uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Christina George
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A.Sc. (Electrical Engineering)
GRADE / DEGREE

School of Information Technology and Engineering
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Revised Syntactic Attributes for Relative Clause Simplication and Relative Pronoun Correction

TITRE DE LA THÈSE / TITLE OF THESIS

N. Japkowicz
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

S.Szpakowicz

F. Oppacher

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Revised Syntactic Attributes for Relative Clause Simplification and Relative Pronoun Correction

Christina George

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
For the MASc degree in Electrical Engineering

Electrical and Computer Engineering
School of Information Technology and Engineering
University of Ottawa

Your file Votre référence
ISBN: 978-0-494-18418-9
Our file Notre référence
ISBN: 978-0-494-18418-9

ABSTRACT

Using previous work for resolving relative clause attachment ambiguities as a starting point, we expanded an existing framework consisting mainly of semantic features to include new syntactic features. We used these new syntactic attributes on their own, as well as with the existing ones in two separate applications. First, we examined sentences containing restrictive and non-restrictive relative clauses and classified them according to whether the relative clauses could be removed to simplify the sentences. In the second application, we corrected the selection of relative pronouns in French to English machine translations. To test our revised syntactic attributes, we devised a two-stage system. In the first stage, we used binary classification to perform an initial screening of sentences, followed by a more detailed classification in the second stage. We then evaluated the performance of each stage as well as the system as a whole. In each application, we made specific adjustments and succeeded in demonstrating that the combination of syntactic and semantic attributes was more effective in these classification tasks than relying on one type of attribute exclusively.

ACKNOWLEDGEMENTS

To my supervisor, Dr. Nathalie Japkowicz, thank you for all your help and encouragement throughout this journey. I am extremely appreciative of all the efforts you made to accommodate my school and work schedule and to offer feedback without hesitation. Without your support, this experience would not have been as rewarding or enjoyable.

To the members of my thesis defence committee, Dr. Stan Szpakowicz and Dr. Franz Oppacher, thank you for your insightful and detailed comments.

To my director, manager, team leaders, and colleagues at the Department of Public Works and Government Services, thank you for your understanding and patience. I consider myself very fortunate to be part of an organization that values education.

To my friends Fiona, Oana, Carrie, Louise, and Francesco, thank you for the pep talks, the words of encouragement, and for being a source of comfort and diversion. To the members of my research group, thank you for allowing me to learn and benefit from your experiences.

To my parents and my loving sister Tammy, no amount of thanks could ever match the unconditional and unwavering support you've always given me. I will continue to strive to for success in my endeavours and know that the examples you have set are my roadmap to achieving it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1: Introduction

Relative clauses and the relative pronouns contained within them are one example of grammatical structure that contributes to the overall complexity of English sentences and presents a challenge when automated tools must process them.

The purpose of relative clauses is to provide additional and sometimes essential information about the person, thing or concept being talked about. Relative clauses present a challenge in terms of understanding the main idea of a sentence because they may consist of requisite or non-essential information. Restrictive relative clauses (RRCs) contain essential information that is required for the overall understanding of a sentence while non-restrictive relative clauses (NRRCs) provide additional information about the person or thing being talked about. With the knowledge that there are two types of relative clauses, the task of distinguishing between them becomes an essential step towards text simplification. Because a NRRC provides extra information pertaining to a person, place or concept, it may be removed from the sentence entirely without "serious injury to the precise understanding of the sentence as a whole" [1].

Contained within relative clauses are relative pronouns whose purpose is to "connect the clause with the rest of the sentence" [1]. The identification of the correct antecedent of a relative pronoun is precisely what renders the selection of relative pronouns difficult, particularly in the context of French to English machine translation. In many cases where the relative pronoun does not immediately follow the antecedent or the antecedent is difficult to identify, the relative pronoun chosen by the machine translation system is incorrect, rendering the English translation difficult to understand.

1.1 System Overview

This research study is broken down into several important building blocks, as illustrated by Figure 1. The starting point for our study was Siddharthan's [2] work for determining local versus wide attachment of relative clauses. We modified his machine learning (ML) framework, consisting primarily of semantic attributes, to include new syntactic attributes we devised from a grammatical analysis of sentences containing relative clauses.

Once our new syntactic attributes were established, we developed a two-stage evaluation system which incorporated ML algorithms. In this two-stage system, the One Rule, Decision Tree and Naïve Bayes algorithms were used to evaluate the performance of our attributes in several applications.

To begin, we tested our revised syntactic attributes in the task of relative clause attachment and from here we developed two new applications. In this process, we considered the possibility that the set of attributes generated for Siddharthan's ML framework could be useful in tasks related to relative clause attachment because they were generated based on his study of a specific grammatical structure of the phrase preceding relative clauses in sentences. This led us to hypothesize that the same set of attributes from Siddharthan's system could also be effective in the identification of relative clause types and the correction of relative pronouns since knowing the antecedent of a relative clause effectively determines the relative pronoun to be used within it, as well as the type of relative clause in certain cases. With this hypothesis, we were able to explore two separate but interconnected applications.

For our first application, we looked at the identification of RRCs and NRRCs and examined the task of removing NRRCs in order to produce a simplified version of the original sentence. In the second application, we looked at the correction of relative pronouns in the context of French to English machine translations. Specifically, we were interested in the identification of mistranslated sentences produced by a machine translation system and the correction of the relative pronouns contained within them.

Using our revised syntactic attributes, we sought to discover what type of attributes, be it semantic attributes, syntactic attributes, or the combination of both types, was most effective in the identification and removal of NRRCs in sentences for simplification purposes, and the correction of relative pronouns in a French to English translation setting.

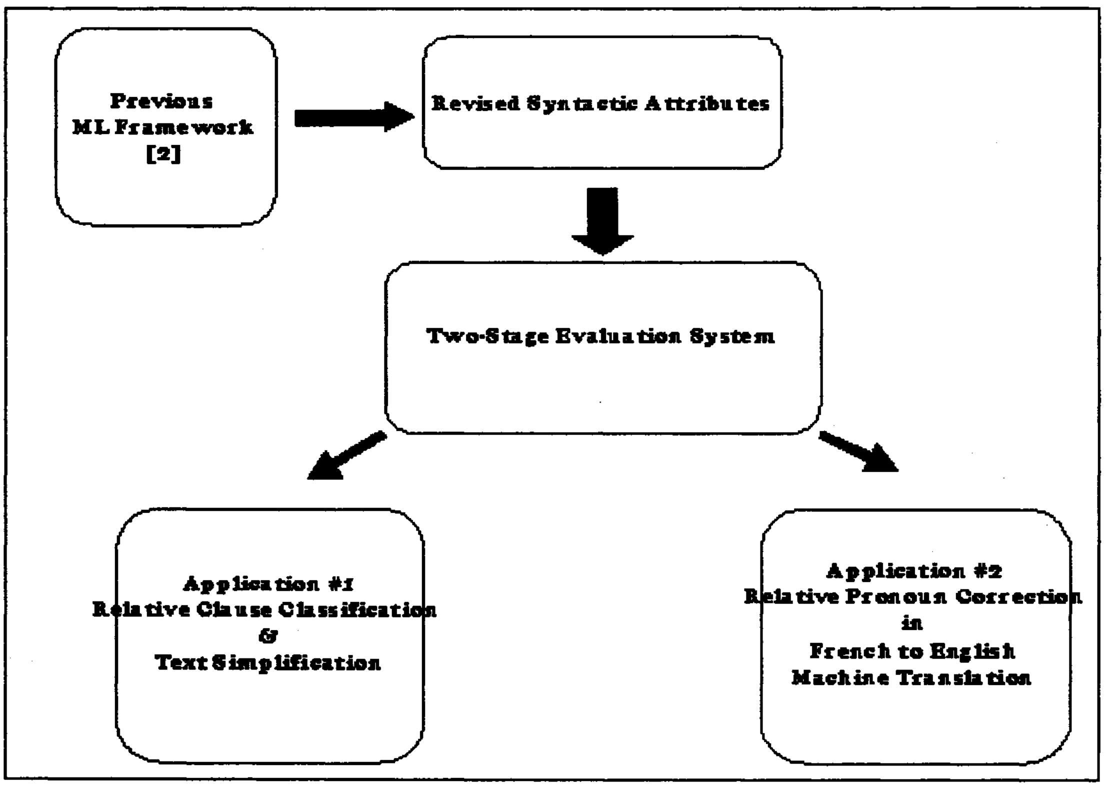

Figure 1 - Complete System Overview

1.2 Research Contributions

The primary contribution of our research was the expansion of Siddharthan's previous work to come up with new syntactic attributes. For our study, we identified certain key aspects of Siddharthan's framework that could be further developed by incorporating additional linguistic knowledge. Our intent was to present a more complete study of Siddharthan's relative clause attachment application. Once we revised the set of attributes and evaluated its performance in determining local versus wide attachment, we devised a two-stage system which made use of ML algorithms. We demonstrated that our revised syntactic attributes and two-stage evaluation process could be used for various applications, including the identification and removal of NRRCs in order to simplify sentences, as well as the identification and correction of relative pronouns in the context of machine translation.

Our two applications have several significant implications. The identification and removal of NRRCs is a direct application of text simplification, which is the stepping stone for a number of natural language processing (NLP) tasks such as parsing, machine translation, information retrieval, summarization, and improving the clarity of text. Our second application deals directly with the improvement of machine translation systems (MTS). In addition to correcting the output of a MTS in the desired target language, our relative pronoun correction system may also be useful in learning about the use of relative pronouns in the source language. By capturing frequent relative pronoun errors and correcting them, it may be possible to trace these errors back to particular relative pronoun and clause structures in the source language. Therefore, our application is not only a method of improving the information relayed in the target language, but also a way to learn more about the source language.

Our detailed research methodology is presented in the following chapters. In Chapter 2, we describe previous research in the areas of relative clause identification, relative pronoun correction, text simplification, and ML, and then go on to explain where our research is situated among the work of other researchers. We then describe our revised syntactic attributes and two-stage evaluation system in detail in Chapter 3. Chapters 4 and 5 are devoted to the two applications which made use of our revised syntactic attributes and two-stage evaluation system. Finally, in Chapter 6, we present our conclusions and recommendations for future work.

Chapter 2: Previous Work

Although we used a common set of attributes for two separate applications, these two applications may be considered rooted in the same previous work because they both deal with the same grammatical construct within sentences. Relative pronouns are an essential component of relative clauses, whose function is to provide additional information regarding some antecedent in a sentence. Therefore, the work done previously in the areas of relative pronoun correction and relative clause identification is often inextricably linked.

The two-stage evaluation system we developed was also common to our two separate applications. This system made use of general ML principles, which have been developed and improved over time. Relevant previous work with regard to relative clause identification, relative pronoun correction, and text simplification are described in the remainder of this chapter.

2.1 Classification of Relative Clauses: Restrictive and Non-Restrictive

From a grammatical standpoint, the distinction between RRCs and NRRCs has been studied extensively by linguists and specific syntactic rules related to their use have been firmly established. Some notable examples of previous work in which the characteristics of NRRCs have been intensely studied include [4], [5]. Both [4] and [5] provide detailed explanations of the differences between RRCs and NRRCs despite the use of different terminology.

2.2 Relative Pronouns in Anaphora Resolution

The correction of relative pronouns may be categorized as a task which falls under the larger umbrella of anaphora resolution. During the late 1970s and early 1980s, significant progress was achieved in anaphora resolution using discourse-based approaches such as [6]. However, the main drawback of approaches is that they were heavily dependent on domain and linguistic knowledge, which posed a problem because the work was difficult to process and required a significant amount of human input. This area of NLP is once again of great interest, but now the focus is on developing methods that require little human input or knowledge. Recent techniques for anaphora resolution make use of NLP resources such as large corpora, part-of-speech (POS) taggers, and ML techniques to do the majority of the work [7]. While it still remains to be seen whether systems which rely exclusively on ML can achieve better results than those that incorporate deep syntactic and semantic knowledge, both techniques have been explored and combined to exploit the benefits of each.

Cardie's [8] corpus-based approach to automatically acquire disambiguation heuristics for locating the antecedents of relative pronouns is one such example of this combined approach, where ML in the form of clustering is used by the system to build the hierarchy of disambiguation heuristics. The MUC-3 (Message Understanding Conference, 1993) corpus of 1500 newspaper articles and the UMass (University of Massachusetts)/ MUC-3 parser are used to determine the low-level constituents in the phrase that precedes the relative clause. To create the training instances, Cardie assigns an attribute-value pair to each of the constituents, namely its syntactic class and its position with respect to the relative pronoun.

To build the hierarchy of disambiguation heuristics, Cardie applies a clustering system called COBWEB [9] which, given a new instance to classify, recovers the most specific concept that describes the instance. The UMass /MUC-3 system generates a set of attribute-value pairs that represent the clause preceding the relative pronoun. These attribute-value pairs are known as probes, which the COBWEB system then uses to retrieve the individual sentence that is most similar to the probe.

Cardie trained her system on 170 sentences containing instances of *who* and tested them on 71 new instances. The correct antecedent was identified for 92% of the test instances, compared to the UMass/MUC-3 system that relied on hard-coded heuristics and was only 87% successful.

Soon et al. [7] used a ML approach for the coreference resolution of noun phrases. Their system uses "a small, annotated corpus and tackles pronouns, proper names and definite descriptions" and "the evaluation is carried out on the MUC-6 and MUC-7 test corpora" [7]. In this research, the authors achieve accuracy results that are comparable to non-learning methods.

The two methods described previously focus specifically on the task of identifying the antecedents of pronouns, which is often viewed as one step in the larger NLP task of text simplification.

2.3 Relative Clause Attachment as a Step toward Text Simplification

One school of thought among researchers with respect to resolving relative clause attachment is to view this task as a stepping-stone in the larger task of text simplification. Text simplification is an important area of research because it results in simpler sentential structures and reduced ambiguity. Therefore, it is an effective pre-processing tool in a variety of tasks such as machine translation, parsing, and text summarization [10].

There are a variety of text simplification techniques that have been developed to date. Some of these involve the removal of entire clauses or descriptive phrases, while others break the original sentence into two simpler ones.

The goal of Jing's [11] automatic sentence reduction system is to remove irrelevant phrases from sentences "without major loss" to the meaning of the sentence. For this purpose, Jing devised a five-step algorithm. First, the input sentences are parsed and then checked to identify the components that must remain for it to be grammatically correct. Step three involves an assessment of the components most related to the main topic of the sentence. This is accomplished using a weight formula, where an importance score for each word is computed based on the number and types of lexical relations it has with other words in the sentence. In the fourth stage, the corpus probabilities are calculated to help capture human practice of phrase removal. Finally, the fifth stage involves final reduction decisions based on the results for the previous four steps. With this system, Jing succeeded in creating a corpus of 500 sentences and their reduced forms, and showed that 81.3% of the reduction decisions made by the system agreed with the reductions achieved through human analysis.

While Jing's sentence reduction system for text simplification involved the removal of extraneous information, it did not specifically address the task of simplifying sentences containing relative clauses. Therefore, we looked at previous research in which the task of simplifying sentences containing relative pronouns was examined.

