

A Resource for Investigating the Impact of Anaphora and Coreference on Inference

Azad Abad¹, Luisa Bentivogli², Ido Dagan³,
Danilo Giampiccolo¹, Shachar Mirkin³, Emanuele Pianta², Asher Stern³

¹CELCT
Trento, Italy

²FBK-irst
Trento, Italy

³Bar-Ilan University
Ramat Gan, Israel

{abad,giampiccolo}@celct.it; {bentivo,pianta}@fbk.eu; {dagan,mirkins}@cs.biu.ac.il; astern7@gmail.com

Abstract

Discourse phenomena play a major role in text processing tasks. However, so far relatively little study has been devoted to the relevance of discourse phenomena for inference. Therefore, an experimental study was carried out to assess the relevance of anaphora and coreference for Textual Entailment (TE), a prominent inference framework. First, the annotation of anaphoric and coreferential links in the RTE-5 Search data set was performed according to a specifically designed annotation scheme. As a result, a new data set was created where all anaphora and coreference instances in the entailing sentences which are relevant to the entailment judgment are solved and annotated. A by-product of the annotation is a new “augmented” data set, where all the referring expressions which need to be resolved in the entailing sentences are replaced by explicit expressions. Starting from the final output of the annotation, the actual impact of discourse phenomena on inference engines was investigated, identifying the kind of operations that the systems need to apply to address discourse phenomena and trying to find direct mappings between these operation and annotation types.

1. Introduction

Discourse phenomena play a major role in text processing tasks. Discourse structure analysis (e.g. RST), anaphora resolution, intra- and cross-document coreference resolution are just some of the tasks which have attracted the attention of NLP researchers in this area. However, so far, relatively little study has been devoted to the relevance of discourse phenomena for inference. As Textual Entailment has become a popular approach for modeling inference (Dagan & Glickman, 2004), investigating discourse-related issues and their potential effects on inference systems under this framework appears to be promising. In particular, the Search Pilot Task proposed in the fifth Recognizing Textual Entailment challenge (Bentivogli et al., 2009b) happens to be particularly suitable for this purpose. In this RTE-5 task, given a corpus about a topic and a textual assertion about the same topic (hypothesis, H), systems are required to find in the corpus each sentence (text, T) from which the hypothesis can be inferred¹, i.e. the cases where *T entails H*. In such a context, where the sentences to be judged for entailment are interpreted in the context of the corpus, solving discourse phenomena becomes crucial in order to correctly assess the entailment relation between the text and the hypothesis². Therefore, we performed an experimental study focusing on two of the most pervasive discourse phenomena, namely anaphora and coreference (henceforth also referred to jointly as *textual relations*).

The first phase of the study consisted of the annotation of anaphoric and coreferential links in the

RTE-5 Search data set. The goal of this annotation was to create a new data set where all the anaphora and coreference instances in the entailing sentences which are relevant for entailment are solved and annotated according to a specifically designed annotation scheme. The final annotated data set is meant to represent a resource useful primarily for RTE system developers, who can use the information provided in the annotation to solve anaphora and coreference in the RTE-5 Search data set, in order to focus on other components of the systems, once issues related to textual relations have been removed. Moreover, as a by-product of the annotation, a new “augmented” data set is produced, where all the referring expressions which need to be resolved in the entailing sentences are substituted by explicit expressions, which are recovered either from their antecedent or from a previous more complete mention of an entity or event. Exploiting such annotated corpus, developers can evaluate the impact of anaphora and coreference on the performances of their systems.

The annotated data set, which will be made publically available, represents also a valuable resource that can be used in different fields of NLP, such as in anaphora and coreference resolution research, and also in general linguistics for research on textual relations. Moreover, although the annotation concerns only a portion of the textual relations present in the corpus - namely those which are relevant to the entailment task - it can still provide useful information on the distribution and frequency of anaphora and coreference phenomena in a corpus, allowing the development of specific strategies to address the most frequent ones.

Based on the final output of the annotation, we also investigated the actual impact of textual relations on inference engines and the kind of operations that these systems need to apply to address them. The idea is to define a set of operations - more specifically, syntactic

¹ For more details see <http://www.nist.gov/tac/2009/RTE>.

