss' Kappa, Krippendorff's alpha, Conger's Kappa, and Brennan Prediger's agreement coefficient

- pa — the percent agreement
- pe — the percent chance agreement
- coeff.val — the estimated value of an agreement coefficient
- coeff.se — the agreement coefficient standard error
- conf.int — the confidence interval

Agreement coefficients

- Percent agreement exceeds 90% in all cases, except for the examples of marking the role of the construction in rhetorical relation with the following discourse fragment (rr-foll).
- Gwet's AC (AC1) values are in the range [0.78; 0.98]
- Fleiss' Kappa, Krippendorff´s Alpha, Conger's Kappa values are in the range [0.73; 0.97]
- Brennan and Prediger's agreement coefficient (AC2) values are in the range [0.78; 0.98]

Agreement coefficients



Conclusions

(Landis and Koch 1977) benchmark scale

Coefficient thresholds:

values falling in the range [0.81; 1.00] could considered as "almost perfect"

values falling in the range [0.61; 0.80] could considered as "substantial"

First conclusion:

Annotators perform markup in a consistent way, the instruction is successfully compiled and does not require significant revision.

Number of categories

The probability of a random response depends **on the number of categories in a coding scheme**.

fields	**categories**
role-prec, role-foll	2
rr-prec, rr-foll (dialogues)	6
rr-prec, rr-foll (monologues)	7

The average percentage agreement for core vs satellite role markup is 94%.
The average percentage agreement for marking rhetorical relations (rr) is 88%.

Number of consistent annotations

fields	N of CA	total	percent agreement
rr-prec (monologues)	12	13	92.31
rr-foll (monologues)	10	13	76.92
rr-prec (dialogues)	114	117	97.44
rr-foll (dialogues)	94	117	80.34

Fisher's exact criterion, results:

- differences in the N of agreed and unagreed responses in the "role" and "rr" fields are statistically significant ($p < 0.001$)
- differences in the N of agreed and unagreed responses when comparing dialogical and monological fragments are statistically insignificant ($p = 0.383$)
- differences in "rr" field annotation results when comparing dialogical and monological contexts are also insignificant ($p = 0.517$)
- comparison of the N of agreed and unagreed responses, categorized by "foll" and "prec", showed statistically significant differences ($p < 0.001$).

Conclusions

1. annotators consistently determine what is the nucleus and what is the satellite,
2. the influence of the number of categories in the annotation scheme on consistency is confirmed by our data (annotators handle the markup in the "role" field significantly better than in the "rr" field),
3. annotators assign tags of rhetorical relations with the preceding discourse (compared to the following one) in a more consistent way.

Lexical markers of rhetorical relations

Observable rhetorical relation markers: potomu *čto* 'because', *no* 'but', *nu* 'well', *prosto* 'just', *vsë-taki* 'still', as well as *i kogda* 'and when', *xotja* 'though', *tak čto* 'so what', *krome togo* 'besides', etc.

Hypothesis: the presence of explicit markers in the analyzed context simplifies the rhetorical relations annotation.

field	N of CA	N of markers
rr-prec	126	46
rr-foll	104	46

The hypothesis for the observed data should be rejected, since the number of markers matches.

General conclusions

1. annotators demonstrate either an almost perfect or substantial agreement, i.e. the annotation guidelines are successfully compiled and do not need to be modified significantly
2. annotators assign the role of tautologies in the rhetorical relation significantly more consistently than the particular rhetorical relations
3. the influence of the number of categories in the annotation scheme on consistency is relevant for our data
4. rhetorical relations with the preceding discourse fragment are annotated in a more consistent way than those with the following fragment

Acknowledgments

The 1st and 2nd sections of our paper were prepared with the support of RSF grant #19-78-10048 "Structures with Lexical Repetitions from the Viewpoint of Contemporary Linguistic Theories".
The 3rd section is supported by St. Petersburg State University, the project #92562973 "Modeling of Russian Megalopolis Citizens' Communicative Behavior in Social, Speech and Pragmatic Aspects Using Artificial Intelligence Methods".

Thanks for your attention!

