

Bridging Corpus for Russian in comparison with Czech

Anna Roitberg
IMPB of IPM RAS
School of Linguistics HSE RSU
Moscow, Russia
aroytberg@lpm.org.ru

Anna Nedoluzhko
Charles University in Prague
Malostranske nam. 25
CZ-11800 Prague
nedoluzko@ufal.mff.cuni.cz

Abstract

In this paper, we present a syntactic approach to the annotation of bridging relations, so-called *genitive bridging*. We introduce the RuGenBridge corpus for Russian annotated with genitive bridging and compare it to the semantic approach that was applied in the Prague Dependency Treebank for Czech. We discuss some special aspects of bridging resolution for Russian and specifics of bridging annotation for languages where definite nominal groups are not as frequent as e.g. in Romance and Germanic languages. To verify the consistency of our method, we carry out two comparative experiments: the annotation of a small portion of our corpus with bridging relations according to both approaches and finding for all relations from the RuGenBridge their semantic interpretation that would be annotated for Czech.

1 Motivation

Anaphora plays an important role in understanding textual cohesion and coherence. Clark (1975) divides anaphoric relations into two classes, distinguishing direct and indirect anaphora. Direct anaphora (coreference) takes place between language expressions referring to the same discourse entity. In the case of indirect interferences (also called *bridging*), the antecedent is not mentioned but associated with some expression in the previous text. These are, for instance, relations between *two people* – *the woman* and *murdered* – *the murderer* in Clark’s Example (1) and (2) below:

- (1) *I met two people yesterday. The woman told me a story.* (Clark, 1975)
- (2) *John was murdered yesterday. The murderer got away.* (Clark, 1975)

Generally speaking, a *bridging relation* can be understood as an inference about two non-coreferential expressions introduced in a text that are related in some particular way that is not explicitly stated, but this relation contributes essentially to the text coherence. This Clark’s definition of bridging relations, as vague as it is, led to different notions of bridging being used in different approaches. For example, in (Clark, 1975), non-identity semantic relations between entities are classified into three groups: indirect reference by association, indirect reference by characterization and rhetorical relations. For the time being, there is no generally accepted classification of bridging relations. The basic principle accepted in most of the existing approaches is that a list of bridging relations is based on types of semantic relations. Thus, typical examples of bridging are anaphoric relations between entities, which at the same time are e.g. in meronymic relations as represented in Example (3).

- (3) *There were some fruits on the table. John took an apple.*

However, such interpretation sounds very vague. If bridging relations are expected to rely on semantics, we have at least two questions to answer before we begin to apply a systematic classification or annotation of language data. First, we have to decide which kinds of relations we are interested in, how

detailed the classification should be and which relations should be ignored. Second, we have to delimit the boundaries between semantic language-based relations and the relations which are recognized based on the world knowledge or extralinguistic context. Both problems are complicated tasks, the final decision mostly depends on the purpose of the analysis, amount of data and the resources available. To avoid these problems, we decided to choose a syntactic approach to bridging relations, instead of the traditional semantic one. We annotate so-called *genitive bridging*: the case where two elements (an anchor/antecedent and a bridging element/anaphor) can form a genitive construction, where the anchor is marked with the genitive case in Russian. In Example (4), the anchor is *dom* [house], the bridging element is *stenah* [walls], and the genitive construction that can be formed is *stenah doma_Gen* [the walls of the house].

- (4) *U nego byl milyj dom s plyuř'om na stenah* (doma). [He has a nice house with ivy climbing the walls (of the house).]

We believe that this approach will improve the consistency of the annotated data and will allow us to create a more reliable corpus for the prepared computational experiments.