² For an analysis of the relevance of discourse references in the production of the RTE-5 Search data set see (Bentivogli et al., 2009a).

manipulations - that can be applied on the original text so that an RTE system would be able to find the required inference. Based on that approach, an innovative research direction would be to find a direct mapping between annotation types and operations that need to be performed by the system. Such mapping will enable inference systems to identify the operation required for inference purposes directly from the output of the anaphora/coreference resolution component.

The remainder of this paper is structured as follows: Section 2 reviews the main anaphora and coreference annotation schemes and reports about existing annotated corpora. Section 3 describes in detail the annotation principles and specifications that were followed in our annotation, while Section 4 investigates the actual impact of anaphora and coreference on inference engines. Finally, Section 5 draws some conclusions.

2. Related Work

Although anaphorically and coreferentially annotated corpora are indispensable to anaphora and coreference resolution and to most NLP tasks or applications, they are not numerous, and not of a large size. Among the anaphorically annotated corpora, beside the Lancaster Anaphoric Treebank (Leech & Garside, 1991), some new corpora have become available in the recent years, such as the Gnome Corpus (Poesio, 2004), the Venex Italian Corpus (Poesio et al., 2004), and the Arrau corpus (Poesio & Artstein, 2008). Recently, a new project, called ANAWIKI, has been initiated to create high-quality, large-scale anaphorically annotated resources (Poesio et al., 2008), by taking advantage of the Web community. The project also offers the possibility to share anaphorically annotated resources at the Anaphoric Bank³.

As for coreferentially annotated corpora, text collections annotated for coreferential links have been produced for the MUC-6 and MUC-7 coreference tasks and for the ACE Entity Detection and Recognition task. The Penn Treebank has been partially annotated with information on coreference (Ge, 1998) and two corpora annotated with coreferential chains have been produced at the University of Wolverhampton (Mitkov et al., 2000; Hasler et al., 2006). More recently, the Catalan and Spanish AnCorra corpus has been annotated at different linguistic levels including coreference (Taulé et al., 2008), and the OntoNotes project is annotating a large corpus in English, Chinese, and Arabic with structural information and shallow semantics, including word sense, linked to coreference (Hovy et al., 2006).

In order to annotate coreference and anaphora phenomena, a number of annotation schemes have been developed. An overview of the most relevant annotation schemes can be found in (Mitkov, 2008). Among the most widely used, the UCREL scheme has been utilized for anaphora annotation of newswire texts (Fligelstone, 1992; Garside et al., 1997). A SGML-based scheme has

been specifically developed for the MUC conference (Hirshman & Chinchor, 1997). The DRAMA scheme (Passonneau & Litman, 1997) has been created to identify anaphors and antecedents, marking coreference relationships between them; meanwhile the MATE scheme has been conceived to annotate coreference in dialogues (Davis et al., 1998).

3. Annotating Anaphora and Coreference in the RTE-5 Search Data Set

3.1. The RTE-5 Search Data Set

The annotation was carried out on the RTE-5 Search Task data set. This data set is based on the data created for the Summarization track of the Text Analysis Conference (TAC)⁴ and is composed of 19 topics, 10 used for the Development Set and 9 for the Test Set. For each topic, the data consist of a set of 10 newswire documents and between 6 and 10 hypotheses. Totally, the data set contains 161 hypotheses and 4,487 sentences, among which 1,156 entail at least one hypothesis.

3.2. Annotation Principles and Scheme

The aim of the annotation is to create a new data set where all the anaphora and coreference instances relevant to the textual entailment task are solved and annotated. According to this goal, not all possible markables in the corpus were annotated. First, we restricted the annotation to entailing sentences, and then, within the entailing sentences we chose as markables only those items relevant to the entailment inference. Consider Example 1:

Ex. 1:

H: The ice is melting in the Antarctic.

T: According to a recent study published in *Geophysical Research Letters*, the discharge rate of three important glaciers still remaining on **the Antarctic peninsula** accelerated [...]. Ice is thinning at the rate of tens of meters per year on **the peninsula** [...], *it* found.

Both the noun phrase “the peninsula” and the pronoun “it” in T are anaphoric, because, to fully understand their meaning, we need to recover the relation of “it” back to “Geophysical Research Letters”, and of “the peninsula” back to “the Antarctic peninsula”. However, only “the peninsula” is annotated, as recovering the antecedent of this phrase is the only one relevant for inference.

Before carrying out the annotation, the data set was re-arranged, creating for each topic corpus a new “document” containing all the hypotheses relevant to it. The hypotheses document was first annotated in order to identify the set of markables to be annotated in the entailing sentences.

