A Preliminary Study of Finding Entailing Texts in a Domain-specific Monolingual Parallel Corpora

Kenston Choi  
De La Salle University  
2401 Taft Avenue  
Manila, Philippines  
kenston.choi@gmail.com

Nathalie Rose Lim-Cheng  
De La Salle University  
2401 Taft Avenue  
Manila, Philippines  
nats.lim@delasalle.ph

ABSTRACT
This paper introduces the possible usages, benefits, and challenges involved in the use of domain-specific monolingual parallel corpora in determining textual entailment (TE). A system that finds entailing text for a given statement is to be developed using monolingual parallel translations of the Bible as corpus as this is one of the most accessible monolingual parallel corpora. Different existing methods for textual entailment are reviewed and related to the use of domain-specific and monolingual parallel corpora.

Keywords
Textual entailment, monolingual parallel corpora, information retrieval, Bible domain

1. INTRODUCTION
Interests in the task of finding entailing texts began to increase following the initial PASCAL Recognizing Textual Entailment (RTE) Challenge in 2005 that covers various text pairs from different domains and datasets used in different natural language processing (NLP) tasks and challenges. A text \( T \) is said to entail a given hypothesis \( H \) when a human reading \( T \) can infer that \( H \) is most likely true [4]. The RTE Main Task focuses on determining whether an isolated \( T-H \) pair is an entailment pair or not. In RTE-5, the Pilot Search Task aims to draw the task even closer to realistic applications by requiring both \( T \) and \( H \) to be interpreted within the context of the corpus, such that systems must find all sentences in a corpus that entail a given \( H \) [13].

Although different sentences in the Pilot Search Task corpus may express the same idea, the redundancy is not intended, since not all sentences have corresponding statements and no alignment information for those redundant sentences are given to the participants. While texts used in the challenges usually come from news articles, topics covered are broad, with no particular domain being focused. These two observations lead us to look at the impacts of domain-specific monolingual parallel corpora in finding entailing texts.

The Christian Bible consisting of the Old and New Testament is one of the most translated and purchased book in history [1]. There have been only a few NLP applications that attempt to use the Bible in NLP research [8]. Such observation may be true when we exclude the available resources like commentaries, lexicons, cross-references, and the various translations of the Bible based on different translation paradigms. Thus, it is necessary to explore the Bible domain and to study the impacts of using domain resources, particularly the monolingual parallel translations in a domain-specific textual entailment search system.

The next section reviews some of the existing approaches in RTE as they are mentioned in the other succeeding sections. Section 3 covers the possible usages and benefits of monolingual parallel corpora in finding entailing texts, while Section 4 discusses the possibilities and advantages when limiting the task to a specific domain. Section 5 mentions some of the challenges and possible issues that need to be addressed in order to fully benefit from the stated possibilities in Section 3 and 4.

2. EXISTING APPROACHES IN RECOGNIZING TEXTUAL ENTAILMENT
Several approaches operating in both shallow and deep levels have been demonstrated during the RTE Challenge. String similarity approaches involve using string metrics to approximate the distance of \( T \) and \( H \). Some systems like in [10] perform scored phrase alignments based on various features before comparing the similarity score to a learned threshold. While such approach works in some instances, some non-entailment pairs with high word overlaps are prone to be marked as entailment. On the other hand, logic-based approaches involving representation of texts to some logic form tend to achieve a higher precision, as demonstrated in [3], but they may also suffer from low recall due to representation issues and the lack of background knowledge.

Different approaches using constituent and/or dependency parse trees have become more common since text can usually be represented by an available parser as a graph. This involves performing comparisons and determining graph similarity or subsumption. Some systems like [9] compute for the tree edit...
distance (TED) or the cost of editing the parse tree of $H$ to make it similar to the parse tree of $T$. Other works like [14] focus on transforming the parse tree of $T$ such that it embeds the parse tree of $H$ within. This approach may cost only a single transformation based on specified rules (e.g., active to passive), while achieving the same change in TED-based approach may be counted as several operations (e.g., insert, delete, move). In [15], cross-pair similarity between two $T$-$H$ pairs that have the same rewrite rules or similar subtree structures are considered, where the undecided pair $U$ is mapped to the decided pair $M$ in the corpus, such that the decision for $U$ can be derived from the decision for $M$.

Some systems like in [5] treated the task as a classification problem where various lexico-semantic features are selected for the machine to learn a classifier. In [7], a set of discourse commitments (propositions) is extracted from a given pair, and a trained classifier processes the set of aligned commitment pairs. Entailment decision is determined based on the number of aligned pairs removed due to $H$ contradicting other commitments extracted from $T$.

While there are many other notable approaches that also use various lexical resources and paraphrase rules, there is no guarantee that one approach will consistently outperform other approaches by huge margin across different datasets. Some submissions to the Pilot Search Task indicate that additional components need to be integrated as part of an existing entailment system designed for the Main Task to improve performance and to cover other linguistic phenomena, as demonstrated in [12]. Hybrid systems combining at least two approaches specializing in certain cases are shown to possibly improve overall performance [11]. For example, tree transformation and string similarity approaches usually performs well for cases that involves only with syntactic and lexical phenomena, respectively. This can be the same motivation or reason when some of the above-mentioned works use string similarity approaches as part of their system.

### 3. TEXTUAL ENTAILMENT USING MONOLINGUAL PARALLEL CORPORA

Having each $H$ paired with each translation text $t$ obtained from monolingual parallel corpora is analogous to using ensemble approaches in a hybrid TE system that combines multiple local decisions into a single decision. In the case of parallel texts, the system will have to combine the multiple decisions for $H$ as paired across different translations that expresses the same idea of a text. We explore some of the advantages and possibilities in considering parallel texts.

#### 3.1 Evading Linguistic Phenomena

In [2], at least five general categories of linguistic phenomena have been identified from the RTE-5 dataset, and this includes: lexical, lexical-synactic, syntactic, discourse, and reasoning. One evident way for TE systems to easily determine the entailment in a pair that typically demonstrate different linguistic phenomena is to rewrite the $T$ involved. Consider the verse Romans 3:25 as $T$ with the parallel texts:

New English Translation (NET)\(^1\): God publicly displayed him at his death as the mercy seat accessible through faith. This was to demonstrate his righteousness, because God in his forbearance had passed over the sins previously committed.

Easy English (EE)\(^2\): God chose to offer Christ as a sacrifice so that he could forgive sinful people. God sent Christ to bleed and to die. And, as a result, God forgives people who believe Christ. God did this to show clearly that he is always completely right and fair. In past times, God did not punish people who were doing wrong things.

Some coreference resolution tools may not be able to resolve all pronouns correctly in the NET translation, like his as Christ for the first instance and as God for the other two instances, even if the previous verse text “But they are justified freely by his grace through the redemption that is in Christ Jesus” is given. Thus, either one of the hypotheses “Christ died,” or “God is just,” may not be correctly matched to $T$ with high confidence due to the discourse issue.

If we consider the rewritten version from EE Bible, then the linguistic phenomenon or the need to resolve the anaphora is no longer present. In this particular phenomenon, resolved anaphora can be compared across all parallel texts. Similarly, the $H$ “God is patient.” being implicitly explained in EE may require certain reasoning, where as it is explicitly stated in NET. Since no tool can easily rewrite the verse without first going back to resolving the anaphora or performing the necessary reasoning, we consider monolingual parallel texts as a necessary alternative collection of rewritten texts.

#### 3.2 Determining Word Senses and Relations

Word alignments in parallel texts can be used to eliminate word senses that are far from what is intended by considering the consistency in the meaning of the word used across different translations. Consider the non-entailment $H$ “Food cannot bring us secretly to God,” being paired with the verse in 1 Corinthians 8:8 as $T$, with the following translation texts:

**NET:** Now food will not bring us close to God. We are no worse if we do not eat and no better if we do.

**EE:** But food does not bring us near to God. If we do not eat the food, that will not make us worse. If we do eat the food, it will not make us better.

The words secretly and close may be related to each other the way the phrases “a secret meeting” and “a close meeting” are related. However, if we consider the word near (used in EE) as the sense being held by the word close, then a non-entailment decision (or unknown decision in three-way decision) can be given in both NET and EE. To some extent, the scope of word senses and the possible word relationships (i.e., close $\rightarrow$ secretive, close $\rightarrow$ near) can be restricted locally in each case.

---

\(^1\) Scripture quoted by permission. Quotations designated (NET) are from the NET Bible® copyright ©1996-2006 by Biblical Studies Press, L.L.C. http://bible.org. All rights reserved.

\(^2\) Quotations designated (EE) are from the Easy English Bible copyright ©1997-2005 by Wycliffe Associates (UK), http://www.easyenglish.info/. All rights reserved.
and globally when considering the domain as a whole (e.g., grace taking the meaning of unmerited favor of God or kindness, and avoiding the meaning of elegance).

3.3 Providing Redundancy and Validation

Using several parallel texts create some form of redundancy as the same idea is expressed using different words and under different sentence structures. Scores for each verse can be combined from scores returned by the IR system for each different Bible translation’s search index. This may improve coverage and relevance ranking depending on the combination method used. Cases where transforming a verse to logic form or parsing a verse result to failure or inaccurate output may be compensated at some rate by other parallel monolingual translation of the same verse.

Approaches like TED and tree transformation involve traceable sequence of operations, and these results can be compared across all parallel texts. In string similarity approaches, resulting alignments can be compared to determine consistency and to perform validation. Baseline experiment may simply be pairing parallel texts with H and processing the pair using any existing approach or system followed by determining the majority decision given among the pairs.

Following the work of [7] on discourse commitments, parallel texts offer additional propositions to match against H. This may increase chances of coverage or confidence during the alignment of the propositions of T and H. Also, if a certain proposition from H is consistently ruled out across different parallel texts, we can conclude that such decision is likely correct. Parallel texts can be used as part of training for entailment cases, and to add more rewrite rules for approach using cross-pair similarities like in [15].

4. TEXTUAL ENTAILMENT IN CLOSED DOMAINS

The Bible being a closed canon or book can itself stand as a domain of interest in NLP. Textual entailment search system is more practical and more necessary in closed-text setting since the possible entailing texts are consistently the same while the relevance of the texts and the application remain valuable over time. In contrast to the RTE Challenge where the set of T’s in the Development Set (text pairs released to participants for training) is different from the T’s in the Test Set, a closed-text TE implies that the set of T’s is shared and pre-determined during both development and testing. We go through some of the possible advantages that can be gained when restricting a TE system to a specific domain.

4.1 Reducing Scope and Ambiguities

Textual entailment involving domain-specific and closed text may minimize some of the concerns in open-domain TE like named-entity recognition and term disambiguation. Proper names for people and places can be exhaustively listed in a gazetteer. A domain-specific lexicon can be provided to match unknown terms and to supply the part-of-speech information. Polysemous words can take the meaning that is closer to the domain as guided by the domain information in some lexical databases. A better approximation of the threshold needed for some of the TE approaches can be made in a closed-text setting; since the set of T’s used during testing may have been similar or used during training. Certain aspects like evaluation of numerical expressions or comparison of quantities important in TE for other domains may not be applicable in this domain. The possibilities in constructing the input hypothesis can be regulated.

4.2 Sharing Background Knowledge

Aside from using lexical and paraphrase resources, some groups in the RTE Challenge added geographical (e.g., capital of a country) and acronym information as part of background knowledge (BK) for their systems; such information can be easily obtained and used. While the RTE Challenge prohibits using knowledge alone as basis for determining entailment [13], it does not imply a restriction to only incorporate linguistic and commonsense knowledge. Consider the H “We are saved through the Gospel” and the T from the verse Romans 10:17:

NET: Consequently faith comes from what is heard, and what is heard comes through the preached word of Christ.

Assuming that the word Gospel has been aligned to the phrase word of Christ, the pair may be decided incorrectly as non-entailment due to lack of knowledge that we are saved by grace through faith as stated in Ephesians 2:8:

NET: For by grace you have been saved through faith. And this is not your own doing; it is the gift of God.

In that sense, we may say that we are saved through the Gospel because one cannot even have saving faith without even first knowing the Gospel. While decisions involving such cases may be subjectively objected (especially when evaluating it as strict entailment) by those who are not familiar with the particular BK, it remains sound to consider the verse (Romans 10:17) in an entailment search system where wider coverage is more important, and the verse being likely relevant to other people.

The BK obtained from other verses can be used to support other pairs. Domain-specific BK can be supplied as part of the set of propositions to be able to find proofs or to derive at a decision. Extracting relationships or assertions from the text using information and relation extraction tools like ReVerb [6] can be used build such collection. Entailment decisions can be initially guessed by determining the common assertions extracted from H and T that are being shared.

4.3 Considering Text Relationships

Establishing relationships between sentences used as T’s may be useful in determining entailment. The grouping of sentences in one Bible translation based on their relatedness in thought can be used to establish confidence in a decision. While this is quite similar to having parallel texts of a specific verse, having other verses being paralleled to a specific verse is different. In the former, parallel translation texts of a verse maintain common expressions and descriptions that are expressed using different words, but in the latter, parallel verses contain new but related information with respect to that specific verse. Consider the H “We must tell the truth” being entailed by the verse in Ephesians 4:25:

NET: Therefore, having laid aside falsehood, each one of you speak the truth with his neighbor, for we are members of one another.

Using available cross-reference information manually made for the Bible, we can further relate that the same H may be entailed by the verse Colossians 3:9:
NET: Do not lie to one another since you have put off the old man with its practices

An entailing \( T \) (i.e., Ephesians 4:25) determined with high confidence based on the score given by a chosen approach can signal that the verse being cross-referenced (i.e., Colossians 3:9) may be considered even if the cross-referenced verse scored lower than expected. Valid application of such cases can be determined by a simplistic approach like string similarity or by the same chosen approach adjusted with a lower threshold, where the involved parts (i.e., speak the truth and do not lie) as determined by \( H \) are both present and related. Thus, Colossians 3:9 can no longer be suggested to entail the \( H \) “We belong to the same body.” despite being entailed by Ephesians 4:25.

5. CHALLENGES

While the necessary data, particularly the corpus, are no longer an issue in this domain, creating a TE search system that considers all of these aspects and possible advantages can be difficult. A method for combining the different results for each parallel text has to be defined, and some weighing schemes may have to be implemented before arriving with a final entailment decision.

It is important to know which translation is more valuable in specific cases, as text containing figures of speech in one translation can be expressed in another translation without the use of figurative language. In that case, the system should prefer the latter text. In addition, specific approaches to use or to combine have to be targeted to the right text. For example, splitting a verse into propositions is expected in EE to be less erroneous (less processing needed) and better in terms of quality than in NET. On the other hand, tree transformation and TED approaches must be able to handle multiple sentences in texts from EE (beyond carrying coreference information), and treat them as related structure rather than isolated trees.

These concerns have to be addressed to reduce the use of parallel texts as merely for hit-or-miss purpose. Proper ablation tests have to be designed and performed to estimate the impacts of every approach and assumption being tested.

6. CONCLUSION AND DIRECTIONS

We slightly differ from the RTE Challenge where our main goal is not necessarily to devise a novel textual entailment approach, but rather to develop a system for a particular domain by putting existing approaches in textual entailment into practice and utilizing different available resources (e.g., monolingual corpora, lexicons) exclusive in a domain. Particular approaches to use or to combine in recognizing textual entailment have to be identified while considering some of the identified challenges.

The use of monolingual parallel corpora may likely contribute to an improvement in performance or as a solution to certain difficulties in a specific domain like the Bible domain. Techniques to be used may also be applied to other domains that have similar types of data sources. Known issues in the RTE Challenge (e.g., lack of background knowledge and difficulties over certain linguistic phenomena) may be minimized when applied in a specific domain and when using alternative texts from monolingual parallel corpora. These assumptions are yet to be backed by further experiments and evaluations as a system capable of finding entailing text in the Bible is developed.

7. REFERENCES