This work describes an ongoing project, with the data annotated within the new syntactic approach (RuGenBrigde corpus). To the best of our knowledge, this approach has not been applied to any large-scale data annotation yet, so we do not have any corroboration of its reliability. To prove the advantages of our approach, it is necessary to (i) provide the empirical verification of the quality of our annotation scheme through double annotation and measuring the inter-annotator agreement and (ii) to compare our annotations to other bridging annotation approaches. This paper addresses the second task. We decided to compare our annotations to bridging relations annotation in the Prague Dependency Treebank (PDT, Bejček et al., (2013)). There are several reasons for this choice:

- PDT is one of a restricted number of corpora with a large-scale annotation of bridging relations;

- The texts in PDT are in Czech, which is a Slavic language with many structural (grammatical and syntactic) similarities, e.g. it has the similar declination system, so the genitive bridgings are expected in the same way as in Russian; like Russian, Czech lacks the grammatical category of definiteness;
- The bridging annotation approach used in PDT is claimed to be purely semantic (Zikánová et al., 2015), thus the comparison is especially interesting;
- The number and the types of bridging relations applied in PDT is an average compared to state-of-the-art bridging approaches applied.

The paper is structured as follows: after observing the related work in 1.1, we present the RuGenBrigde corpus for Russian (Section 2) and bridging annotation in Czech (Section 3); we compare the annotation schemes in Section 4. Further, in Section 5, we describe two experiments that have been carried out on the Russian texts: (i) the application of Czech and Russian annotation schemes on the same texts and (ii) the annotation of all pairs from RuGenBrigde corpus with possible PDT bridging relations marks. We discuss the results in Section 6.

1.1 Related work

There are two main annotation approaches. The first (and more popular) is based on semantic constraints on bridging relations. This approach is close to Clark's reference by association. Such bridging interpretation is used in the studies of Asher and Lascarides (1998), applying the segmented discourse representation theory to bridging relations, corpus annotations by M. Poesio for English and Italian (cf. Poesio (2000), Poesio et al. (2004); Poesio and Arstein (2008)), M. Recasens (Recasens et al., 2007) for Spanish and Catalan, Zikánová et al. (2015) for Czech; the semantic approach is also used in Lungen (2008), Gardent (2003) and so on. The typical relations of the semantic approach are meronymic part-whole and set-subset relations, co-hyponymy (*mother – father* as family members), relations of belonging (e.g. a person and his/her clothes), relations between the situation and its par-

ticipants (*murder – murderer*), some symptomatic relations (*fever – illness*) and so on.

Alternatively, there are a few corpora, where there are no strict semantic constraints on bridging relations, and all types of "associative" relations between nominal groups are taken into account. This approach is realized e.g. in (Hou et al., 2013).

It should also be noted that usually the term bridging relation is used for definite nominal groups, see e.g. (Löbner, 1998) or (Poesio and Artstein, 2008). However, the same kind of implicit anaphoric linking is also possible with indefinite or quantifying or even generic nominal groups, cf. distribution statistics in (Hou et al., 2013). For instance, in Example (5), a bridging relation can be observed between the Czech generic nominal group *nový VW Golf* [*the new VW Golf*]¹ and an indefinite nominal group *jedním novým golfem* [*one of the new Golfs*] (one arbitrary car of this category).

- (5) *Nový VW Golf je vybaven motorem o síle 110 kW... Dostali jsme možnost se jedním novým golfem projet.* (PDT, cit. from (Zikánová et al., 2015) [*The new VW Golf is equipped with an engine power 110 kW... We had an opportunity to ride in one of the new Golfs.*]

2 Annotation of bridging relations in Russian

Here, we present a new corpus RuGenBrigde, the first corpus annotated with bridging relations for Russian. We develop this corpus for training and testing automatic bridging detection and resolution systems. In the present stage of the project, RuGenBrigde consists of 207 news texts² (35,841 tokens), most texts contain 100 - 250 words each. The corpus was annotated with automatic part of speech tagging by FreeLing.³ The bridging cases were annotated manually using BRAT annotation tool⁴.

2.1 Bridging in genitive constructions

Unlike most approaches defining bridging relations in terms of semantic and pragmatic categories, in the

Russian corpus, we use rather syntactic than semantic criteria. We focus mainly on the cases of bridging in genitive construction, so-called, *genitive bridging*. This is the case where the dependent nominal group of the construction is marked with the genitive case in Russian, the head NP has no case restrictions. For instance, in Example (6), there is a genitive bridging relation between *voditel'* [*driver*] and *avtobus* [*bus*], because it is understood as *voditel' avtobusa_Gen* [*the driver of the bus*].