The annotation was carried out according to a scheme defined considering the relevant literature (see Mitkov, 2003; Huang, 2000) and other existing annotation schemes used in anaphora/coreference resolution tasks

³ <http://anawiki.essex.ac.uk/anaphoricbank/>

⁴ <http://www.nist.gov/tac/2009/Summarization/>

(see Section 2).

3.2.1. Anaphora versus Coreference

Although anaphora and coreference are not always distinguished in computational linguistics, they are treated as different phenomena in our annotation. From a theoretical point of view, two expressions are in *coreferential* relation if they refer to the same entity or event in the real world, whereas two expressions are in *anaphoric* relation if one of them can be interpreted only through the recognition of a textual relation with the other. For instance, we say that the expression “she” is an *anaphor*, because we can assign a referent to it only by recognizing that it is textually related to its *antecedent* (e.g. “Camilla Parker Bowles”).

In many cases two expressions can be both in coreferential and anaphoric relations. However, in some cases the two notions do not overlap. On the one hand, two textual descriptions can be coreferring, and yet be completely independent from the point of view of the interpretation of the reference, as there is no need to link one to the other to interpret them (e.g. “Camilla Parker Bowles” and “Prince Charles’ wife”). On the other hand, there can be anaphora cases where the two elements involved in the relation do not corefer, as in the so-called associative or bridging anaphora, which expresses non-identity relations such as “part of”, “member of” and “participant in event”. For instance, consider the following short text: “I entered the room (*antecedent*). The ceiling (*anaphor*) was very high”. To interpret the expression “the ceiling” we need to recognize that it is the ceiling “of the room” mentioned in the previous sentence. But the ceiling and the room do not refer to the same entity.

Given the above definitions, in most cases deciding whether two expressions are in anaphoric and/or coreferential relation is relatively easy. It is clear for instance that all pronouns are both anaphoric and coreferential. Also, if in order to fully interpret an expression, we need to link it back to an antecedent, and the two expressions do not refer to the same entity, again we have a (bridging) anaphora and no coreference. An anaphoric relation can hold also between two full nominal expressions. Consider, for instance, the following short text: “The ambassador is going to meet the king of Spain during the next week. The arrival of the king is planned for tomorrow”. The expression “the king” in the second sentence is clearly related to the expression “the king of Spain” in the first sentence, through a textual mechanism which can be labeled as partial repetition; to correctly interpret the reference of the second nominal expression (what king are we talking about?), we need to recover the link back to the first more complete description.

However, when the first and the second nominal descriptions of the same individual are quite distant in the text, it may be difficult to decide whether there is some textual relation between them, and whether one is dependent on the other for its own interpretation. In other

words, it may be difficult to decide whether they are only in a coreference or also in an anaphoric relation.

The difference between anaphora and generic coreference can be understood in cognitive terms but has also computational implications. In fact, anaphora relations are based on textual structural mechanisms that are assumed to facilitate the recovery of the relation between the anaphor and the antecedent both for humans and machines. For example, the search for the antecedent of a pronoun can be limited to the sentences immediately preceding the pronoun, possibly exploiting morphological agreement constraints or syntactic restrictions. Or consider the short text mentioned above: “The ambassador is going to meet the king of Spain. The arrival of the king is planned for tomorrow”. Given that “the king” comes immediately after a sentence mentioning “the king of Spain” (partial repetition), we can straightforwardly assume that they are in anaphoric relation and thus corefer. If the two phrases were far away from each other in the same text, or even more if they belonged to two different texts, deciding whether the two expressions corefer could require complex inferential steps.

3.2.2. The Annotation Scheme

As none of the existing schemes (see Section 2) seemed to fully fit our requirements, we developed a partially new annotation scheme, which attends to the needs of textual relations resolution for inference purposes while reflecting the distinction between anaphora and coreference as explained above. In order to facilitate such distinction during the annotation, the following principles have been followed:

- Pronouns are always anaphoric and coreferential - by their own nature
- Complete named entities (e.g. “Camilla Parker Bowles”) are never considered anaphoric and can only be part of a coreferential relation
- A noun phrase (NP) in an identity relation with another textual element, about which it is difficult to decide whether the relation is only coreferential or also anaphoric (see section 3.2.1), is annotated according to the following operational rule:
 - if the closest element coreferring with the NP is in the previous sentence, the relation between the two items is classified also as anaphora
 - otherwise, the two items are considered only coreferential

Two additional rules have been followed in annotating anaphora and coreference:

- for each anaphor, the annotation is not limited to the anaphor/antecedent pair, but an anaphoric chain is created up to the point where the information relevant to inference can be retrieved, for example:

*Ex. 2: [...] infrastructures that are built on **permafrost**. If **it** stays hard you can build on **it**. But **it** starts to melt [...]*

- for all the markables in coreferential relations, the complete coreference chains in the text are annotated.

