# Bridging Resolution: A Survey of the State of the Art

**Hideo Kobayashi** and **Vincent Ng**
Human Language Technology Research Institute
University of Texas at Dallas
Richardson, TX 75083-0688
`{hideo,vince}@hlt.utdallas.edu`

## Abstract

Bridging reference resolution is an anaphora resolution task that is arguably more challenging and less studied than entity coreference resolution. Given that significant progress has been made on coreference resolution in recent years, we believe that bridging resolution will receive increasing attention in the NLP community. Nevertheless, progress on bridging resolution is currently hampered in part by the scarcity of large annotated corpora for model training as well as the lack of standardized evaluation protocols. This paper presents a survey of the current state of research on bridging reference resolution and discusses future research directions.

## 1 Introduction

Bridging resolution is an anaphora resolution task that involves identifying and resolving bridging/associative anaphors, which are anaphoric references to non-identical associated antecedents. To better understand the difficulty of the task, consider the following sentences:

> Even if *baseball* triggers losses at CBS – and he doesn't think it will – "I'd rather see **the games** on our air than on NBC and ABC," he says.

In this example, a bridging link exists between the anaphor *the games* and its antecedent *baseball*, as the definite description cannot be interpreted correctly unless it is associated with *baseball*.

Bridging anaphora resolution is arguably more difficult than entity coreference resolution, the task of determining which entity mentions in a text refer to the same real-world entity. For entity coreference resolution, there are well-defined linguistic constraints at the grammatical (e.g., gender and number agreement), syntactic (e.g., binding theory), semantic (e.g., semantic class agreement), and discourse (e.g., centering) levels. Oftentimes, the antecedent of an anaphor can be identified by comparing its lexical similarity with the anaphor. In contrast, there are typically no clear syntactic or other surface clues for identifying the antecedent of a bridging anaphor. It is not uncommon that resolution requires the use of context as well as commonsense inference. Furthermore, while antecedents in entity coreference are noun phrases (NPs), antecedents in bridging can also be non-NPs such as verb phrases (VPs) or clauses, which considerably increase the possible number of candidate antecedents for each anaphor.

Bridging resolution is comparatively less studied than entity coreference resolution. Progress on bridging resolution is currently hampered in part by the scarcity of large annotated corpora for model training as well as the lack of standardized evaluation protocols. More specifically, while there are a few bridging corpora that are used more extensively than the others for evaluation purposes, many of which do not have standard train-test partitions. Moreover, these corpora were annotated with somewhat different definitions of bridging, so good performance on one corpus does not necessarily translate to good performance on another. Worse still, resolvers were evaluated under different settings. For instance, different researchers employ different, sometimes undocumented strategies for filtering bridging anaphors and candidate antecedents, while others employ gold annotations (e.g., syntactic parses, coreference) for feature computation. Above all, many of the implementations of bridging resolvers have not been

| Corpora | Domain Type | Size | | | Number of anaphors | Anaphor type | Antecedent type |
|---|---|---|---|---|---|---|---|
| | | Docs | Tokens | Mentions | | | |
| ISNotes | WSJ news | 50 | 40292 | 11272 | 663 | All NPs | entity, event |
| BASHI | WSJ news | 50 | 57709 | 18561 | 459 | All NPs | entity, event |
| ARRAU RST | news | 413 | 228901 | 72013 | 3777 | All NPs | entity |
| ARRAU GNOME | medical, art history | 5 | 21458 | 6562 | 692 | All NPs | entity |
| ARRAU PEAR | spoken narratives | 20 | 14059 | 4008 | 333 | All NPs | entity |
| ARRAU TRAINS | dialogues | 114 | 83654 | 16999 | 710 | All NPs | entity |
| SciCorp | scientific text | 14 | 61045 | 9407 | 1366 | Definite NPs | entity |

Table 1: Comparison of commonly used English corpora for bridging resolution.

made publicly available. The lack of a standard evaluation protocol has made it somewhat difficult to track research progress on this task. To some extent, this is reminiscent of the state of affairs with entity coreference research prior to the CoNLL 2011 and 2012 shared tasks on entity coreference resolution.

As significant advances have been made on entity coreference resolution, we believe that bridging reference resolution will gain increasing attention in the years to come. Our goal in this paper is to provide a timely survey of the current state of research on bridging anaphora resolution.

## 2 Task Definition: Some Historical Perspectives

The definition of bridging has evolved over the years, particularly with respect to (1) the types of relations bridging should cover, and (2) the types of linguistic expressions that can serve as bridging anaphors. In this section, we take a closer look at these two issues.

First, *what types of relations should bridging cover*? As a linguistic phenomenon, bridging has been studied extensively by linguists (e.g., Clark (1975), Prince (1981), Gundel et al. (1993)). Clark (1975), who started this area of research, introduced a broad concept of bridging that includes coreference (i.e., the identity relation). Coreference, however, is gradually being excluded from bridging over time. For instance, while some early studies still included the difficult cases of coreference where two coreferent mentions do not share the same head as bridging (Poesio and Vieira, 1998; Vieira and Poesio, 2000; Bunescu, 2003)[1], most of the recent studies focus on non-identity cases of bridging, which is the closest to Hawkins's (1978) concept of associative anaphora. Among the non-identity relations, bridging covers various types of semantic relations. While early studies typically restrict themselves to predefined relations such as part-of, subset, set membership, and possession relations (Poesio and Vieira, 1998; Poesio et al., 2004b), recent studies claim that bridging is a diverse phenomenon that cannot be simply captured with a limited set of predefined relations (Markert et al., 2012; Rösiger, 2018a).

Second, *what types of linguistic expressions can serve as bridging anaphors*? Many traditional studies (Hawkins, 1978; Poesio and Vieira, 1998; Lassalle and Denis, 2011; Rösiger, 2016) limited bridging anaphors to definite expressions, excluding indefinite expressions since they generally introduce new information that can be interpreted without the discourse context. However, Löbner (1998) claimed that bridging anaphors can also be indefinite because these indefinite expressions can have semantic relations with preceding expressions. Recent studies therefore allow both definite and indefinite expressions to serve as bridging anaphors (Poesio and Artstein, 2008; Markert et al., 2012; Rösiger, 2018a).

## 3 Corpora

This section provides an overview of existing corpora used for bridging research, with a focus on four widely-used English corpora, namely ISNotes (composed of 50 WSJ articles in OntoNotes) (Markert et al., 2012) , BASHI (The Bridging Anaphors Hand-annotated Inventory, composed of another 50 WSJ articles in OntoNotes) (Rösiger, 2018a), ARRAU (composed of articles from four domains, RST, GNOME, PEAR, and TRAINS) (Poesio and Artstein, 2008; Uryupina et al., 2020), and SciCorp (The Scientific Corpus, composed of scientific articles from computational linguistics and genetics) (Rösiger, 2016).

