On the Role of Lexical and World Knowledge in RTE3

Peter Clark[1], William R. Murray[1], John Thompson[1], Phil Harrison[1], Jerry Hobbs[2], Christiane Fellbaum[3]
[1]Boeing Phantom Works, The Boeing Company, Seattle, WA 98124
[2]USC/ISI, 4676 Admiralty Way, Marina del Rey, CA 90292
[3]Princeton University, NJ 08544
{peter.e.clark,william.r.murray,john.a.thompson,philip.harrison}@boeing.com, hobbs@isi.edu, fellbaum@clarity.princeton.edu

Abstract

To score well in RTE3, and even more so to create good justifications for entailments, substantial lexical and world knowledge is needed. With this in mind, we present an analysis of a sample of the RTE3 positive entailment pairs, to identify where and what kinds of world knowledge are needed to fully identify and justify the entailment, and discuss several existing resources and their capacity for supplying that knowledge. We also briefly sketch the path we are following to build an RTE system (Our implementation is very preliminary, scoring 50.9% at the time of RTE). The contribution of this paper is thus a framework for discussing the knowledge requirements posed by RTE and some exploration of how these requirements can be met.

1 Introduction

The Pascal RTE site defines entailment between two texts T and H as holding "if, typically, a human reading T would infer that H is most likely true" assuming "common human understanding of language as well as common background knowledge." While a few RTE3 entailments can be recognized using simple syntactic matching, the majority rely on significant amounts of this "common human understanding" of lexical and world knowledge. Our goal in this paper is to analyze what that knowledge is, create a preliminary framework for it, and explore a few available sources for it. In the short term, such knowledge can be (and has been) used to drive semantic matching of the T and H dependency/parse trees and their semantic representations, as many prior RTE systems perform, e.g., (Hickl et al., 2006). In the long term, computers should be able to perform deep language understanding to build a computational model of the scenario being described in T, to reason about the entailment, answer further questions, and create meaningful justifications. With this longer term goal in mind, it is useful to explore the types of knowledge required. It also gives a snapshot of the kinds of challenges that RTE3 poses.

The scope of this paper is to examine the underlying lexical/world knowledge requirements of RTE, rather than the more syntactic/grammatical issues of parsing, coreference resolution, named entity recognition, punctuation, coordination, typographical errors, etc. Although there is a somewhat blurry line between the two, this separation is useful for bounding the analysis. It should be noted that the more syntactic issues are themselves vast in RTE, but here we will not delve into them. Instead, we will perform a thought experiment in which they have been handled correctly.

2 Analysis

Based on an analysis of 100 (25%) of the positive entailments in the RTE3 test set, we have divided the knowledge requirements into several rough categories, which we now present. We then summarize the frequency with which examples in this sample fell into these categories. The examples below are fragments of the original test questions, abbreviated and occasionally simplified.

2.1 Syntactic Matching

In a few cases, entailment can be identified by syntactic matching of T and H, for example:

Proceedings of the Workshop on Textual Entailment and Paraphrasing, pages 54–59,
Prague, June 2007.

489.T "The Gurkhas come from Nepal and…"
489.H "The Gurkhas come from Nepal."

Other examples include 299, 489, and 456. In some cases, the syntactic matching can be very complex, e.g., examples 152, 724.

2.2 Synonyms

Synonymy is often needed to recognize entailment,

648.T "…go through … licencing procedures..."
648.H "…go through the licencing processes."

Other examples include 286 ("dismiss"/"throw out"), 37 ("begin"/"start"), 236 ("wildfire"/"bush fire"), and, arguably, 462 ("revenue"/"proceeds").

2.3 Generalizations (Hypernyms)

Similarly, subsumption (generalization) relationships between word senses need to be recognized (whether or not a fixed set of senses are used), eg.

148.T "Beverly served...at WEDCOR"
148.H "Beverly worked for WEDCOR."

Others include 178 ("succumbed" as a kind of "killed"), and 453 ("take over" as a kind of "buy").

2.4 Noun Redundancy

Sometimes a noun in a compound can be dropped:

607.T "single-run production process..."
607.H "Single-run production..."

Other examples include 269 ("increasing prevalence of" → "increasing"), 604 ("mini-mill process" → "mini-mill"), and (at the phrase level) 668 ("all segments of the public" → "the public").

2.5 Noun-Verb Relations

Often derivationally related nouns and verbs occur in the pairs. To identify and justify the entailment, the relationship and its nature is needed, as in:

480 "Marquez is a winner..." →"Marquez won..."

Other examples include 286 ("pirated", "piracy"), and 75 ("invent", "invention"). In some cases, the deverbal noun denotes the verb's event, in other cases it denotes one of the verb's arguments (e.g., "winner" as the subject/agent of a "win" event).

2.6 Compound Nouns

Some examples require inferring the semantic relation between nouns in a compound, e.g.,

168 "Sirius CEO Karmazin" → "Karmazin is an executive of Sirius"
583 "physicist Hawking" → "Hawking is a physicist"
In some cases this is straightforward, others require more detailed knowledge of the entities involved.

2.7 Definitions

Although there is somewhat of a fuzzy boundary between word and world knowledge, we draw this distinction here. Some examples of RTE pairs which require knowing word meanings are:

667 "… found guilty..." → "…convicted..."
328 "sufferers of coeliac disease..." → "coeliacs..."

The second example is particularly interesting as many readers (and computers) will not have encountered the word "coeliacs" before, yet a person can reasonably infer its meaning on the fly from context and morphology - something challenging for a machine to do. Definitions of compound nouns are also sometimes needed, e.g., "family planning" (612) and "cough syrup" (80).

2.8 World Knowledge: General

A large number of RTE pairs require non-definitional knowledge about the way the world (usually) is, e.g.,:

273 "bears kill people" → "bears attack people"

People recognize this entailment as they know (have heard about) how people might be killed by a bear, and hence can justify why the entailment is valid. (They know that the first step in the bear killing a person is for the bear to attack that person.) Some other examples are:

499 "shot dead" → "murder"
705 "under a contract with" → "cooperates with"
721 "worked on the law" → "discussed the law"
731 "cut tracts of forest" → "cut trees in the forest"
732 "establishing tree farms"
    → "trees were planted"
639 "*X*'s plans for reorganizing"
    → "*X* intends to reorganize"
328 "the diets must be followed by <*person*>"
    → "the diets are for <*person*>"
722 *X* and *Y* vote for *Z* → *X* and *Y* agree to *Z*.

All these cases appeal to a person's world knowledge concerning the situation being described.

2.9 World Knowledge: Core Theories

In addition to this more specific knowledge of the world, some RTE examples appeal to more general, fundamental knowledge (e.g., space, time, plans, goals). For example

6.T "Yunupingu is one of the clan of..."
6.H "Yunupingu is a member of..."

appeals to a basic rule of set inclusion, and 10 (a negative entailment: "unsuccessfully sought election" → not elected) appeals to core notions of goals and achievement. Several examples appeal to core spatial knowledge, e.g.:

491.T "...come from the high mountains of Nepal."
491.H "...come from Nepal."

178.T "...3 people in Saskatchewan succumbed to the storm."
178.H "...a storm in Saskatchewan."

491 appeals to regional inclusion (if X location Y, and Y is in Z, then X location Z), and 178 appeals to colocation (if X is at Y, and X physically interacts with Z, then Z is at Y). Other spatial examples include 236 ("around Sydney" → "near Sydney"), and 129 ("invasion of" → "arrived in").

2.10 World Knowledge: Frames and Scripts

Although loosely delineated, another category of world knowledge concerns stereotypical places, situations and the events which occur in them, with various representational schemes proposed in the AI literature, e.g., Frames (Minsky 1985), Scripts (Schank 1983). Some RTE examples require recognizing the implicit scenario ("frame", "script", etc.) which T describes to confirm the new facts or relationships introduced in H are valid. A first example is:

358.T "Kiesbauer was target of a letter bomb..."
358.H "A letter bomb was sent to Kiesbauer."

A person recognizes H as entailed by T because he/she knows the "script" for letter bombing, which includes sending the bomb in the mail. Thus a person could also recognize alternative verbs in 358.H as valid (e.g., "mailed", "delivered") or invalid (e.g., "thrown at", "dropped on"), even though these verbs are all strongly associated with words in T. For a computer to fully explain the entailment, it would need similar knowledge.

As a second example, consider:

538.T "...the O. J. Simpson murder trial..."
538.H "O. J. Simpson was accused of murder."

Again, this requires knowing about trials: That they involve charges, a defendant, an accusation, etc., in order to validate H as an entailed expansion of T. In this example, there is also a second twist to it as the noun phrase in 538.T equally supports the hypothesis "O. J. Simpson was murdered." (e.g., consider replacing "O. J." with "Nicole"). It is only a reference elsewhere in the T sentence to "Simpson's attorneys" that suggests Simpson is still alive, and hence couldn't have been the victim, and hence must be the accused, that clarifies 538.H as being correct, a highly complex chain of reasoning.

As a third example, consider:
736.T "In a security fraud case, Milken was sentenced to 10 years in prison."
736.H "Milken was imprisoned for security fraud."

This example is particularly interesting, as one needs to recognize security fraud as Milken's crime, a connection which not stated in T. A human reader will recognize the notion of sentencing, and thus expect to see a convict and his/her crime, and hence can align these expectations with T, validating H. Thus again, deep knowledge of sentencing is needed to understand and justify the entailment.

Some other examples requiring world knowledge to validate their expansions, include 623 ("narcotics-sniffing dogs" → "dogs are used to sniff out narcotics"), and 11 ("the Nintendo release of the game" → "the game is produced by Nintendo").

2.11 Implicative Verbs

Some RTE3 examples contain complement-taking verbs that make an implication (either positive or negative) about the complement. For example:

668 "A survey shows that *X*..." → "*X*..."
657 "...*X* was seen..." → "...*X*..."
725 "...decided to *X*..." → "...*X*..."
716 "...have been unable to *X*..." → "...do not *X*"

In the first 3 the implication is positive, but in the last the implication is negative. (Nairn et al, 2006) provide a detailed analysis of this type of behavior. In fact, this notion of implicature (one part of a sentence making an implication about another part) extends beyond single verbs, and there are some more complex examples in RTE3, e.g.:

453 "...won the battle to *X*..." → "...*X*..."

784.T "X reassures Russia it has nothing to fear..."
784.H "Russia fears..."

In this last example the implication behavior is quite complex: (loosely) If X reassures Y of Z, then Y is concerned about not-Z.

2.12 Metonymy/Transfer

In some cases, language allows us to replace a word (sense) with a closely related word (sense):
708.T "Revenue from stores funded..."
708.H "stores fund..."

Rules for metonymic transfer, e.g., (Fass 2000), can be used to define which transfers are allowed. Another example is 723 "…pursue its drive towards *X*" → "…pursue *X*".

2.13 Idioms/Protocol/Slang

Finally, some RTE pairs rely on understanding idioms, slang, or special protocols used in language, for example:

12 "Drew served as Justice. Kennon returned to claim Drew's seat" → "Kennon served as Justice."
486 "*name*, 1890-1970" → "*name* died in 1970"
408 "takes the title of" → "is"
688 "art finds its way back" → "art gets returned"

The phrases in these examples all have special meaning which cannot be easily derived compositionally from their words, and thus require special handling within an entailment system.

2.14 Frequency Statistics

Table 1 shows the number of positive entailments in our sample of 100 that fell into the different categories (some fell into several). While there is a certain subjectivity in the boundaries of the catego-

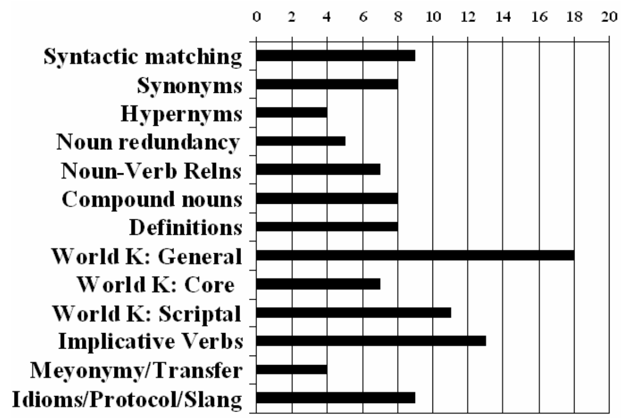

Table 1: Frequency of different entailment phenomena from a sample of 100 RTE3 pairs.

ries, the most significant observation is that very few entailments depend purely on syntactic manipulation and simple lexical knowledge (synonyms, hypernyms), and that the vast majority of entailments require significant world knowledge, highlighting one of the biggest challenges of RTE. In addition, the category of general world knowledge -- small, non-definitional facts about the way the world (usually) is -- is the largest, suggesting that harvesting and using this kind of knowledge should continue to be a priority for improving performance on RTE-style tasks.

3 Sources of World Knowledge

While there are many potential sources of the knowledge that we have identified, we describe three in a bit more detail and how they relate to the earlier analysis.

3.1 WordNet

WordNet (Fellbaum, 1998) is one of the most extensively used resources in RTE already and in computational linguistics in general. Despite some well-known problems, it provides broad coverage of several key relationships between word senses (and words), in particular for synonyms, hypernyms (generalizations), meronyms (parts), and semantically ("morphosemantically") related forms. From the preceding analysis, WordNet does contain the synonyms {"procedure","process"}, {"dismiss","throw out"}, {"begin","start"}, but does not contain the compound "wild fire" and (strictly correctly) does not contain {"revenue","proceeds"} as synonyms. In addition, the three hypernyms mentioned in the earlier analysis are in WordNet. WordNet also links (via the "morphosemantic" link) the 3 noun-verb pairs mentioned earlier (win/winner, pirated/piracy, invent/invention) – however it does not currently distinguish the nature of that link (e.g., agent, result, event). WordNet is currently being expanded to include this information, as part of the AQUAINT program.

Two independently developed versions of the WordNet glosses expressed in logic are also available: Extended WordNet (Moldovan and Rus, 2001) and a version about to be released with WordNet3.0 (again developed under AQUAINT). These in principle can help with definitional knowledge. From our earlier analysis, it turns out "convicted" is conveniently defined in WordNet as "pronounced or proved guilty" potentially bridging the gap for pair 667, although there are problems with the logical interpretation of this particular gloss in both resources mentioned. WordNet does have "coeliac", but not in the sense of a person with coeliac disease[1].

In addition, several existing "core theories" (e.g., TimeML, IKRIS) that can supply some of the fundamental knowledge mentioned earlier (e.g., space, time, goals) are being connected to WordNet under the AQUAINT program.

3.2 The DIRT Paraphrase Database

Paraphrases have been used successfully by several RTE systems (e.g., Hickl et al., 2005). With respect to our earlier analysis, we examined Lin and Pantel's (2001) paraphrase database built with their system DIRT, containing 12 million entries. While there is of course noise in the database, it contains a remarkable amount of sensible world knowledge as well as syntactic rewrites, albeit encoded as shallow rules lacking word senses.

Looking at the examples from our earlier analysis of general world knowledge, we find that three are supported by paraphrase rules in the database:

273: X kills Y → X attacks Y
499: X shoots Y → X murders Y
721: X works on Y → X discusses Y

And one that could be is not, namely:

705: X is under a contract with Y → X cooperates with Y (not in the database)

Other examples are outside the scope of DIRT's approach (i.e., "X $pattern_1$ Y" → "X $pattern_2$ Y"), but nonetheless the coverage is encouraging.

3.3 FrameNet

In our earlier analysis, we identified knowledge about stereotypical situations and their events as important for RTE. FrameNet (Baker et al, 1998) attempts to encode this knowledge. FrameNet was used with some success in RTE2 by Burchardt and Frank (2005). FrameNet's basic unit - a Frame - is a script-like conceptual schema that refers to a situation, object, or event along with its participants (Frame Elements), identified independent of their syntactic configuration.

We earlier discussed how 538.T "...the O. J. Simpson murder trial..." might entail 538.H "O. J. Simpson was accused of murder." This case applies to FrameNet's Trial frame, which includes the Frame Elements Defendant and Charges, with Charges being defined as "The legal label for the crime that the Defendant is accused of.", thus stating the relationship between the defendant and the charges, unstated in T but made explicit in H, validating the entailment. However, the lexical units instantiating the Frame Elements are not yet disambiguated against a lexical database, limiting full semantic understanding. Moreover, FrameNet's world knowledge is stated informally in English descriptions, though it may be possible to convert these to a more machine-processable form either manually or automatically.

3.4 Other Resources

We have drawn attention to these three resources as they provide some answers to the requirements identified earlier. Several other publicly available resources could also be of use, including VerbNet (Univ Colorado at Boulder), the Component Library (Univ Texas at Austin), OpenCyc (Cycorp), SUMO, Stanford's additions to WordNet, and the Tuple Database (Boeing, following Schubert's 2002 approach), to name but a few.