The annotation has been carried out using the CLARK system⁵ and following the annotation specifications which are shown in Table 1.

The extent of the markables has been determined as follows:

- for noun phrases, the markable includes the head, its determiners, and its left- and right modifiers;
- in verb phrases, only the head is marked.

From the point of view of the actual annotation, the information about coreference and anaphora is inserted in the corpus by adding the element <REF> to the markables. The <REF> element contains several attributes, some used to annotate anaphora and others to annotate coreference.

All markables have the attributes MENT_ID, which identifies it univocally; SYNT_TYPE, which indicates its syntactic type; and ENT_SEM, which indicates its semantic type (e.g. entity, event, etc.).

When a markable is in an anaphoric relation, the following attributes are added:

- ANTE_ID, indicating the identifier of its antecedent;
- REF_TYPE, specifying the anaphora type – either *coreferential*, *bound* or *bridging*;
- REL_SEM, indicating the semantic relation between the entity/event referred to by the anaphor and its antecedent. In case of coreferential anaphora, the relation is always *Identity*, meanwhile in case of bridging anaphora it can vary (*Part-of*, *Participant*, etc.);
- ENTITIES_SEM, indicating the semantic type of the entities involved in the anaphora;
- SCOPE, indicating whether the antecedent is in the same sentence as the anaphor, or in a different part of the text;
- DIRECTION, indicating whether it is an anaphora or a cataphora;

Ex. 3:

```
<REF [...] MENT-ID="815">The A380</REF>,
will roll ...
<REF MENT_ID="816"
SYNT-TYPE="NP"
ENT_SEM="Event"
ANTE_ID="815"
REF_TYPE="Bridging anaphora" REL_SEM="Event-of"
ENTITIES-SEM="Event-Entity"
SCOPE="Extra" DIRECTION="Backward"
REPLACEMENT="sales of Airbus A380" >
Sales</REF> have beat expectations [...]
```

When a markable is in a coreferential relation, the attribute ENT_ID, which univocally identifies the entity or the event in the real world to which it refers, is added:

Ex. 4:

```
The modern collapse of the <REF MENT_ID="199"
SYNT_TYPE="PROPNAME"
ENT_SEM="ENTITY"
ENT_ID="LARSEN B"
```

Larsen B iceshelf</REF> is a unique event [...]

Finally, the REPLACEMENT attribute is added to all markables and is used to create the augmented data set (see Section 3.3).

3.3. The Augmented RTE-5 Data Set.

Based on the annotation, a new “augmented” data set is generated, where the implicit information relevant to inference is made explicit by automatically replacing the markable with the value of its REPLACEMENT attribute.

The REPLACEMENT value is created in different ways, depending on the type of the textual relation in which the markable is involved. In the cases of simple coreference and of anaphoric/coreferential relation, the replacement string contains the most explicit mention in the anaphoric/coreferential chain. For instance, in Example 5 below, the replacement string for “the iceshelf” is not its closest antecedent “it” (nor, in the same sentence, “LIS-B”), as the most explicit mention in the anaphoric/coreferential chain is chosen, i.e. “The Larsen B iceshelf”.

Ex. 5:

S1: The modern collapse of **the Larsen B iceshelf** is a unique event...S2: **The LIS-B** thinned to the point where **it** succumbed.

S3: **The iceshelf** disintegrated in February 2002.

S-3 REPLACEMENT="THE LARSEN B ICESHELF"

S3-augmented: THE LARSEN B ICESHELF disintegrated in February 2002.

In case of bridging anaphora, the replacement string puts together the information expressed by both the anaphor and the most explicit mention(s) in the coreferential/anaphoric chain starting from the antecedent.

Ex. 6:

S1: **Prince Charles** announced his official engagement to **Camilla Parker Bowles**.

S2: **Charles** and **Camilla** will get married next month...