Table 1 compares these corpora along five dimensions: (1) the domain type, (2) the size (in terms of the number of documents, tokens, and mentions), (3) the number of bridging anaphors, (4) the types of

---

[1]These difficult cases of coreference are later being referred to as *lenient bridging* (Hou et al., 2018).

anaphor, and (5) the types of antecedent. While early corpora limited anaphors to definite NPs and pre-defined relations (Poesio and Vieira, 1998; Poesio et al., 2004b), many of these newer corpora do not. For instance, ISNotes and BASHI include both definite and indefinite expressions as anaphors and both entity and event mentions as antecedents; moreover, they do not restrict bridging relations to predefined relations. Also, all of these corpora contain coreference in addition to bridging annotations. In addition to the differences shown in Table 1, there are several notable differences among these corpora:

**Referential vs. lexical bridging.** Rösiger et al. (2018b) introduced the notions of referential bridging and lexical bridging as a way to explain a key difference between ARRAU and the other corpora. Referential bridging refers to the cases in which the bridging anaphor cannot be interpreted without the antecedent (e.g., *the window* in *Tim walked into the room. The window was broken*), whereas lexical bridging refers to the cases where the reference can be interpreted independently of the antecedent (e.g., *Tokyo* in *The capital of Japan is Tokyo*). While ISNotes, BASHI and SciCorp are composed of referential bridging references, ARRAU contains both referential and lexical bridging references, with lexical bridging references being the majority.

**Information status.** Information status (IS), a linguistic notion that is related to bridging, describes the extent to which a discourse entity is available to the hearer/reader. At a coarse level, a discourse entity's IS is (1) OLD to the hearer if it is known to the hearer and has previously been referred to; (2) NEW if it is unknown to her and has not been previously referred to; and (3) MEDIATED if it is newly mentioned but the hearer can infer its identity from a previously-mentioned entity or world knowledge. By definition, bridging is a subcategory of MEDIATED. While BASHI does not contain IS annotations, ISNotes has eight IS classes ("new", "old", and six subclasses of mediated (one of them is bridging)), ARRAU has three ("new", "old", and non-referring), and SciCorp has eight (one of them is bridging).

**Predefined relations.** Some corpora provide the semantic relation type of each bridging link. In IS-Notes, a link is labeled with one of the following relation types: part-of/attribute-of, set, and other (including encyclopedic and frame relations). The RST domain of ARRAU also has annotations of pre-defined relations, which include possessive, subset, element, comparative (labeled as "other"), everything else (labeled as "underspecified"), as well as the inverse of each of these relation types.

**Comparative anaphora.** A comparative anaphor is a non-identity anaphor that is compared to another mention (Modjeska, 2003). In ISNotes, comparative anaphors are excluded from the bridging category because such anaphors often have surface indicators, containing modifiers such as "other" and "another" (Markert et al., 2012). In contrast, BASHI and ARRAU consider them as a subcategory of bridging.

Several non-English bridging corpora exist. DIRNDL (Björkelund et al., 2014), SemDok (Bärenfänger et al., 2008), and GRAIN (Schweitzer et al., 2018) are in German. DEDE (Gardent and Manuélian, 2005) and PAROLE (Gardent et al., 2003) are in French. Caselli/Prodanof (Caselli and Prodanof, 2006) and Italian Live Memories Corpus (Rodríguez et al., 2010) are in Italian. Prague Dependency Treebank (Hajič et al., 2018) is in Czech. CESS-ECE (Recasens et al., 2007) is in Spanish. COREA (Hendrickx et al., 2008) is in Dutch. The Sasano-Kurohashi corpus (Sasano and Kurohashi, 2009) is in Japanese. RuGenBridge (Roitberg and Nedoluzhko, 2016) is in Russian.

Parallel bridging corpora are also available. For instance, Copenhagen Dependency Treebank is a parallel corpus involving Danish, English, Italian, German, and Spanish (Korzen and Buch-kromann, 2011), and CorefPro is a parallel corpus involving German, English, and Russian (Grishina, 2016).

While not widely used, GUM is an ever-expanding English corpus annotated with bridging links by students at Georgetown University (Zeldes, 2017).

## 4 Evaluation Issues

As mentioned before, an issue surrounding bridging resolution research concerns the lack of a standard-ized evaluation protocol. In this section, we take a look at current evaluation practices.

**Evaluation settings.** Bridging resolvers operate in one of three settings.

In *end-to-end bridging resolution*, a system is given a raw document as input. The goal is to iden-tify the bridging anaphors (a subtask known as bridging *recognition*) and resolve each of them to its

antecedent. Since this setting is considered very challenging, none of the existing bridging resolvers are evaluated in an end-to-end fashion. In *full bridging resolution*, a system is given as input not only a document but also the *gold* (i.e., hand-annotated) mentions in the document. The goal is to identify the subset of the gold mentions that are bridging anaphors and resolve them to their antecedents, which are also chosen from the gold mentions. In principle, gold mentions are mentions that can participate in a bridging relation. In practice, gold mentions are typically much smaller than the set of possible mentions. Since full bridging resolution constrains the selection of anaphors and antecedents to those that are gold mentions, it is less challenging than the end-to-end setting. Gold mentions are defined slightly differently in different corpora. In ISNotes, gold mentions include NPs, possessive nouns/pronouns, premodifiers, and verbs. In BASHI, gold mentions are assumed to be all and only those NPs that can be extracted from gold parse trees. In ARRAU, gold mentions include all NPs, possessive pronouns, and a subset of premodifiers. In SciCorp, gold mentions are definite NPs, which include definite descriptions, named entities, and pronouns. Finally, in *bridging resolution*, a system is given as input not only a document and the *gold* mentions it contains, but also the gold anaphors. The goal is to resolve each gold anaphor to its antecedent, which is chosen from the given set of gold mentions. This setting is the least challenging of the three, as it focuses solely on resolution and does not require bridging anaphors to be identified.

**Evaluation metrics.** For full bridging resolution, results are reported for both recognition and resolution in terms of precision, recall, and F-score. For recognition, recall is defined as the fraction of gold anaphors that are correctly identified, whereas precision is defined as the fraction of anaphors identified by the system that are correct. For resolution, recall and precision can be defined in a similar fashion. For bridging resolution, since gold anaphors are given, results are reported in terms of resolution accuracy, which is the fraction of gold anaphors that are correctly resolved.