In his methodology, Chandrasekar [12] simplifies text using a Formal Syntactic Grammar (FSG) in which sentences are broken up into word groups called chunks. These chunks may be split at articulation points such as the beginning and end of sentences, punctuation marks, coordinating conjunctions, and relative pronouns. The words in each chunk are tagged and identified using the FSG and ordered using simplification rules. The rule that simplifies sentences containing a relative pronoun is given below:

X:NP, Rel_Pron Y, Z → X:NP Z. X:NP, Y.

In words, this rule may be interpreted as: "If a sentence starts with a noun phrase (**X: NP**), and is followed by a phrase with a relative pronoun of the form (**, Rel_Pron Y,**), followed by some **Z**, where **Y** and **Z** are arbitrary sequences of words, then the sentence may be simplified into two sequences, namely the sequence **X** followed by **Z** and **X** followed by **Y**" [12]. This simplification rule is illustrated by the following example.

John, who loves Mary, is sad. → John loves Mary. John is sad.

Looking ahead to using Chandrasekar's FSG for text simplification as a future task of his work, Siddharthan explored an important issue related to relative clauses: local versus wide attachment. This study, described in detail in Section 3.1, served as the starting point from which our research was developed.

2.4 Situating our Research among Previous Work

Situating our revised syntactic attributes and applications among previous work presents an interesting challenge because our study incorporates several aspects of NLP and ML. Initially, we focused on Siddharthan's machine learning framework because our intent was to present a more complete study of relative clause attachment by incorporating additional linguistic knowledge. By analyzing the grammatical structure of sentences containing relative pronouns, we were able to add new syntactic attributes to the existing machine learning framework consisting primarily of semantic attributes.

We then explored applications related to relative clause attachment in which to use our revised syntactic attributes. For our first application, where sentences containing relative clauses were identified and then reduced by removing NRRCs to simplify the sentence, we developed additional syntactic attributes related to the characteristics of relative clauses. While several studies including [4] and [5] describe the characteristics of relative clauses, we were able to apply these characteristics as attributes to describe the instances in a ML context. As far as the removal of entire NRRCs in sentences as a means of text simplification is concerned, this method could present a benefit to users who seek a shorter, simplified version of text.

For our second application, we found that the correction of relative pronouns in French to English machine translation was another area in which improvements could be sought. Machine translation, particularly statistical machine translation (SMT), is an area of computational linguistics under constant evolution. The use of SMT models may be attributed in large part to the team of Brown et al. [13], who developed a statistical approach for the translation of single sentences. Since that time, the SMT approach has been modified and expanded by several researchers including Marcu and Wong [14], who developed a phrase-based translation model, and Yamada and Knight [15] who proposed a syntax-based SMT model to include structural and syntactic aspects of language into the translation model. Although these SMT approaches have proven to be effective, the common thread throughout these methods is that they are global. Unlike these systems, our revised syntactic attributes work exclusively on the output of machine translation systems and should therefore be regarded as a local, post-processing approach.

While the majority of attributes for Siddharthan's system were semantic ones derived from Word Net hierarchies, we revised and expanded his study to include syntactic attributes related to relative pronouns and clauses. By selecting the attributes to correspond to our applications, we created versatile syntactic attributes which may be selected and varied in order to suit the application.

Chapter 3: Revised Syntactic Attributes

3.1 Siddharthan's Machine Learning Framework for Resolving Relative Clause Attachment Ambiguities

3.1.1 Purpose and Description

The basis of our revised syntactic attributes is Siddharthan's [2] method for resolving relative clause attachment ambiguities with Word Net and machine learning techniques.

The purpose of Siddharthan's work was to correctly assign local or wide attachment to relative clauses in sentences in which the relative clause is preceded by the noun phrase construct, **NP1 Prep NP2**. Local attachment is assigned to relative clauses that refer to the noun phrase NP2, while wide attachment is assigned to those referring to the noun phrase NP1. For example, in the sentence,

> ***"The pace of life was slower in those days', says [51-year-old Cathy Tinsall] from [South London], who had five children" [2],***

NP1 is "51-year-old Cathy Tinsall", NP2 is "South London", and the relative clause is 'who had five children'. This sentence also displays a case of wide attachment because the relative clause refers to NP1.

To accomplish the goal of correctly learning relative clause attachment, Siddharthan limits his treatment of relative pronouns to two in particular, namely *who* and *which* because previous analysis of the Penn Treebank revealed that 21% of relative clauses containing *who* and 27% of relative clauses containing *which* were preceded by the **NP1 Prep NP2** phrase structure [16]. Siddharthan's ML framework consists of attributes rooted in lexicalization over prepositions and Word Net hierarchies. First, Siddharthan studied the prepositions in the preceding noun phrases of his dataset to determine if they favour local or wide attachment. Although he found that certain prepositions did show a preference for a particular attachment, the top five most commonly occurring prepositions did not. From here, Siddharthan went on to develop attributes for his machine learning framework based on Word Net hierarchies. Word Net is "an online lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives, and adverbs are organized into synonym sets, each representing one underlying lexical concept" [17]. To distinguish between *who* and *which*, Siddharthan relied on the fact that *who* normally refers to something with personality, whereas *which* refers to something without [4]. When he applied this knowledge to Word Net, Siddharthan logically concluded that the relative pronoun *who* could only refer to hyponyms of the human, group or animal classes while *which* could not refer to humans. At this point, Siddharthan defined the attributes for his machine learning framework.

3.1.2 Siddharthan's Attributes and Machine Learning Approach

Siddharthan's machine learning framework consisted of a vector of thirty-five binary attributes for each sentence in his dataset consisting of a restrictive or non-restrictive relative clause, and containing one of the relative pronouns *who* or *which* preceded by the phrase structure **NP1 Prep NP2**. Each instance in Siddharthan's dataset was encoded in an attribute-value vector, which also included the class of each instance. Siddharthan's list of binary attributes is given in Table 1.

Table 1 - Siddharthan's List of Attributes [2]

0:	Target (wide attachment)
1:	Target (local attachment)
2:	Restrictive Clause (defined by absence of comma)
3:	NP1 is a *person*
4:	NP1 is a *group*
5:	NP1 is an *animal*
6:	NP1 is a *possession*
7:	NP1 is an *entity*
8:	NP1 is an *act*
9:	NP1 is an *abstraction*
10:	NP1 has no WordNet class
11:	NP1 is a proper noun
12:	NP1 is a definite NP (presence of definite determiner)
13:	NP1 has no determiner
14-17:	Presence of top 4 prepositions
18:	Prep favours local attachment
19:	Prep favours wide attachment
20:	NP2 is a *person*
21:	NP2 is a *group*
22:	NP2 is an *animal*
23:	NP2 is a *possession*
24:	NP2 is an *entity*
25:	NP2 is an *act*
26:	NP2 is an *abstraction*
27:	NP2 has no WordNet class
28:	NP2 is a proper noun
29:	NP2 is a definite NP (presence of definite determiner)
30:	NP2 has no determiner
31:	Verb selects for singular subject
32:	Verb selects for plural subject
33:	NP1 is singular
34:	NP2 is singular

From Table 1, it can be seen that attributes 3 to 13 and 20 through 30 are parallel because they apply to noun phrases NP1 and NP2 respectively. In addition, Siddharthan uses his study of the most commonly occurring prepositions and his findings for preposition preferred attachment for attributes 14 to 17, 18, and 19.

To train his system in the decision between local and wide attachment, Siddharthan used the SNoW machine learning package [18] and achieved results with the WINNOW algorithm that were significantly higher than the baseline results in which local attachment was always selected. Although Siddharthan's results suggest that the use of Word Net semantic classes is extremely helpful in the disambiguation of relative clause attachment, upon closer examination of his attributes, we hypothesized that these results could be improved by expanding the machine learning framework with several important modifications. These changes form the basis for our revised syntactic attributes described in the following section.

3.2 Incorporating Linguistic Knowledge and Additional Syntactic Attributes

In order to further develop Siddharthan's system for resolving relative clause attachment ambiguities, it was necessary to identify and expand on several important aspects of the task.

The first important difference between Siddharthan's machine learning framework and our expanded version is that our system attempts to correct sentences containing the relative pronouns *whom*, *whose*, and *that*, in addition to *who* and *which*. The relative pronouns *who*, *whom* and *whose* are defined as those which have personal references, although *whose* may also have a non-personal reference [4]. A personal reference is usually human, but can also be non-human inanimate [5]. While *who* is used in the subjective case, *whom* and *whose* are used in the objective and genitive cases respectively [4]. The sentences (a), (b), and (c) illustrate the use of these relative pronouns.

(a) The student ***who*** scored the highest grade received the prize.
→ In this case, *who* refers to 'student'

(b) The patient ***whom*** the doctor had treated recovered from the car accident.
→ In this case, *whom* refers to 'patient'

(c) The girl ***whose*** dog was lost adopted a new pet from the shelter.
→ In this case, *whose* refers to 'girl'

The relative pronoun *which* may only have a non-personal reference [4]. An example of its use is given in sentence (d).

(d) The library ***which*** had been destroyed in the fire was rebuilt last month.
→ In this case, *which* refers to 'library'

Finally, the pronoun *that* may have a personal or non-personal reference and replaces *who*, *which*, and *whom*, when it introduces restrictive relative clauses [4]. Examples (e), (f) and (g) illustrate the use of the relative pronoun *that* in place of *who*, *which* and *whom*.

(e) The man ***that*** is flying the plane is the pilot.
→ In this case, *that* refers to 'man' and may be used in place of *who*

(f) The paper that the graduate student read was very helpful.
→ In this case, *that* refers to 'paper' and may be used in place of *which*

(g) The woman that the firefighter rescued was extremely thankful.
→ In this case, *that* refers to 'woman' and may be used in place of *whom*

Although *who* and *which* make up the majority of relative pronouns used in sentences containing relative clauses, we deliberately selected a broad set of relative pronouns to work with in order to expand the existing framework.

The second aspect of Siddharthan's framework we examined was the syntactic structure of the phrase preceding the relative pronoun. Our goal in studying the phrase preceding the relative pronoun was to broaden the framework so that it could handle a variety of phrase structures, in addition to **NP1 Prep NP2.** This is important because the structure of the preceding phrase is variable. Although it is reasonable to expect the **NP1 Prep NP2** phrase structure to precede a relative containing *who* or *which* because noun phrases are used to describe persons, places or things, this is not always the case. For example, previous research [16] indicates that this construct preceded only 25% of sentences containing *who* or *which* in the Penn Treebank. Therefore, we felt that it was important to consider sentences with a variety of phrase structures. Some examples of preceding phrase structures handled by our system are: **NP1 VP2**, **VP1 NP2**, and **NP1 VP NP2**. We have modified the list of attributes to account for these clause structures.

In addition to exploring various grammatical structures of the phrase preceding a relative pronoun, another aspect of Siddharthan's framework we sought to improve was the resolution of relative clause attachment in cases where the attachment decision must be made between two personal or non-personal references. The sentence given in (h) illustrates this type of situation.

(h) [The story] in [the book] **which was published** was written by a Canadian author.

In (h), both NP1, 'the story', and NP2, 'the book', are non-personal references and therefore, the decision of local versus wide attachment of the relative pronoun *which* becomes more complex. Keeping in mind this type of situation, we devised new syntactic attributes which would capture the elements contained within the preceding phrase as well as within the relative clause to distinguish between NP1 and NP2 and thus determine the appropriate relative pronoun attachment.

As a result of these considerations, we were able to revise to Siddharthan's existing framework and add new syntax-based attributes to the existing attribute vector. The decision to perform further syntactic analysis of sentences was motivated by the fact that the majority of attributes Siddharthan selected were based on semantic information obtained from Word Net regarding the categories of the preceding noun phrases. Using the UCREL CLAWS part-of-speech (POS) C5 tag set [19], a varying window of one word up to a maximum of five words before and after the relative pronoun was tagged (tags n±1, n±2, n±3, n±4, n±5, where n is the relative pronoun). The tags obtained for this window of words were then used as attributes for the system. The C5 POS tag set is given in Table 2. Our intention in using POS tags as attributes was to identify and exploit syntactic structures contained in the preceding phrases and relative clauses. These syntactic structures are described in the following paragraphs.

Table 2 – UCREL CLAWS C5 Tag set [19]

AJ0	Adjective
AJC	comparative adjective
AJS	superlative adjective
AT0	Article
AV0	adverb (unmarked)
AVP	adverb particle
AVQ	wh-adverb
CJC	coordinating conjunction
CJS	subordinating conjunction
CJT	the conjunction THAT
CRD	cardinal numeral
DPS	possessive determiner form
DT0	general determiner
DTQ	wh-determiner
EX0	existential THERE
ITJ	interjection or other isolate
NN0	noun (neutral for number)
NN1	singular noun
NN2	plural noun
NP0	proper noun
NULL	the null tag
ORD	ordinal

PNI	indefinite pronoun
PNP	personal pronoun
PNQ	wh-pronoun
PNX	reflexive pronoun
POS	the possessive (or genitive morpheme) 'S or '
PRF	the preposition OF
PRP	preposition (except for OF)
PUL	punctuation - left bracket
PUN	punctuation - general mark (i.e. . ! , : ; - ? ...)
PUQ	punctuation - quotation mark (i.e. ` ' ")
PUR	punctuation - right bracket (i.e.) or])
TO0	infinitive marker TO
UNC	"unclassified" items which are not words of the English lexicon
VBB	the "base forms" of the verb "BE" (except the infinitive)
VBD	past form of the verb "BE"
VBG	-ing form of the verb "BE"
VBI	infinitive of the verb "BE"
VBN	past participle of the verb "BE"
VBZ	-s form of the verb "BE"
VDB	base form of the verb "DO" (except the infinitive)
VDD	past form of the verb "DO"
VDG	-ing form of the verb "DO"
VDI	infinitive of the verb "DO"
VDN	past participle of the verb "DO"
VDZ	-s form of the verb "DO"
VHB	base form of the verb "HAVE" (except the infinitive)
VHD	past tense form of the verb "HAVE"
VHG	-ing form of the verb "HAVE"
VHI	infinitive of the verb "HAVE"
VHN	past participle of the verb "HAVE"
VHZ	-s form of the verb "HAVE"
VM0	modal auxiliary verb
VVB	base form of lexical verb (except the infinitive)
VVD	past tense form of lexical verb
VVG	-ing form of lexical verb
VVI	infinitive of lexical verb
VVN	past participle form of lex. verb
VVZ	-s form of lexical verb
XX0	the negative NOT or N'T
ZZ0	alphabetical symbol

The first pattern we attempted to catch by tagging the words in the relative clause is the **Preposition + Relative Pronoun** combination that may also be followed by an infinitive in the relative clause [4]. Example sentences with this grammatical structure are given in (i) and (j).

(i) She has a lot of friends ***from whom*** to seek advice.

(j) The employee required a larger office ***in which*** to work on the project.

The POS tags for the five words preceding the relative pronoun (tags n-1, n-2, n-3, n-4 and n-5, where n is the relative pronoun) were also carefully selected as attributes to exploit the grammatical patterns that exist in this part of the sentence. Because the relative pronoun *whom* is used in the objective case, this suggests that the noun subject(s) should be contained in the preceding phrase. Similarly, because the pronoun *whose* is used in the genitive case, this also suggests that the referent noun is located in the preceding phrase. The fact that many relative pronouns are preceded by prepositions also suggests a pattern in the word directly preceding the relative pronoun (tag n-1). Finally, the words *where*, *when*, and *how* can also be used as relative pronouns in place of *which*. However, in order to do so, the structure of the relative clause must be made up of the combination, **Subject + Finite Verb,** as seen in example (k) [20]. This is another pattern that we wanted to detect using tags as attributes to describe each sentence.