S3: **The wedding** will take place at Windsor Castle.

S3-REPLACEMENT="THE WEDDING OF PRINCE CHARLES AND CAMILLA PARKER BOWLES"

S3-augmented: THE WEDDING OF PRINCE CHARLES AND CAMILLA PARKER BOWLES will take place at Windsor Castle.

⁵ <http://www.bultreebank.org/clark/index.html>

Attribute name	Description	Examples
MENT_ID	Unique numerical identifier of the markable	
SYNT_TYPE	Syntactic type of the markable, e.g. NP, VP, PRON, ZERO, PROPNAME; APNAME (abbreviated proper name).	
ENT_SEM	Semantic type of the markable. Possible values are: <i>entity, event, time, location, main location</i>	
ANTE_ID	Unique numerical identifier of the anaphor's antecedent	
REF_TYPE	Anaphora type. Possible values:	
	– <i>Co-Referential anaphora</i>	Prince Charles and Camilla will get married [...]. Finally they (<i>anaphor</i>) have [...]
	– <i>Bound-variable anaphora</i>	Every man has his own agenda.
REL_SEM	– <i>Bridging anaphora</i>	Charles and Camilla will get married [...]. The wedding (<i>anaphor</i>) will [...]
	Semantic type of the relation between the entity or the event referred to by the anaphor and its antecedents. Possible values are:	
	– <i>Identity (the anaphor and its antecedent refer to the same entity/event)</i>	Indigenous people [...] urged European countries to step up the fight against global warming , saying it (<i>anaphor</i>) is threatening
	– <i>Participant (the anaphor refers to a participant of the event referred to by the antecedent)</i>	
	– <i>Part-of (the anaphor refers to a part of the entity referred to by the antecedent)</i>	Disney officials consulted [...] experts before building Hong Kong Disneyland , making such changes as [...] “no fire” zones in kitchens (<i>anaphor</i>).
	– <i>Sub-event (the anaphor refers to an event which is part of the event referred to by the antecedent)</i>	
	– <i>Event-of (the anaphor refers to the event of which the entity referred to by the antecedent is a participant)</i>	Steve Fossett's GPS malfunctioned. The plane flight (<i>anaphor</i>) is [...].
	– <i>Space-Located (the anaphor refers to the event which takes place in the location referred to by the antecedent)</i>	[...] temperatures in the Arctic will increase [...]. The softening permafrost (<i>anaphor</i>) has [...].
– <i>Time-Located (the anaphor refers to the event which takes place at the time referred to by the antecedent).</i>		
ENTITIES_SEM	Semantic type of the entities involved in the anaphora, expressed by a binary combination between: entity, event, time, location (e.g. <i>entity-entity, entity-event</i>).	
SCOPE	Indication of whether the anaphora elements are in the same sentence (<i>Value: Intra</i>) or in a different segment of the text (<i>Value: Extra</i>)	
DIRECTION	Indication of whether the antecedent appears before its anaphor (<i>Value: Backward</i>), or after (<i>Value: Forward</i>)	
ENT_ID	Unique textual identifier of the entity or event in the world to which the markable refers	
REPLACEMENT	Substitution phrase that replaces the markable to make it explicit (in the “augmented” data set)	
COMM	Field to add comments to the annotation	

Table 1: Annotation specification.

As Example 6 shows, the replacement string combines the anaphor (“the wedding”) and the most specific mentions (i.e. “Prince Charles” and “Camilla Parker Bowles”) found in the chain starting from the antecedents (i.e. “Charles” and “Camilla”). In such cases, as the relation is of non-identity and the information comes from different parts of the text, morpho-syntactic adjustments are needed in order to obtain a well-formed augmented sentence expressing the relation of the bridging anaphora. For instance, in Example 6 the conjunction “and” and the preposition “of” are added in order to generate a correct sentence in English, expressing the information which is implicit in the original sentence. The need to respect grammaticality in the augmented sentence has a direct impact on the extent of the markable. In fact, there are cases in which replacing the markable with the value of the REPLACEMENT attribute would not generate a well-formed augmented sentence.

