

A Semantic Approach to Recognizing Textual Entailment

Marta Tatu and Dan Moldovan

Language Computer Corporation

Richardson, TX 75080, USA

marta,moldovan@languagecomputer.com

Abstract

Exhaustive extraction of semantic information from text is one of the formidable goals of state-of-the-art NLP systems. In this paper, we take a step closer to this objective. We combine the semantic information provided by different resources and extract new semantic knowledge to improve the performance of a *recognizing textual entailment* system.

1 Recognizing Textual Entailment

While communicating, humans use different expressions to convey the same meaning. Therefore, numerous NLP applications, such as, Question Answering, Information Extraction, or Summarization require computational models of language that recognize if two texts semantically overlap. Trying to capture the major inferences needed to understand equivalent semantic expressions, the PASCAL Network proposed the *Recognizing Textual Entailment* (RTE) challenge (Dagan et al., 2005). Given two text fragments, the task is to determine if the meaning of one text (the entailed hypothesis, H) can be inferred from the meaning of the other text (the entailing text, T).

Given the wide applicability of this task, there is an increased interest in creating systems which detect the semantic entailment between two texts. The systems that participated in the Pascal RTE challenge competition exploit various inference elements which, later, they combine within statistical models, scoring methods, or machine learning

frameworks. Several systems (Bos and Markert, 2005; Herrera et al., 2005; Jijkoun and de Rijke, 2005; Kouylekov and Magnini, 2005; Newman et al., 2005) measured the word overlap between the two text strings. Using either statistical or WordNet's relations, almost all systems considered lexical relationships that indicate entailment. The degree of similarity between the syntactic parse trees of the two texts was also used as a clue for entailment by several systems (Herrera et al., 2005; Kouylekov and Magnini, 2005; de Salvo Braz et al., 2005; Raina et al., 2005). Several groups used logic provers to show the entailment between T and H (Bayer et al., 2005; Bos and Markert, 2005; Fowler et al., 2005; Raina et al., 2005) and some of them made use of world knowledge axioms to increase the logic prover's power of inference (Bayer et al., 2005; Bos and Markert, 2005; Fowler et al., 2005).

In this paper, we describe a novel technique which employs a set of *semantic axioms* in its attempt to exhaustively extract semantic knowledge from texts. In order to show the contribution that our semantic information extraction method brings, we append it as an additional module to an already existing system that participated in the RTE challenge. Our system (Fowler et al., 2005), first, transforms the text T and the hypothesis H into semantically enhanced logic forms, and, then, the integrated logic prover tries to prove or disprove the entailment using a set of world-knowledge axioms (*die of blood loss* \rightarrow *bleed to death*), linguistic rewriting rules which break down complex syntactic structures, like coordinating conjunctions, and WordNet-based lexical chains axioms (*buy/VB/I* \rightarrow *pay/VB/I*).

2 Approach

We believe that a logic-based semantic approach is highly appropriate for the RTE task¹. Text T semantically entails H if its meaning logically implies the meaning of H . Because the set of semantic relations encoded in a text represents its meaning, we need to identify all the semantic relations that hold between the constituents of T and, subsequently, between the constituents of H to understand the meaning of each text. It should be noted that state-of-the-art semantic parsers extract only some of the semantic relations encoded in a given text. To complete this information, we need semantic axioms that augment the extracted knowledge and, thus, provide a better coverage of the text's semantics. Once we gather this information, we state that text T entails hypothesis H if and only if we find similar relations between a concept from T and a semantically analogous concept from H . By analogous concepts, we mean identical concepts, or words connected by a chain of SYNONYMY, HYPERNYMY or morphological derivation relations in WordNet.

Because the set of semantic elements identified by a semantic parser does not necessarily convey the complete meaning of a sentence, we shall use a set of semantic axioms to infer the missing pieces of information. By combining two semantic relations or by using the FrameNet's frame elements identified in a given text, we derive new semantic information.

In order to show if T entails H , we analyze their meanings. Our approach to semantic entailment involves the following stages:

1. We convert each text into logic form (Moldovan and Rus, 2001). This conversion includes part-of-speech tagging, parse tree generation, and name entity recognition.
2. Using our semantic parser, we identify some of the semantic relations encoded in the analyzed texts. We note that state-of-the-art semantic parsers cannot discover all the semantic relations conveyed implicitly or explicitly by the text. This problem compromises our system's performance. To obtain the complete set of semantic relations that represents the meaning of the given texts, we introduce a new step in our algorithm.

¹After all, the *entailment*, *inference*, and *equivalence* terms originated from logic.