- (6) *V avtobuse nachalsya pozhar. Voditel' (avtobusa) sam potušil ogon'.* [*The fire broke out in the bus. The driver (of the bus) put out the fire by himself.*]

In fact, we capture bridging relations in genitive constructions if an anaphor of bridging pair may have a dependent NP in genitive case, but it is mostly not expressed in the sentence because the potential dependent NP was used recently earlier in the text and it is still actualized in the mind of the reader. For example, by *driver* in Example (6), an addressee can easily infer that the driver of *the bus* mentioned in the previous sentence is meant.

The most typical semantics of Russian genitive constructions is the 'part-whole' relation in a broad sense, where the whole is marked by genitive case (*glaza ubijcy_Gen* [*eyes of the murderer*]). Other frequent cases are expressions, where a head is a deverbal noun with a genitive participant *vybory prezidenta_Gen* [*elections of the president*], measure nouns *barel' nefi_Gen* [*barrel of oil*] etc.

2.2 Annotation scheme for Russian

Nouns or nominal groups are subjects to annotation. We adhere to the principle of the minimum possible markable: if possible, annotators tag a bare noun, the whole noun phrase is annotated only in the case when it is the minimum possible name of the entity. Thus, in *my beautiful dog*, the markable *dog* is annotated, but in *The Ministry of Justice*, the whole phrase is annotated as a markable because all the words compose the name of the organization.

In RuGenBridge, the following types of bridging relations are annotated:

1. Bridging relations in genitive constructions (type BRIDGE). See Examples (4) and (6)

¹Golf is a type of car made by Volkswagen.

²News from www.polit.ru site

³<http://nlp.lsi.upc.edu/freeling/>

⁴<http://brat.nlplab.org>

above. There are 362 cases of type BRIDGE in our corpus.

2. We also annotate some cases which are very close to the genitive bridging, but genitive construction in Russian is not possible there, for purely syntactic reasons. We use NON-GEN mark for such pairs, see the relation between *Russian Federation* and *Syberia* in Example (7). This type is especially common with the named entities. There are just 8 cases of NON-GEN in our corpus.

- (7) *Pravitel'stvo Rossijskoj Federacii vneslo na rassmotrenie (...) Etot proekt takze sposobstvuet razvitiyu Sibiri. [The Government of the Russian Federation brings a bill (...) This bill also promotes the development of Siberia.]*

It should be noted, that our annotation scheme is oriented on the language properties of Russian, primarily on the properties of Russian genitive constructions. The ability to form a genitive construction is a very important criterion for the annotators by marking bridging relations in the corpus; in fact, they are guided by it. So, in Example (8) below⁵ we annotate the bridging relation *bag – mum*, because *sumka mamy_Gen [mum's bag]* is grammatical in Russian; the link *bag – supermarket* would not be annotated, because **sumka supermarketa_Gen* [**supermarket's bag*] is ungrammatical.

- (8) *The mum came from a supermarket and got lost in her Facebook. The bag is still in the doorway.*

The statistics of the annotated types in RuGenBrigde is presented in Table 2 in Section 5.2, together with the results of comparison experiments.

Apart from the annotation of bridging relations, three most frequent types of annotated NPs are manually marked with special labels in our corpus: (i) GEO (157 cases) for all geographic names (*Moscow, Atlantic Ocean, Thailand* etc.), (ii) ORG (35 cases) for official organizations, both proper and current names (*ministry, policy, LifeNews* etc.) and (iii) POST (22 cases) for political positions (*pres-*

ident, deputy etc.). The ORG mark in RuGen-Bridge is comparable to the NORP (Nationality, Organizations and Political organizations) category in OntoNotes (Stoyanov et al., 2011), but we do not include Nationalities, e.g. we do not mark *Swedes* as ORG in Example (9).