(k) I study in a school ***which*** has three floors.
→I study in a school ***where*** there are three floors.

Once our syntactic analysis was complete, we were able to include several new syntactic attributes in the existing machine framework. The list of new attributes, along with the possible values for each is given in Table 3.

Table 3 – Revised Syntactic Attributes

New Attribute	Possible Values
VP1 exists	0, 1
VP2 exists	0, 1
n+1 tag / n-1 tag n+2 tag / n-2 tag n+3 tag / n-3 tag n+4 tag / n-4 tag n+5 tag / n-5 tag	AJ0, AJC, AJS, AT0, AV0, AVP, AVQ, CJC, CJS, CJT, CRD, DPS, DT0, DTQ, EX0, ITJ, NN0, NN1, NN2, NP0, NULL, ORD, PNI, PNP, PNQ, PNX, POS, PRF, PRP, PUL, PUN, PUQ, PUR, TO0, UNC, VBB, VBD, VBG, VBI, VBN, VBZ, VDB, VDD, VDG, VDI, VDN, VDZ, VHB, VHD, VHG, VHI, VHN, VHZ, VM0, VVB, VVD, VVG, VVI, VVN, VVZ, XX0, ZZ0

Having developed and expanded Siddharthan's machine learning framework with our syntactic study, we devised six attribute sets, with Siddharthan's attributes making up the baseline set. For the remaining five sets, we varied the number of new syntactic attributes in order to determine which combination of semantic and syntactic features was the most effective for several tasks. A description of each attribute set is given in Table 4.

Table 4 - Attribute Sets

Attribute Set	*Description*
a	**Baseline set: Siddharthan's Attributes**
b	Siddharthan's Attributes, POS tag for first word after relative pronoun (tag n+1)
c	Siddharthan's Attributes, POS tags for five words after relative pronoun (tags n+1, n+2, n+3, n+4, n+5)
d	POS tags for five words after relative pronoun (tags n+1, n+2, n+3, n+4, n+5)
e	POS tags for five words before and after relative pronoun (tags n±1, n±2, n±3, n±4, n±5)
f	Siddharthan's Attributes, POS tags for five words before and after relative pronoun (tags n±1, n±2, n±3, n±4, n±5)
Note: 'n' is the position of the relative pronoun	

Upon examination of Table 3, it can be seen that for the new syntactic attributes we introduced, the number of possible values was very large. Having such a large number of values for these attributes could weaken the performance of the machine learning algorithms we used because this meant that the branching factors were also very large. However, an analysis of the distribution of the attribute values revealed that this was not the case. As seen in Table 4, the maximum number of POS tags used to describe any instance in our datasets at one time was ten and this occurred when attribute sets (e) and (f) were used. By counting the number of occurrences of each POS tag, we found that out of the sixty-two possible attribute values, sixteen were never assigned to any of the instances in our datasets. In addition, twenty-six of these values occurred fewer than fifteen times. As a result, the majority of instances were described by the remaining twenty values of this attribute and thus, the branching factor was significantly less than hypothesized. Furthermore, because we were interested in the performance of attribute sets (b)-(f) compared to that of the baseline attribute set (a), the high branching factor of the new syntactic attributes could be disregarded because it was consistent throughout the new attribute sets and did not affect the results obtained for attribute set (a).

3.3 A General Two-Stage System

Once we had revised Siddharthan's framework to include new syntactic attributes, we focused our attention on the development of new applications. For our study, we looked at the problems of relative clause identification and simplification, as well as relative pronoun correction in French to English machine translations. It became apparent that a common two-stage system could be used for both tasks because they required the same approach, where we performed an initial screening of the data in the first stage, followed by a more detailed examination in the second stage to correct the problem at hand. In the following sections, this two-stage process is described for each application, first in general terms, followed by a machine learning perspective.

3.3.1 General Description of the Two-Stage System

The initial screening performed in stage 1 served to extract the data which required further analysis and correction in stage 2. For the task of relative clause identification and simplification, the purpose of stage 1 was to distinguish between sentences containing non-restrictive and restrictive relative clauses. Then, in stage 2, we focused exclusively on sentences containing non-restrictive relative clauses because our goal was to determine whether these clauses could then be removed in order to simplify the sentences.

For the task of relative pronoun correction in French to English machine translations, stage 1 served to distinguish between correct and incorrect sentences, where the latter contained an inappropriately translated relative pronoun. In stage 2, we focused exclusively on the incorrect sentences in order to fix the relative pronoun they contained.

Although we utilized the same two-stage approach for both applications, the importance of each stage varied in each case. For relative clause identification and simplification, stage 1 was the more crucial step because it was here that we separated the sentences containing non-restrictive clauses from those containing restrictive clauses. Because further simplification of the sentences in stage 2 could only be done on sentences containing non-restrictive relative clauses, this meant that the initial separation of sentences according to clause types was essential to the overall success of the system.

By contrast, for the task of relative pronoun correction, the detailed analysis of the data in stage 2 was more crucial to the overall success of this application because it was here that we were able to correct the relative pronoun errors contained in the sentences. Although the initial screening in stage 1 separated the incorrect sentences from the correct ones, the ability of the system to correct relative pronoun errors in stage 2 did not depend on the success of this initial screening process.

The previous description illustrates the two stages of our system in general terms. However, we implemented our system using machine learning techniques and algorithms. As such, we considered a number of factors, including the setup of each stage, machine learning algorithms, and evaluation methodology.

3.3.2 Detailed Description of the Two-Stage System

For the first stage of our two-stage system, binary classification was used, where each instance was classified into one of two classes using our revised syntactic attributes. In a two-class problem, there are four potential outcomes when learning algorithms are used to predict the class of an instance. The first two outcomes are correct classifications and are referred to as **true positive (TP)** and **true negative (TN)**. True positives are defined as positive instances that are correctly classified as positive while true negatives are negative instances that are correctly classified as negative. The other two possible prediction outcomes in binary classification are **false positive (FP)** and **false negative (FN)**, where the former refers to negative instances incorrectly classified as positive and the latter refers to positive instances incorrectly classified as negative [21]. The formulae to calculate the four possible prediction outcomes are as follows:

$$\begin{aligned} TPRate &= \frac{TP}{TP+FN} \\ FPRate &= \frac{FP}{FP+TN} \\ TNRate &= \frac{TN}{TN+FP} \\ FNRate &= \frac{FN}{FN+TP} \end{aligned} \quad [21]$$

This terminology is often depicted as a "two-dimensional confusion matrix with a row and column for each class. Each matrix element shows the number of test examples for which the actual class is the row and the predicted class is the column; good results correspond to large numbers down the main diagonal and small, ideally zero, off-diagonal elements" [21]. The confusion matrix for a two-class problem such as the one in stage 1 can be seen in Figure 2.

		Predicted Class	
		Positive	**Negative**
Actual Class	**Positive**	*TP*	*FN*
	Negative	*FP*	*TN*

Figure 2 - Confusion Matrix

In stage 1, the instances from our datasets were classified as positive or negative, where the negative instances were then carried forward into stage 2. For the task of relative clause identification and simplification, negative instances referred to sentences containing non-restrictive relative clauses and for relative pronoun correction, negative instances referred to sentences containing incorrect relative pronouns. As a result, the ability of our stage 1 system to correctly classify positive and negative instances was an important evaluation criterion of our system. To evaluate the performance of our system in stage 1, we relied on two performance measures, namely recall and precision.

Recall measures the number of negative instances in our dataset that were correctly classified as negative. The calculation for recall is equivalent to the TN Rate given above [21].

Considered the counterpart of recall and defined as "a measure of the proportion of selected items that the system got right" [22], the precision of stage 1 corresponds to the number of instances classified as negative that are in fact negative. Precision is calculated using the following formula:

$$precision = \frac{TN}{TN + FN} \quad [21]$$

For the second stage, we branched out into a larger classification task, where the true negatives and false negatives from the first stage were then classified according to classes specific to each application. In the first application, the classes used were based on the ability of the non-restrictive relative clauses to be simplified, while the correct relative pronouns were the classes for our second application. As with stage 1 of our system, the syntactic and semantic attribute sets were used to classify the instances. A schematic of the two-stage machine learning system is given in Figure 3.

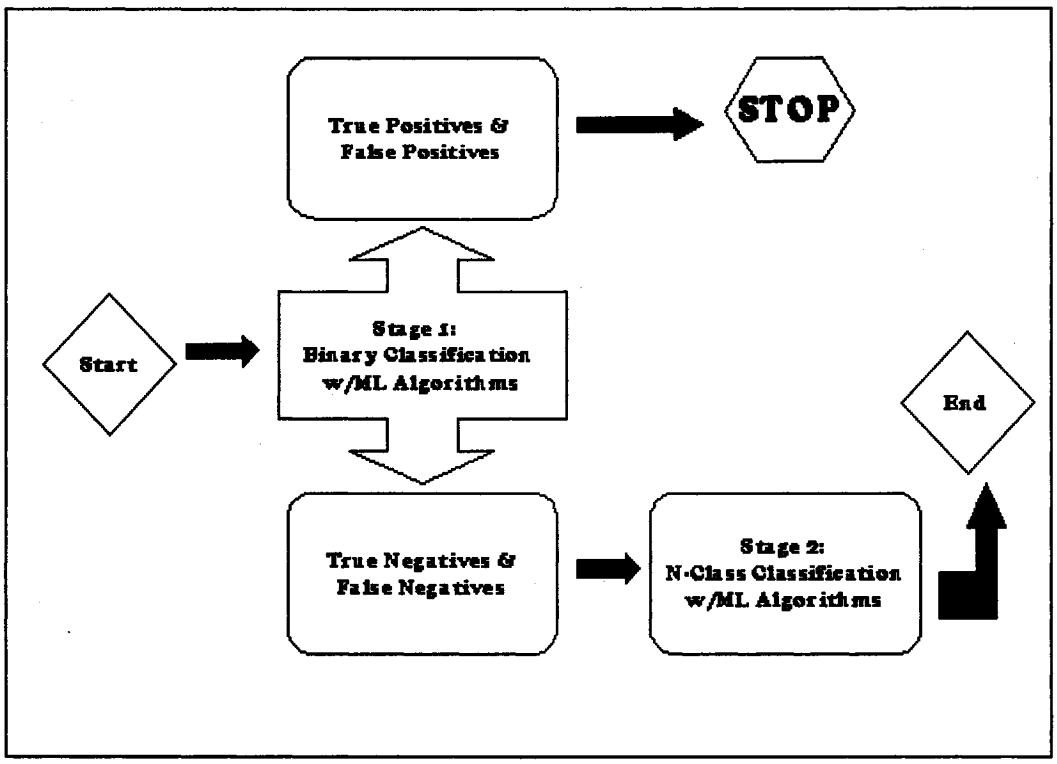

Figure 3 - Two-Stage Machine Learning System

3.4 Machine Learning Algorithms

To train and test our two-stage system, we used the WEKA [21] implementations of the One Rule, Decision Tree and Naïve Bayes algorithms.

The One Rule algorithm was selected because despite its simplicity, it is often effective in coming up with rules for classifying instances. In essence, the One Rule algorithm "uses a single attribute as the basis for its decisions and chooses the one that works best" [21]. It is commonly known as a one-level decision tree.

Decision Trees are commonly known as a 'divide and conquer' scheme for machine learning and consist of nodes, where the occurrence of a node involves testing a particular instance. Typically, the test at a particular node compares an attribute value with a constant, but the comparison may also be between two attributes. The term 'leaf node' is applied to nodes that give a classification for an instance. A decision tree uses the following strategy for classifying an instance: "It is routed down the tree according to the values of the attributes tested in successive nodes, and when a leaf is reached the instance is classified according to the class assigned to the leaf" [21]. The Decision Tree algorithm is the more complicated extension of the One Rule. For this reason, we were interested in comparing the results achieved by the two algorithms to determine whether a complex algorithm is in fact needed to achieve the same or better results.

For its part, the Naïve Bayes algorithm is based on Bayes' rule for conditional probability given by:

$$\Pr(A \mid B) = \frac{\Pr(B \mid A) \cdot \Pr(A)}{\Pr(B)} \quad [22]$$

where Pr (A) is the probability of event A, Pr (A | B) is the probability of event A conditional on event B, and Pr (B) is the probability of event B. Using Bayes' rule for conditional probability, the most likely, or maximum a posteriori (MAP) class for an instance given its attributes is calculated using the following formula:

$$\begin{aligned} v_{MAP} &= \underset{v_j \in V}{\arg\max}\, P(v_j | a_1, \ldots a_n) \\ &= \underset{v_j \in V}{\arg\max} \frac{P(a_1, \ldots a_n | v_j) P(v_j)}{P(a_1, \ldots a_n)} \\ &= \underset{v_j \in V}{\arg\max}\, P(a_1, \ldots a_n | v_j) P(v_j) \end{aligned} \quad [23]$$

Because the simplifying assumption of the Naïve-Bayes classifier is that the attribute values are conditionally independent given the class value, this formula may be reduced to the following:

$$v_{NB} = \underset{v_j \in V}{\arg\max}\, P(v_j) \prod_i (a_i | v_j) \quad [23]$$

where $a_i, \ldots a_n$ are attributes and v_{NB} is the target output value of the classifier. Although the Naïve-Bayes algorithm is simplistic and assumes independence between attributes, it is still effective in classification tasks.

In terms of the parameters for the learning algorithms, the default values for One Rule, Decision Trees and Naïve-Bayes were chosen because the primary purpose of these experiments was not to evaluate the performance of each algorithm individually by altering its parameters, but to compare their performance relative to one another. Therefore, the parameters of each algorithm had to remain consistent in order to observe their relative performance.

3.5 Experimental Setup

To evaluate the two stages of our system, we used a slightly different evaluation methodology in each case. In stage 1, we used the holdout method, where half our data was set aside for testing leaving the other half for training [21]. Because we were interested in evaluating the performance of the two-stage system as a whole, we deliberately kept the test set large to ensure that the highest number of true negatives would be passed on to stage 2. By the same token, we were also interested in making sure that the lowest number of false negatives was propagated to stage 2.

To evaluate the classification task in stage 2, we used tenfold cross-validation. For tenfold cross-validation, the dataset, consisting of the true negatives and false negatives from stage 1, was randomly divided into ten parts or folds, where nine of these parts were used for training and the tenth part for testing. For each test fold, the percentage of correctly classified instances was calculated. Because the process is 10-fold, it was repeated ten times, using a different test portion each time and then the average of these ten repetitions was calculated [21].