**Entity- vs. mention-based evaluation.** A resolver needs to resolve an anaphor to an antecedent chosen from a set of *candidate* antecedents. For (full) bridging resolution, the candidate antecedents can simply be taken to be the set of gold mentions that appear in all of the previous sentences or a fixed sentence window (Poesio et al., 2004a). Slightly more sophisticated candidate selection strategies have been employed. For instance, the window size can be tuned for each rule in rule-based systems (Hou et al., 2014; Rösiger, 2018b; Rösiger et al., 2018a; Rösiger et al., 2018b). The top $k$ salient mentions can be used in addition to those from the fixed window (Hou et al., 2013b; Hou, 2018b; Hou, 2018a). Moreover, Hou et al. (2013b) and Hou et al. (2018) have proposed an entity-based evaluation method where an anaphor is resolved to a preceding *entity* rather than a preceding *mention*. The idea is to first use *gold* coreference information to group the candidate antecedents of an anaphor into coreference clusters, and then extract cluster/entity-level features for encoding each of the resulting clusters/entities. The goal of the resolver is to resolve the anaphor to one of these clusters/entities based on the extracted features. Note that the resolution task is simplified when an anaphor is resolved to a cluster/entity as opposed to a candidate antecedent, because the number of clusters/entities is smaller than the number of candidate antecedents. Moreover, the use of gold coreference chains to produce entities and extract cluster-level features also makes it unfair to compare these entity-based evaluation results against other results.

**Anaphor filtering.** Several kinds of bridging anaphors are excluded from evaluation. One filtering rule says that any bridging anaphor whose closest antecedent is coreferent with it should be excluded from evaluation (Hou et al., 2014). This is understandable as these anaphors should be resolved by a coreference resolver instead. Another rule excludes a bridging anaphor from consideration as long as one of its antecedents is coreferential with it (Rösiger, 2018b). We believe that this rule is rather unmotivated, and may remove bridging links that cannot otherwise be recovered from other mentions.

Some rules filter bridging anaphors that are "problematic". Rösiger (2018b) enumerates exactly what is being filtered (i.e., anaphors with multiple antecedents, antecedents spanning more than one sentence, empty antecedents and discontinuous markables). In contrast, Hou (2018a) simply says that "a few problematic cases on each corpus" are filtered out without even mentioning why they are problematic. The lack of such details may make it difficult to replicate her results.

**Antecedent filtering.** Besides anaphor filtering, there have also been attempts to filter candidate antecedents prior to resolution in order to improve resolution performance. For instance, Hou et al. (2013b)

exclude candidate antecedents that are coreferent with a bridging anaphor. The motivation is that by definition, these candidate antecedents cannot serve as the antecedents of a bridging anaphor. However, to ensure a fair comparison between systems that employ filtering and those that do not, we believe that predicted, rather than gold, coreference information should be used in the filtering process.

**Data splits.** While ARRAU has a standard train-dev-test split, the other corpora do not. In the absence of a standard data split, resolvers are evaluated via $k$-fold cross validation, which makes a head-to-head comparison of their results difficult.

## 5 Bridging Resolution Approaches

### 5.1 Rule-based Approaches

Virtually all early ruled-based resolvers operate in the least challenging setting, i.e., bridging resolution. Vieira and Teufel (1997) use a heuristic to resolve bridging anaphors based on synonymy, hyponymy, and meronymy relations from WordNet 1.6. Poesio et al. (1997) improve this system by limiting the use of some WordNet relations and improving the antecedent search strategy. For further improvement of this system, Poesio et al. (2002) complement WordNet coverage with another lexical resource of meronymy relations, which is acquired by querying syntactic patterns such as *NP of NP* and *NP's NP* in the British National Corpus. To have a large corpus as a resource for acquiring semantic relations, Markert et al. (2003) use the Web to extract meronymy and hyponymy relations.

Following an early rule-based bridging system (Vieira and Poesio, 2000)[2], all recently-developed rule-based bridging systems are composed of rules that perform recognition and resolution at the same time. For instance, Hou et al. (2014) propose a system consisting of eight rules for ISNotes. While Rösiger et al. (2018b) manage to apply Hou et al.'s system to BASHI with just one additional rule, Rösiger (2018b) finds that her system is not directly applicable to ARRAU: ARRAU consists mostly of lexical bridging while ISNotes, for which Hou et al.'s rules are developed, contains only referential bridging. Consequently, Rosiger et al. retain only the three rules in Hou et al.'s ruleset that still capture common patterns that appear both in ISNotes and ARRAU, and add eight rules that are designed specifically for ARRAU. One disadvantage of rule-based bridging resolvers, which is also true for rule-based systems in general, is that new rules may need to be designed for a new corpus annotated with a different scheme.

Table 2 shows the rules designed by Hou et al. (2014) and Rösiger et al. (2018a) for full bridging resolution in ISNotes. The rules are sorted by precision and should be applied in the order in which they are presented in the table. Each rule is composed of two conditions: one on the anaphor and the other on the antecedent. If the two mentions satisfy these conditions, the rule will posit a bridging link between them. In the table, each rule is expressed in terms of its name, the condition on the anaphor, the condition on the antecedent, the motivation behind its design, and its recognition and resolution recall and precision on ISNotes (I), BASHI (B), and ARRAU RST (A). As we can see, these are mostly low-recall rules: many bridging anaphors cannot be recognized or resolved using these rules. Moreover, each rule has different performances (in terms of recognition and resolution) on different corpora, meaning that these rules, which are designed for ISNotes, do not generalize across corpora.

### 5.2 Learning-based Approaches

We divide existing learning-based approaches into three categories.

**Feature-based approaches.** In these approaches, a pairwise classifier, known as the mention-pair model in the coreference resolution literature (Soon et al., 2001; Ng and Cardie, 2002), is trained to determine whether two mentions has a bridging relation. Each training instance therefore corresponds to two mentions, one of which is a bridging anaphor and the other is its candidate antecedent. If the candidate antecedent is its correct antecedent, the instance is labeled as POSITIVE; otherwise, it is labeled as NEGATIVE. Table 3 shows the list of features that have been used to train the mention-pair model.

The mention-pair model works well if a resolver is given gold anaphors as input. To perform full bridging resolution, in which gold mentions are given, we need to first train a "recognition" classifier

---

| Rule | Condition on anaphor | Condition on antecedent | Motivation | Recognition | | Resolution | |
|---|---|---|---|---|---|---|---|
| | | | | P (%) | R (%) | P (%) | R (%) |
| Set: Percent-age | Percentage NPs in subject position | Closest NP modifying another percentage NP via "of" (e.g., 22% of the firms) | Percentage expressions can indicate set bridging | I: 100 B: 0.0 A: 100 | I: 0.8 B: 0.0 A: 0.2 | I: 100 B: 0.0 A: 100 | I: 0.8 B: 0.0 A: 0.2 |
| Building part[3] | Common NPs whose head is a building part without nominal pre-modifications | NP with the strongest semantic connectivity to the anaphor | Building part is often involved in meronymy | I: 80.0 B: 25.0 A: 0.0 | I: 1.2 B: 0.21 A: 0.0 | I: 60.0 B: 0.0 A: 0.0 | I: 0.9 B: 0.0 A: 0.0 |
| Set: Number or indefinite pronoun | Number expressions (e.g., two dogs) or indefinite pronouns (e.g., some ...) | Closest plural NP in subject position. If not found, closest plural NP in object position | Numbers or indefinite pronouns can indicate set bridging | I: 90.0 B: 41.7 A: 50.0 | I: 1.4 B: 1.1 A: 0.2 | I: 70.0 B: 25.0 A: 0.0 | I: 1.1 B: 0.7 A: 0.0 |
| Argument taking NPs 1 | NPs with high argument ratio[4] and without nominal/adjective pre-modifications or indefinite determiners | 1. take all nominal modifiers of NPs whose head is same as anaphor's head. 2. closest NP that is a realization of these modification | Different instances of the same noun predicate likely maintain the same argument fillers indicated by nominal modifiers | I: 72.7 B: 42.9 | I: 1.2 B: 1.3 | I: 54.6 B: 36.7 | I: 0.9 B: 1.1 |
| Relative person | Non-generic NPs whose head is a relative without nominal/adjective pre-modifications | Closest non-relative person NP | Handles relative nouns, which tend to be bridging | I: 71.4 B: 25.0 | I: 0.8 B: 1.1 | I: 57.1 B: 20.0 | I: 0.6 B: 0.9 |
| GPE job title | Job titles with country pre-modifications (e.g., Italian mayor) | Most salient GPE (e.g., Italy) | Some job title NPs implicitly refer to the globally salient GPE | I: 77.3 B: 57.1 | I: 2.6 B: 0.9 | I: 63.6 B: 42.9 | I: 2.1 B: 0.7 |
| Profess-ional role | Professional role NPs (e.g., professor) | Most salient organization name | A more general rule than "Relative person" and "GPE job title" | I: 84.0 B: 40.0 | I: 3.2 B: 1.7 | I: 64.0 B: 25.0 | I: 2.4 B: 1.1 |
| Argument taking NPs 2 | NPs in subject position with high argument ratio and without nominal/adjective pre-modifications | NP with the strongest semantic connectivity to the anaphor | A NP in subject position that is likely to take arguments tends to be bridging anaphor | I: 66.7 B: 16.7 | I: 2.4 B: 0.9 | I: 37.5 B: 4.2 | I: 1.4 B: 0.2 |
| Meronymy relation | Unmodified definite NPs | NP classified as meronym with the anaphor by a relation classifier trained using WordNet | Handles meronym bridging | I: 49.1 B: 37.1 | I: 4.1 B: 7.2 | I: 25.5 B: 14.6 | I: 2.1 B: 2.8 |

Table 2: Rules for resolving bridging anaphors in ISNotes. The first eight rules are proposed by Hou et al. (2014) and the last rule is proposed by Rösiger et al. (2018a). 'I', 'B', and 'A' refer to ISNotes, BASHI, and ARRAU RST respectively.

to identify the bridging anaphors from the gold mentions and then pass the resulting anaphors to the mention-pair model for resolution. While in principle a binary classifier can be trained to determine whether a gold mention is an anaphor or not, previous work has trained classifiers for determining the IS of a mention and assumed that those mentions that are classified as "bridging" are bridging anaphors. Table 4 enumerates the features that have been proposed to train a classifier for determining the IS of a mention (Nissim, 2006; Rahman and Ng, 2011; Cahill and Riester, 2012; Markert et al., 2012; Rahman and Ng, 2012a; Hou et al., 2013a; Hou, 2016; Hou et al., 2018).

---

[3] A building part is a noun that denotes a part of a building (e.g., "window", "door").

[4] Argument ratio is a measure designed by Hou et al. (2014) to quantify how likely a NP is to take arguments.

| Feature | Description | Paper |
|---|---|---|
| **Lexical features** | | |
| Head match | whether $m_i$ and $m_j$ have the same head | Hou et al. (2013b) |
| Compound premodification | whether $m_j$ is a compound noun and $m_i$'s head is premodifiying $m_j$ | Hou et al. (2013b) |
| **Syntactic features** | | |
| Co-argument | whether $m_i$ and $m_j$ are the subject and the object of the same verb respectively | Hou et al. (2013b) |
| Parallel structure | whether $m_i$ has the same syntactic role and is in the same sentence (but not the same clause) as $m_j$ | Hou et al. (2018) |
| Closest modifier | whether $m_i$'s syntactic head is a modifier of one or more of the occurrences of the lemma of $m_j$'s head in the associated text | Hou et al. (2018) |
| **Semantic features** | | |
| WordNet query | whether $m_i$ and $m_j$ have a "part-of" relation in WordNet | Hou et al. (2013b) |
| Google distance | number of hits of the query "*the X of the Y*" returned by Google, where *X* is the $m_j$'s head and *Y* is $m_i$'s head | Poesio et al. (2004a) |
| WordNet distance | (the inverse value of) the shortest path length between $m_i$'s head and $m_j$'s head among all synset combinations | Poesio et al. (2004a) Hou et al. (2018) |
| Verb pattern (relative) | the semantic compatibility (expressed in PMI) between $m_i$ and $m_j$'s governing verb | Hou et al. (2018) |
| Verb pattern (top) | whether $m_i$ is the candidate antecedent that has the highest semantic compatibility with $m_j$'s governing verb | Hou et al. (2013b) |
| Preposition pattern (relative) | hit count (converted into the Dunning Root Log Likelihood association measure) obtained by querying the pattern *X prep Y* where *X* is $m_j$, *Y* is $m_i$, and *prep* is one of the three prepositions most frequently associated with *X* | Hou et al. (2013b) |
| Preposition pattern (top) | whether $m_i$ is the candidate antecedent that has the highest Dunning Root Log Likelihood association measure with $m_j$ using the aforementioned preposition patterns | Hou et al. (2013b) |
| GPE role person match | whether $m_j$ is a GPE role person and $m_i$ is a GPE | Hou et al. (2018) |
| OTHER role person match | whether $m_j$ is a role person (other than GPE role person) and $m_i$ is an organization | Hou et al. (2018) |
| Relative person match | whether $m_j$ is a relative person and $m_i$ is a person | Hou et al. (2013b) |
| Both spatial/temporal | whether $m_i$ and $m_j$ are both spatial mentions or temporal mentions | Hou et al. (2013b) |
| **Salience features** | | |
| Utterance distance | the distance between the utterance containing $m_j$ and the utterance containing $m_i$ | Poesio et al. (2004a) |
| First mention (local) | whether $m_i$ has been realized in the first position within the previous five sentences of $m_j$ | Poesio et al. (2004a) |
| First mention (global) | whether $m_i$ has been realized in the first position anywhere | Poesio et al. (2004a) |
| Entity frequency (relative) | the span of text (measured in sentences) in which $m_i$ is mentioned divided by the number of sentences in the document | Hou et al. (2018) |
| Entity frequency (top) | whether $m_i$ is the candidate antecedent that has the highest entity frequency (see the previous feature) | Hou et al. (2013b) |
| Global candidate | the sum of the semantic connectivity scores of $m_i$ with all possible anaphors in the document | Hou et al. (2013b) |
| **Sibling anaphor features** | | |
| Similar anaphors | whether the two anaphors are linked via conjunction, have the same head, have high similarity score, or are syntactically parallel. | Hou et al. (2013b) Hou et al. (2018) |
| Set-relation | whether the two anaphors are syntactically parallel sibling anaphors that are set expressions | Hou et al. (2018) |

Table 3: Features for bridging resolution. In this table, $m_j$ is a bridging anaphor, and $m_i$ is a candidate antecedent of $m_j$. The "Paper" column shows the papers that proposed the corresponding features.


**Embedding approaches.** Hou (2018b) observes that commonly used word representations such as GloVe (Pennington et al., 2014) capture "genuine" similarity and relatedness, but bridging resolution requires lexical association knowledge instead of semantic similarity information between synonyms or hypernyms. This motivates her to train task-specific embeddings for bridging resolution. To do so, she first observes that the prepositional (i.e., *X of Y*) and possessive structures (i.e., *Y's X*) of NPs encode a variety of bridging relations between anaphors and their antecedents. For example, *the window of the room* implies a part-of relation between *the window* and *the room*, and in *Japan's prime minister*, there is a bridging relation between *Japan* and *prime minister*. Then she extracts noun pairs involved in these syntactic structures from the parsed Gigaword corpus and uses them as distant supervision signals to train

| Feature | Description | Paper |
|---|---|---|
| **String-Matching features** | | |
| Exact match | whether $m_j$ has an exact match with a preceding mention | Nissim (2006) |
| Head match | whether $m_j$'s head has an exact match with a preceding mention's head | Nissim (2006) |
| Content word match | whether $m_j$ and a preceding mention has any content word overlap | Markert et al. (2012) |
| **Statistical features** | | |
| Length | the length of $m_j$ in words | Nissim (2006) |
| Exact match frequency | number of preceding mentions that have the same string as $m_j$ | Nissim (2006) |
| Head match frequency | number of preceding mentions that have the same head as $m_j$ | Markert et al. (2012) |
| **Lexical features** | | |
| Unigrams | whether $m_j$ includes a certain unigram | Rahman&Ng (2011) |
| **Syntactic features** | | |
| Determiner | $m_j$'s determiner | Nissim (2006) |
| NP type | whether $m_j$ is a common noun, proper noun, pronoun, or other | Nissim (2006) |
| Grammatical role | subject, object, predicate, PP, or other | Nissim (2006) |
| Change verb | whether $m_j$ is a number that is the object of an increase/decrease verb | Hou et al. (2013a) |
| **Semantic features** | | |
| Argument ratio | how likely $m_j$ takes arguments | Hou et al. (2014) |
| WordNet relational noun | whether $m_j$'s head is a relational noun in WordNet | Hou et al. (2013a) |
| Professional role term | whether $m_j$ is a professional role term in General Inquirer | Hou et al. (2013a) |
| Country modifier | whether $m_j$ include a country modifier (e.g. Italian) | Hou et al. (2013a) |
| Nesting mention | whether $m_j$ contains another mention | Hou et al. (2013a) |
| Building part nouns | whether $m_j$ is a building part noun in General Inquirer | Hou et al. (2013a) |
| Set element | whether $m_j$'s head is a number or an indefinite pronoun, or is modified by "each" or "one" | Hou et al. (2013a) |
| Year | whether $m_j$ is a year expression | Hou et al. (2013a) |
| Spatial or temporal | whether $m_j$ contains a spatial or temporal modifier in General Inquirer | Hou et al. (2013a) |
| Semantic class | $m_j$'s semantic class | Markert et al. (2012) |
| Comparative markers | whether $m_j$ is pre-modified by markers indicating a comparative anaphor | Markert et al. (2012) |
| **Generic NP features** | | |
| Number of a mention | whether $m_j$ is singular or plural | Markert et al. (2012) |
| Verb POS | the verb form of the clause where $m_j$ is present | Hou et al. (2013a) |
| If clause | whether $m_j$ appears in an "if" clause | Hou et al. (2013a) |
| Frequent generic NPs | whether $m_j$ is a generic NP from the ACE-2 corpus | Hou et al. (2013a) |
| Frequent proper names | whether $m_j$ is a proper name in at least 100 docs in the Tipster corpus | Hou et al. (2018) |
| World knowledge | whether $m_j$ is a NP annotated as World Knowledge in the training set | Hou et al. (2013a) |
| General Quantifiers | whether $m_j$ is modified by quantifiers such as "all" and "no" | Hou et al. (2013a) |
| **Discourse features** | | |
| Coherence | captures the coherence of the sentence in which $m_j$ appears to the previous sentences via string matching, presence of pronouns, etc. | Hou et al. (2013a) |
| 1st mention (sentence) | whether $m_j$ is the first mention of the sentence it appears in | Hou et al. (2013a) |
| 1st mention (document) | whether $m_j$ is the first mention in the document it appears in | Hou et al. (2013a) |
| **Relational features** | | |
| Has child relations | whether $m_j$ has a relation with its child mentions via possessive, prepositional, or coordination phrases | Markert et al. (2012) |
| Precedence relations | whether $m_j$ is in a precedence relation with another mention | Markert et al. (2012) |
| Coordination-joined relations | whether $m_j$ and another mention are children of the same coordination | Hou et al. (2018) |

Table 4: Features for bridging recognition. In this table, $m_j$ denotes a candidate bridging anaphor. The "Paper" column shows the papers that proposed the corresponding features.

an embedding model, embeddings_PP. The resulting embeddings can be used to select an antecedent for a bridging anaphor by calculating the vector similarity between the anaphor's head and a candidate antecedent's head. Moreover, she combines embeddings_PP, which covers only a subset of nouns, and the GloVe embeddings so that both non-nouns and additional nouns are covered (Hou, 2018a).

**Neural models.** Yu and Poesio (2020) propose the first neural model for full bridging resolution, leveraging a *span-based* neural model originally developed for entity coreference resolution by Kantor and Globerson (2019). Kantor and Globerson's span-based model is a *mention-ranking* model (Denis and Baldridge, 2008), meaning that it is trained to rank the candidate antecedents of an anaphor so that the correct antecedent has the highest rank. Key to the success of this and other span-based coreference models is their ability to learn *text spans* corresponding to entity mentions as well as their *representations* so

| System | Approach | Gold coref? | Dataset | | |
|---|---|---|---|---|---|
| | | | ISNotes | BASHI | ARRAU |
| Hou et al. (2018) | Feature based | Feature extraction | 50.7 | - | - |
| Hou (2018a) | Embedding | No | 39.5 | 27.4 | 32.4 |
| Yu and Poesio (2020) | MTL | No | 40.7 | 34 | 49.3 |
| Hou (2020) | QA | No | 50.1 | 38.7 | - |

Table 5: Accuracies of state-of-the-art bridging resolvers.

| System | Approach | Gold coref? | Dataset | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | ISNotes | | BASHI | | ARRAU | |
| | | | Rec | Res | Rec | Res | Rec | Res |
| Rösiger et al. (2018b) | Rule based | Anaphor filtering | 29.3 | 20.4 | 28.7 | 14.1 | 30.8 | 19.5 |
| Hou et al. (2018) | Feature based | Feature extraction | 46.1 | 21.6 | - | - | - | - |
| Yu and Poesio (2020) | MTL | Anaphor filtering | 43.6 | 23.2 | 27.2 | 14.4 | 36.7 | 24.0 |

Table 6: Recognition (Rec) and resolution (Res) F-scores of state-of-the-art *full* bridging resolvers.

that the representations of coreferent mentions are similar to each other. Yu and Poesio present two extensions to Kantor and Globerson's model. First, they provide gold mentions as input to the model, meaning that the model needs to learn the span representations but not the span boundaries. Second, and more importantly, they propose to train the model to perform coreference and bridging in a multi-task learning (MTL) framework. In this framework, the span representation layer is shared by the two tasks, so that information learned from one task can be utilized when learning the other task. Unlike feature-based approaches, where feature engineering plays a critical role in performance, this neural model employs only two features, the length of a mention and mention-pair distance.

More recently, Hou (2020) has proposed a neural approach to bridging resolution based on question answering (QA). Given a gold anaphor, the idea is to (1) create a question from the anaphor in the form of "anaphor *of what?*", (2) create candidate answers from the candidate antecedents (i.e., the preceding mentions that appear in a fixed sentence window), and (3) use a BERT-based QA system pretrained on the SQuAD corpus (Joshi et al., 2020) to choose the most probable answer (i.e., the antecedent). An appealing aspect of this approach is that it does not require any gold or system mention information as the antecedent candidates. Hou further hypothesizes that the results could be improved if the model were pretrained on a bridging corpus rather than a QA corpus. However, as mentioned before, all existing bridging corpora are too small to train an effective neural model. To address this problem, Hou employs a distant supervision method (see the embedding approaches above) to generate an automatically labeled bridging corpus, and subsequently shows that the model pretrained on this bridging corpus offers better performance than the one pretrained on SQuAD.

## 6 The State of the Art

To better understand the state of the art, we show the best results achieved on bridging resolution and full bridging resolution by different approaches on three commonly used datasets (ISNotes, BASHI, and ARRAU RST) in Tables 5 and 6 respectively. Note that (1) these results are taken verbatim from the respective papers, and (2) not all of them are directly comparable, as some rely on gold coreference chains to compute cluster-level features or perform anaphor filtering.

As we can see in Table 5, solid progress is being made for bridging resolution, even though the best accuracy is only around 50%. *Full* bridging resolution results are shown in Table 6. Note that Hou et al.'s (2018) resolver cannot be directly compared with the other two because it uses gold coreference information in a different way. Comparing Rösiger et al.'s (2018b) rule-based system with Yu and Poesio's (2020) MTL model, we see that the neural model has achieved solid improvements in both recognition and resolution on ARRAU and ISNotes. Additional experiments are needed to understand why similar improvements are not observed on BASHI. Nevertheless, the best resolution F-score is below 25%. Overall, these results suggest that both bridging resolution and full bridging resolution are far from being solved.

# 7 Concluding Remarks

We have presented a survey of the current state of research on bridging resolution, a task that is far from being solved. Given that significant advances have been made on coreference resolution recently, we believe that bridging resolution will be the anaphora resolution task that will receive increasing attention in the near future. We conclude this paper with several recommendations on future research directions.

**Resources and evaluation.** The CoNLL 2011 and 2012 shared tasks (Pradhan et al., 2011; Pradhan et al., 2012) have played a crucial role in the accelerated progress on entity coreference resolution in the past few years by providing a standardized evaluation protocol (e.g., standard evaluation corpus and metrics) that facilitates performance comparisons of different resolvers. Progress on current bridging resolution research, which is reminiscent of that of entity coreference research in the pre-CoNLL era, is hindered in part by the lack of such standardization. As we move forward, it is imperative that a common evaluation framework be established for bridging research. Part of this effort should include the development of an annotated corpus that is much larger than those currently available. While the use of distant supervision (to produce automatically labeled data) and pretrained models (to transfer knowledge from other tasks) may have reduced the amount of annotated data needed for model training, we believe having a large annotated corpus will not only stimulate interest in the task among researchers in other areas (e.g., by allowing them to develop complex models) but also provide task-specific linguistic insights. Corpora that contain other discourse-level annotations (e.g., the discourse relations in the Penn Discourse Treebank (Prasad et al., 2008)) would be ideal choices, as they can facilitate the development of joint models that enable the study of the potential interactions between bridging and other discourse phenomena.

**Cross-area collaboration.** While bridging has been studied primarily by discourse researchers, the task is so broad that it covers many semantic relations studied by researchers in lexical semantics and information extraction, such as meronymy (Hearst, 1998; Berland and Charniak, 1999; Girju et al., 2006), hyponymy (Hearst, 1992; Cederberg and Widdows, 2003; Pantel and Pennacchiotti, 2006), and set-membership. While space limitations have precluded a detailed discussion of this line of related work, it is important to note that some of the ideas we have seen in this paper have also been explored in research on extracting these specific relations. For instance, like Hou (2018b), Hearst (1992) and Modjeska et al. (2003) have explored the use of lexico-syntactic patterns for automatically harvesting hyponyms and antecedents of *other*-anaphora, respectively. As another example, Girju et al.'s (2003) decision tree approach to part-whole relation extraction has employed many of the features that are also used in feature-based approaches to bridging resolution. Rather than reinventing the wheels, we encourage researchers in these different areas work together on bridging resolution. In particular, it is conceivable that the best approach to bridging resolution may involve first decomposing the task into different types of relations and then have researchers from different areas address each type of relations.

**Models.** Existing bridging resolvers have all assumed as input either gold anaphors or gold mentions, making them unusable in practice where such gold annotations are not available. Consequently, we believe that time is ripe for *end-to-end* bridging resolution. Though it is a very challenging evaluation setting, we believe that researchers should examine whether the successes of end-to-end neural models developed for various NLP tasks can be transferred to bridging resolution. In addition, researchers may consider *hybrid* rule-based and learning-based models for bridging resolution. So far, rule-based and learning-based approaches have been viewed as distinct approaches in bridging research, but it is worth investigating whether they have complementary strengths. Finally, while much of the work on bridging has been conducted for English, we believe it is time to examine multilingual bridging resolution. Since we have bridging-annotated data for multiple languages including parallel corpora, it would be interesting to see if multilingual bridging resolution can be addressed using projection-based techniques (Yarowsky et al., 2001; Postolache et al., 2006; Rahman and Ng, 2012b; Grishina and Stede, 2015; Martins, 2015).

# References

Maja Bärenfänger, Daniela Goecke, Mirco Hilbert, Harald Lüngen, and Maik Stührenberg. 2008. Anaphora as an indicator of elaboration: A corpus study. *Journal for Language Technology and Computational Linguistics*, 23(2):49–73.

Matthew Berland and Eugene Charniak. 1999. Finding parts in very large corpora. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, pages 57–64.

Anders Björkelund, Kerstin Eckart, Arndt Riester, Nadja Schauffler, and Katrin Schweitzer. 2014. The extended dirndl corpus as a resource for automatic coreference and bridging resolution. In *Proceedings of the 9th International Conference on Language Resources and Evaluation*, pages 3222–3228.

Razvan Bunescu. 2003. Associative anaphora resolution: A web-based approach. In *Proceedings of the EACL Workshop on The Computational Treatment of Anaphora*, pages 47–52.

Aoife Cahill and Arndt Riester. 2012. Automatically acquiring fine-grained information status distinctions in german. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 232–236.

Tommaso Caselli and Irina Prodanof. 2006. Annotating bridging anaphors in Italian: in search of reliability. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.

Scott Cederberg and Dominic Widdows. 2003. Using lsa and noun coordination information to improve the recall and precision of automatic hyponymy extraction. In *Proceedings of the Seventh Conference on Natural Language Learning*, pages 111–118.

Herbert H. Clark. 1975. Bridging. In *Theoretical Issues in Natural Language Processing*.

Pascal Denis and Jason Baldridge. 2008. Specialized models and ranking for coreference resolution. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 660–669.

Claire Gardent and Hélène Manuélian. 2005. Création d'un corpus annoté pour le traitement des descriptions définies. *Traitement Automatique des Langues*, 46(1):115–140.

Claire Gardent, Hélène Manuélian, and Eric Kow. 2003. Which bridges for bridging definite descriptions? In *Proceedings of 4th International Workshop on Linguistically Interpreted Corpora*.

Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, pages 80–87.

Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2006. Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1):83–135.

Yulia Grishina and Manfred Stede. 2015. Knowledge-lean projection of coreference chains across languages. In *Proceedings of the Eighth Workshop on Building and Using Comparable Corpora*, pages 14–22.

Yulia Grishina. 2016. Experiments on bridging across languages and genres. In *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes*, pages 7–15.

Jeanette K Gundel, Nancy Hedberg, and Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, pages 274–307.

Jan Hajič, Eduard Bejček, Alevtina Bémová, Eva Buráňová, Eva Hajičová, Jiří Havelka, Petr Homola, Jiří Kárník, Václava Kettnerová, Natalia Klyueva, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Petr Pajas, Jarmila Panevová, Lucie Poláková, Magdaléna Rysová, Petr Sgall, Johanka Spoustová, Pavel Straňák, Pavlína Synková, Magda Ševčíková, Jan Štěpánek, Zdeňka Urešová, Barbora Vidová Hladká, Daniel Zeman, Šárka Zikánová, and Zdeněk Žabokrtský. 2018. Prague dependency treebank 3.5. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

John A Hawkins. 1978. Definiteness and indefiniteness: A study in reference and grammaticality prediction. *Journal of Linguistics*, 27:405–442.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 15th International Conference on Computational Linguistics: Volume 2*, pages 539–545.

Marti A Hearst. 1998. Automated discovery of wordnet relations. *WordNet: an electronic lexical database*, 2:132–152.

Iris Hendrickx, Gosse Bouma, Frederik Coppens, Walter Daelemans, Veronique Hoste, Geert Kloosterman, Anne-Marie Mineur, Joeri Van Der Vloet, and Jean-Luc Verschelde. 2008. A coreference corpus and resolution system for Dutch. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*.

Yufang Hou, Katja Markert, and Michael Strube. 2013a. Cascading collective classification for bridging anaphora recognition using a rich linguistic feature set. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, pages 814–820.

Yufang Hou, Katja Markert, and Michael Strube. 2013b. Global inference for bridging anaphora resolution. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 907–917.

Yufang Hou, Katja Markert, and Michael Strube. 2014. A rule-based system for unrestricted bridging resolution: Recognizing bridging anaphora and finding links to antecedents. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 2082–2093.

Yufang Hou, Katja Markert, and Michael Strube. 2018. Unrestricted bridging resolution. *Computational Linguistics*, 44(2):237–284.

Yufang Hou. 2016. Incremental fine-grained information status classification using attention-based LSTMs. In *Proceedings of the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1880–1890.

Yufang Hou. 2018a. A deterministic algorithm for bridging anaphora resolution. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 1938–1948.

Yufang Hou. 2018b. Enhanced word representations for bridging anaphora resolution. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 1–7.

Yufang Hou. 2020. Bridging anaphora resolution as question answering. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1428–1438.

Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S Weld, Luke Zettlemoyer, and Omer Levy. 2020. SpanBERT: Improving pre-training by representing and predicting spans. *Transactions of the Association for Computational Linguistics*, 8:64–77.

Ben Kantor and Amir Globerson. 2019. Coreference resolution with entity equalization. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 673–677.