- (9) *Swedes usually drink coffee in the morning.*

3 Annotation of bridging relations in Czech

Bridging relations in Czech are annotated on the Prague Dependency Treebank (PDT). This is a large-scale annotation on ca. 50000 sentences of news texts. Apart from bridging, other textual phenomena (syntactic structure, ellipsis, coreference, discourse relations, information structure, etc.) are annotated, see (Poláková et al., 2013). The classification of bridging relations in PDT is based on semantic and pragmatic principles. The annotation preserves distinctions between the following groups: (1) meronymy relations between a part and a whole (subtypes PART-WHOLE and WHOLE-PART, as e.g. *in face – eyes*), (2) the relation between a set and its subsets or elements of the set (subtypes SUBSET-SET and SET-SUBSET, as in *a group of students – some students – a student*), (3) the relation between an entity and a singular function on this entity (subtypes P-FUNCT and FUNCT-P, as in *company – director*) (4) the relation between coherence-relevant discourse opposites (type CONTRAST, as in *black flags – white flags*), (5) non-coreferential explicit anaphoric relation (type ANAPH, as in *first world war – at that time*) and (6) further underspecified group REST consisting of six other bridging subtypes (e.g. relations between family members, event – argument, locality – inhabitant, etc.).

Unlike in RuGenBrigde, bridging relations in PDT connect not only the individual nominal groups but the whole coreference chains. Thus, once postulating a bridging relation between two elements of different coreference chains, it should not be marked again for coreferential expressions later in text. Another significant distinction is the principle of the maximum possible markables (all dependency subtrees of antecedent and bridging elements are considered to be markables).

⁵We thank our reviewer for this example.

4 Comparison of annotation schemes

The difference of annotation schemes is immediately related to the scope and nature of the corpora. Our goal here is not to compare the corpora: it is useless to compare a big and richly annotated corpus with a small and a focused one, which is still in the early stage of its development. Thus, the comparison concerns only the relevant points.

4.1 Characteristics of markables

The first relevant point concerns properties of markables and the scope. The scope is different: RuGenBrigde chooses the minimum and PDT the maximum scope of the markables. On the other hand, both approaches consider some referential adjectives as markables, first of all those which are derived from locational nouns (e.g. *USA – American*). Only referential and abstract nouns can be annotated in the Czech corpus, non-referential nouns are not concerned. For example, such nouns as measures, points etc. are considered to be non-referential, so bridging relations can not be marked in pairs like *barrel – oil* or *point – share price*. Opposite to this, in the Russian corpus, both referential and non-referential nouns can take part in bridging relations. So, in examples like (10) below, bridging relations will be marked in Russian corpus and will not be marked in Czech corpus.

- (10) *Oil futures contracts rose by 1.79% and settled at \$45.54 per barrel (of oil) on Friday.*

4.2 Inventory of relations

Bridging annotation in the Czech corpus is a part of discourse level annotation, the semantics of relations was taken into account (Nedoluzhko and Mírovský, 2011) and the corpus is meant to be multi-purpose. The Russian corpus is primarily aimed to create training and testing data for an automatic resolution system. For this reason, semantic classification of relations is not so important. Nevertheless, syntactic constraints inevitably produce some semantic constraints. For instance, the Russian genitive construction is typically used for marking possessive relations (broadly defined). The most common examples of this construction include: *sumka mamy [the mom's bag]*, *stena doma [the wall of the house]*, *hvorost kota [cat's tail]* and so on. On the other hand,

examining the list of the most frequent genitive construction examples in Russian National Corpus⁶, we can observe three groups of non-possessive cases: (i) the first group consists of expressions with nouns derived from verbs: *uvol'nenie nachal'nika [termination of the boss]*, *pohorony aktera [funeral of an actor]*; (ii) the second group contains expressions with measure words, e.g. *liter, kilogram*; and (iii) the third group represents mostly government positions, such as *ministr inostrannyh del [foreign secretary]*, lit. *minister of foreign affairs*.

Hence, we cannot say that the genitive constraint is identical to the possessive constraint: there is a finite list of semantic relations between the anchor and the bridging element. Moreover, this set of possible semantic relations seems to be comparable to some types of Czech bridging relations. The bridging anaphora of types PART-WHOLE (WHOLE-PART) and SET-SUB (SUB-SET) are often the cases of general possession, and FUNCT-P (P-FUNCT) are often the cases of government positions.