3. We add *semantic axioms* to the already created set of world knowledge, NLP, and WordNet-based lexical chain (Moldovan and Novischi, 2002) axioms that assist the logic prover in its search for proofs. We developed semantic axioms that show how two semantic relations can be combined. This will allow the logic prover to combine, whenever possible, semantic instances in order to infer new semantic relationships. The instances of relations that participate in semantic combinations can be either provided by the text or annotated between WordNet synsets. We also exploit other sources of semantic information from the text. For example, the frames encoded in the text sentence provide information which complements the meaning given by the semantic relations. Our second type of axioms derive semantic relations between the frame elements of a given FrameNet frame.

We claim that the process of applying the semantic axioms, given the semantic relations detected by a semantic parser, will capture the complete semantic information expressed by a text fragment. In this paper, we show the usefulness of this procedure for the RTE task, but we are convinced that it can be used by any system which plans to extract the entire semantic information from a given text.

4. We load the COGEX logic prover (Moldovan et al., 2003) which operates by "reductio ad absurdum" with H 's negated form and T 's predicates. These clauses are weighted in the order in which they should be chosen to participate in the search. To ensure that H will be the last clause to participate, we assign it the largest value. The logic prover searches for new inferences that can be made using the smallest weight clauses. It also assigns a value to each inference based on the axiom it used to derive it. This process continues until the set of clauses is empty. If a refutation is found, the proof is complete. If a contradiction cannot be found, then the predicate arguments are relaxed and, if the argument relaxation fails, then predicates are dropped until a proof by refutation is found. Its score will be computed by deducting points for each argument relaxation and predicate removal. If this value falls below a threshold, then T does not entail H . Otherwise, the (T, H) pair is a true entailment.

We present a textual entailment example to show the steps of our approach. This proof will not

T	John and his son, George, emigrated with Mike, John's uncle, to US in 1969.
\mathcal{LF}_T	$\text{John}(x_1) \wedge \text{son}(x_2) \wedge \text{George}(x_3) \wedge \text{ISA}(x_3, x_2) \wedge \text{KIN}(x_1, x_3) \wedge \text{emigrate}(e_1) \wedge \text{AGT}(x_1, e_1) \wedge \text{AGT}(x_2, e_1) \wedge \text{Mike}(x_4) \wedge \text{uncle}(x_5) \wedge \text{ISA}(x_4, x_5) \wedge \text{KIN}(x_1, x_4) \wedge \text{US}(x_6) \wedge \text{LOC}(e_1, x_6) \wedge 1969(x_7) \wedge \text{TMP}(e_1, x_7)$.
$T_{\text{Departing}}$	[John and his son, George,] Theme.fe emigrated [with Mike, John's uncle,] Cotheme.fe to [US] Goal.fe in [1969] Time.fe .
T_{Kinship}	[John] Ego.fe and his son, [George,] Alter.fe emigrated with [Mike] Alter.fe , [John] Ego.fe 's uncle, to US in 1969.
T_{Axiom_1}	$\text{KIN}(w_1, w_2) \leftrightarrow \text{KIN}(w_2, w_1)$ $\text{KIN}(x_1, x_3) \rightarrow \text{KIN}(x_3, x_1)$ ($\text{KIN}(\text{John}, \text{George}) \rightarrow \text{KIN}(\text{George}, \text{John})$)
T_{Axiom_2}	$\text{KIN} \circ \text{KIN} = \text{KIN}$ ($\text{KIN}(w_1, w_2) \wedge \text{KIN}(w_2, w_3) \rightarrow \text{KIN}(w_1, w_3)$) $\text{KIN}(x_3, x_1) \wedge \text{KIN}(x_1, x_4) \rightarrow \text{KIN}(x_3, x_4)$ ($\text{KIN}(\text{George}, \text{Mike})$)
T_{Axiom_3}	$\text{DEPARTING.F} \rightarrow \text{LOC}(\text{Theme.fe}, \text{Goal.fe})$ ($\text{LOC}(\text{John}, \text{US}) \wedge \text{LOC}(\text{George}, \text{US})$)
T_{Axiom_4}	$\text{DEPARTING.F} \rightarrow \text{LOC}(\text{Cotheme.fe}, \text{Goal.fe})$ ($\text{LOC}(\text{Mike}, \text{US})$)
$T_{\text{Semantics}}$	$\text{KIN}(\text{John}, \text{George})$, $\text{KIN}(\text{John}, \text{Mike})$, $\text{KIN}(\text{George}, \text{Mike})$, $\text{LOC}(\text{John}, \text{US})$, $\text{LOC}(\text{George}, \text{US})$, $\text{LOC}(\text{Mike}, \text{US})$, $\text{TMP}(\text{emigrate}, 1969)$, $\text{AGT}(\text{John}, \text{emigrate})$, $\text{AGT}(\text{George}, \text{emigrate})$
H	George and his relative, Mike, came to America.
\mathcal{LF}_H	$\text{George}(x_1) \wedge \text{relative}(x_2) \wedge \text{Mike}(x_3) \wedge \text{ISA}(x_3, x_2) \wedge \text{KIN}(x_1, x_3) \wedge \text{come}(e_1) \wedge \text{AGT}(x_1, e_1) \wedge \text{AGT}(x_2, e_1) \wedge \text{America}(x_4) \wedge \text{LOC}(e_1, x_2)$
H_{Arriving}	[George and his relative, Mike,] Theme.fe came to [America] Goal.fe .
H_{Kinship}	[George] Ego.fe and his relative, [Mike] Alter.fe , came to America.
H_{Axiom_1}	$\text{ARRIVING.F} \rightarrow \text{LOC}(\text{Theme.fe}, \text{Goal.fe})$ ($\text{LOC}(\text{George}, \text{America}) \wedge \text{LOC}(\text{Mike}, \text{America})$)
$H_{\text{Semantics}}$	$\text{KIN}(\text{George}, \text{Mike})$, $\text{LOC}(\text{George}, \text{America})$, $\text{LOC}(\text{Mike}, \text{America})$