Consider, for instance, Example 7:

Ex: 7:

```
<REF [...] MENT_ID="306"> China</REF>'s mines are
said to be the most dangerous in the world [...] The
government says it
<EXT_MENT_ID="600">
<REF [...] ANTE_ID="306">
REF_TYPE="Bridging anaphora"
REPLACEMENT="has shut down hundreds of unsafe mines in
China"
TO_REPLACE="600">has shut down</REF> hundreds of
unsafe mines</EXT_MENT>.
```

In this example, the information “in China” needs to be included in the replacement string, as relevant to the inference. However, if the markable “has shut down” (i.e. the head of VP) were substituted with the replacement string “has shut down in China”, the resulting augmented sentence would be ungrammatical, splitting the verb and its object complement “hundreds of unsafe mines”.

In order to deal with such cases, the markable and all the other text elements which are to be substituted in the augmented sentence are marked up with the element <EXT_MENT>, containing a unique ID. The attribute TO_REPLACE is then added to the element <REF> of the markable, pointing to the ID of the element <EXT_MENT> and indicating that the portion of text to be substituted is not the markable itself, but the entire segment marked up with the <EXT_MENT> element. The value of the attribute REPLACEMENT is set accordingly, including all the text elements to be replaced according to the correct syntax.

3.4. The Corpus Up to Now

At the moment, one annotation of the corpus has been carried out. Table 2 shows some statistics based on a first round of annotation⁶. Before releasing the annotated corpus, the coreference chains - which have not been fully annotated yet - will be completed, and a

cross-annotation will be carried out, providing data about the inter-annotator agreement. It is interesting to note that about 80% of the entailing sentences (919 out of 1156) are annotated as containing anaphora and/or coreference, which confirms the relevance of discourse references for Textual Entailment.

Sentences in the data set	4,648	
Sentences containing at least one markable	1,661	
Sentences entailing at least one H	1,156	
Entailing sentences containing at least one markable	919	
Markables in the data set	3,515	
Anaphors	Coreferential anaphors	1,181
	Bridging anaphors	204
	Total	1,385

Table 2: Annotation statistics.

The annotated RTE-5 dataset and the augmented dataset will be publically available upon the publication of this paper, together with the annotation guidelines.

4. Considering Discourse Information by Inference Systems

Based on the annotation presented in Section 3, and considering the various phenomena specified in the annotation, we have investigated the kinds of discourse-induced operations that inference systems should take into account. We suggest that there are three main kinds of operations incorporating discourse relations that an inference engine should be able to perform. By applying these operations on text sentences we wish to automatically generate resolved sentences, similar to the “augmented sentences” generated through the manual reformulation in the annotated data set, but without requiring the manual specification of a REPLACEMENT attribute. A resolved sentence requires no further discourse-resolution operations, but can rather be used independently to infer the hypothesis. The goal is that an entailment system that can apply such operations will be able to generate such a resolved sentence based on the output of a discourse resolver.

Table 3 presents examples of these operations based on sentences and hypotheses from the RTE-5 development set. In these examples, a sentence (T) which entails the corresponding hypothesis (H), is modified on the basis of the information found in another sentence (T’), to generate the resolved sentence (T*). Note that other types of non-discourse inference operations may still be required to infer H from T*. For instance, in Example 3 in the Table, after textual relations have been resolved in T*, it is still necessary to infer that from a Russian point of view, the United States’ assistance is considered “international help”. The three operation types are presented below:

1) Substitution: This is the simplest and most common operation that corresponds to most nominal coreference and anaphora relations. In this operation, a mention of an entity or event in T (the anaphor) is

⁶ Final numbers may vary following cross-annotation.

Operation	Text		Hypothesis
(1) Substitution	T'	If the terrorists' main aim in London last week was simply to kill people with bombs on public transport, their attacks were a grim success.	The attack occurred in London
	T	Mayor Ken Livingstone, rode the Underground to work on Monday, four days after bomb attacks on the British capital's transit system killed at least 52 people.	
	T*	Mayor Ken Livingstone, rode the Underground to work on Monday, four days after bomb attacks on London's transit system killed at least 52 people.	
(2) Merge	T'	The AS-28 submarine got entangled in fishing nets.	The AS-28 mini submarine was trapped underwater
	T	The mini submarine had become attached to an as yet unidentified object during a routine exercise in the Bay of Berezovaya.	
	T*	The AS-28 mini submarine had become attached to an as yet unidentified object during a routine exercise in the Bay of Berezovaya.	
(3) Merge	T'	The Russian navy called for international help to rescue the AS-28.	Russia requested international help to rescue the AS-28
	T	Russia's Pacific Fleet commander was in talks with top U.S. Navy officers over how the United States might help .	
	T*	Russia's Pacific Fleet commander was in talks with top U.S. Navy officers over how the United States might help rescue the AS-28 .	
(4) Insertion	T'	If Arctic temperatures continue to rise, many of the lakes that are now ubiquitous in high northern latitudes could eventually disappear.	Ice is melting in the Arctic
	T	If permafrost continues to melt , it could also affect everything from oil platforms to landing strips.	
	T*	If permafrost continues to melt in the Arctic , it could also affect everything from oil platforms to landing strips.	
(5) Insertion	T'	The National Transportation Safety Board said it wants all 50 states to ban those with learner's permits from using cell phones while driving .	Some places ban the use of cell phones by drivers with learner's permits.
	T	New Jersey and Maine are the only two that have passed such laws .	
	T*	New Jersey and Maine are the only two that have passed such laws which ban those with learner's permits from using cell phones while driving .	