5 The experiments - application of the PDT and RuGenBrigde schemes

Starting the annotation of Russian corpus, we supposed that elements of annotated pairs will form semantic relational classes similar to those annotated within semantic approaches to bridging relations as the result. So, we expected to catch such cases as part-whole (*krysha doma_Gen [roof of the house]*) or possessive (*sumka mamy_Gen [mom's bag]*) relations. With such a result, systematic syntactic approach could reflect the semantic aspect of bridging relations. To test this hypothesis, we decided to apply a semantically oriented annotation scheme to the Russian texts. For the reasons stated in Section 1, we have chosen the PDT annotation scheme. In what follows, we describe two experiments in application of the PDT scheme for Russian.

5.1 Experiment 1: application of PDT and RuGenBrigde schemes for a subset of RuGenBrigde

In the first experiment, we have annotated 8 documents in Russian with the PDT and RuGenBrigde

⁶<http://www.ruscorpora.ru>

schemes in parallel with two annotators. One annotator used the PDT semantic approach and another annotator used the syntactic approach of the RuGenBrigde corpus. Contrary to the expectative closeness of semantics between the relation sets, there is a very low coincidence between the annotated pairs. The results are shown in Table 1.

Czech annotation scheme		Russian annotation scheme	
TOTAL	69	TOTAL	22
CONTRAST	6	BRIDGE	22
FUNCT-P	3		
P-FUNCT	11		
PART-WHOLE	3		
SET-SUB	5		
SUB-SET	13		
WHOLE-PART	18		
REST	10		

Table 1: Comparison of Russian and Czech annotation schemes on 8 documents from RuGenBrigde

We have 69 bridging pairs with the PDT annotation scheme and only 22 with the Russian one. Furthermore, there are only 7 coincidence cases where anchor and bridging element of the pair are the same, notably that 3 (of the 7) cases belong to one sentence (man and his body parts).

One of the reasons for such difference is that genitive bridging in Russian corpus is allowed in only one direction, where the bridging element (to which the genitive form of the anchor can be potentially added) follows the anchor in text. In Czech, both directions (e.g. PART-WHOLE and WHOLE-PART) are possible. The second reason is that relations between proper names (e.g. *Washington – USA*) are allowed in the PDT scheme and are very seldom in RuGenBrigde. There is only one class of these relations: names of regions may be linked with name of countries, as in case of *Moscow region – Russia* described in Section 2.2 above).

The results of the experiment evidence that the semantic approach is more broad-based than our genitive syntactic approach. However, we believe that syntactic approach could be more systematic and clear for annotation, thus presenting a more reliable data for automatic resolution systems. To test this statement we are intending to annotate our data with more annotators in the near future. On the other

hand, this experiment displayed that syntactic approach brights out specific bridging relation types, which are not identified in the semantic approach. For this reason, we decided to conduct another experiment and to examine the cases, which seem to be difficult to catch while applying a semantic annotation scheme.

5.2 Experiment 2: application of the PDT types to all possible RuGenBrigde genitive bridgings

This experiment is aimed to find out which semantic relations are more frequent among the cases of genitive bridging. As another task, we want to distinguish and classify the cases that are not overlapped by the set of PDT semantic relations.

In this experiment, we checked out all cases of genitive bridging in the RuGenBrigde corpus, and for all pairs where it was possible, we added the relations that would be annotated within the PDT annotation framework. As the result, for 430 bridging pairs annotated with genitive bridging (types BRIDGE, COREF-BRIDGE or NON-GEN), we have 165 pairs annotated with the PDT tags and 265 pairs remained without the PDT tags. Table 2 shows the numerical results of the experiment.