Table 1: Entailment proof example. Table 2 lists the semantic relations and their abbreviations. Sections 3.2 and 4.1 will detail the semantics behind the axioms T_{Axiom_1} , T_{Axiom_2} , T_{Axiom_3} , T_{Axiom_4} , and H_{Axiom_1} .

make use of any world knowledge axioms. Let the text T be *John and his son, George, emigrated with Mike, John's uncle, to US in 1969* and the entailed hypothesis H *George and his relative, Mike, came to America*. Our system transforms each text into its corresponding semantically enhanced logic form (\mathcal{LF}_T and \mathcal{LF}_H in Table 1). Then, the logic prover uses the newly added semantic axioms to derive extra semantic information from T and H (for example, *George* and *Mike* are relatives, but T does not explicitly specify this), after another preprocessing step which identifies the frame elements of each frame encoded in the two texts ($T_{\text{Departing}}$, T_{Kinship} , H_{Arriving} , H_{Kinship}). In our example, the axioms T_{Axiom_1} and T_{Axiom_2} denote the *symmetry* and the *transitivity* of the KINSHIP relation. T_{Axiom_3} , T_{Axiom_4} and H_{Axiom_1} are the frame-related axioms used by the logic prover. The $T_{\text{Semantics}}$ and $H_{\text{Semantics}}$ rows (Table 1) summarize the meaning of T and H . We note that half of these semantic instances were extracted using the semantic axioms. Once the lexical chains between the concepts in T and the ones from H are computed, the entailment becomes straightforward. We represented, graphically, the meaning of the two texts in Figure 1. We also show the links between the analogous concepts that help prove the entailment.

In the coming sections of the paper, we detail the process of semantic axiom generation. We start with a summary of the axioms that combine two semantic relations.

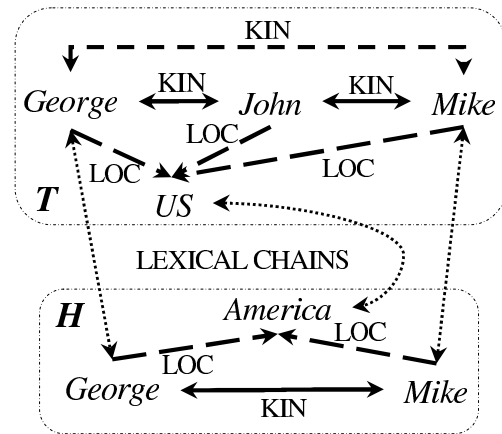


Figure 1: $T_{\text{Semantics}}$ and $H_{\text{Semantics}}$. The solid arrows represent the relations identified by the semantic parser. The dotted arrows symbolize the lexical chains between concepts in T and their analogous concepts in H (US_T and $America_H$ belong to the same WordNet synset). The dash arrows denote the relations inferred by combining two semantic relations. The long dash arrows indicate the relations between frame elements.

3 Semantic Calculus

3.1 Semantic relations

For this study, we adopt a revised version of the semantic relation set proposed by (Moldovan et al., 2004). Table 2 enumerates the semantic relations that we consider².

²See (Moldovan et al., 2004) for definitions and examples.