We validated the stage 2 results for the percentage of correctly classified instances by determining the statistical significance of these results. For our study, we used the paired t-test, which is used to calculate a percentage confidence interval estimate [23]. For our experiments, we used the paired t-test to calculate the percentage confidence interval between the attribute set among our revised sets which achieved the highest accuracy (one of sets (b), (c), (d), (e), or (f)), and Siddharthan's set of attributes (set (a)). In this case, we selected a desired confidence interval of N =95% because this is the value typically used in tenfold cross validation [21]. The percentage confidence interval estimate is given by:

$$\delta \pm t_{N,k-1} s_\delta \quad [23]$$

where δ is the average accuracy difference between the two attribute sets over ten folds, k -1 is the number of degrees of freedom, s_δ is an estimate of the standard deviation of the distribution governing δ, and $t_{N,k-1}$ is a constant chosen based on the desired confidence level. For a confidence interval of 95%, tables were used to determine a confidence limit of 1.83 for the Student Distribution with nine degrees of freedom, which is the corresponding number for 10-fold cross-validation [21]. The value of s_δ was calculated using the equation,

$$s_\delta = \sqrt{\frac{1}{k(k-1)} \cdot \sum_{i=1}^{k} (\delta_i - \delta)^2} \qquad [23]$$

where δ_i represents the accuracy difference between the two attribute sets for each fold of cross-validation. Using these formulae, we verified the hypothesis that the error of the baseline attribute set exceeded the error of the attribute set from our revised framework by a given confidence percentage.

With the two-stage evaluation system in place, we were able to develop applications with which we could evaluate the performance of our revised syntactic attributes. The first 'mini-application' is discussed in Section 3.6 and our full applications pertaining to relative clause identification and relative pronoun correction are discussed in Chapters 4 and 5 respectively.

3.6 Mini-application: Local vs. Wide Attachment to illustrate Theoretical Framework

As the first step in demonstrating the effectiveness of our revised syntactic attributes, we repeated Siddharthan's experiments for assigning local or wide attachment to relative clauses. Our primary goal in doing so was to prove our hypothesis that the use of new syntactic attributes together with the existing ones would be more effective than relying solely on Siddharthan's attributes to classify the relative clauses. For these experiments, there were a variety of considerations, including data collection and the experimental setup for evaluating the revised syntactic attributes.

3.6.1 Training and Test sets

The first step in verifying the validity of our hypothesis that the combination of semantic and syntactic attributes was better suited for the task of assigning local or wide attachment to relative clauses was to collect a dataset of sentences in which the relative pronoun was preceded by the **NP1 Prep NP2** phrase construct described in Section 3.1.1. A dataset of 148 sentences containing relative pronouns preceded by this particular phrase construct was collected from articles contained in [24]. Each of the 148 instances in the dataset was encoded in a vector containing the value for each attribute described in Section 3.1.2 and the attribute sets from Table 4 were used. This means that the experiments were repeated six times, using a different combination of attributes to describe the instances each time. For these experiments, the class of each instance was the correct attachment, local or wide, as determined from an analysis of each instance by the author.

3.6.2 Evaluation Methodology

To perform an initial evaluation of our revised syntactic attributes, we repeated the experiments from [2]. In this case, we used only the stage 2 evaluation procedure described in Section 3.3.2, where the instances were classified according to local or wide attachment and the percentage of correctly classified instances was calculated.

3.6.3 Experimental Results

To evaluate the performance of our revised syntactic attributes in the classification of sentences according to local or wide attachment, various experiments were conducted. For experiment #1, we included the relative pronoun in each sentence as an attribute, along with the attributes we devised based on our grammatical analysis from Section 3.2. Table 5 shows the percentage of correctly classified instances for each attribute set using the three learning algorithms.

Table 5 - Experiment #1: ***% of Correctly Classified Instances***

Experiment #1				
Attribute Set	*One Rule (1R)*	*Decision Tree (DT)*	*Naïve-Bayes (NB)*	*Average*
a	***55.59***	***69.08***	***65.22***	***63.30***
b	55.12	70.76	65.91	**63.93**
c	65.11	63.84	68.07	**65.67**
d	65.11	65.52	59.76	**63.46**
e	65.11	65.52	67.50	**66.04**
f	65.11	63.78	68.73	**65.87**

For experiment #2, the relative pronoun in each sentence was removed from the list of attributes, as it did not form part of the revised syntactic attributes we developed and only served as a means of keeping track of the instances in each experiment. The removal of this attribute was necessary in order to accurately evaluate the performance of the system. The results for experiment #2 are given in Table 6.

Table 6 - Experiment #2: ***% of Correctly Classified Instances***

Experiment #2				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a'	***55.59***	***69.50***	***65.55***	***63.55***
b'	55.12	72.32	65.30	**64.25**
c'	65.11	64.66	67.65	**65.81**
d'	65.79	65.29	58.98	**63.35**

e'	65.11	65.52	67.44	**66.02**
f'	65.11	64.46	68.53	**66.03**

From the results of Tables 5 and 6, it can be seen that our revised attribute sets outperform Siddharthan's attribute set in four out of five cases. In experiment #1, the percentage of correctly classified instances is higher in all cases when our syntactic attributes are used in combination with the baseline set of attributes or by themselves. In experiment #2, the percentage of correctly classified instances obtained for attribute set (d') falls slightly below Siddharthan's attribute set. This is likely due to the fact that only syntactic attributes were used in this attribute set. However, this result supports our hypothesis that in order to achieve the highest results, our revised syntactic attributes must be used *in conjunction with* the existing attributes.

To calculate the percentage confidence interval estimate, we performed a paired t-test comparing the best result for the baseline attribute set (a') and the best result achieved among our revised attribute sets. From Table 6, it can be seen that these values corresponded to attribute sets (a') and (b') where the Decision Tree algorithm was used in both cases.

Using the formulae given in Section 3.5, we calculated the percentage confidence interval estimate to be 2.82 ± 1.89%. This meant that we could accept the hypothesis that the error of the baseline attribute set (a') exceeded the error of our revised attribute set (b') with a confidence, $\delta = 2.82\%$.

With the knowledge that four out of five of our revised syntactic attribute sets outperformed Siddharthan's attribute set in the task of classifying relative clauses according to local versus wide attachment, we explored additional applications related to relative clauses and relative pronouns in which to test these attributes.

Chapter 4: Application #1 - Classification and Simplification of Sentences Containing Relative Clauses

4.1 Problem Description

The first application to which we applied our revised syntactic attributes was the classification and simplification of sentences containing relative clauses.

In English, there are two main types of relative clauses: restrictive relative clauses and non-restrictive relative clauses. *Restrictive relative clauses* (RRCs) are "closely connected to their antecedent" [5] and are not typically set off by a comma in sentences. By contrast, *non-restrictive relative clauses* (NRRCs) are "parenthetic comments which usually describe, but do not further define their antecedent" [5]. Unlike RRCs, NRRCs are always separated from the other elements in a sentence by commas. The examples in Table 7 illustrate the use of each type of relative clause.

Table 7 - Sentences containing RRC and NNRC

Sentence containing RRC	The man <u>who robbed the bank</u> was arrested.
Sentence Containing NRRC	My grandmother, <u>who was born in Tanzania</u>, lives in New York.

Grammars such as [3], [4] and [5] have concurred that the primary syntactic means of distinguishing between types of relative clauses is a comma to separate a NRRC from the remainder of a sentence. However, the true semantic distinction between RRCs and NRRCs in English text remains a challenge to ascertain. This distinction is often clearer in spoken English because the speaker is able to include non-verbal cues such as a pauses or intonation breaks [27]. Although a comma is one of the widely accepted relative markers in English text, it is not the only one. In this application, we were interested in exploiting additional syntactic characteristics of relative clauses in order to perform text simplification.

Because RRCs contain essential information about the person or thing being described, their omission from the sentence will result in a severe loss of meaning and thus reduce the overall comprehension of the text. On the other hand, NRRCs contain additional information about the person or thing being described in the sentence. As such, it may be possible to remove a NRRC from the sentence and still retain the essential meaning. The effect of removing a relative clause from a sentence can be seen in Table 8, where the examples from Table 7 were simplified. The sentence previously containing a RRC, although still grammatically correct, has lost the essential information required for understanding, while the basic understanding of the sentence previously containing a NRRC has been retained despite the removal of the relative clause.

Table 8 - Simplified Sentences (RRC and NNRC Removed)

Simplified Sentence (RRC removed)	The man was arrested.
Simplified Sentence (NRRC removed)	My grandmother lives in New York.

Although the sentence in Table 8 in which the NRRC has been removed retained its essential meaning, it is worth noting that in a complete text, the true meaning of a sentence is often derived from its surrounding context. Therefore, although a sentence in which a NRRC has been removed remains comprehensible, it may still be difficult to fully grasp the intended meaning if the NRRC contained information related to another sentence in the text. However, with our application, our intent is to present a simplification method, which does not take surrounding context into account. The focus of this study is to present simple and direct methods for the simplification of individual sentences.

4.2 Two-Stage System Overview

To apply our revised syntactic attributes, we employed the two-stage system described in Section 3.3. The purpose of the first stage was to classify the sentences according to the type of relative clause they contained: restrictive or non-restrictive. From here, the true negatives and false negatives were passed on to the second stage. For this application, true negatives refer to sentences containing a NRRC correctly identified as containing a NRRC, and true positives referred to sentences containing a RRC, correctly identified as containing a RRC. In addition, a sentence containing a NRRC falsely classified as one containing a RRC was considered a false positive and a sentence containing a RRC falsely classified as one containing a NRRC was labeled a false negative. The goal of the second stage was to automatically classify the sentences containing NRRCs according to text simplification rankings using our syntactic and semantic attribute sets. The rankings used as classes in stage 2 were values of 0 and 1, where these values represented a recommendation for simplification. For sentences ranked '0', this meant that simplification was not recommended as the removal of the NRRC would result in the loss of too much essential information and render the sentence incomprehensible. A ranking of '1' was given to sentences in which the removal of the NRRC was recommended because it did not result in the loss of any vital information and thus was a good candidate for further simplification. A schematic of the overall system for this application is given in Figure 4.

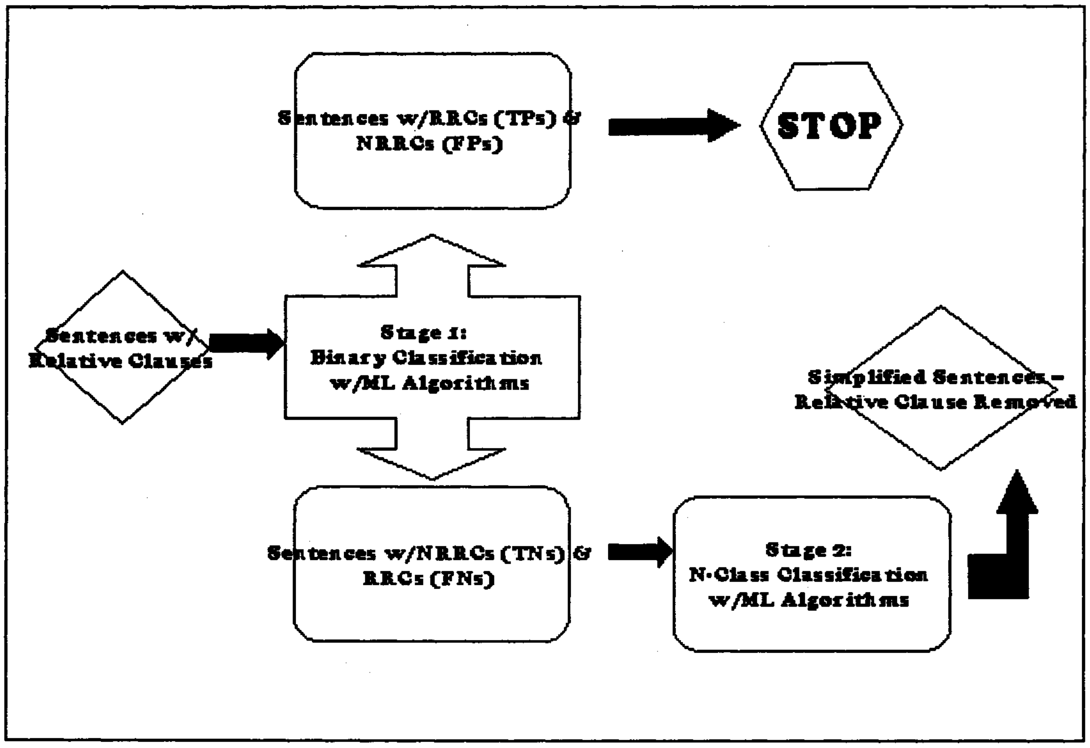

Figure 4 – Architecture for Application #1

4.3 Categories of Relative Clauses

For this application, a dataset of 395 sentences was collected from [24], each containing either a RRC or NRRC. The number of occurrences of each type of relative clause is given in Table 9.

Table 9 - Breakdown of Sentences according to Relative Clause Type

Relative Clause Type	Number Of Occurrences
Restrictive	251
Non-Restrictive	144
Total	***395***

4.4 Training and Test Sets

To train the system, a dataset of 199 instances was randomly selected from the complete set of instances described in Section 4.3. To test the system, the remaining 196 instances were used. Each instance in the training and test sets was encoded using the attribute-value vector from our revised attribute sets. The class value of each instance was the type of relative clause it contained, where the presence of a comma served to make the distinction between a RRC and NRRC.

4.5 Experimental Results: Standard Case

With the two-stage system in place, we were able to perform experiments and evaluate the performance of our revised attribute sets in the classification and simplification of sentences containing relative clauses. We began with a basic case and subsequently modified it to see how the results could be further improved.

In our basic experimental case, we included the relative pronoun of the sentence in the attribute vector describing each instance. For stage 1, the positive and negative recall rates for each attribute set obtained using our three learning algorithms are given in Tables 10 and 11. The recall rates correspond to the true positive and true negative rates.

Table 10 - Stage 1 Experimental Results: Positive Recall Rate

Stage 1 Experiment: ***Positive Recall Rate***			
Attribute Set	*1R*	*DT*	*NB*
a	0.86	**0.87**	0.87
b	0.86	0.86	**0.88**
c	**0.86**	0.86	0.83
d	0.93	**1.00**	0.80
e	**0.95**	0.89	0.90
f	**0.95**	0.86	0.89

From Table 10, it can be seen that our revised attribute sets were very effective in the identification of positive instances, achieving results between 86% and 100%. This is an important result because it meant that very few false negatives were carried into the second stage and thus reduced the likelihood that a sentence containing a RRC would be incorrectly classified as an instance in which removal of the relative clause was recommended.

Table 11 - Stage 1 Experimental Results: Negative Recall Rate

Stage 1 Experiment: *Negative Recall Rate*			
Attribute Set	*1R*	*DT*	*NB*
a	0.57	0.47	**0.58**
b	0.57	0.53	**0.62**
c	**0.57**	0.55	0.57
d	0.24	0.00	**0.47**
e	0.46	0.41	**0.53**
f	0.46	0.56	**0.58**

By contrast, the results from Table 11 show an initially poor performance by our revised attribute sets in correctly classifying negative instances, that is, sentences containing a NRRC. For the three learning algorithms, the negative recall rate ranged between 0 and 57%, which meant that very few negative instances from our initial dataset were correctly identified as negative and passed on to the second stage. In particular, when the Decision Tree algorithm was used on attribute set (d), the system was unsuccessful in classifying any of the negative instances correctly. However, the highest negative recall rate achieved by the three learning algorithms for each attribute set (seen in bold in Table 11) was passed on to the second stage. This meant that the datasets used in the second stage classification task consisted of, on average, only 55% of the negative instances from our test set. The second stage experiments were run, where the true negatives and false negatives were classified according to whether they could be simplified or not. The stage 2 experimental results are given in Table 12. In experiment #1, the relative pronoun in each relative clause was used as an attribute, but it was removed for experiment #2.