References

1. Lubensky, S.: Russian–English Dictionary of Idioms. Random House, New York (1995). 2. Mann, W. C., Thompson S.: Rhetorical Structure Theory: Toward a functional theory of text organization. Text 8, 243–281 (1988). 3. Polanyi, L.: A formal model of the structure of discourse. Journal of Pragmatics 12: 601–638 (1988). 4. Zeevat, H.: Rhetorical relations. In: Maienborn C., Von Heusinger K., Portner P. (eds.) Semantics: An International Handbook of Natural Language and Meaning, pp. 946–970. Walter de Gruyter, Berlin (2011). 5. Lascarides, A., Asher, N.: Temporal interpretation, discourse relations and commonsense entailment. Linguistics and Philosophy 16, 437–493 (1993). 6. Asher, N., Lascarides, A.: Logics of conversation: Cambridge University Press, Cam-bridge (2003). 7. Asher, N., Lascarides, A.: Strategic Conversation. Semantics and Pragmatics 6 (2), 1–62 (2013).
8. Jasinskaja, K., Karagjosova E.: Rhetorical Relations. In: Gutzmann D., Matthewson L., Meier C., Rullmann H., Zimmermann T. (eds.) The Blackwell Companion to Semantics. Wiley-Blackwell, Hoboken, N.J. (2021), https://www.researchgate.net/publication/346089248_Rhetorical_Relations, last accessed 06/04/2022.
9. Jasinskaja K., Zeevat H.: Explaining conjunction systems: Russian, English, German. In: Riester A., Solstad T. (eds.) Proceedings of Sinn und Bedeutung, vol. 13, pp. 231–246. University of Stuttgart, Stuttgart (2008).
10. Jasinskaja, K.: Corrective contrast in Russian, in contrast. Oslo Studies in Language 2 (2): 433–466 (2010).
11. Umbach, C.: On the Notion of Contrast in Information Structure and Discourse Structure. Journal of Semantics 21 (2): 155–175 (2004).
12. Wierzbicka, A.: Cross-cultural pragmatics: the semantics of human interaction. Mouton de Gruyter, Berlin and New York (1991).
13. Bulhof, J., Gimbel, S.: Deep tautologies. Pragmatics and Cognition 9 (2): 279-291 (2001).
14. Levinson, S.: Pragmatics. Cambridge University Press, Cambridge (1983).
15. Miki, E.: Evocation and Tautologies. Journal of Pragmatics 25 (5): 635–648 (1996).
16. Snider, T.: Using Tautologies and Contradictions. In Csipak E., Zeijlstra H. (eds.) Proceedings of Sinn und Bedeutung 19, pp. 610–627. LinG, Göttingen (2015).
17. Artstein, R. Inter-annotator Agreement. In: Ide N., Pustejovsky J. (eds) Handbook of Linguistic Annotation. Springer, Dordrecht (2017). 18. Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., Taylor, P., Martin, R., Van Ess-Dykema, C., and Meteer, M.: Dialogue act modeling for automatic tagging and recognition of conversational speech. Computational Linguistics 26(3), 339–374. (2000).

References

19. Ostendorf, M, Price, PJ, Shattuck-Hufnagel, S. The Boston University Radio News Corpus. Boston Univ; Boston, MA: Mar. Tech. Rep. ECS-95-001 (1995).
20. Ghaddar, A., Langlais, P. WikiCoref: An English Coreference-annotated Corpus of Wikipedia Articles. In: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pp. 136–142, Portorož, Slovenia (2016).
21. Bayerl, P. S., Paul, K. I.: What determines inter-coder agreement in manual annotations? A meta-analytic investigation. Computational Linguistics 37(4), 699–725 (2011).
22. Carlson, L., Marcu, D., Okurowski, M. E. Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory. In: van Kuppevelt, J., Smith, R.W. (eds) Current and New Directions in Discourse and Dialogue. Text, Speech and Language Technology, vol 22. Springer, Dordrecht (2003).
23. Meyer, C. M. A Brief Tutorial on Inter-Rater Agreement, `https://dkpro.github.io/dkpro-statistics/inter-rater-agreement-tutorial.pdf`, last accessed 2022/04/06.
24. Gut, U., Bayerl, P. S. Measuring the reliability of manual annotations of speech corpora. In Proceedings of the Speech Prosody, pp. 565–568, Nara, Japan. (2004).
25. Carletta, J.: Assessing Agreement on Classification Tasks: The Kappa Statistic. Computational Linguistics, 22(2), 249–254 (1996).
26. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. Computational Linguistics. 34 (4), 555–596 (2008).
27. Siegel, S, Castellan, NJ. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill (1988).
28. Di Eugenio, B., Glass, M.: The kappa statistic: A second look. Computational Linguistics, 30(1), 95–101 (2004).
29. Artstein, R., Poesio, M. Bias decreases in proportion to the number of annotators, In: Proceedings of the 10th conference on Formal Grammar and the 9th Meeting on Mathematics of Language, pp. 141–150 (2005).
30. Kilem, L. G. irrCAC: Computing Chance-Corrected Agreement Coefficients (CAC). R package version 1.0, `https://CRAN.R-project.org/package=irrCAC`, last accessed 2022/04/06.
31. Klein, D.: Implementing a General Framework for Assessing Interrater Agreement in Stata. The Stata Journal, 18(4), 871–901 (2018).
32. Landis, J. R., Koch, G. G.: The measurement of observer agreement for categorical data. Biometrics 33, 159–174 (1977).