POSSESSION (POS)	MAKE-PRODUCE (MAK)	RECIPIENT (REC)	THEME-PATIENT (THM)
KINSHIP (KIN)	INSTRUMENT (INS)	FREQUENCY (FRQ)	RESULT (RSL)
PROPERTY-ATTRIBUTE (PAH)	LOCATION-SPACE (LOC)	INFLUENCE (IFL)	STIMULUS (STI)
AGENT (AGT)	PURPOSE (PRP)	ASSOCIATED WITH (OTH)	EXTENT (EXT)
TEMPORAL (TMP)	SOURCE-FROM (SRC)	MEASURE (MEA)	PREDICATE (PRD)
DEPICTION (DPC)	TOPIC (TPC)	SYNONYMY-NAME (SYN)	CAUSALITY (CSL)
PART-WHOLE (PW)	MANNER (MNR)	ANTONYMY (ANT)	JUSTIFICATION (JST)
HYPERNYMY (ISA)	MEANS (MNS)	PROBABILITY OF EXISTENCE (PRB)	GOAL (GOL)
ENTAIL (ENT)	ACCOMPANIMENT (ACC)	POSSIBILITY (PSB)	BELIEF (BLF)
CAUSE (CAU)	EXPERIENCER (EXP)	CERTAINTY (CRT)	MEANING (MNG)

Table 2: The set of semantic relations

3.2 Combinations of two semantic relations

Our goal is to devise semantic axioms for combinations of two relations, R_1 and R_2 , by observing the semantic connection between the w_1 and w_3 words for which there exists at least one other word, w_2 , such that $R_1(w_1, w_2)$ and $R_2(w_2, w_3)$ hold true³.

Harabagiu and Moldovan (1998) tackled the problem of semantic combinations, for the first time. Their set of relations included the WordNet1.5 annotations and 12 relationships derived from the WordNet glosses⁴. In our research, unlike (Harabagiu and Moldovan, 1998), the semantic combinations use the relations identified in text with a rather minimal contribution from the WordNet relations.

Harabagiu and Moldovan (1998) also investigate the number of possible semantic combinations. Based on their properties, we can have up to eight combinations between any two semantic relations and their inverses, not counting the combinations between a semantic relation and itself⁵. For instance, given an asymmetric relation and a symmetric one which share the same part-of-speech for their arguments, we can produce four combinations. $ISA \circ ANT$, $ISA^{-1} \circ ANT$, $ANT \circ ISA$, and $ANT \circ ISA^{-1}$ are the four possible distinct combinations between HYPERNYMY and ANTONYMY. “ \circ ” symbolizes the semantic composition between two relations compatible with respect to the part-of-speech of their arguments: for any two concepts, w_1 and w_3 , $(R_i \circ R_j)(w_1, w_3)$ if and only if $\exists w_2$, a third concept, such that $R_i(w_1, w_2)$ and $R_j(w_2, w_3)$ hold. By R^{-1} ,

³ $R(x, y)$ indicates that relation R holds between x and y .

⁴This set includes the AGENT, OBJECT, INSTRUMENT, BENEFICIARY, PURPOSE, ATTRIBUTE, REASON, STATE, LOCATION, THEME, TIME, and MANNER relations.

⁵Harabagiu and Moldovan (1998) lists the exact number of possible combinations for several WordNet relations and part-of-speech classes.

we denote the inverse of relation R : if $R(x, y)$, then $R^{-1}(y, x)$.

While analyzing the combinations, we observed some regularities within the semantic composition process. For example, $R_1^{-1} \circ R_2^{-1} = (R_2 \circ R_1)^{-1}$ for any, not necessarily distinct, semantic relations R_1 and R_2 ⁶. If one of the relations is symmetric ($R^{-1} = R$), the statement is still valid. Using $(R^{-1})^{-1} = R$ and the previous equality, we can reduce by half the number of semantic combinations that we have to compute for $R_1 \neq R_2$.

We plan to create a 40×40 matrix with all the possible combinations between any two semantic relations from the set we consider. Theoretically, we can have up to 27,556 semantic combinations, but only 25.79% of them are possible⁷ (for example, $MNR(r, v)$ and $SYN(n, n)$ cannot be combined). Many combinations are not semantically significant either because they are very rare, like, $KIN(n, n) \circ TMP(n, v)$, or because they do not result into one of the 40 relations, for instance, $PAH(a, n) \circ AGT(n, v)$ ⁸. We identified two approaches to the problem mentioned above. The first tries to fill one matrix cell at a time in a consecutive manner. The second approach tries to solve the semantic combinations we come upon in text corpora. As a result, we analyzed the RTE development corpus and we devised rules for some of the $R_i \circ R_j$ combinations that we encountered. We validated these axioms by man-

⁶The equality holds only if the two composition terms exist.

⁷On average, each semantic relation has 2.075 pairs of arguments. For example, SRC can connect two nouns (*US investor*), or an adjective and a noun (*American investor*) and, depending on its arguments, SRC will participate in different combinations. Out of the 27,556 combinations, only 7,109 are syntactically possible.

⁸ n, v, a , and r stand for *noun*, *verb*, *adjective*, and *adverb*, respectively. As an example, $R(n, n)$ means that relation R can connect two nouns.