Czech annotation marks		Russian annotation marks	
AllPDT	152	AllRuGenBridge	370
PART-WHOLE	73	BRIDGE	362
P-FUNCT	55	NON-GEN	8
SET-SUBSET	8		
REST	16		

Table 2: All relation marks for Russian RuGenBridge pairs

As shown in Table 2, PART-WHOLE (*house – roof*) and FUNCT-P (*Russia – prime-minister*) are significantly more frequent than other PDT relations. 218 pairs which were not annotated with PDT marks can be further sub-classified into the following groups:

1. Anchors are geographical names, bridging elements (56 pairs) can be further divided into two subclasses:
 - (a) something is located in this geographic object (*Moscow – hospitals*), or

- (b) something is concerned with this geographic object (*Russia – budget*).
2. Among the rest 162 pairs, we detected the following types:
- (a) object – its possessor (*flat – landlord*),
 - (b) object – something belonging to this object, but not the part of the object (*aerodrome – airplane*),
 - (c) expressions with the names of measures (*oil – barrel*),
 - (d) collocations, mostly deverbative nouns (*rates – increase*).

The measure group (2c) reflects the Russian language-specific feature: measure words require genitive dependents, so this bridging relation can be really considered as purely syntactic. Most of examples in the last group (2d) are of syntactic nature (more detail in Section 6).

6 Discussion

Let us now analyze some characteristics of bridging relations which make the output of the syntactic annotation approach. Looking at the semantics of expressions taking part in the bridging relations in ReGenBridge, we can see that there is a significant number of antecedents referring to geographic names: among the total of 370 cases, 135 antecedents (36,5% of all bridging relations in the corpus) are marked with the GEO label. These are mostly names of the countries and the relations can be often interpreted as part-whole bridging relations in the Prague annotation scheme (e.g. country – part of this country, region, etc. make up 41 cases, or 31% of GEO antecedents). Another frequent correlation between the GEO antecedents in the PDT bridging types is the type FUNCT, these are often relations between the name of the country and some unique function on it, e.g. *USA - ministry of foreign affairs* (34 cases, or 25% of GEO antecedents in our corpus). We note that the relations where GEO labels in genitive bridging annotation correspond to PART-WHOLE and FUNCT-P in PDT primarily reflect the world knowledge. This speaks against the purity of the Prague semantic annotation. On the

other hand, as mentioned above, the borderline between the world knowledge and semantics is quite fuzzy.

The remainder cases of the GEO antecedents (59 cases) could not be annotated with any of the PDT bridging types.⁷ Looking at these pairs in more detail, we can see that the anaphoricity between the entities is not given by a semantic relation, but rather by a textual structure and referent activation practices. Typically for the news genre, events are located in a specific place that is introduced once and remains activated for the whole description (e.g. once introducing *Russia*, we speak about budget, hospitals and schools there without repetitions like *hospitals of Russia*, *schools of Russia* and so on). In this way, redundant repetitions are avoided, and this also speaks for the discourse origin of bridging anaphora.

Considering genitive bridgings which did not find any semantic interpretation within the PDT annotation scheme (218 out of 370 cases), we notice that anaphoricity of many pairs is given by situational relationships within the texts. These relations do not have semantic nature, so they can be hardly included in any dictionaries of ontologies. On the other hand, they are not purely pragmatic. They are text-given: being introduced at the beginning, they are further used as known and accepted. In this respect, the general phenomenon of bridging relations (and our genitive bridging is a subset of them) can be considered to be deictic and may be related to the category of definiteness or contextual boundness of expressions in text, where the variables are also introduced and further used in text as known.

An additional interesting point that is given by the comparison of genitive bridging approach to the PDT annotation scheme is that it gives the possibility to test the consistency of the PDT annotation. In some cases, we found problematic the borderlines between SET-SUBSET and FUNCT bridging types (cf. the problematic point of uniqueness in *parliament – deputies* (no annotation to parliament) *parliament – premier* (P-FUNCT to *parliament*, because premier in the parliament is unique)). Also, we met a number of cases which were not annotated

⁷These are mostly the cases of multiple objects located in a place marked with a GEO antecedent, e.g. *Russia – schools*, *banks*, *hospitals*, *parks*, etc.

in PDT, however they could be interpreted in terms of PDT semantic relations, for example the pair *defendant – criminal case* was not annotated although it can be considered as ‘event – argument’ and annotated within the REST subtype. Additionally, we found that the pairs ‘a geographic name – something located there’ are very common, but such cases were not included in the PDT annotation scheme.