Table 12 - Stage 2 Experimental Results: ***% Correctly Classified Instances***

Stage 2: Experiment #1				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a	***68.67***	***60.00***	***75.33***	***68.00***
b	**72.00**	**58.50**	**70.00**	**66.83**
c	45.50	54.00	60.50	53.33
d	51.00	63.33	63.25	59.19
e	44.67	63.33	54.33	54.11
f	56.67	62.00	61.83	60.17
Experiment #2				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a'	***57.33***	***63.17***	***71.67***	***64.06***
b'	**72.00**	**58.50**	**70.00**	**66.83**
c'	45.50	54.00	54.50	51.33
d'	51.00	63.33	63.33	59.22
e'	44.67	63.33	54.33	54.11
f'	55.67	61.67	63.50	60.28

The experimental results for stage 2 indicated an initially weak performance by our attribute sets in this text simplification task. However, the removal of the relative pronoun from the attribute sets in experiment #2 did not have an adverse impact on the results. In general, with the exception of attribute set (b') in experiment #2, Siddharthan's attribute set outperforms the attribute sets (c'), (d'), (e'), and (f') which incorporate revised syntactic attributes in the second stage classification task.

The results obtained for our basic test case indicated that further development of our revised syntactic attributes was necessary in order to improve the overall performance of the system. Because the success of our system required that the highest number of negative instances be propagated to stage 2, we decided that a revision of our syntactic attributes was necessary in order to increase the percentage of true negatives.

4.6 Incorporating Additional Syntactic Attributes

Since the task of classifying relative clauses differed from the original task of determining local or wide attachment, we decided to add more syntactic attributes to our revised attribute sets which better reflected the grammatical construction of RRCs and NRRCs. The purpose of doing so was to improve the ability of our system to correctly classify sentences containing NRRCs so that they could be carried forward into the stage 2 classification task for text simplification. It should be noted however, unlike the syntactic attributes from Table 3 that we devised originally, the values for the additional syntactic attributes described in the following three sub-sections were determined manually because our intention was to assess the ability of these attributes in the stage 1 task of correctly classify sentences containing NRRCs.

4.6.1 Trial #1

In our first attempt to improve the stage 1 results, we introduced two attributes rooted in grammatical rules for relative clauses. The first binary attribute we added to revised attribute sets dealt with the position of the relative clause in the sentence. In a previous study which used the Lancaster/IBM Spoken English Corpus [25], it was found that because NRRCs often contain longer and more complex information than RRCs, they are more likely to appear at the end of a sentence. As such, it was possible to distinguish between NRRCs and RRCs using this characteristic. For our first additional binary attribute, the instances in our dataset in which the relative clause appeared at the end of the sentence were assigned a value of '1', while those instances in which the relative clause appeared elsewhere in the sentence received a value of '0'.

Unlike the previous attribute, the second syntactic attribute we added to our existing attribute sets in an attempt to improve the percentage of true negatives in stage 1 was a numeric one. We used the number of words contained within the relative clause, including the relative pronoun, as an attribute to describe each instance. The basis for this attribute was [26], in which the author studied the difference between RRCs and NRRCs in terms of how much information each one offers. One of the fundamental hypotheses of this work is that the most basic difference between a RRC and a NRRC is the number of information units each one involves. As such, the number of words contained in its relative clause further described each instance in our dataset.

4.6.2 Trial #2

For our second attempt at improving the stage 1 results of our relative clause classification task, we incorporated two more attributes to be used with the existing syntactic attributes. These attributes were based on [27], in which the authors sought to identify NRRCs in English data. Although for the purposes of their study, the authors focused on spoken English, some of the results of their study were applicable to our study. For their research, the authors analyzed 692 sentences containing NRRCs and determined that the use of NRRCs was often accompanied by the occurrence of specific verb forms and types.

The authors found that the majority of the NRRCs in their study had the following construct: **which + modal expression + 'to be'**. Breaking down this construction, the authors reported that the preferred syntactic configuration of NRRCs appeared to be a relative pronoun, followed by an optional modal expression, and some form of the verb 'to be'. A sentence containing this syntactic construction is given in example (l).

(l) The chemistry exam, which should have been the most difficult, was the easiest.
→**which** + modal expression **(should)** + form of 'to be' **(have been)**

Modal expressions are verbs, but there are several important differences that distinguish them from regular verbs. Some of these are: modals do not take an '-s' in the third person, the negation 'not' is used to make a modal verb negative, and lastly, many modal verbs cannot be used in the past or future tenses [28]. Examples of frequently used modal verbs are given in Table 13.

Table 13- Common Modal Verbs [29]

Modal Verbs
Can
Will
May
Would
Should
Must
Need
Might

With the knowledge of this syntactic construct in NRRCs, we were able to devise two additional binary attributes. Instances in which the relative clause contained a modal verb, or a form of the verb 'to be', or both immediately following the relative pronoun were assigned values of '1' and '0' otherwise.

4.6.3 Trial #3

For our third attempt at improving the percentage of true negatives in stage 1 of our application, we incorporated a fifth syntactic attribute into our system. This attribute deals with the antecedent of relative clauses. In [30], the author contends that RRCs are more likely to modify noun antecedents, while NRRCs can have nouns, verbs, adverbs, and even clauses as antecedents. In earlier work [31], the restrictions on relative clauses were examined and it was found that NRRCs are more likely to modify proper nouns. Relying on the results of these studies, the values for our final syntactic attribute that were assigned to each sentence in our dataset were based on the part of speech of the relative clause antecedent, namely noun, proper noun, or clause.

In total, we came up with five new syntactic attributes to include in our existing attribute sets. With these new attributes, we performed three experimental trials in which we gradually increased the number of new attributes in our system and re-tested our revised attribute sets on stage 1 of the relative clause application.

4.6.4 Experimental Results

From the results of Trials 1 to 3 for positive recall rate in Table 14, it can be seen that the progressive incorporation of additional syntactic attributes had very little effect on the positive recall rate from the initial test case given in Table 10. The positive recall rate remained consistently high for all three learning algorithms across all attribute sets, which meant that the number of positive instances that were passed onto the second stage test sets was minimal. This result demonstrates a strong aspect of the system, which is that very rarely was a positive instance misclassified in stage 1 and treated as a negative example in stage 2. Using the results from the third trial, an average of only 7% of positive instances were incorrectly identified as negative instances and carried into stage 2.

Table 14 - Stage 1 Experimental Results: Positive Recall Rate

Stage 1 Experiment: ***Positive Recall Rate***			
Attribute Set	*1R*	*DT*	*NB*
a1	0.86	0.90	0.86
a2	0.86	0.90	0.85
a3	**0.93**	0.89	0.84
b1	0.86	0.87	0.89
b2	0.85	0.87	0.86
b3	**0.93**	0.91	0.88
c1	0.86	0.86	0.83
c2	0.85	0.86	0.83
c3	**0.93**	0.91	0.87

d1	0.93	1.00	0.81
d2	0.91	1.00	0.81
d3	**0.93**	**0.93**	0.87
e1	0.95	0.89	0.90
e2	0.95	0.88	0.91
e3	**0.93**	**0.93**	0.92
f1	0.95	0.86	0.89
f2	0.95	0.86	0.89
f3	**0.93**	0.91	0.90

With respect to the classification of negatives instances, we were able to significantly improve the negative recall results of our basic experiments with the incorporation of additional syntactic attributes into our revised attribute sets. The negative recall results are given in Table 15. In particular, the negative recall rate obtained for attribute set (d) in the third trial increased dramatically with the incorporation new attributes. With the improvements to our system, an average of 60% of the negatives instances in our test set continued on into stage 2, which represented an increase of 5% from our initial experimental case.

Table 15 - Stage 1 Experimental Results: Negative Recall Rate

Stage 1 Experiment: ***Negative Recall Rate***			
Attribute Set	*1R*	*DT*	*NB*
a1	0.57	0.48	0.58
a2	0.57	0.47	0.58
a3	0.53	0.51	**0.60**
b1	0.57	0.53	0.60
b2	0.56	0.53	0.61
b3	0.53	0.49	**0.61**
c1	0.57	0.56	0.56
c2	0.56	0.55	0.57
c3	0.53	0.49	**0.64**

d1	0.24	0.0	0.44
d2	0.25	0.0	0.46
d3	0.52	0.57	**0.57**
e1	0.46	0.40	0.51
e2	0.46	0.41	0.52
e3	0.43	**0.57**	0.56
f1	0.46	0.54	0.57
f2	0.46	0.54	0.58
f3	0.43	0.50	**0.58**

At this point, the two experiments for stage 2 repeated, first including the relative pronoun contained in each instance as an attribute and then without. The stage 2 results based on our expanded attribute sets are seen in Table 16.

Table 16 - Stage 2 Experimental Results: ***% Correctly Classified Instances (true and false negatives from stage 1)***

Stage 2: Experiment #1				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a	***46.26***	***59.02***	***63.48***	***56.25***
b	61.29	48.05	62.43	57.26
c	54.62	60.57	70.83	62.01
d	60.50	54.50	62.83	59.28
e	44.67	50.60	62.67	52.47
f	65.33	63.29	64.12	64.25
Experiment #2				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a'	***48.71***	***59.05***	***63.67***	***57.14***
b'	61.29	47.41	63.83	57.51
c'	54.62	61.00	70.43	62.02
d'	60.50	54.50	64.10	59.70
e'	44.67	51.40	60.57	52.21
f'	65.33	63.29	65.02	64.55

While there was very little difference between the results for experiments 1 and 2, the most significant aspect of these results was that with the exception of attribute set (e'), all of our attribute sets outperform the baseline attribute set (a'). Our results for attribute set (e') also suggest that the use of a unique type of attribute, be it syntactic or semantic, is not the most appropriate for this classification task.

To validate our stage 2 results, we calculated the percentage confidence interval estimate using the paired t-test. This time, as seen in Table 16, the baseline attribute set (a') achieved the highest percentage of correctly classified instances when the Naïve Bayes algorithm was used, and the strongest result by our revised attribute sets was for attribute set (c') for the Naïve Bayes algorithm as well.

The confidence interval estimate was calculated to be 6.76 ± 4.50%. As a result, we accepted the hypothesis that the error of the baseline attribute set (a') exceeded the error of our revised attribute set (c') with a confidence, $\delta = 6.76\%$.

4.7 Application #1 Summary

The overall results for the two-stage relative clause identification and simplification system are summarized in Table 17. These results were calculated using the average values obtained from the six attribute sets across the three machine learning algorithms from experiment #2 of Table 16.

Table 17 - Overall Results for Application #1

Stage 1	
Precision	62.53%
Recall	53.89%
Stage 2	
% sentences simplified appropriately	58.86%
% sentences simplified inappropriately	41.14%

For this application, we used our revised syntactic attributes in the classification of relative clauses, followed by the simplification of sentences through the removal of these clauses. In our basic experiment, the results achieved in stage 1 were mixed. While the revised syntactic attributes were very effective in correctly classifying sentences consisting of RRCs, the percentage of correctly classified negative instances containing NRRCs was considerably lower. Because the purpose of the second stage was then to further classify the true negative instances, we deemed it necessary to devise additional syntactic attributes to improve the stage 1 results. Relying on previous studies of relative clauses, we were able to devise five additional attributes that were incorporated into our existing sets. We conducted three additional experimental trials, in which we gradually incorporated the new syntactic attributes. The third trial, where all five new attributes were included in the attribute sets, showed a distinct improvement in the stage 1 negative recall rate and a consistently high positive recall rate. Although there is still room for improvement in the stage 1 negative recall rate, the increase in the number of correctly identified sentences containing a NRRC may be attributed to the inclusion of additional application-specific syntactic attributes. With these results, we proceeded to stage 2, where the true negatives and false negatives from stage 1 were classified according to whether the relative clause could be removed for overall simplification of the sentence. In our stage 2 experiments, all but one attribute set outperformed the baseline set. This result demonstrated the effectiveness of our revised syntactic attributes in the classification and simplification of sentences containing relative clauses.

In addition to comparing the average performance of our attribute sets with that of the baseline set of attributes across the three machine learning algorithms, we also conducted statistical significance tests to assess the validity of the results achieved by our strongest attribute set. Because the results of the paired t-test revealed a statistically significant difference in the performance of the attribute sets containing new syntactic attributes compared to the baseline set, this served to emphasize the effectiveness of our new syntactic attributes in the task of relative clause identification and text simplification.

Chapter 5: Application #2 - Relative Pronoun Correction in French to English Machine Translation

5.1 Problem Description

For the second application of our revised syntactic attributes, we considered the problem of relative pronoun correction in the context of French to English translation. Upon examination of English translations of French text produced by a machine translation system (MTS), it is obvious that there is room for much improvement, particularly with regard to the translation of sentences containing relative pronouns. Table 18 gives an example French sentence obtained from [32], and the English translation obtained from Babel Fish, a well-known MTS [33]. We elected to use Babel Fish for several reasons. First, we were interested in using a well-known and easily accessible MTS. The other essential feature of Babel Fish that distinguishes it from more state-of-the-art machine translation tools is that the user is not required to supply the translation or language model. This was an important consideration because we were interested in working with the output of the MTS, not the system itself.

Table 18 - Example of French to English Machine Translation

Original French Sentence	Le fonds servira aux étudiants de la Faculté des sciences *qui* auront su garder une moyenne de 8,5 sur 10 et *qui* sont inscrits à un programme de doctorat pour une période de trois à cinq ans.[1]
English Sentence translated using Machine Translator	The funds will be used to the students of the Faculty of Science ***which*** will have known to keep an average of 8,5 out of 10 and ***which*** is registered with a programme of doctorate for one period from three to five years.

As suggested by the sentence in Table 18, when a French text is translated into English using a MTS, the relative pronoun that begins a relative clause may be incorrectly

[1] The official English translation of the sentence reads: "The funds will support Faculty of Science students who maintain a minimum average of 8.5 on 10 and are registered in a doctoral program for three to five years. [24]

translated. The goal of our revised syntactic attributes was to correct these errors and thus produce grammatically correct English text. Our attribute sets were applied exclusively to the output of a MTS and should therefore be regarded as a post-processing approach which allows for the rapid identification and correction of relative pronoun translation errors that affect the ease of comprehension for users.

5.2 Two-Stage System Overview

To apply revised syntactic attribute sets, the general two-stage system described in Section 3.3 was used. The purpose of the first stage was to take the output of a MTS and to identify incorrect English sentences containing mistranslated relative pronouns. Each instance was classified as correct or incorrect using our attribute sets. From here, the true negatives and false negatives were passed on to the second stage. For this application, true negatives refer to sentences containing an incorrect relative pronoun that were correctly identified as incorrect, and true positives refer to sentences containing a correct relative pronoun that were correctly identified as correct. In addition false positives refer to incorrect sentences that were inappropriately identified as correct and false negatives referred to correct sentences that were inappropriately classified as incorrect. The goal of the second stage was to automatically classify the sentences, using our syntactic and semantic attribute sets once again, but this time, according to the appropriate relative pronoun. A visual representation of the system architecture for this application can be seen in Figure 5.

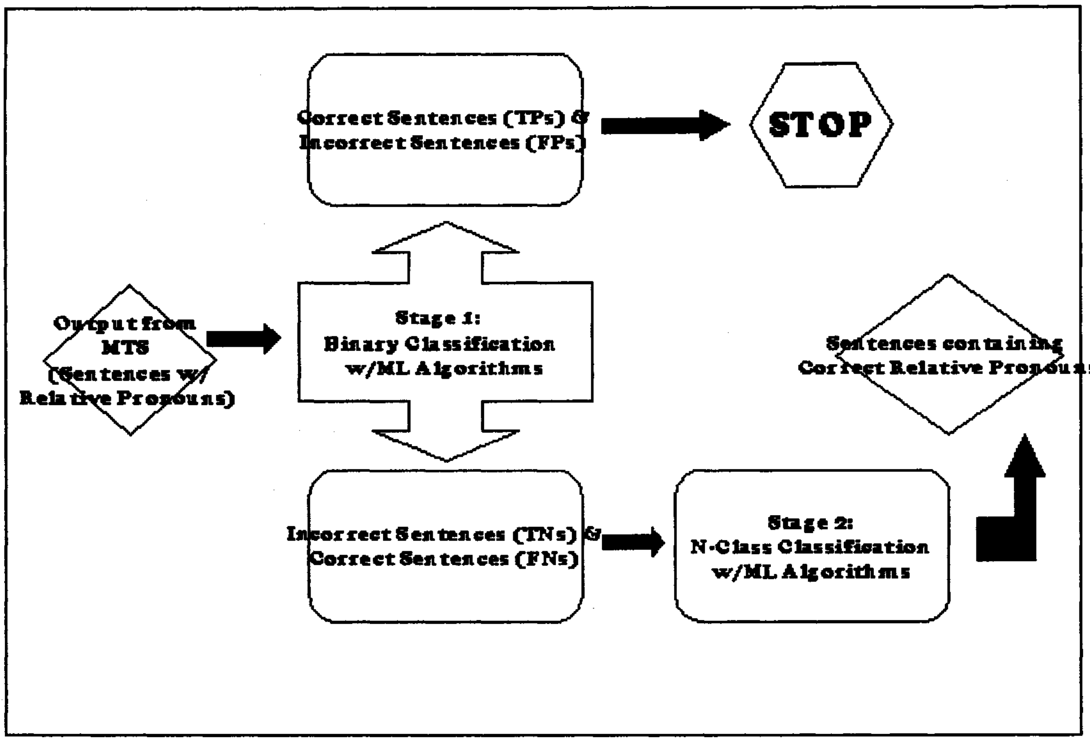

Figure 5 - Architecture for Application #2

5.3 Categories of Relative Pronoun Errors

To determine the types of errors that occur with French to English relative pronoun translations, a dataset of 401 translated sentences was collected using [33]. From the dataset, 200 sentences were labeled “correct” because they contained one or more relative pronouns that were correctly translated. The remaining 201 “incorrect” sentences contained at least one mistranslated relative pronoun. Although it was possible for a sentence to contain more than one appropriate and inappropriate relative pronoun, we did not consider more than one relative pronoun from the same sentence in our dataset. It should also be noted that the majority of sentences in the dataset contained other types of translation errors such as the mistranslation of prepositions and collocations. The translation errors contained in the 201 incorrect sentences were broken down into several categories as shown in Table 19, with the incorrect relative pronoun produced by the MTS and the correct relative pronoun determined from grammatical analysis by the author using [4] and [5]. In the majority of cases, the MTS selected the wrong relative pronoun, but there were cases in which the output of the MTS was not a relative pronoun when it should have been, or where the input of the MTS was not a relative pronoun but the output contained one.

Table 19 - Breakdown of Relative Pronoun Translation Errors

Incorrect Pronoun From MTS	Correct Pronoun From Human Analysis	Number Of Occurrences
Whom	Which	4
Who	Which	11
Which	Who	131
Which	That	12
Which	Whose	2
What	Which	15
What	That	1
That	What	2
That	Than	5
That	Who	1
Than	That	1
That	Whom	1
Which	Whom	5
Whose	Which	2
Which	What	3
Whereas	What	1
Whom	That	3
Who	That	1
Total		**201**

5.4 Training and Test Sets

To train the system, a dataset of 200 instances, consisting of 100 correct sentences and 100 incorrect sentences randomly selected from the complete set of instances described in Section 5.3. To test the system, the remaining 201 instances were used, where 100 of these sentences were correct and 101 were incorrect. Each instance in the training and test sets was encoded using our attribute-value vector.

5.5 Experimental Results: Standard Case

To evaluate the performance of our revised attribute sets which incorporated both semantic and syntactic attributes, we performed a number of experiments in which the percentage of correctly classified instances was calculated. Similar to the first application, we began with a basic case and subsequently modified it to see how the results could be further improved.

In our basic experimental case, we included the relative pronoun produced by the MTS in the attribute vector describing each instance. For stage 1, the recall rate for each attribute set using the three learning algorithms described in Section 3.4 is given in Table 20. The recall rate is equivalent to the true negative rate. For each attribute set, the highest recall rate obtained among the three learning algorithms (shown in bold in Table 20) was used such that the maximum number of true negative instances was propagated to stage 2.

Table 20 - Stage 1 Experimental Results: Recall Rate

Stage 1 Experiment: ***Recall Rate***			
Attribute Set	*1R*	*DT*	*NB*
a	**0.92**	0.82	0.63
b	0.76	**0.82**	0.64
c	**0.92**	0.82	0.63
d	0.42	**0.92**	0.63

e	0.52	**0.92**	0.61
f	0.52	**0.82**	0.62

From Table 20, it can be seen that our revised attribute sets were very effective in the identification of English sentences containing incorrect relative pronouns, particularly in the case of the Decision Tree algorithm, where recall rose from 82% using Siddharthan's baseline set of attributes to 92% with our revised attribute sets. Using the One Rule algorithm, the recall rates for attribute sets (b), (d), (e), and (f) were considerably lower than the baseline attribute set (a). For the Naïve Bayes algorithm, the recall rates remained fairly consistent for the six attribute sets. In general, between 82% and 92% of the negative instances from our test set were propagated to the second stage.

In the second stage, the true negatives identified in the first stage were then classified according to their correct relative pronoun. Two different experiments were conducted in order to determine the percentage of correctly classified instances for each set of attributes using the learning algorithms. In experiment #1, we included the relative pronoun chosen by the MTS as an attribute and the system achieved fairly high results. However, this attribute was removed in experiment #2 in order to eliminate the bias it introduces into the system. This bias relates to the specific MTS that was used. Because the language model of the system is built using specific sentence structures and templates, it is possible that certain relative pronouns will always be selected for specific phrase structures and thus, their occurrence may not be based on linguistic rules, but rather on pre-determined patterns. The stage 2 results are given in Table 21.

Table 21 - Stage 2 Experimental Results: ***% Correctly Classified Instances***

Stage 2				
Experiment #1				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a	***84.12***	***83.92***	***79.91***	***82.65***
b	**87.51**	**89.13**	**83.42**	**86.82**
c	79.37	83.04	82.91	81.77
d	79.88	84.12	82.52	82.17
e	79.88	84.12	79.66	81.22
f	**82.43**	**88.19**	**81.19**	**83.94**
Experiment #2				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a'	***78.69***	***77.03***	***75.67***	***77.13***
b'	78.40	80.96	81.31	**80.22**
c'	79.37	77.17	83.27	**79.94**
d'	79.88	78.80	84.14	**80.94**
e'	79.88	78.80	80.67	**79.78**
f'	75.68	79.97	81.19	**78.95**

From Table 21, it can be seen that the system yielded the highest percentage of correctly classified instances in experiment #1 for attribute sets (b) and (f), in which POS tags were incorporated into the attribute set. The combination of semantic and syntactic attributes achieved nearly 4% better results than the baseline set of attributes in set (a), which indicates that expanding the baseline set of attributes is beneficial to the correction of relative pronoun translations.

In experiment #2, the pronoun generated by the MTS was removed from the attribute sets and the results for the One Rule and Decision Tree algorithms on the six attribute sets were approximately 5% lower than previously. However, the results remained the same or better with the Naïve Bayes algorithm. By removing the pronoun assigned by the MTS from the attribute sets, we removed a potentially redundant attribute and attempted to ensure independence between attributes, which is required in order for the Naïve Bayes algorithm to be successful [21]. In experiment #2, all five of the attribute sets outperformed the baseline set (a'). While attribute sets (d') and (e') consisted of only the syntactic attributes we devised, sets (b'), (c') and (f') combined syntactic attributes and the baseline semantic attributes.

Two important aspects of our system were demonstrated by the results of the stage 2 experiments. First, that the system is capable of learning the correct relative pronouns for these sentences using syntactic descriptors in conjunction with semantic attributes and second, that the system is capable of learning from incorrect instances.

The results obtained for this two-stage system were very encouraging and seemed to suggest that there is a genuine advantage in combining semantic and syntactic attributes for the purpose of identifying incorrect English sentences and correcting relative pronouns. However, once again keeping in mind the realistic performance of our system, we recognized that in stage 1 of our application, it was possible to catch not only true negatives, but to mistakenly catch false negatives as well. Therefore, the precision rate achieved for each attribute set in stage 1 was examined, along with the false negative rate obtained for each attribute set. These results are given in Tables 22 and 23.

Table 22 - Stage 1 Experimental Results: Precision Rate

Stage 1 Experiment: *Precision Rate*			
Attribute Set	*1R*	*DT*	*NB*
a	0.60	0.87	0.68
b	0.55	0.87	0.66
c	0.60	0.87	0.67
d	0.47	0.60	0.60
e	0.60	0.60	0.67
f	0.60	0.87	0.69

Table 23 - Stage 1 Experimental Results: False Negative Rate

Stage 1 Experiment: *False Negative Rate*			
Attribute Set	*1R*	*DT*	*NB*
a	0.64	**0.13**	0.30
b	0.65	**0.13**	0.34
c	0.64	**0.13**	0.31
d	0.49	0.64	**0.43**
e	0.35	0.64	**0.30**
f	0.35	**0.13**	0.28

The precision rates given in Table 22 reflect the percentage of instances in the test set that were classified as negative that are in fact negative. From these results, it can be seen that for all three learning algorithms, the majority of instances classified as negative were in fact negative. However, there are cases, such as for attribute set (d) in which the One Rule algorithm was used, where the number of false negatives carried into stage 2 exceeded the number of true negatives.

The stage 2 experiments were repeated, this time carrying forth both the true negatives and false negatives from stage 1. For each attribute set, the lowest false negative rate obtained from among the three learning algorithms (shown in bold in Table 23) was used such that the minimum number of false negatives was propagated to stage 2. Table 24 shows the revised stage 2 results.

Table 24 - Stage 2 Experimental Results: ***% Correctly Classified Instances (true and false negatives from Stage 1)***

Stage 2: Experiment #1				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a	***80.45***	***78.47***	***76.40***	***78.44***
b	69.84	71.56	71.77	71.06
c	71.91	74.60	73.93	73.48
d	57.97	63.55	65.40	62.31
e	60.15	70.18	66.56	65.63
f	63.23	80.30	75.33	72.95

Experiment #2				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a'	***75.82***	***74.13***	***72.75***	***74.23***
b'	71.12	70.94	72.63	**71.56**
c'	71.91	70.80	74.58	**72.43**
d'	57.97	59.91	64.81	60.90
e'	60.15	66.18	66.97	64.43
f'	61.33	**77.02**	**74.92**	**71.09**

When both the true and false negatives were propagated to stage 2, the results did change significantly. Although the average results for attribute sets (b'), (c') and (f') came within 5% of the results for the baseline attribute set (a'), they did not perform as well as when only true negatives were used in stage 2. The lowest results were obtained with sets (d') and (e'), in which only syntactic attributes were used. In an indirect manner, this complements our assertion that the best results may be achieved with a combination of semantic and syntactic attributes rather than using one type of attribute exclusively. Looking at the performance of the attribute sets with respect to each learning algorithm in experiment #2, it can also be seen that the highest results for the Decision Tree and Naïve-Bayes algorithms were obtained using attribute set (f'). This result suggests that perhaps the One Rule algorithm may be too simplistic for this task. Because we are aiming to determine the optimal combination of semantic and syntactic attributes to correctly classify the instances, it is important for the learning algorithm to take all attributes into account and not disregard all but one, which is the case with the One Rule algorithm.

The initial results of the two-stage application were encouraging because they showed that it is possible to outperform the baseline set of attributes using a combination of semantic and syntactic features. However, the incorporation of false negatives into the second stage contributed to a weaker performance by the system. As a result, we made several effective modifications to our attribute sets in order to improve these results.

5.6 Modification #1: Grammatically Correct Sentences

5.6.1 Problem Description

In applying our revised syntactic attributes to relative pronoun correction in the context of French to English translation, there was one aspect of our study that diverged from Siddharthan's previous work in order to suit the task. This had to do with the fact that in his machine learning framework, Siddharthan relied exclusively on grammatically correct English sentences. That is, all the sentences in his dataset, in addition to containing one of the relative pronouns *who* or *which* and the **NP1 Prep NP2** phrase structure, did not contain any other grammatical or orthographical mistakes. Due to the nature of this application, we did not make the same assumption. While the translations produced by a machine translation tool may give the user a basic understanding of the text, no tool achieves perfect results in all cases and as such, the translation may often still contain a number of other errors. For this reason, our dataset consisted of translated English sentences, which, at minimum, contained one incorrect relative pronoun. In the majority of cases, the sentence contained other grammatical errors and we attempted to correct the pronoun translation in the presence of these errors.

Based on the results obtained for the relative pronoun correction task detailed in Section 5.5, we were interested in seeing if the performance of our system was affected by the fact that our system did not utilize grammatically correct sentences. To verify this hypothesis, we manually modified the translated English sentences so that every element was correct, but we did not alter the relative pronoun produced by the machine translation tool. Table 25 contains an example of the unaltered English translation, along with its corrected version and demonstrates how the corrections made to the English sentences pertained to word order, word sense disambiguation, punctuation, selection of prepositions and collocations, and insertion or omission of words. Wherever possible, we retained the words from the original translation and maintained the given word order. Although the position of the relative pronoun was modified in certain cases, we did not alter the relative pronoun selected by the machine translator.

Table 25 - Example of French to English Machine Translation and Revision

Original French Sentence	Le Service de la protection a embauché un étudiant en géographie à l'automne 2002, Yannick Lanthier, dont la responsabilité première était de voir à la mise sur pied d'un tel système au sein du Service [32].
English Sentence produced by Machine Translator	The Service of protection engaged a student in geography with the autumn 2002, Yannick Lanthier, for which the responsibility first was to see with the setting-up of such a system within the Service [33].
Revised Translation determined by Author	The Service of Protection hired a geography student in autumn 2002, Yannick Lanthier, **which** first responsibility was to oversee the setting up of such a system within the Service.

Once all the sentences in the dataset were corrected, they were once again encoded using our revised attribute sets. The same experiments were repeated to verify the validity of our hypothesis.

5.6.2 Experimental Results

We evaluated the performance of our modified framework on the two-stage application by repeating the experiments described in Section 3.5. The results are given in Tables 26-29.

Table 26 - Stage 1 Experimental Results: Recall Rate

Stage 1 Experiment: *Recall Rate*			
Attribute Set	*1R*	*DT*	*NB*
a	**0.92**	0.81	0.62
b	**0.92**	0.78	0.62
c	**0.92**	0.81	0.65

d	0.38	**0.92**	0.63
e	0.54	**0.92**	0.67
f	0.54	**0.78**	0.64

From Table 26, it can be seen that with further modification to the system, it remained very effective in the stage 1 identification of English sentences containing incorrect relative pronouns. For the majority of cases, the recall rate remained the same as in the previous case where the instances in the test set contained other grammatical errors in addition to the incorrect relative pronoun. However, the recall achieved for attribute set (d) with the One Rule algorithm was slightly lower than previously, which indicates that exclusive use of syntactic attributes in the stage 1 classification task was not the most effective, particularly in this attribute set which contained only five new syntactic attributes. In general, with the correction of all instances in the dataset except the relative pronoun, between 78% and 92% of negatives instances were propagated into the second stage.

Table 27 - Stage 1 Experimental Results: Precision Rate

Stage 1 Experiment: ***Precision Rate***			
Attribute Set	*1R*	*DT*	*NB*
a	0.60	0.83	0.69
b	0.60	0.83	0.64
c	0.60	0.83	0.65
d	0.51	0.60	0.60
e	0.71	0.60	0.67
f	0.71	0.83	0.69

The precision rates achieved by the system in stage 1, seen in Table 27, were also improved when we modified the system. With the exception of the Decision Tree algorithm on attribute sets (a), (b) and (c), the number of instances classified as negative that were in fact negative, increased. This meant that there were fewer false negatives carried into stage 2, as seen in Table 28. The false negative rate ranged between 16% and 37%, which meant that no more false negatives were carried into the second stage than previously.

Table 28 - Stage 1 Experimental Results: False Negative Rate

Stage 1 Experiment: *False negative rate*			
Attribute Set	*1R*	*DT*	*NB*
a	0.64	**0.17**	0.29
b	0.64	**0.16**	0.35
c	0.64	**0.17**	0.36
d	**0.37**	0.64	0.42
e	**0.23**	0.64	0.34
f	0.23	**0.16**	0.29

For stage 2, we re-ran the experiments, first carrying forth only the true negatives from stage 1 and second, with both the true and false negatives from stage 1. In both cases, the percentage of correctly classified instances was calculated with and without the inclusion of the relative pronoun as an attribute. The results are shown in Tables 29 and 30.

Table 29 - Stage 2 Experimental Results: % Correctly Classified Instances *(true negatives only from stage 1)*

Stage 2 **Experiment #1**				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a	***82.02***	***80.92***	***78.94***	***80.63***
b	84.86	82.02	85.80	84.23
c	84.86	82.02	87.86	84.91
d	84.86	82.02	83.98	83.62
e	84.86	82.02	81.32	82.73
f	87.43	88.68	85.43	87.18
Experiment #2				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a'	***76.74***	***78.27***	***75.62***	***76.88***
b'	84.86	76.74	84.19	81.93
c'	84.86	76.74	88.06	83.22
d'	84.86	76.74	84.74	82.11
e'	84.86	76.74	81.23	80.94
f'	87.43	82.38	86.46	85.42

Table 30 - Stage 2 Experimental Results: % Correctly Classified Instances *(true and false negatives from stage 1)*

Stage 2: Experiment #1				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a	***76.36***	***74.91***	***74.09***	***75.12***
b	77.30	74.98	77.91	76.73
c	75.92	73.91	80.68	76.84
d	61.46	68.46	71.69	67.20
e	80.41	79.37	78.19	79.32
f	70.99	78.14	77.00	75.38
Experiment #2				
Attribute Set	*1R*	*DT*	*NB*	*Average*
a'	***71.82***	***71.18***	***72.55***	***71.85***
b'	77.30	70.66	77.46	75.14
c'	75.92	69.48	80.23	75.21
d'	61.46	64.62	69.85	65.31
e'	80.41	75.06	77.85	77.77
f'	70.99	72.84	76.48	73.44

Although the results from Table 29 support our previous results in which all the revised attribute sets outperform the baseline set, Table 30 shows the true improvement of the system because this case includes both the true negatives and false negatives from stage 1. Similar to previous experiments, the results between Table 29 and 30 show an overall drop in the percentage of correctly classified instances for the One Rule and Decision Tree algorithms, but remain consistent for the Naïve Bayes algorithm. More importantly, all attribute sets with the exception of set (d') outperform the baseline system. The average percentage of correctly classified instances for the three learning algorithms on attribute sets (b'), (c'), (e'), and (f') solidly exceeded the results obtained for the baseline set of attributes. This demonstrates that the presence of other errors in our sentences did have an adverse effect on the results obtained using our attribute sets in the initial experimental case. By correcting all errors in the sentences, we achieved a higher average for four out of five attribute sets than the baseline attribute set. Although attribute set (d') continued to produce a lower percentage of correctly classified instances than the baseline set (a') with the presence of both true and false negatives in stage 2, it does support the notion that relying exclusively on one particular type of attribute, be it syntactic or semantic, is not effective in achieving the goal of improving the baseline results. This result also indicates that the number of attributes used in the classification schemes plays an important role in achieving a high percentage of correctly classified instances.

In keeping with the experimental procedures described in Section 3.5, we once again calculated the percentage confidence interval estimate using the paired t-test to compare the baseline attribute set (a') and the best of our revised attribute sets. This time, as seen in experiment #2 of Table 30, the baseline attribute set (a') achieved the highest percentage of correctly classified instances when the Naïve Bayes algorithm was used, and the strongest result achieved by our revised attribute sets was for attribute set (e') with the One Rule algorithm.

In this case, the confidence interval estimate between these two results was calculated to be 7.86 ± 3.51%. This confidence interval was similar to the one achieved for our first application and signified that we could accept the hypothesis that the error achieved by the baseline attribute set (a') was greater than the error of our revised attribute set (e') with a confidence of 7.86%.

In general, the correction of all the grammatical errors in our dataset did improve the percentage of corrected relative pronoun errors and demonstrate the effectiveness of our revised syntactic attributes. However, we attempted one more modification to build on these results.

5.7 Modification #2: Attribute Selection

5.7.1 Problem Description / Theoretical Background

In our final attempt to improve the results of our revised syntactic attributes applied to relative pronoun correction, we examined the ability of our three machine learning algorithms to select attributes for predicting the classes of our instances. Attribute selection is the process by which "a subset of the original attribute set is collected such that the classifier will perform *at least as well* on this subset as on the original set of attributes [23].

So-called divide-and-conquer algorithms such as the C4.5 Decision Tree and Rule-Based algorithms such as One Rule may both be detrimentally affected by the inclusion of random irrelevant attributes. In both cases, the amount of data used to make classification decisions is gradually reduced, which means that eventually, the irrelevant attribute may be selected to classify the instances. According to previous experiments performed using the C4.5 decision tree classifier, the addition of a random binary attribute to standard datasets caused the results to deteriorate by nearly ten percent. This is because at each node in a decision tree, an attribute is selected to split on and continue branching. The problem occurs as we delve deeper into the tree, there are fewer attributes from which to choose. As a result, the likelihood that the random irrelevant attribute will be selected and thus affect the results unfavourably is significantly increased [21].

Because the fundamental assumption of the Naïve-Bayes algorithm is attribute independence, it is able to ignore random irrelevant attributes. However, this same assumption makes the performance of the classifier very sensitive to the influence of redundant attributes. For example, if a new attribute were added to the set with the same values as an existing attribute, this would increase the influence of this existing attribute in a multiplicative fashion. Simply put, the higher the number of attributes added to the attribute list with the same values as an existing attribute, the higher the influence of this attribute on the classification decision [21]. Because the three learning algorithms we used for our experiments are sensitive to the inclusion of relevant and irrelevant attributes, we opted to use attribute selection in attempt to improve our experimental results.

There are three main approaches commonly used for attribute selection. The first and best method is manual selection, which involves the selection of relevant attributes "based on a deep understanding of the learning problem and what the attributes actually mean" [21]. However, there are also several effective automatic methods. The first of these automatic methods is known as the filter method, where the data alone is used to select the subset of attributes. The wrapper, or scheme-specific, selection method is the second of the automatic methods. In this case, the attribute selection process is linked to the classifier used on the data [21].

5.7.2 Experimental Results: Attribute Selection

For our experiments, we performed attribute selection on 200 instances from our dataset. We then partitioned the remaining 201 instances and performed training on half of this data, leaving the other half for testing.

To perform attribute selection, we used the Information Gain evaluator, which "evaluates the worth of an attribute by measuring the information gain with respect to the class" [21]. We also selected the Ranker search method in WEKA, where the attributes are ranked based on their individual evaluations, and selected a value of zero for the evaluation threshold. With this experimental setup, we performed attribute selection on our six sets of attributes. For sets (a), (b), (c), and (f), the same four attributes ranked below the prescribed threshold. As a result, the NP1_isact, NP1_isanimal, NP2_isact and NP2_isanimal attributes were eliminated from the attribute sets. It is important to note that all of these attributes belonged to the original set of Word Net semantic-based attributes from Siddharthan's work. For attribute sets (d) and (e) which contained only the new syntactic attributes we devised, the rankings for all attributes fell within the pre-determined threshold so none of the attributes in these datasets were removed. Since there were no changes to attribute sets (d) and (e), the performance of the machine learning algorithms on these attribute sets remained unchanged and therefore, there was no need to give them further consideration for this portion of experiments.

5.7.3 Experimental Results: Relative Pronoun Correction

Once the four semantic attributes were removed from sets (a), (b), (c) and (f), we repeated the two-stage experiments. The recall results are given in Table 31. When these results are compared to those of Table 26, it can be seen that the recall rate for attribute sets (a) and (b) remained unchanged while the recall rate for sets (c) and (f) decreased. This meant that the number of negative instances propagated to stage 2 was reduced for these two attribute sets.

Once again, taking into account the realistic performance of the system, we examined the number of false negatives carried forward into stage 2. Table 32 shows the precision rates achieved for the four attribute sets by our three machine learning algorithms. With the exception of the One Rule algorithm on set (c), the precision rate remained unchanged from the previous experiments in all cases. This indicates that the number of false negatives carried into stage 2 remained the same for all cases except for attribute set (c).

Table 31 - Stage 1 Experimental Results: Recall Rate

Stage 1 Experiment: *Recall Rate*			
Attribute Set	*1R*	*DT*	*NB*

a	**0.92**	0.75	0.62
b	**0.92**	0.75	0.62
c	0.38	**0.75**	0.65
f	0.54	**0.75**	0.66

Table 32 - Stage 1 Experimental Results: Precision Rate

Stage 1 Experiment: ***Precision Rate***			
Attribute Set	*1R*	*DT*	*NB*
a	0.60	0.83	0.69
b	0.60	0.83	0.64
c	0.51	0.83	0.65
f	0.71	0.83	0.69

The stage 2 experiments were repeated on attribute sets (c) and (f) since these were the only two attribute sets for which the stage 1 results differed significantly from previous experiments. The stage 2 results obtained using attribute selection can be seen in Table 33.

Table 33 - Stage 2 Experimental Results: % Correctly Classified Instances *(true and false negatives from stage 1)*

Stage 2: **Experiment #1**				
Attribute Set	*1R*	*DT*	*NB*	*Average*
c	69.81	78.97	79.44	76.07
f	69.81	78.97	76.97	75.25
Experiment #2				
Attribute Set	*1R*	*DT*	*NB*	*Average*
c'	69.81	73.34	80.54	74.56
f'	69.81	73.34	77.62	73.59

Comparing the results of experiments 1 and 2 from Table 33 to our previous results in Table 30, it can be seen that on average, using attribute selection did not help to exceed our previous results. However, looking specifically at the performance of the Decision Tree algorithm, the percentage of correctly classified instances increased in all cases, which demonstrates that the removal of irrelevant attributes was an effective strategy for improving the results obtained by this learning scheme because it reduced the chances that an irrelevant attribute would be selected to classify our instances.

5.8 Application #2 Summary

The overall results for the two-stage relative pronoun automatic correction system are summarized in Table 34. Once again, these results were calculated by taking the average values across each attribute set and machine learning algorithm in experiment #2 of Table 30.

Table 34 - Overall System Results for Application #2

Stage 1	
Precision	67.72%
Recall	68.28%
Stage 2	
% sentences corrected appropriately	73.12%
% sentences corrected inappropriately	26.88%

For this application, we re-used our attribute sets for the correction of relative pronouns in the context of French to English machine translation. For our two-stage system, sentences were identified as correct or incorrect in the first stage, and then the incorrect sentences were corrected by classification according to the correct relative pronoun in stage 2. Although the results from stage 1 demonstrated the effectiveness of the revised syntactic attributes in achieving a high recall rate and precision, the performance of the system in stage 2 was not as effective. In this case, none of our attribute sets achieved a higher percentage of correctly classified instances than the baseline set, which indicated that we needed to revise our methodology. To modify our system, we corrected the sentences in our datasets for all grammatical errors except the selection of relative pronoun and re-ran our experiments for the two stages. In stage 1, the recall and precision rates remained high and with these results, we proceeded to the second stage, where the true negatives and false negatives from stage 1 were classified according to the correct relative pronoun. In our stage 2 experiments, all but one attribute set outperformed the baseline set. The consistently low performance of attribute set (d') with the One Rule algorithm indicates the tendency of this learning scheme to gravitate towards highly branching attributes, which in the worst case have a different value for each instance. While the selection of a highly branching attribute yields strong results in training, the attribute is often very unsuccessful in classifying test instances, which appears to be the case with our experiments [23].

We attempted once again to improve on our two-stage results for relative pronoun correction, this time using attribute selection to remove redundant or irrelevant attributes. Attribute selection led to the removal of four semantic attributes from attribute list, which were present in attribute sets (a), (b), (c) and (f). We repeated the stage 1 and 2 experiments and found that in stage 1, the results obtained for attribute sets (a) and (b) remained unchanged. As a result, the stage 2 experiments were conducted on attribute sets (c) and (f) only. The difference between the previous results and those obtained after attribute selection was minimal, but there was a significant improvement to those obtained using the Decision Tree algorithm.

With the modifications we made to our two-stage system, we were able to improve the initial results and outperform the baseline set of attributes in four out of five cases. The consistently high performance of attribute sets (b), (c), (e), and (f) is significant because in this application, we went from a two-class problem in stage 1 to a six-class problem in stage 2. In the first stage, with a two-class problem, there was a 50% probability of correctly classifying sentences as grammatically correct or incorrect. However, in stage 2, we branched out into a multi-class problem, where the probability of correctly classifying a sentence according to the appropriate relative pronoun decreased to less than 20%. This makes the improvement of our revised attribute sets all the more considerable because we were able to outperform the baseline attribute by approximately 5%. In doing so, we validated the theory that the combination of revised syntactic along with the existing attributes is beneficial to the task of relative pronoun correction in French to English machine translation.

Chapter 6: Conclusions & Future Work

6.1 Summary

In this study, we devised new syntactic attributes to be used in a two-stage machine learning system. Using Siddharthan's work for resolving relative clause attachment ambiguity as a starting point, we studied the grammatical structure of sentences in order to create new syntactic attributes. The syntactic attributes we incorporated took the form of POS tags because our intention was to detect structural patterns in sentences using machine learning algorithms. Once the syntactic attributes were developed, we formulated five attribute sets, each containing a different number of new syntactic attributes along with Siddharthan's attributes. We were interested in comparing the performance of these attribute sets with the baseline attribute set which contained attributes from the previous system exclusively, in various classification tasks. To start, we applied our revised syntactic attributes to Siddharthan's study in which relative clauses were classified according to local or wide attachment. In this classification task, four out of five revised attribute sets outperformed the baseline set.

To further validate our hypothesis, we then applied our revised syntactic attributes to two separate applications. In the first application, we looked at the classification of sentences according to the type of relative clause they contained. From here, we attempted to simplify the sentences by removing the relative clause from the sentence. We explored the correction of relative pronouns in the context of French to English machine translation for our second application. The incorrect English relative pronouns produced by a MTS were identified and subsequently corrected.

To evaluate these applications, we devised a common two-stage system, in which we applied three machine learning algorithms on our six attribute sets. In the first stage, we used binary classification to classify the instances as positive or negative. For application #1, the first stage served to distinguish between sentences containing RRCs and NRRCs, whereas for application #2, the purpose of the first stage was to distinguish between sentences containing correct and incorrect relative pronouns. From here, the true negative and false negative instances continued to the second stage for a more complex classification task. For application #1, instances were classified according to whether relative clauses could be removed for simplification purposes and for application #2, instances were classified according to the correct relative pronoun.

In initial testing, the attribute sets in which we incorporated our new syntactic attributes did not outperform the baseline attribute set. However, once we made several effective modifications to the system, four out of the five attribute sets in which we combined syntactic attributes with the existing ones produced higher results than the baseline attribute set alone.

6.2 Conclusions

There are a number of conclusions we can draw from the results we achieved using our revised syntactic attributes in two separate applications.

The first conclusion we can draw based on our experimental results is that the choice of machine learning algorithms is essential to our classification tasks because each one has its own benefits and drawbacks. In our two applications, the results achieved using the One Rule algorithm remained on par with the Decision Tree and Naïve Bayes classifiers. This result is surprising because One Rule is a very simplistic algorithm in which only one attribute is used to determine the class of an instance. In our experiments, the consistently strong performance of this rudimentary learning scheme may be due to the fact that our applications were also very simplistic in the sense that deep semantic knowledge of the sentences in our datasets was not required. By performing a detailed syntactic analysis of the sentences, we were able to extract and incorporate relevant syntactic attributes into the attribute sets which were effective enough to produce results with the One Rule algorithm that were comparable to the Decision Tree and Naïve Bayes algorithms. Despite the fact that One Rule is a very basic algorithm, the performance achieved by this classifier using our attribute sets demonstrated that inexpensive methods may achieve effective rules for characterizing data.

For its part, the Naïve Bayes algorithm may also be susceptible to errors under certain conditions. For instance, the addition of redundant attributes may influence the ability of this algorithm to classify instances because the effect is multiplicative. As a result, as more attributes with the same values were added to our attribute sets, the influence of a particular attribute on this classifier grew as well.

Finally, the Decision Tree algorithm was the most complex of our three classifiers. This classifier is preferred for simpler classification tasks because it "splits the training set into smaller and smaller subsets. This makes correct generalization harder, since there may not be enough data for reliable prediction, and incorrect generalization easier, since smaller sets have accidental regularities that don't generalize" [22].

In general, the three machine learning algorithms used in our two-stage evaluation system achieved fairly similar results despite the range of complexity between them.

The second conclusion we can draw from the results of our study is that it is in fact possible to apply revised syntactic attributes and a two-stage evaluation methodology to two independent but related NLP tasks. In addition to demonstrating the versatility of our revised syntactic attributes, we evaluated its effectiveness using a common evaluation system rooted in machine learning techniques. Although the classification tasks in the two stages differed for each application, we were able to apply the same system architecture for both relative clause simplification and relative pronoun correction.

The third conclusion which can be drawn from this study is that we have developed a flexible and adaptable set of syntactic attributes. In our application in which relative clauses were classified according to clause type, we incorporated five new syntactic attributes into our attribute sets based on grammar rules for relative clauses. We re-ran the same experiments and the results showed a marked improvement in correctly classifying negative instances. These results strongly suggest that tailoring the attribute sets to the specific task is an effective way to improve the system results.

Lastly, based on the results achieved by the three learning algorithms on the six attribute sets in our two applications, we can conclude that a combination of our revised syntactic attributes and the existing attributes is more effective than the baseline set of attributes for our classification tasks. The improvement achieved by our revised attribute sets compared to the baseline set was especially important in the relative pronoun application because in this case, we progressed from a two-class problem in stage 1 to a multi-class problem in stage 2. As such, the probability of correctly classifying the instances in stage 2 was significantly reduced, but we were able to outperform the baseline set in four out of five cases.

In both of our applications, four out of five attribute sets outperformed the baseline attribute set. However, in both applications, the attribute sets which consisted solely of syntactic attributes never succeeded in producing higher results than the baseline set. This means that neither semantic nor syntactic attributes alone are as effective as a combination of both types.

We have expanded previous work for resolving relative clause attachment ambiguity to include new syntactic attributes and applied these attributes in two separate but related applications using a two-stage system which makes use of machine learning algorithms and evaluation techniques. Four out of five new attribute sets which combine semantic and syntactic attributes successfully outperformed the baseline set of attributes. While the results show a significant improvement from previous work, it is still possible to build on our system.

6.3 Future Work

The first future task will be to consolidate the values that can be taken by the new syntactic attributes we introduced into Siddharthan's machine learning framework. As seen in Table 3 and discussed in Section 3.2, the number of possible values for the POS tag attributes was very high. Although sixteen of these sixty-two values were not assigned to the instances in our datasets and the majority of instances were described with only twenty of the POS tags, it may be argued that with a better distribution of the attribute values, the performance of our machine learning algorithms would improve as well. By consolidating the values of this attribute by reducing the number of forms for each part of speech, we will achieve a better attribute-value distribution and thus, the performance of the machine learning algorithms may be improved.

In terms of building on our revised syntactic attributes, there are also several interesting tasks which could be explored.

For our first application in which non-restrictive relative clauses were identified in the first stage and then recommended for removal in the second stage, the input of the system was an English sentence while the output was a simplified version of the original English sentence with the relative clause was removed. One important task would be to make use of a tool such as [34] in order to evaluate how the summarized versions of the sentences produced by our two-stage system compare to the original sentence. In [34], there is "Skip Bi-gram" evaluation tool which allows for the comparison of bi-grams in the original sentence with any pair of words in the reduced version while allowing for arbitrary gaps.

In a similar vein, for our second application in which relative pronouns were corrected in English translations of French text, it would be worthwhile to determine how important having a correct relative pronoun is to the overall understanding of English sentences. One method we could use to quantify the significance of our results would be to take our dataset of sentences containing incorrect relative pronouns and run the evaluation process from [35] on them. Then, we could repeat the same evaluation process on the sentences in which the relative pronouns were corrected using our two-stage machine learning system. The difference between the two results would be one measure of the effectiveness of our system.

A third task which could be undertaken is to combine our two applications. Starting with relative pronoun correction on English sentences translated from French text, we could then take these corrected sentences and identify the type of relative clauses contained within them. From here, depending on the type of relative clause, it could be removed for text simplification purposes. This means that the input to the system would be complete French sentences, while the output would consist of simplified English translations. We could then compare the original French sentences and the simplified English output using a combination of [34] and [35] because this process involves both translation and simplification.

Once the value of our two applications has been ascertained using the quantitative measures described above, it will be necessary to consider the practical utility of our revised syntactic attributes by exploring the construction of tools to automatically extract the values of these attributes. In particular, the five attributes described in Sections 4.6.1 to 4.6.3 were developed specifically to improve the first stage of the relative clause identification and text simplification application. We tested our revised syntactic attributes manually and our results demonstrated that they were effective in two different applications. Therefore, the development of automated software to assign values to our syntactic attributes is a crucial step in preparing our system for real-world use.

By performing these future tasks, it will be possible to ascertain the benefit of our revised syntactic attributes and make the step towards implementing our two-stage evaluation system.

REFERENCES

[1] J.O. Jespersen, A *Modern English Grammar on Historical Principles: Part III.* London: Allen & Unwin, 1949.

[2] A. Siddharthan, "Resolving Relative Clause Attachment Ambiguities using Machine Learning Techniques and WordNet Hierarchies", in *Proceedings of the 5th National Colloquium for Computational Linguistics in the UK, (CLUK 2002),* 2000, pp. 45-49.

[3] M. Newbrook, "Which Way? That Way? Variation and Ongoing Changes in the English Relative Clause", *World Englishes*, vol. 17, no.1, pp43-59, 1998.

[4] R. Quirk et al., *A Comprehensive Grammar of the English Language*. London: Longman, 1985.

[5] R. Huddleston and G. Pullum, *The Cambridge Grammar of the English Language.* New York: Cambridge University Press, 2002.

[6] D. Carter, *Interpreting Anaphors in Natural Language Texts*. U.K: Ellis Horwood, 1987.

[7] W. Soon, H. Ng, and D. Lim, "A Machine Learning Approach to Coreference Resolution of Noun Phrases", *Computational Linguistics,* vol. 27, no. 4, pp. 521-544, Dec. 2001.

[8] C. Cardie, "Corpus-Based Acquisition of Relative Pronoun Disambiguation Heuristics", in *Proceedings of the 30th Annual Conference of the Association for Computational Linguistics*, 1992, pp. 216-223.

[9] D. H. Fisher, "Acquisition Via Incremental Conceptual Clustering", *Machine Learning*, pp. 139-172.

[10] R. Chandrasekar, C. Doran, and B. Srinivas, "Motivations and Methodologies for Text Simplification", in *Proceedings of the Sixteenth International Conference on Computational Linguistics (COLING '96)*, 1996, pp. 1041-1044.

[11] H. Jing, "Sentence Reduction for Automatic Text Summarization", in *Proceedings of the First Annual Meeting of the North American Chapter of the Association for Computational Linguistics NAACL-2000*, 2000, pp. 310–315.

[12] R. Chandrasekar, "A Hybrid Approach to Machine Translation Using Man Machine Communication", Ph. D. dissertation, Tata Institute of Fundamental Research/University of Bombay, Bombay, 1994.

[13] P. Brown et al., "A Statistical Approach to Machine Translation", *Computational Linguistics*, vol. 16, no. 2, pp.79-85, 1990.

[14] D. Marcu and W. Wong, "A Phrase-Based Joint Probability Model for Statistical Machine Translation", in *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, July 2002, pp. 133-139.

[15] Kenji Yamada and Kevin Knight, "A Syntax-based Statistical Translation Model", in *Proceedings of the 39th Annual meeting of the Association for Computational Linguistics,* 2001, pp 523-530.

[16] M. Marcus, B. Santorini, and M. Marcinkiewicz, "Building a Large Natural Language Corpus of English: The Penn Treebank, *Computational Linguistics*, vol. 19, pp. 313-330, 1993.

[17] Princeton University, "WordNet: A lexical database for the English language", *Cognitive Science Laboratory, Princeton University*, 2005. [Online]. Available: http://wordnet.princeton.edu/perl/webwn. Accessed: 17 January 2005].

[18] A. Carlson et al, "The SNoW learning architecture", UIUC Computer Science Department, Tech. Rep. UIUCDCS-R-99-2101, 2000.

[19] University Centre for Computer Corpus Research on Language, "Free CLAWS WWW Trial Service", UCREL, Date unknown. [Online]. Available: http://www.comp.lancs.ac.uk/ucrel/claws/trial.html. [Accessed December 2005].

[20] Edict Incorporated, "Relative Clauses", *Edict Incorporated*, Date unknown. [Online]. Available: http://www.edict.com.hk/vlc/clauses/relativeclauses.htm. [Accessed 16 May 2005].

[21] I. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Ttechniques*, 2nd Edition. San Francisco: Morgan Kaufmann Publishers, 2005.

[22] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge: The MIT Press, 1999.

[23] T. Mitchell, *Machine Learning*. New York: McGraw Hill, 1997.

[24] University of Ottawa, "Gazette", *Various Authors*, 2004. [Online]. Available: http://www.gazette.uottawa.ca/archiveindex_e.html. [Accessed: October-December 2005].

[25] J. Yamashita, "An Analysis of Relative clauses in the Lancaster/IBM Spoken-English Corpus", *English Studies*, vol. 75, no. 1, pp. 73-84, 1994.

[26] I. Depraetere, "Factors Requiring, Promoting and Excluding the Use of a (Non-) Restrictive Relative Clause", *Leuvense Bijdragen*, vol. 83, no. 4, pp. 433-458, 1995.

[27] H. Tao and M. J. McCarthy, "Understanding Non-Restrictive Which-Clauses in Spoken English, Which is not an Easy Thing", *Language Sciences*, vol. 23, pp. 651-677, 2001.

[28] Language Dynamics, "Modal Verb Tutorial", Language Dynamics, 2004. [Online]. Available: http://www.englishpage.com/modals/modalintro.html. [Accessed 15 April 2006].

[29] D. Jurafsky and J. Martin, *Speech and Language Processing – An introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.* New Jersey: Prentice-Hall Inc., 2000.

[30] N. Fabb, "The difference between English Restrictive and Non-Restrictive Relative Clauses", *Journal of Linguistics*, vol. 26, pp. 57-77, 1990.

[31] A. Cornilescu, "Nonrestrictive Relative Clauses: an essay in semantic description", *Revue Roumaine de Linguistique*, vol. 26, no. 1, pp. 41-67, 1981.

[32] University of Ottawa, "Gazette", *Various Authors*, 2004. [Online]. Available: http://www.gazette.uottawa.ca/archiveindex_f.html. [Accessed: October-December 2005].

[33] Alta Vista, "Babel Fish Translation", *Alta Vista*, 2006. [Online]. Available: http://babelfish.altavista.com. [Accessed October-December 2005].

[34] C. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries", in *Proceedings of the Workshop on Text Summarization Branches Out (WAS 2004)*, July 25-26, 2004.

[35] K. Papineni, S. Roukos, T. Ward, and W. Zhu, "BLEU: A Method for Automatic Evaluation of Machine Translation, IBM Research Report RC22176 (W0109-022).

ADDITIONAL RESOURCES

A. Siddharthan, "Resolving Attachment and Clause Boundary Ambiguities for Simplifying Relative Clause Constructs", in *Proceedings of the Student Research Workshop, 40th Meeting of the Association for Computational Linguistics (ACL 2002)*, 2002, pp.60-65.

B. Levin, *English Verb Classes and Alternations: A preliminary investigation.* London: The University of Chicago Press Ltd., 1993.

Murdoch University, "How To Cite References – IEEE Style", Murdoch University Library, January 2006. [Online]. Available: http://wwwlib.murdoch.edu.au/find/citation/ieee.html. [Accessed 2 May 2006].

Online English Grammar, "Relative Clauses", *Edufind*, Date unknown. [Online]. Available: http://www.edufind.com/english/grammar/rel2.cfm. [Accessed 27 February 2006].