$\text{LOCATION} \circ \text{PART-WHOLE} = \text{LOCATION} (\text{LOCATION}(x, l_1) \wedge \text{PART-WHOLE}(l_1, l_2) \rightarrow \text{LOCATION}(x, l_2))$ <i>Example:</i> John lives in Dallas, Texas. $\text{LOCATION}(\text{John}, \text{Dallas})$ and $\text{PART-WHOLE}(\text{Dallas}, \text{Texas})$ imply that $\text{LOCATION}(\text{John}, \text{Texas})$.
$\text{ISA} \circ \text{ATTRIBUTE} = \text{ATTRIBUTE} (\text{ISA}(x, y) \wedge \text{ATTRIBUTE}(y, a) \rightarrow \text{ATTRIBUTE}(x, a))$ <i>Example:</i> Mike is a rich man. If $\text{ISA}(\text{Mike}, \text{man})$ and $\text{ATTRIBUTE}(\text{man}, \text{rich})$, then $\text{ATTRIBUTE}(\text{Mike}, \text{rich})$.
Similar statements can be made for other “attributes”: LOCATION, TIME, SOURCE, etc. $\text{ISA} \circ \text{LOCATION} = \text{LOCATION} (\text{ISA}(x, y) \wedge \text{LOCATION}(y, l) \rightarrow \text{LOCATION}(x, l))$ <i>Example:</i> The man in the car, George, is an old friend of mine. $\text{ISA}(\text{George}, \text{man})$ and $\text{LOCATION}(\text{man}, \text{car}) \rightarrow \text{LOCATION}(\text{George}, \text{car})$
$\text{KINSHIP} \circ \text{KINSHIP} = \text{KINSHIP} (\text{KINSHIP}(x, y) \wedge \text{KINSHIP}(y, z) \rightarrow \text{KINSHIP}(x, z))$ See example in Section 2.
$\text{THEME} \circ \text{ISA}^{-1} = \text{THEME} (\text{THEME}(e, y) \wedge \text{ISA}(x, y) \rightarrow \text{THEME}(e, x))$ <i>Example:</i> Yesterday, John ate some fruits: an apple and two oranges. $\text{THEME}(\text{eat}, \text{fruit}) \wedge \text{ISA}(\text{apple}, \text{fruit}) \rightarrow \text{THEME}(\text{eat}, \text{apple})$
$\text{THEME} \circ \text{PART-WHOLE}^{-1} = \text{THEME} (\text{THEME}(e, y) \wedge \text{PART-WHOLE}(x, y) \rightarrow \text{THEME}(e, x))$ <i>Example:</i> Five Israelis, including two children, were killed yesterday. $\text{THEME}(\text{kill}, \text{Israeli}) \wedge \text{PART-WHOLE}(\text{child}, \text{Israeli}) \rightarrow \text{THEME}(\text{kill}, \text{child})$
Similar statements can be made for all the thematic roles: AGENT, EXPERIENCER, INSTRUMENT, CAUSE, LOCATION, etc. $\text{AGENT} \circ \text{ISA}^{-1} = \text{AGENT} (\text{AGENT}(e, y) \wedge \text{ISA}(x, y) \rightarrow \text{AGENT}(e, x))$ $\text{AGENT} \circ \text{PART-WHOLE}^{-1} = \text{AGENT} (\text{AGENT}(e, y) \wedge \text{PART-WHOLE}(x, y) \rightarrow \text{AGENT}(e, x))$

Table 3: Examples of semantic combination axioms

ually checking all the LA Times corpus (w_1, w_3) pairs which satisfy $(R_i \circ R_j)(w_1, w_3)$. We have identified 64 semantic axioms that show how semantic relations can be combined. These axioms use relations such as PART-WHOLE, ISA, LOCATION, ATTRIBUTE, or AGENT. We listed several example rules in Table 3. The 64 axioms can be applied independent of the concepts involved in the semantic composition. We have also identified rules that can be applied only if the concepts that participate satisfy a certain condition or if the relations are of a certain type. For example, $\text{LOC} \circ \text{LOC} = \text{LOC}$ only if the LOC relation shows inclusion ($\text{John is in the car in the garage} \rightarrow \text{LOC}(\text{John}, \text{garage})$). $\text{John is near the car behind the garage} \not\rightarrow \text{LOC}(\text{John}, \text{garage})$.

4 FrameNet Can Help

The Berkeley FrameNet project⁹ (Baker et al., 1998) is a lexicon-building effort based on the theory of *frame semantics* which defines the meanings of lexical units with respect to larger conceptual structures, called frames. Individual lexical units point to specific frames and establish a binding pattern to specific elements within the frame. FrameNet describes the underlying frames for different lexical units and examines sentences related to the frames using the BNC corpus. The result is an XML database that

⁹<http://framenet.icsi.berkeley.edu>

contains a set of frames, a set of frame elements for each frame, and a set of frame annotated sentences.

4.1 Frame-based semantic axioms

With respect to a given target, the frame elements contribute to the understanding of the sentence. But they only link each argument to the target word (for example, $\text{THM}(\text{theme}, \text{target})$ or $\text{AGT}(\text{theme}, \text{target})$, $\text{LOC}(\text{place}, \text{target})$, etc.). Often enough, we can find relations between the frame elements of a given frame. These new instances of semantic relations take as arguments the frame elements of a certain frame, when they are expressed in the text. For example, given the DEPARTING frame, we can say that the origin of the *theme* is the *source* ($\text{SRC}(\text{theme}, \text{source})$) and that the new location of the *theme* is the *goal* frame element ($\text{LOC}(\text{theme}, \text{goal})$). Moreover, if the text specifies the *cotheme* frame element, then we can make similar statements about it ($\text{SRC}(\text{cotheme}, \text{source})$ and $\text{LOC}(\text{cotheme}, \text{goal})$). These new relation instances increase the semantic information that can be derived from text.

So far, we manually inspected 54 frames and analyzed the relationships between their frame elements by examining their definitions and the annotated corpus provided with the FrameNet data. For each frame, we retained only the rules independent of the

CLOTHING_PARTS_F \rightarrow PW(<i>subpart, clothing</i>)
CLOTHING_PARTS_F \rightarrow PW(<i>material, subpart</i>) <i>Example:</i> “Hello, Hank” they said from the depths of the [fur] _{Material} [collars] _{Subpart,Target} of [their] _{Wearer} [coats] _{Clothing} . PW(<i>fur, collar</i>) and PW(<i>collar, coat</i>)
CLOTHING_F \rightarrow PAH(<i>descriptor, garment</i>) \vee PAH(<i>descriptor, material</i>) <i>Example:</i> She didn’t bring heels with her so she decided on [gold] _{Descriptor} [leather] _{Material} [flip-flops] _{Garment,Target} . PAH(<i>gold, leather</i>) \vee PAH(<i>gold, flip – flop</i>)
KINSHIP_F \rightarrow KIN(<i>ego, alter</i>) <i>Example:</i> The new subsidiary is headed by [Rupert Soames] _{Alter} , [son] _{Target} [of the former British Ambassador to France and EC vice-president] _{Ego} . KIN(<i>Rupert Soames, the former British Ambassador to France and EC vice-president</i>)
GETTING_F \rightarrow POS(<i>recipient, theme</i>)
GETTING_F \rightarrow \neg POS(<i>source, theme</i>) (only if the <i>source</i> is a person) <i>Example:</i> In some cases, [the BGS libraries] _{Recipient} had [obtained] _{Target} [copies of theses] _{Theme} [from the authors] _{Source} [by purchase or gift] _{Means} , and no loan records were available for such copies. POS(<i>the BGS libraries, copies of theses</i>) and \neg POS(<i>authors, copies of theses</i>)
GETTING_F \rightarrow SRC(<i>theme, source</i>) (if the <i>source</i> is not a person) <i>Example:</i> He also said that [Iran] _{Recipient} [acquired] _{Target} [fighter-bomber aircraft] _{Theme} [from countries other than the USA and the Soviet Union] _{Source} . SRC(<i>fighter-bomber aircraft, countries other than the USA and the Soviet Union</i>)

Table 4: Frames-related semantic rules

frame’s lexical units. We identified 132 semantic axioms that hold in most cases¹⁰. We show some examples in Table 4.

4.2 Context importance

There are cases when the rules that we identified should not be applied. Let’s examine the sentence *John intends to leave the kitchen*. If we consider only the DEPARTING frame and its corresponding rules, without looking at the context, then our conclusions (\neg LOC(*John, kitchen*) and SRC(*John, kitchen*)) will be false. This sentence states an intention of motion, not the actual action. Therefore, our semantic axioms apply only when the context they are in, allows it. To overcome this problem, we do not apply the axioms for target words found in planning contexts, contexts related to beliefs, intentions, desires, etc. As an alternative, we keep track of plans, intentions, desires, etc. and, if, later on, we confirm them, then we apply the semantic axioms. Also, when we analyze a sentence, the frame whose rules we apply needs to be chosen carefully. For example, in the sentence [A boat]_{Agent} [carrying]_{Target} [would-be Moroccan illegal emigrants]_{Theme} [from UK]_{Path_start} [to Spain]_{Path_end} sank in the Strait of Gibraltar on June 8, the CARRYING frame’s axioms do not apply. The boat nor the emigrants reach Spain (the *path_end* of

the motion) because the boat sank. Here, the rules given by *sink.v*’s frame should be given priority over the *carry.v*’s rules. We can generalize and conclude that, given a sentence that contains more than one target (therefore, maybe multiple frames), the dominant frame, the one whose rules should be applied, is the frame given by the predicative verb. In the previous sentence, the dominant frame is the one given by *sink.v* and its rules should be applied before the axioms of the CARRYING frame. It should be noted that some of the axioms semantically related to the CARRYING frame still apply (for example, SRC(*emigrants, UK*) or SRC(*boat, UK*)). Unlike LOC(*emigrants, Spain*), the previous relations do not conflict with the semantics given by *sink.v* and its location (*the Strait of Gibraltar*).

5 Experimental Results

5.1 The RTE data

The benchmark corpus for the RTE task consists of seven subsets with a 50%-50% split between the positive entailment examples and the negative ones. Each subgroup corresponds to a different NLP application: Information Retrieval (IR), Comparable Documents (CD), Reading Comprehension (RC), Question Answering (QA), Information Extraction (IE), Machine Translation (MT), and Paraphrase Acquisition (PP). The RTE data set includes 1367 English (*T, H*) pairs from the news domain (political, eco-

¹⁰Section 4.2 lists some exception cases.

Semantic Axioms	CD		IE		IR		MT		PP		QA		RC	
	T	F	T	F	T	F	T	F	T	F	T	F	T	F
Test data (%)														
applied to all T s	13.33	21.33	26.66	10	6.66	4.44	11.66	10	8	0	15.38	7.69	21.43	17.14
applied to all H s	1.33	9.33	5	10	0	0	1.66	1.66	8	0	1.53	0	0	1.43
solution for (T, H)	9.33	0	20	0	4.44	0	10	1.66	0	0	10.77	1.53	10	5.71
Development data (%)														
applied to all T s	22	27.08	34.28	5.71	8.57	8.57	18.51	18.51	5.12	9.3	28.88	0	9.61	9.8
applied to all H s	4	8.33	5.71	2.85	0	2.85	7.4	3.7	0	0	2.22	0	0	1.96
solution for (T, H)	10	2.08	22.85	5.71	5.71	2.85	18.51	3.7	2.56	0	20	0	7.69	0

Table 5: The impact of the semantic axioms on each NLP application data set. **T** and **F** stand for *True* and *False* entailments, respectively.

nomical, etc.). The development set consists of 567 examples and the test set contains the remaining 800 pairs.

5.2 Semantic axiom applicability

We measured the applicability of our set of semantic rules, by counting the number of times they extract new semantic information from text. Table 6 shows, in percentages, the coverage of the semantic axioms when applied to the texts T and the hypotheses H . We also show the number of times the semantic rules solve a (T, H) entailment without employing any other type of axioms.

Semantic Axioms	True	False	Overall (True and False)
Test data (%)			
applied to all T s	15.75	11.75	13.75
applied to all H s	2.00	3.74	2.87
both T s and H s	8.87	7.75	8.31
solution for (T, H)	10.25	1.50	5.87
Development data (%)			
applied to all T s	18.02	11.26	14.64
applied to all H s	2.47	2.81	2.65
both T s and H s	10.25	7.04	8.64
solution for (T, H)	12.36	1.76	7.05

Table 6: Applicability on the RTE data

Clearly, because the texts T convey much more information than H , they are the ones that benefit the most from our semantic axioms. The hypotheses H are more straightforward and a semantic parser can extract all their semantic information. Also, the rules tend to solve more positive (T, H) entailments. Because there are seven subsets corresponding to different NLP applications that make up the RTE data, we analyzed the contribution of our semantic axioms to each of the seven tasks. Table 5 shows the axioms' impact on each type of data. The

logic-based approach proves to be useful to tasks like Information Extraction, Reading Comprehension, or Comparable Documents, and it doesn't seem to be the right choice for the more lexical-orientated applications like Paraphrase Acquisition, Machine Translation, and Information Retrieval.

5.3 RTE performance

To show the impact of our semantic axioms, we measured the contribution they bring to a system that participated in the RTE challenge. The ACC and F columns (Table 7) show the performance of the system before and after we added our semantic rules to the list of axioms needed by the logic prover.

Task	Original		Enhanced	
	ACC	F	ACC	F
Test-IR	.478	.472	.5	.505
Test-CD	.78	.736	.847	.819
Test-RC	.514	.558	.6	.636
Test-QA	.485	.481	.523	.537
Test-IE	.483	.603	.575	.687
Test-MT	.542	.444	.567	.49
Test-PP	.45	.585	.44	.576
Test	.551	.561	.604	.621
Development	.63	.619	.718	.714

Table 7: The accuracy(ACC) and f-measure(F) performance values of our system

The results show that richer semantic connectivity between text concepts improve the performance of a semantic entailment system. The overall accuracy increases with around 5% on the test data and almost 8% on the development set. We obtained performance improvements for all application settings, except for the Paraphrase Acquisition task. For this application, we obtained the smallest axiom coverage (Table 5). The impact of the semantic axioms on each NLP application data set correlates with the

improvement that the addition of the rules brought to the system's accuracy.

Our error analysis showed that the system did not take full advantage of our semantic axioms, because the semantic parser did not identify all the semantic relations needed as building blocks by the axioms. We noticed a significant decrease in the logic prover's usage of world-knowledge axioms.

6 Conclusion

In this paper, we present a logic-based semantic approach for the *recognizing textual entailment* task. The system participating in the RTE competition used a set of world-knowledge, NLP, and lexical chain-based axioms and an in-house logic prover which received as input the logic forms of the two texts enhanced with semantic relation instances. Because the state-of-the-art semantic parsers cannot extract the complete semantic information encoded in text, the need for *semantic calculus* in NLP became evident. We introduce *semantic axioms* that either combine two semantic instances or label relations between the frame elements of a given frame. Preliminary statistical results show that incorporating semantic rules into the logic prover can double the semantic connectivity between the concepts of the analyzed text. Our process of identifying more semantic instances leads to a smaller dependency of the logic-based RTE system on world knowledge axioms, while improving its overall accuracy.

References

- Collin Baker, Fillmore Charles, and John Love. 1998. The Berkeley FrameNet project. In *Proceedings of COLING/ACL*.
- Samuel Bayer, John Burger, Lisa Ferro, John Henderson, and Alexander Yeh. 2005. MITRE's Submissions to the EU Pascal RTE Challenge. In *Proceedings of the PASCAL RTE Challenge*.
- Johan Bos and Katja Markert. 2005. Combining Shallow and Deep NLP Methods for Recognizing Textual Entailment. In *Proceedings of the PASCAL RTE Challenge*.
- Ido Dagan, Oren Glickman, and Bernardo Magnini. 2005. The PASCAL Recognising Textual Entailment Challenge. In *Proceedings of the PASCAL RTE Challenge*.
- Rodrigo de Salvo Braz, Roxana Girju, Vasin Punyakanok, Dan Roth, and Mark Sammons. 2005. An Inference Model for Semantic Entailment in Natural Language. In *Proceedings of the PASCAL RTE Challenge*.
- Abraham Fowler, Bob Hauser, Daniel Hodges, Ian Niles, Adrian Novischi, and Jens Stephan. 2005. Applying COGEX to Recognize Textual Entailment. In *Proceedings of the PASCAL RTE Challenge*.
- Sanda Harabagiu and Dan Moldovan. 1998. Knowledge Processing on Extended WordNet. In *WordNet: an Electronic Lexical Database and Some of its Applications*.
- Jess Herrera, Anselmo Peas, and Felisa Verdejo. 2005. Textual Entailment Recognition Based on Dependency Analysis and WordNet. In *Proceedings of the PASCAL RTE Challenge*.
- Valentin Jijkoun and Maarten de Rijke. 2005. Recognizing Textual Entailment Using Lexical Similarity. In *Proceedings of the PASCAL RTE Challenge*.
- Milen Kouylekov and Bernardo Magnini. 2005. Recognizing Textual Entailment with Tree Edit Distance Algorithms. In *Proceedings of the PASCAL RTE Challenge*.
- Dan Moldovan and Adrian Novischi. 2002. Lexical Chains for Question Answering. In *Proceedings of COLING*.
- Dan Moldovan and Vasile Rus. 2001. Logic Form Transformation of WordNet and its Applicability to Question Answering. In *Proceedings of ACL*.
- Dan Moldovan, Christine Clark, Sanda Harabagiu, and Steve Maiorano. 2003. COGEX A Logic Prover for Question Answering. In *Proceedings of the HLT/NAACL*.
- Dan Moldovan, Adriana Badulescu, Marta Tatu, Daniel Antohe, and Roxana Girju. 2004. Models for the Semantic Classification of Noun Phrases. In *Proceedings of HLT/NAACL, Computational Lexical Semantics workshop*.
- Eamonn Newman, Nicola Stokes, John Dunnion, and Joe Carthy. 2005. UCD IIRG Approach to the Textual Entailment Challenge. In *Proceedings of the PASCAL RTE Challenge*.
- Rajat Raina, Aria Haghighi, Christopher Cox, Jenny Finkel, Jeff Michels, Kristina Toutanova, Bill MacCartney, Marie-Catherine de Marneffe, Christopher Manning, and Andrew Ng. 2005. Robust Textual Inference using Diverse Knowledge Sources. In *Proceedings of the PASCAL RTE Challenge*.