7 Conclusion

In this paper, we introduced the syntactic approach to bridging annotation and presented some preliminary investigations on its semantic interpretation. The comparison has shown that genitive bridging provides an opportunity to find out new functional types of bridging relations with respect to textual structure. We believe that this approach is more consistent than semantic annotation of bridging, because it is based on formal criteria and it does not require fixing a borderline between semantics and the world knowledge. However, the paper presents the ongoing research which is in the first half of its development. Our immediate goals for the future work are (i) to annotate the existing corpus with two annotators and a supervisor, and to measure the inter-annotator agreement, (ii) to extend the corpus and analyze bridging cases attested and (iii) to develop a system for genitive bridging resolution based on the information in the corpus.

Acknowledgments

The study was supported by the Russian Foundation for Basic research (grant No. 15-07-09306), the Grant Agency of the Czech Republic (grant 16-05394S) and Ministry of Education, Youth and Sports (grant LH14011).

References

Nicholas Asher and Alex Lascarides. 1998. Bridging. *Journal of Semantics*, 15(1):83–113.

Eduard Bejček, Eva Hajičová, Jan Hajič, Pavlína Jínová, Václava Kettnerová, Veronika Kolářová, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Jarmila Panevová, Lucie Poláková, Magda Ševčíková, Jan Štěpánek, and Šárka Zikánová. 2013. Prague Dependency Treebank 3.0. Data/software. [cit. 2015_07_22].

Herbert H Clark. 1975. Bridging. In *Proceedings of the 1975 workshop on Theoretical issues in natural language processing*, pages 169–174. Association for Computational Linguistics.

Claire Gardent, H  l  ne Manu  lian, and Eric Kow. 2003. Which bridges for bridging definite descriptions. In *Proceedings of the EACL 2003 Workshop on Linguistically Interpreted Corpora*, pages 69–76.

Yufang Hou, Katja Markert, and Michael Strube. 2013. Global inference for bridging anaphora resolution. In *HLT-NAACL*, pages 907–917.

Sebastian L  bner. 1998. Definite associative anaphora. *manuscript*) <http://user.phil-fak.uni-duesseldorf.de/~loebner/publ/DAA-03.pdf>.

Harald Lungen. 2008. RRSet-Taxonomy of rhetorical relations in SemDok. *Interne Reports der DFG-Forschergruppe*, 437.

Anna Nedoluzhko and Jiří Mírovský. 2011. Annotating extended textual coreference and bridging relations in the Prague Dependency Treebank. *Annotation manual. Technical report*, (44).

Massimo Poesio and Ron Artstein. 2008. Anaphoric Annotation in the ARRAU Corpus. In *LREC*.

Massimo Poesio, Rahul Mehta, Axel Maroudas, and Janet Hitzeman. 2004. Learning to resolve bridging references. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, page 143. Association for Computational Linguistics.

Massimo Poesio. 2000. Annotating a corpus to develop and evaluate discourse entity realization algorithms: Issues and preliminary results. In *LREC*.

Lucie Poláková, Jiří Mírovský, Anna Nedoluzhko, Pavlína Jínová, Sárka Zikánová, and Eva Hajičová. 2013. Introducing the Prague Discourse Treebank 1.0. In *IJCNLP*, pages 91–99.

Marta Recasens, M Antonia Martı, and Mariona Taul  . 2007. Text as scene: Discourse deixis and bridging relations. *Procesamiento del lenguaje natural*, 39:205–212.

Veselin Stoyanov, Uday Babbar, Pracheer Gupta, and Claire Cardie. 2011. Reconciling ontonotes: Unrestricted coreference resolution in ontonotes with reconcile. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pages 122–126. Association for Computational Linguistics.

Šárka Zikánová, Eva Hajičová, Barbora Hladká, Pavlína Jínová, Jiří Mírovský, Anna Nedoluzhko, Lucie Poláková, Kateřina Rysová, Magdal  na Rysová, and Jan V  cl. 2015. *Discourse and Coherence. From the Sentence Structure to Relations in Text*, volume 14 of *Studies in Computational and Theoretical Linguistics*. Charles University in Prague, Praha, Czechia.