Table 3: Examples for the discourse-based operations in the annotated data set. T and T' refer to original text sentences in the RTE-5 development set and T* is the resolved sentence generated based on the resolution of reference relations. The referring mentions are shown in boldface; some sentences were simplified for brevity.

completely replaced by a referring mention (the antecedent) found in another sentence, usually a previous one. For instance, in Example 1 in Table 3, the phrase “the British capital” is entirely substituted by the coreferring mention of the entity “London” in T'.

2) **Merge**: In this operation, applicable to phrases with identical or coreferring heads, both sentences T and T' contribute information to the resolved sentence. In Example 2 in Table 3, the coreferring mentions of the submarine in T' (“AS-28 submarine”) and in T (“the mini submarine”) are merged to generate an unambiguous phrase identifying the submarine as it appears in H: “the AS-28 mini submarine”. This operation also seems to be particularly useful for coreferring predicates. It should be applied, for example, when some of the predicate's arguments are mentioned in one sentence and some in other. Example 3 in the Table is such a case where the subject of the predicate “help” is provided in T, while T' provides a complement for the predicate.

3) **Insertion**: In the above two operations, all the information in the resolved sentences explicitly appears in the original sentences containing the references, and is used in the resolved sentence in the same syntactic relation as in the original sentences. However, sometimes, the resolved sentence needs to be complemented with relational information not appearing explicitly in any of the original sentences. Examples 4 and 5 in Table 3 demonstrate this operation: in Example 4, the location

“Arctic”, mentioned in T' is connected to the event “melt” in T by the preposition “in” inserted into T*; in example 5, the description of “such laws”, mentioned in T', is connected to this phrase in the resolved sentence using a clause marker “which”.

With respect to the actual correlation between the annotation of textual relations and the inference operations that need to be performed, we observed that coreference relations and anaphoric identity relations typically activate the substitution operation, especially if the relation is between noun phrases. Merge operations seem particularly useful for coreferring predicates, or to noun phrases with multiple modifiers, meanwhile the insertion operation typically corresponds to bridging anaphora. Specifically it can be applied, for example, to space-located or time-located relations using spatial or temporal prepositions, respectively, and to the Part-of relation using genitive markers such as “'s” or “of” (e.g. “the government of Turkey”).

In general, it is desired to find out whether different discourse phenomena, as reflected by different annotation types, correspond to distinct operations to be applied by the inference system. Identifying such correlation is operationally valuable, as it enables the inference system to determine its action based on the discourse phenomenon, as identified by a resolver of coreference and anaphora relations. Further research is required to investigate the extent of such correlation between coreference and anaphora relations and

inference operations. For that purpose, the annotated RTE-5 data set and in particular the related augmented data set constitute invaluable resources.

5. Conclusions

In this paper we present an experimental study aimed at assessing the relevance of anaphora and coreference for Textual Entailment, and consisting of a comprehensive annotation of inference-oriented discourse phenomena in the RTE-5 Search data set. This annotation does not only help outline the different coreference and anaphora phenomena to be addressed while dealing with inference needs, but provides also information on the distribution of the different phenomena, allowing system developers to prioritize them by their importance. Moreover, using the information provided in the annotation, a new “augmented” data set has been automatically created, where the implicit information relevant to inference is made explicit. This data set can be used by inference system developers to assess the performance of other components of their systems, after the impact of inference-oriented discourse issues has been removed.