[1] This seems to be an accidental gap; WordNet contains many interlinked disease-patient noun pairs, incl. "diabetes-diabetic," "epilepsy-eplileptic," etc.

4 Sketch of our RTE System

Although not the primary purpose of this paper, we briefly sketch the path we are following to build an RTE system able to exploit the above resources. Our implementation is very preliminary, scoring 50.9% at the time of RTE and 52.6% now (55.0% on the 525 examples it is able to analyze, guessing "no entailment" for the remainder). Nevertheless, the following shows the direction we are moving in

Like several other groups, our basic approach is to generate logic for the T and H sentences, and then explore the application of inference rules to elaborate T, or transform H, until H matches T. Parsing is done using a broad coverage chart parser. Subsequently, an abstracted form of the parse tree is converted into a logical form, for example:

299.H "Tropical storms cause severe damage."
subject(_Cause1, _Storm1)
sobject(_Cause1, _Damage1)
mod(_Storm1, _Tropical1)
mod(_Damage1, _Severe1)
input-word(_Storm1, "storm", noun)
[same for other words]

Entailment is determined if every clause in the semantic representation of H semantically matches (subsumes) some clause in T. Two variables in a clause match if their input words are the same, or some WordNet sense of one is the same as or a hypernym of the other. In addition, the system searches for DIRT paraphrase rules that can transform the sentences into a form which can then match. The explicit use of WordNet and DIRT results in comprehensible, machine-generated justifications when entailments are found, , e.g.,:

T: "The Salvation Army operates the shelter under a contract with the county."
H: "The Salvation Army collaborates with the county."
Yes! Justification: I have general knowledge that:
IF X works with Y THEN X collaborates with Y
Here: X = the Salvation Army, Y = the county
Thus, here:
I can see from T: the Salvation Army works with the county (because "operate" and "work" mean roughly the same thing)
Thus it follows that:
The Salvation Army collaborates with the county.

We are continuing to develop our system and expand the number of knowledge sources it uses.

5 Summary

To recognize and justify textual entailments, and ultimately understand language, considerable lexical and world knowledge is needed. We have presented an analysis of some of the knowledge requirements of RTE3, and commented on some available sources of that knowledge. The analysis serves to highlight the depth and variety of knowledge demanded by RTE3, and contributes a rough framework for organizing these requirements. Ultimately, to fully understand language, extensive knowledge of the world (either manually or automatically acquired) is needed. From this analysis, RTE3 is clearly providing a powerful driving force for research in this direction.

Acknowledgements

This work was performed under the DTO AQUAINT program, contract N61339-06-C-0160.

References

Collin Baker, Charles Fillmore, and John Lowe. 1998. "The Berkeley FrameNet Project". Proc 36th ACL

Aljoscha Burchardt and Anette Frank. 2005. Approaching Textual Entailment with LFG and FrameNet Frames. in 2nd PASCAL RTE Workshop, pp 92-97.

Dan Fass. 1991. "Met*: A Method for Discriminating Metonymy and Metaphor by Computer". In Computational Linguistics 17 (1), pp 49-90.

Christiane Fellbaum. 1998. "WordNet: An Electronic Lexical Database." The MIT Press.

Andrew Hickl, Jeremy Bensley, John Williams, Kirk Roberts, Bryan Rink, and Ying Shi. "Recognizing Textual Entailment with LCC's Groundhog System", in Proc 2nd PASCAL RTE Workshop, pp 80-85.

Dekang Lin and Patrick Pantel. 2001. "Discovery of Inference Rules for Question Answering". Natural Language Engineering 7 (4) pp 343-360.

Marvin Minsky. 1985. "A Framework for Representing Knowledge". In Readings in Knowledge Repn.

Dan Moldovan and Vasile Rus, 2001. Explaining Answers with Extended WordNet, ACL 2001.

Rowan Nairn, Cleo Condoravdi and Lauri Karttunen. 2006. Computing Relative Polarity for Textual Inference. In the Proceedings of ICoS-5.

Len Schubert. 2002. "Can we Derive General World Knowledge from Texts?", Proc. HLT'02, pp84-87.