Iørn Korzen and Matthias Buch-kromann. 2011. Anaphoric relations in the Copenhagen Dependency Treebanks. In *Proceedings of Beyond Semantics: Corpus-based Investigations of Pragmatic and Discourse Phenomena*, pages 83–98.

Emmanuel Lassalle and Pascal Denis. 2011. Leveraging different meronym discovery methods for bridging resolution in French. In *The 8th Discourse Anaphora and Anaphor Resolution Colloquium*, pages 35–46.

Sebastian Löbner. 1998. Definite associative anaphora. `http://user.phil-fak.uniduesseldorf.de/~loebner/publ/DAA-03.pdf`.

Katja Markert, Malvina Nissim, and Natalia N Modjeska. 2003. Using the Web for nominal anaphora resolution. In *Proceedings of the EACL Workshop on the Computational Treatment of Anaphora*, pages 39–46.

Katja Markert, Yufang Hou, and Michael Strube. 2012. Collective classification for fine-grained information status. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 795–804.

André F. T. Martins. 2015. Transferring coreference resolvers with posterior regularization. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1427–1437.

Natalia N. Modjeska, Katja Markert, and Malvina Nissim. 2003. Using the web in machine learning for other-anaphora resolution. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, pages 176–183.

Natalia Nygren Modjeska. 2003. *Resolving Other-Anaphora*. Ph.D. thesis, University of Edinburgh.

Vincent Ng and Claire Cardie. 2002. Improving machine learning approaches to coreference resolution. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 104–111.

Malvina Nissim. 2006. Learning information status of discourse entities. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, pages 94–102.

Patrick Pantel and Marco Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 113–120.

Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543.

Massimo Poesio and Ron Artstein. 2008. Anaphoric annotation in the ARRAU corpus. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation*, pages 1170–1174.

Massimo Poesio and Renata Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216.

Massimo Poesio, Renata Vieira, and Simone Teufel. 1997. Resolving bridging references in unrestricted text. In *Proceedings of the ACL/EACL Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*.

Massimo Poesio, Tomonori Ishikawa, Sabine Schulte im Walde, and Renata Vieira. 2002. Acquiring lexical knowledge for anaphora resolution. In *Proceedings of the Third International Conference on Language Resources and Evaluation*.

Massimo Poesio, Rahul Mehta, Axel Maroudas, and Janet Hitzeman. 2004a. Learning to resolve bridging references. In *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, pages 143–150.

Massimo Poesio, Rosemary Stevenson, Barbara Di Eugenio, and Janet Hitzeman. 2004b. Centering: A parametric theory and its instantiations. *Computational linguistics*, 30(3):309–363.

Oana Postolache, Dan Cristea, and Constantin Orasan. 2006. Transferring coreference chains through word alignment. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation*.

Sameer Pradhan, Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel, and Nianwen Xue. 2011. CoNLL-2011 shared task: Modeling unrestricted coreference in OntoNotes. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning: Shared Task*, pages 1–27.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning: Shared Task*, pages 1–40.

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, and Bonnie Webber. 2008. The Penn discourse TreeBank 2.0. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*.

Ellen Prince. 1981. Toward a taxonomy of given-new information. In P. Cole, editor, *Radical Pragmatics*, page 223–255. Academic Press, New York, N.Y.

Altaf Rahman and Vincent Ng. 2011. Learning the information status of noun phrases in spoken dialogues. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1069–1080.

Altaf Rahman and Vincent Ng. 2012a. Learning the fine-grained information status of discourse entities. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 798–807.

Altaf Rahman and Vincent Ng. 2012b. Translation-based projection for multilingual coreference resolution. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 720–730.

Marta Recasens, M Antònia Martí, and Mariona Taulé. 2007. Where anaphora and coreference meet. annotation in the spanish cess-ece corpus. In *Proceedings of Recent Advances in Natural Langage Processing*, pages 504–509.

Kepa Joseba Rodríguez, Francesca Delogu, Yannick Versley, Egon W. Stemle, and Massimo Poesio. 2010. Anaphoric annotation of Wikipedia and blogs in the live memories corpus. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation*, pages 157–163.

Anna Roitberg and Anna Nedoluzhko. 2016. Bridging corpus for russian in comparison with czech. In *Proceedings of the Workshop on Coreference Resolution Beyond OntoNotes*, pages 59–66.

Ina Rösiger, Maximilian Köper, Kim Anh Nguyen, and Sabine Schulte im Walde. 2018a. Integrating predictions from neural-network relation classifiers into coreference and bridging resolution. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 44–49.

Ina Rösiger, Arndt Riester, and Jonas Kuhn. 2018b. Bridging resolution: Task definition, corpus resources and rule-based experiments. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 3516–3528.

Ina Rösiger. 2016. SciCorp: A corpus of English scientific articles annotated for information status analysis. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pages 1743–1749.

Ina Rösiger. 2018a. BASHI: A corpus of wall street journal articles annotated with bridging links. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*.

Ina Rösiger. 2018b. Rule-and learning-based methods for bridging resolution in the arrau corpus. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 23–33.

Ryohei Sasano and Sadao Kurohashi. 2009. A probabilistic model for associative anaphora resolution. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 3*, pages 1455–1464.

Katrin Schweitzer, Kerstin Eckart, Markus Gärtner, Agnieszka Falenska, Arndt Riester, Ina Rösiger, Antje Schweitzer, Sabrina Stehwien, and Jonas Kuhn. 2018. German radio interviews: The grain release of the sfb732 silver standard collection. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation*.

Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. 2001. A machine learning approach to coreference resolution of noun phrases. *Computational linguistics*, 27(4):521–544.

Olga Uryupina, Ron Artstein, Antonella Bristot, Federica Cavicchio, Francesca Delogu, Kepa J Rodriguez, and Massimo Poesio. 2020. Annotating a broad range of anaphoric phenomena, in a variety of genres: The ARRAU corpus. *Natural Language Engineering*, 26(1):95–128.

Renata Vieira and Massimo Poesio. 2000. An empirically based system for processing definite descriptions. *Computational Linguistics*, 26(4):539–593.

Renata Vieira and Simone Teufel. 1997. Towards resolution of bridging descriptions. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Chapter of the Association for Computational Linguistics*, pages 522–524.

David Yarowsky, Grace Ngai, and Richard Wicentowski. 2001. Inducing multilingual text analysis tools via robust projection across aligned corpora. In *Proceedings of the First International Conference on Human Language Technology Research*.

Juntao Yu and Massimo Poesio. 2020. Multi-task learning based neural bridging reference resolution. `arXiv:2003.03666 [cs.CL]`.

Amir Zeldes. 2017. The GUM corpus: Creating multilayer resources in the classroom. *Language Resources and Evaluation*, 51(3):581–612.