The study shows also that different discourse relations may correspond to different operations, thus suggesting a practical way to address some discourse issues within inference systems.

6. References

- Bentivogli, L., Dagan, I., Dang, H.T., Giampiccolo, D., Lo Leggio, M., Magnini, B. (2009a). Considering Discourse References in Textual Entailment Annotation. In *Proceedings of the 5th International Conference on Generative Approaches to the Lexicon (GL 2009)*, Pisa, Italy.
- Bentivogli, L., Dagan, I., Dang H.T., Giampiccolo, D., Magnini, B. (2009b). The Fifth PASCAL Recognizing Textual Entailment Challenge. In *Proceedings of the TAC Workshop*, Gaithersburg, MD, USA.
- Dagan, I., Glickman, O. (2004). Probabilistic textual entailment: Generic applied modeling of language variability. In *PASCAL Workshop on Learning Methods for Text Understanding and Mining, Grenoble, France*.
- Davies S., Poesio M., Bruneseaux F., Romary, L. (1998). *Annotating Coreference in Dialogues: Proposal for a Scheme for MATE*. Available at www.cogsci.ed.ac.uk/~poesio/mate/anno_manual.html.
- Fligelstone, S. (1992). Developing a Scheme for Annotating Text to Show Anaphoric Relation. In G. Leitner (Ed.), *New direction in English Language Corpora: Methodology, Results, Software Developments*, Berlin.
- Garside, R., Fligelstone, S., Botley, S. (1997). Discourse Annotation: Anaphoric Relations in Corpora. In R. Garside, G. Leech & A. McEnery (Eds.), *Corpus Annotation: Linguistic Information from Computer Text Corpora*, London.
- Ge, N. (1998). *Annotating the Penn Treebank with Coreference Information*. Internal report, Department of Computer Science, Brown University.
- Hasler, L., Orasan, C., Naumann K. (2006). NPs for Events: Experiments in Coreference Annotation. In *Proceedings of the 5th edition of the International Conference on Language Resources and Evaluation (LREC2006)*, Genoa, Italy.
- Hirshman, L., Chinchor, N. (1997), *MUC-7 Coreference Task Definition*. Version 3. Available at http://www-nlpir.nist.gov/related_projects/muc/proceedings/muc_7_toc.html.
- Huang, Y. (2000). *Anaphora. A Cross-linguistic Study*. Oxford.
- Leech, G., Garside, R. (1991). Running a Grammar Factory: The Production of Syntactically Analysed Corpora or Treebanks. In: S. Johansson and A-B. Stenström (Eds.), *English Computer Corpora: Selected Papers and Research Guide*, Berlin.
- Mitkov, R. (2003). Anaphora Resolution. In R. Mitkov (Ed.), *The Oxford Handbook of Computational Linguistics*, Oxford.
- Mitkov, R. (2008). Corpora for anaphora and coreference resolution. In A. Lüdeling & M. Kytö (Eds.), *An International Handbook*, Berlin.
- Mitkov, R., Evans, R., Orasan, C., Barbu, C., Jones, L., Sotirova, V. (2000). Coreference and Anaphora: Developing Annotating Tools, Annotated Resources and Annotation Strategies. In *Proceedings of the Discourse, Anaphora and Reference Resolution Conference (DAARC2000)*, Lancaster, UK.
- Passonneau, R., Litman, D. (1997). Discourse Segmentation by Human and Automated Means. In *Computational Linguistics* 23(1), 3--139.
- Poesio, M. (2004). Discourse Annotation and Semantic Annotation in the GNOME Corpus. In *Proceedings of the 2004 ACL Workshop on Discourse Annotation*, Barcellona.
- Poesio, M., Artstein, R. (2008). Anaphoric Annotation in the ARRAU Corpus. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.
- Poesio, M., Delmonte, R., Bristot, A., Chiran, L., Tonelli, S. (2004). *The Venex corpus of anaphoric information in spoken and written Italian*. Essex NLE Technical Note 2003-1. Available at <http://cswww.essex.ac.uk/staff/poesio/publications/VENEX04.pdf>.
- Poesio, M., Kruschwitz, U., Chamberlain, J. (2008). ANAWIKI: Creating Anaphorically Annotated Resources through Web Cooperation. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.
- Taulé, M., Martí, M.A., Recasens, M. (2008) *Ancora: Multilevel Annotated Corpora for Catalan and Spanish*. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco.